Neural Mechanisms Underlying the Emergence of Rhythmic and Stereotyped Vocalizations in Juvenile Songbirds

by

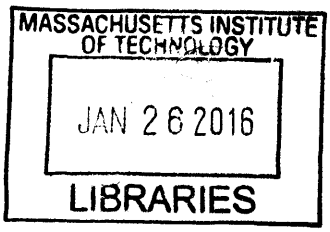
Tatsuo Okubo

B.S., Mathematical Engineering and Information Physics, University of Tokyo, 2006
M.S., Information Physics and Computing, University of Tokyo, 2008

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN NEUROSCIENCE
at the
Massachusetts Institute of Technology

February 2016



ARCHIVES

Signature of Author: Signature redacted

Department of Brain and Cognitive Sciences
Sept 21, 2015

Certified by: Signature redacted

Michale S. Fee
Glenn V. and Phyllis F. Dorflinger Professor of Neuroscience
Thesis Supervisor

Accepted by: Signature redacted

Matthew A. Wilson
Sherman Fairchild Professor of Neuroscience and Picower Scholar
Director of Graduate Education for Brain and Cognitive Sciences

Neural Mechanisms Underlying the Emergence of Rhythmic and Stereotyped Vocalizations in Juvenile Songbirds

by

Tatsuo Okubo

Submitted to the Department of Brain and Cognitive Sciences on September 21, 2015 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Neuroscience.

Abstract

Complex motor behaviors in humans, such as speech, are not innate, but instead are learned. How does the brain construct neural circuits that generate these motor behaviors during learning? To understand the neural mechanisms underlying learned motor skills, I use vocal learning in songbirds as a model. While previous studies have shown that a premotor area in the songbird brain, HVC, is important for stereotyped adult song, the role of HVC in juvenile song is less known. This thesis characterizes how activity in HVC develops during song learning in juvenile birds.

Early in song learning, temporal structure emerged in HVC. During the earliest vocalization of juvenile birds (subsong), HVC neurons exhibit bursts of action potentials. However, only half of the neurons show bursts that are temporally aligned to syllables, and most of these bursts are clustered around onsets of subsong syllables. Over several days, as the bird starts producing the earliest stereotyped vocalization called protosyllables, HVC neurons start exhibiting rhythmic bursts at 5-10 Hz. These rhythmic bursts are aligned to protosyllables, and bursts from different neurons are active at different latencies relative to protosyllables. Thus, as a population, HVC neurons start forming a rhythmic neural sequence.

As the bird matures, multiple distinct syllable types emerge from a protosyllable. During this process, some neurons are active only during a specific syllable type ('specific neurons') while others are active during both syllable types ('shared neurons'). These shared neurons are active at similar latencies for both syllable types, and therefore form a shared neural sequence. Over development, fraction of shared neurons decrease and more neurons become specific. These results demonstrate that splitting of a neural sequence into multiple sequences underlies the emergence of a multiple syllable types. Moreover, this sequence splitting is observed during different song learning strategies, suggesting that this is a fundamental neural mechanism for song learning.

This work demonstrates how the growth of a rhythmic neural sequence and its subsequence splitting gives rise to complex vocalization in songbirds. This may be a general neural mechanism in which the brain constructs neural circuits during learning of a complex motor behavior.

Thesis supervisor: Michale S. Fee
Title: Glenn V. and Phyllis F. Dorflinger Professor of Neuroscience

Table of Contents

Acknowledgments

I would like to start by thanking my thesis adviser Michale Fee for his continuous guidance, support, and advice. I have learned an enormous amount by watching him identify a scientific problem and tackling it with full force. When I first read Michale's paper on HVC as an undergraduate student in Tokyo, I was amazed by how his lab was able to selectively record from a specific type of neurons in HVC and uncovered a beautiful pattern of activity during singing. This encouraged me to switch from engineering to biology. Thus, it was fascinating to be able to work on HVC and record these signals myself. Moreover, it was a great honor for me to be able to take part in his productive research program of trying to obtain a mechanistic understanding of how songbirds learn their song. I will truly miss the exciting lab environment he has created.

I would also like to thank my thesis committee members Matt Wilson, Sebastian Seung, and Ofer Tchernichovski. Matt Wilson provided his expertise on *in vivo* electrophysiology in behaving animals and on how to connect neural sequences with learning and memory. Sebastian Seung has developed an influential model of HVC and taught me the art of critical thinking: how to identify important questions, and how to present the results clearly. I am also delighted to have Ofer Tchernichovski on my committee since he is both a pioneer and an expert on song development in zebra finches. It was a wonderful experience to go back to his papers published ten years ago and see beautiful correspondences between what he has observed at the behavioral level and the neural activity that I have recorded. I would also like to thank the following people for their support: Ann Graybiel who served as a qual exam committee and gave me encouragement; Mehrdad Jazayeri and Charles Jennings for their valuable feedback on the work; Hironobu Sakaguchi and Kazuhiro Wada for encouraging me to work on songbirds. I would also like to thank the Nakajima Foundation and the Schoemaker fellowship for their support during my graduate studies.

Thank you to both previous and current lab members for their generous support and daily discussions: Michael Long, Jesse Goldberg, Tim Gardner, Liora Las, Yael Mandelblat-Cerf, Natalia Denissenko, Anusha Narayan, Maya Bronfeld, Ben Scott, Lena Veit, Jakob Foerster, Dan Rubin, Tim Currier, Anton Konovchenko, Shelbi Ferber, Collyn Messier, and Margo O'Leary. Special thanks to amazing previous graduate students, Aaron Andalman and Dmitriy Aronov who taught me everything from neural recordings to scientific thinking. Michael Stetner, a fellow graduate student, made an essential contribution to the work presented here by providing critical feedback during the numerous discussion sessions held in the student office. I have collaborated with two graduate students in the lab: Emily Mackevicius, who developed the mechanistic model of HVC presented in Chapter 5 and Galen Lynch, who developed statistical tools to analyzed HVC activity presented in Chapter 4. Their hard work and enthusiasm helped move the project forward, and I had great fun working with them. Hannah Payne, a graduate student in Raymond lab at Stanford, has also contributed significantly by developing the HVC model presented in Chapter 5, and has given me useful feedback on the manuscript.

I am fortunate to have friends both within and outside the department to whom I could talk about interesting things that are happening around the world. I would especially like to thank my BCS 2008 classmates, Nikhil Bhatla, Andrew Bolton, and Shin Kira. I also had tremendous fun hanging out with members of Japanese Association of MIT (JAM), and BCS Japan, a group

of Japanese research scientists and post-docs in the department. There are too many of them to list their names here individually, but I am grateful to each one of them.

Last but not least, I would like to thank my family members for their continuous support: my parents Yoshio and Miharu, my sister Saya, and my brother and his family Tomoo, Kanako, and Alyssa. I am glad that all of them had a chance to visit the lab and see the research in action. Finally, I would like to express my sincere gratitude to my wife Ni Ji for her generous support spanning a wide range of contributions from reading the manuscript to organizing fun trips. Thank you very much.

Chapter 1

Introduction

Martha Argerich plays the finale of Bartok's Third Piano Concerto with lightning speed. Koshiro Matsumoto, a distinguished Kabuki actor, performs a complex sequence of movements as he exits the stage at the end of a famous piece Kanjincho[1]. Andrés Iniesta swiftly moves through multiple defenders while maintaining control over the soccer ball until he finds an opportunity for a perfectly-timed pass. These are just a few examples of how professional musicians, actors, and athletes fascinate us through their finely executed skills. What are the neural mechanisms that allow these professionals to learn and perform skills composed of highly complex movements? This is the main topic of this thesis.

I begin this chapter by introducing the topic of motor learning and explaining the rationale behind using songbird as a model. I next summarize the literature and point out what is unknown in the field. I end this chapter by outlining the organization of the thesis.

Development of complex motor behavior

Development of complex motor behaviors in humans

Many complex motor behaviors in humans are not innate but are learned during development (Thelen, 1995; Forssberg, 1999; Graybiel, 2008). During this learning process, the behavior transitions from highly variable to stereotyped. This transition is exemplified in a task where infants learned to jump in a Jolly Jumper, an infant bouncer that consists of a harness with a linear spring (Goldfield et al., 1993). When 6-8 month old infants were first placed in a Jolly Jumper, kicking was irregular and variable in period, but after the exploration phase, infants were able to jump higher by producing a rhythmic kicking characterized by a large decrease in the variability of the period (Goldfield et al., 1993).

[1] This signature disappearing act is called Tobi-roppo, which is an exaggerated form of walking that expresses complex emotions inside the main character's mind as he leaves the stage.

Another example of learning complex motor behavior is speech acquisition, which requires temporally precise movements of vocal apparatus, and auditory feedback (Doupe and Kuhl, 1999). Infants go through multiple stages of vocal development in their first year of life. While the boundaries between these stages are not distinct, and there is variability across individuals, these stages typically consist of cry and fussing (0-6 weeks), cooing (6-16 weeks), vocal play (16-30 weeks) and babbling (6-14 month) after which single words are produced (Oller, 1980; Stark, 1980); this is followed by a rapid increase in the vocabulary (Oller, 1980; Stark, 1980) and a gradual acquisition of transitions between syllable types (Lipkind et al., 2013).

Development of complex motor behaviors in other animals

Development of complex motor behaviors is not unique to humans, and animals can also be trained to produce stereotyped motor sequences in the laboratory. For example, macaque monkeys can be trained to perform sequential movements composed of pushing, pulling, and turning a handle (Tanji, 2001), sequential button press (Hikosaka et al., 1999), or a sequential saccade (Isoda and Tanji, 2002). In these examples, animals were initially trained by visual cues indicating which movement to perform, and later learned to organize these movements into a sequence independent of the visual cue. On the other hand, sequence of movements can also develop through a gradual trial-and-error process without explicit instructions (Desrochers et al., 2010; Jin and Costa, 2010; Kawai et al., 2015).

While tasks mentioned above allow researchers to study neural mechanisms underlying sequential movements (Hikosaka et al., 1999; Tanji, 2001), it is difficult to follow the change in neural activity during the acquisition of such tasks given the long training period required to train

animals in such tasks. An alternative strategy is to examine natural behaviors that are composed of complex sequential movements. Although there are numerous examples of complex motor behaviors in animals, such as mating behavior in sticklebacks (Tinbergen, 1951), stereotyped sequence of grooming in rodents (Aldridge and Berridge, 1998) and in flies (Seeds et al., 2014), most of these behaviors are innate. In contrast, vocal learning in songbirds is an excellent example of a natural behavior consisting of complex movements that are learned (Konishi, 1985; Doupe and Kuhl, 1999), and this will be the next topic.

Songbird as a model for the development of complex motor behavior

Advantages of using songbirds

Birdsong is a prime example of a complex motor behavior composed of a sequence of multiple movements (Fee et al., 2004; Goller and Cooper, 2004; Suthers and Zollinger, 2004) since singing requires controlling the muscles of the vocal organ and the respiratory system on a millisecond timescale. In oscines, songs are not innate but instead learned by imitating the song of the tutor (Catchpole and Slater, 2008; Zeigler and Marler, 2008). Therefore, vocal learning in songbird has been an important model of a learning of an ethologically-relevant motor behavior (Zeigler and Marler, 2008; Mooney, 2009; Fee and Scharff, 2010; Brainard and Doupe, 2013; Mello, 2014).

Using songbird as a model is highly advantageous because of the availability of numerous techniques: one can record the entire song of an individual juvenile bird during song learning (Tchernichovski et al., 2000; Tchernichovski et al., 2001; Tchernichovski and Mitra, 2002; Liu et al., 2004), manipulate song by perturbing the auditory feedback (Leonardo and Konishi, 1999; Tumer and Brainard, 2007; Andalman and Fee, 2009), record from individual

neurons during singing (Fee and Leonardo, 2001; Okubo et al., 2014), lesion or pharmacologically inactivate certain brain regions to examine its effect on song (Nottebohm et al., 1976; Olveczky et al., 2005; Aronov et al., 2008), and generate transgenic animals (Agate et al., 2009; Scott et al., 2010; Abe et al., 2015).

Similarity between speech acquisition in humans and vocal learning in songbirds

Many studies have shown that speech acquisition in humans and vocal learning in songbirds show similarities (Marler, 1970; Doupe and Kuhl, 1999; Lipkind et al., 2013). Both the behavior and the neural pathway important for vocalization and vocal learning are similar between humans and songbirds. Even though the structure of the pallium differs between birds and mammals—birds have nuclei while mammals have cortical layers (Reiner et al., 2005)—molecular profiling and gene expression studies show that there are many similarities between the pallium of birds and mammals (Jarvis et al., 2005; Jarvis et al., 2013; Pfenning et al., 2014). In addition, cell types in basal ganglia are highly conserved between birds and mammals (Farries and Perkel, 2002; Doupe et al., 2005; Goldberg et al., 2010; Goldberg and Fee, 2010) further reinforcing the idea that study of vocal learning in songbirds will help us understand the neural mechanisms of speech acquisition in humans.

The song learning process

It has long been known that songbirds imitate the song of their tutors (Immelmann, 1969; Marler, 1970; Konishi, 1985). Multiple theories on how birds learn their song have been proposed: these include models based on reinforcement learning (Doya and Sejnowski, 1995, 1998, 2000; Fiete et al., 2007; Fiete and Seung, 2008), models that form an internally generated

prediction of the sensory feedback (Troyer and Doupe, 2000a, b; Troyer and Bottjer, 2001), and models that allows birds to perform one-shot learning using the inverse model (Hahnloser and Ganguli, 2013; Hanuschkin et al., 2013).

Recently, it has become possible to analyze this song learning process in detail by recording the entire vocal ontogeny of a juvenile bird (Tchernichovski et al., 2001; Tchernichovski and Mitra, 2002). Here, I summarize the results of these studies by describing the song learning process in a developmental order (Figure 1.1).

Subsong. A male juvenile zebra finch starts out singing highly variable vocalization at around 30 days post hatch (dph) (Immelmann, 1969; Zann, 1996). This variable vocalization is called 'subsong,' and is akin to human babbling (Marler, 1970; Doupe and Kuhl, 1999). Both the acoustic structure and temporal structure of subsong is variable. Distribution of syllable durations in subsong follows an exponential distribution, suggesting that this is a random process (Aronov et al., 2008; Aronov et al., 2011). Gap duration is also variable, although some gaps that consist of inspiratory events are stereotyped in duration (Aronov et al., 2011; Veit et al., 2011).

Protosyllable stage. After several days of subsong, juvenile birds start singing syllables and gaps that are stereotyped in duration (Tchernichovski et al., 2001; Liu et al., 2004; Aronov et al., 2011; Veit et al., 2011), which are called 'protosyllables' and 'protogaps', respectively. Protosyllables and protogaps are the first stereotyped vocalizations that appear in juvenile song, and their appearance can be observed as a development of a peak in syllable or gap duration distribution (Aronov et al., 2011; Veit et al., 2011). Moreover, during this stage, birds tend to produce a back-to-back repetition of protosyllables separated apart by protogaps (Aronov et al., 2011; Veit et al., 2011); this gives rise to the appearance of rhythmicity in the song (Tchernichovski et al., 2001; Liu et al., 2004; Saar and Mitra, 2008). Thus, transition from

subsong to protosyllable stage is characterized by the emergence of stereotyped syllables and gaps, as well as the acquisition of song rhythmicity.

Multi-syllable stage. Following the protosyllable stage, multiple distinct syllable types, characterized by distinct acoustic structure or syllable durations, start to emerge (Tchernichovski et al., 2001; Tchernichovski and Mitra, 2002; Liu et al., 2004). The appearance of multiple distinct syllable types can be observed in the spectrogram (Figure 1.1), and as an emergence of distinct clusters in the acoustic feature space (e.g. Figure 3.2b).

Motif stage. The bird then assembles the multiple syllable types into a 'motif', a sequence of 3-7 syllables sung in a fixed order. During this stage, acoustic structures of syllables continue to become more similar to that of the tutor (Figure 1.1). Finally, at around 90 dph, the song crystalizes and both the acoustic and temporal structure of the song become highly stereotyped (Zann, 1996).

It is important to note that while most birds follow the developmental trajectory outlined above, there are individual variations in how juvenile birds learn their song (Liu et al., 2004). For example, some birds learn all the syllables in a motif nearly simultaneously without changing the syllable order ('motif strategy')(Liu et al., 2004). Birds can also add new syllable types at the beginning or end of bouts (Lipkind and Tchernichovski, 2011; Lipkind et al., 2013). In Chapter 4, I examine the neural mechanisms underlying these diverse strategies for song learning.

The song system: a network of inter-connected nuclei important for song

One benefit of using songbird as a model for neuroscience research is that the brain areas necessary for song production and learning often exist as nuclei that are cyto-architectually distinct from the surrounding areas (Nottebohm et al., 1976; Nottebohm et al., 1982; Feenders et

al., 2008). Most neurons in these nuclei show singing-related activity and lesions of these nuclei affect song, but not other behaviors.

These nuclei are interconnected and form the 'song system' (Figure 1.2). Two important pathways in the song system are the 'vocal motor pathway', and the 'anterior forebrain pathway' (AFP). Vocal motor pathway consists of a premotor area HVC (used as a proper name,) which project to the primary motor area, robust nucleus of the acropallium (RA). RA in turn innervates midbrain and brainstem areas that control the syringeal and respiratory muscles (Nottebohm et al., 1976; Nottebohm et al., 1982; Wild, 2004) and is homologous to the layer 5 of the primary motor cortex (Jarvis, 2004; Pfenning et al., 2014). On the other hand, AFP indirectly connects HVC and RA via three nuclei: song-related portion of the basal ganglia (Area X), dorsolateral portion of the medial thalamus (DLM), and lateral magnocellular nucleus of the anterior nidopallium (LMAN). AFP is analogous to the basal ganglia-thalamocortical-loop in mammals (Farries and Perkel, 2002; Doupe et al., 2005), and is important for song learning (Bottjer et al., 1984; Scharff and Nottebohm, 1991; Doupe et al., 2005; Andalman and Fee, 2009; Fee and Goldberg, 2011)

HVC—a premotor cortical area important for song production and learning

The premotor cortical area HVC is one of the most important brain areas given its unique position in the song system. Sometimes called the 'high vocal center,' HVC gives rise to both the vocal motor pathway and AFP (Figure 1.2). Moreover, auditory and motor signals converge in HVC (Figure 1.2). HVC receives projections from at least four other nuclei in the brain: motor thalamus nucleus uvaeformis (Uva), a cortical area medial magnocellular nucleus of the anterior

nidopallium (MMAN), and at least two auditory cortical areas, nucleus interface (NIf) and nucleus avalanche (Av).

HVC is composed of at least two types of projection neurons: neurons that project to RA (HVC_{RA}) and neurons that project to Area X (HVC_X) (Dutar et al., 1998; Kubota and Taniguchi, 1998; Mooney, 2000; Daou et al., 2013). HVC also projects to Av (Nottebohm et al., 1982; Akutagawa and Konishi, 2010), but less is known about this projection and it will not be discussed further. HVC_{RA} and HVC_X are distinct types of projection neurons since retrograde tracer injections in RA and Area X show no overlapping cell bodies in HVC (Wild et al., 2005). These neurons also show different physiological properties (Dutar et al., 1998; Kubota and Taniguchi, 1998; Mooney, 2000; Daou et al., 2013). HVC also contains local interneurons that are inhibitory (Dutar et al., 1998; Kubota and Taniguchi, 1998; Mooney, 2000; Wild et al., 2005; Kosche et al., 2015).

Role of HVC in song production and learning

Role of HVC in adult song

The first demonstration of the importance of HVC in generating stereotyped vocalizations came from lesion studies. Bilateral HVC lesion in canaries and zebra finches leads to a complete loss of the stereotyped pre-lesion song (Nottebohm et al., 1976; Simpson and Vicario, 1990). Interestingly, however, these birds with HVC lesions could still produce subsong-like vocalizations (Aronov et al., 2008; Veit et al., 2011). Moreover, transection of fibers connecting HVC to RA recapitulates the HVC lesion results (Aronov et al., 2008; Olveczky et al., 2011), suggesting that HVC controls stereotyped vocalizations through the vocal motor pathway.

Local cooling, which allows the manipulation of neural dynamics within a brain area without disrupting them (Aronov and Fee, 2011; Fee and Long, 2011), demonstrated the importance of HVC in controlling song timing (Long and Fee, 2008; Andalman et al., 2011; Aronov et al., 2011; Goldin et al., 2013). Bilateral cooling of HVC in adult birds slowed down the song and stretched the durations of both syllables and gaps uniformly (Long and Fee, 2008); this effect was mostly attributable to stretching of the expiratory pulses and elongation of the inspiratory pressure pulses[2] (Andalman et al., 2011). Moreover, unilateral cooling of HVC stretched some parts of the song, while cooling the HVC on the other side stretched a different part of the song (Long and Fee, 2008). In contrast, cooling downstream area RA did not slow down the song (Long and Fee, 2008). These results suggest that dynamics within HVC are important for the control of temporal structure in the song.

A premotor role of HVC was also suggested by experiments using transient perturbation of neural activity during singing using electrical stimulation. Brief unilateral stimulation of HVC lead to several effects: truncation of an ongoing syllable (Ashmore et al., 2005; Wang et al., 2008), termination of a motif (Vu et al., 1994; Vu et al., 1998), and change in syllable sequence (Vu et al., 1994).

HVC activity during adult song

Electrophysiological recordings of the HVC activity during singing provide insight on how HVC contributes to song production. Single-unit recordings in adult birds showed that HVC_{RA} neurons exhibit high-frequency burst of spikes tightly locked to a particular point in the motif and are silent elsewhere. Different neurons are active at different points in the motif thus

[2] Compared to a uniform stretching of the expiratory pulses, the effect of HVC cooling on inspiratory pulses was more complex; while few inspiratory pulses stretched uniformly, most inspiratory pulses elongated due to a later truncation of an inspiratory event (Andalman et al., 2011).

forming a sequence (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007) (Figure 1.3a). HVC_X also exhibit sparse bursts activity locked to song, but the number of bursts per motif was slightly higher than that of HVC_{RA} neurons (Kozhevnikov and Fee, 2007; Vallentin and Long, 2015). Furthermore, intercellular recordings of HVC neurons during singing showed that these neurons depolarize rapidly before they burst consistent with the possibility that these sparse bursts are driven by a synaptically-connected neuronal chain (Long et al., 2010). These results suggest that HVC_{RA} neurons form a sequence that codes for the stereotyped song in adults (Figure 1.3b).

While HVC_X neurons are not necessary for singing in adults (Scharff et al., 2000), they might nonetheless carry important timing information to AFP. This timing information could be utilized during song learning to make changes in the song at specific times (Fiete et al., 2004; Charlesworth et al., 2011; Fee and Goldberg, 2011; Ravbar et al., 2012). In Bengalese finches, some HVC_X neurons are selectively active during a specific syllable transition or show activity that are correlated with the number of repetitions of a syllable (Fujimoto et al., 2011).

In contrast to the projection neurons that exhibit sparse bursts, putative HVC interneurons in adult birds show elevated activity during singing with stereotyped modulations in firing rate (Yu and Margoliash, 1996; Kozhevnikov and Fee, 2007; Amador et al., 2013).

Based on these results, previous studies have hypothesized that sparse bursts in HVC control the vocalizations on a moment-by-moment basis and that these bursts cover the entire motif (Fee et al., 2004; Leonardo and Fee, 2005). However, a recent work suggested an alternative possibility that HVC bursts might only occur at transitions between gestures, or gesture trajectory extrema (GTE), but not during gestures (Amador et al., 2013). Testing these hypotheses based on recordings from many HVC projection neurons is presented in Chapter 4.

Role of HVC in juvenile song

Less is known about the role of HVC in young birds, but lesions performed early in vocal development suggest its involvement in juvenile singing. HVC lesion during subsong had no effect on syllable duration demonstrating that HVC is not necessary for subsong production (Aronov et al., 2008; Aronov et al., 2011). Instead, production of subsong requires another premotor area LMAN (Aronov et al., 2008). HVC lesion in the protosyllable stage eliminated both protosyllables and protogaps (Aronov et al., 2011), and disrupted the coordination between breathing and singing (Veit et al., 2011). Consistent with these lesion results, local cooling of HVC stretched protosyllables and protogaps, but not the subsong syllables (Aronov et al., 2011).

Multi-unit recordings in the HVC of juvenile birds showed that HVC exhibits signing-related activity (Crandall et al., 2007). However, these recordings did not include single-unit recordings, thus making it difficult to determine the activity of individual projection neurons and interneurons. Therefore, even though the studies mentioned above all suggest the importance of HVC in juvenile song, our understanding of HVC is currently limited due to a lack of a comprehensive characterization of the single-unit activity of HVC neurons throughout song learning. This is the issue addressed in this thesis.

Computational models of HVC function

Models on the burst propagation in HVC of adult birds

Sequence of sparse bursts in HVC projection neurons during adult singing has inspired many computational models on how stereotyped motor behavior is generated in neural circuits (Drew and Abbott, 2003; Li and Greenside, 2006; Jin et al., 2007; Fiete and Seung, 2008; Gibb

et al., 2009; Long et al., 2010; Bertram et al., 2014). Most of these models are based on the idea that synchronous activity propagates along group of neurons that are connected to the next group in a feedforward fashion, often referred to as the 'synfire chain' (Abeles, 1991). Achieving a stable propagation in such a network is difficult (Diesmann et al., 1999), but HVC seems to be achieving stable propagation by using calcium bursts (Jin et al., 2007; Long et al., 2010).

Chains of unidirectional excitatory connections between HVC_{RA} neurons, an important assumption underlying these computational models, has not been characterized in detail. Paired recordings of HVC projection neurons in slice revealed excitatory connections between them, but they were rare (Mooney and Prather, 2005; Kosche et al., 2015). An alternative approach to test the excitatory connections between HVC_{RA} is to use 'connectomics,' or a dense-reconstruction neural circuits at the ultrastructure level (Seung, 2009; Lichtman and Denk, 2011; Denk et al., 2012). While the work described in this thesis does not directly measure the connectivity among HVC neurons, their activity patterns during juvenile song suggests a particular network architecture, as discussed in Chapter 5.

Models on the development of HVC network

While models above focused on how activity propagates in the network of HVC neurons in adults, they do not explain how the network in HVC is constructed and modified during song learning. To address this issue, several models have been put forward (Jun and Jin, 2007; Fiete et al., 2010). These models rely on spike-timing dependent plasticity (STDP), a synaptic plasticity rule that depends on the precise timing between presynaptic and postsynaptic firing (Markram et al., 1997; Bi and Poo, 1998; Dan and Poo, 2006). STDP is effective for developing a network that can generate sequential activity since the asymmetry inherent in the plasticity rule could

impose directionalities in the connection between neurons. To prevent STDP from driving the network toward synchronous activity ('temporal bunching'), or having a few neurons dominate the activity of the network ('spatial bunching'), these models have an additional plasticity rule that enforces competitions between multiple synapses of a single neuron such as axonal pruning (Jun and Jin, 2007) or hetereosynaptic long-term depression (Fiete et al., 2010).

Interestingly, these models make different predictions on how multiple sequences associated with multiple syllable types develop. The model of Jun and Jin predicts that sequence for each syllables are assembled one after another by a process in which additional neurons are recruited at the end of the chain (Jun and Jin, 2007). On the other hand, the model of Fiete et al. assumes that multiple sequences of different length are learned simultaneously (Fiete et al., 2010). These predictions are experimentally tested in our recording of HVC activity in juvenile birds (Chapters 3, 4). While neither model was able to capture our results, simple modifications to these models were sufficient to explain our observations (Chapter 5).

Organization of the thesis

This thesis addresses how singing-related activity in HVC changes during song learning in juvenile birds. The following chapters provide a comprehensive characterization of the activity of individual neurons in HVC at various stages of song learning.

Chapter 2 describes the activity pattern of HVC neurons during subsong and how this pattern changes during the transition from subsong to protosyllable stage. This chapter provides evidence that a growth of rhythmic neural sequence in HVC is the key neural mechanism underlying the emergence of protosyllables.

Chapter 3 examines how HVC activity changes during the emergence of two syllable types from a single protosyllable. Based on the observation of 'shared neurons,' this chapter provides evidence that splitting of neural sequence is important for the emergence of new syllable types.

Chapter 4 illustrates how neural sequence splitting introduced in Chapter 3 can be used in different learning strategies of syllable formation suggesting that there is a common neural mechanism underlying the diverse song learning strategies observed at the behavioral level.

Chapter 5 summarizes all the findings in Chapters 2-4 and proposes a working model on the developmental progression of HVC activity in juvenile birds. Numerical simulations, future experimental tests, and relation to other studies are discussed.

Appendix A describes the methods used for recording single-unit activity in HVC during singing, a key technique employed in Chapters 2-4.

Appendix B describes the detailed methods for the neural network simulation discussed in Chapters 5.

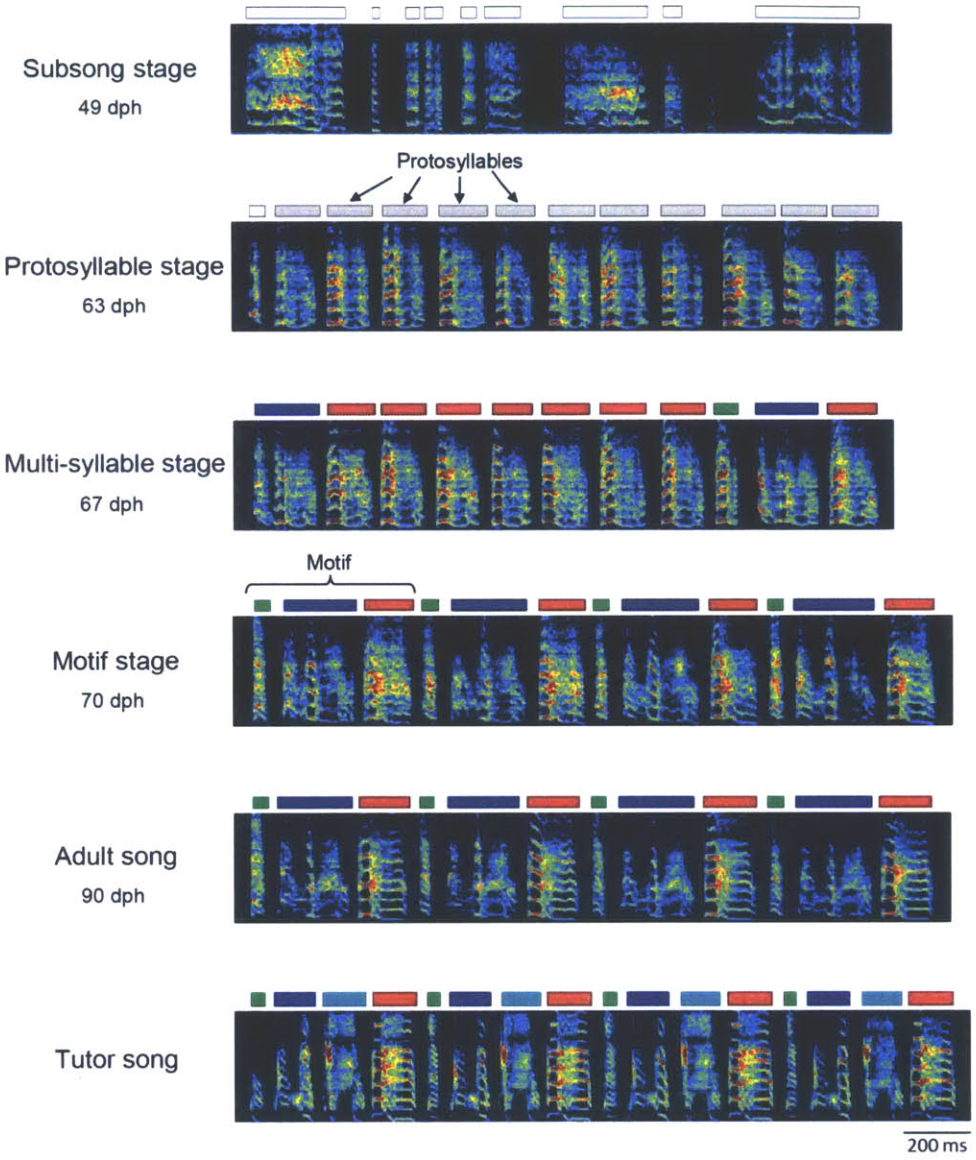

Subsong stage
49 dph
Protosyllables
Protosyllable stage
63 dph
Multi-syllable stage
67 dph
Motif
Motif stage
70 dph
Adult song
90 dph
Tutor song
200 ms

Figure 1.1

Figure 1.1: Song learning process

Example song recordings from a single male zebra finch tracked from subsong to adult stage. Song spectrograms and segmented syllables (top) are shown for each song stage. Subsong is characterized by a highly variable vocalization with syllables of different durations. Protosyllables, in contrast, have stereotyped durations. The multi-syllable stage is characterized by the appearance of multiple distinct syllable types shown in different colors. The motif stage is characterized by the appearance of a sequence of syllables that are sung in a relatively fixed order. Acoustic structures of the syllables are refined to give rise to the adult song. Note the similarity of the adult song to the tutor song, although the blue and cyan syllables in the tutor song was merged into a single syllable (blue) in the pupil song.

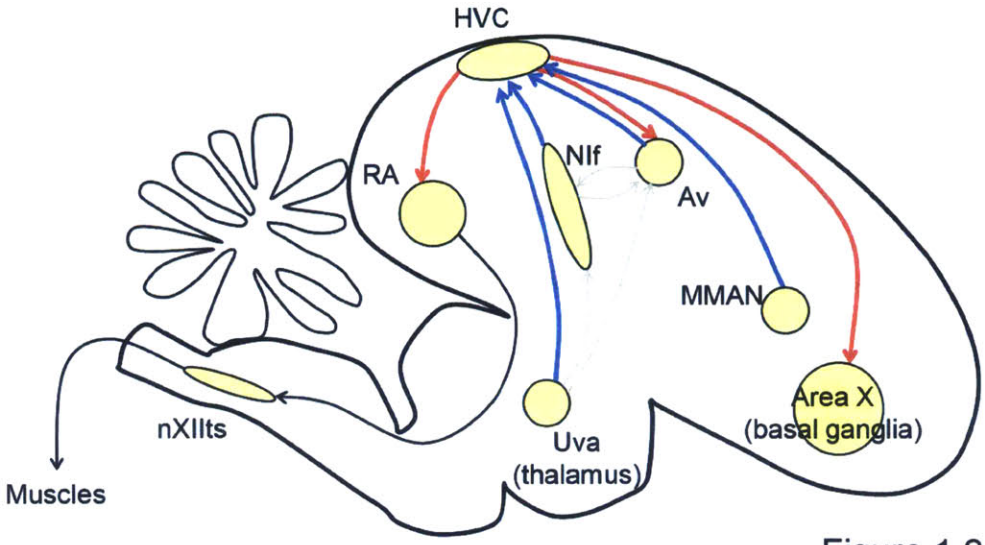

HVC
RA
NIf
Av
MMAN
Area X
(basal ganglia)
Uva
(thalamus)
nXIIts
Muscles

Figure 1.2

Figure 1.2: Anatomy of the song system with a special emphasis on HVC.

Schematic diagram of a sagittal section of a male zebra finch brain. Abbreviations are HVC (used as a proper name), RA (robust nucleus of the arcopallium), nXIIts (tracheosyringeal portion of the hypoglossal nucleus), Area X (used as a proper name), MMAN (medial magnocellular nucleus of the anterior nidopallium), NIf (nucleus interface), Uva (nucleus uvaeformis), Av (nucleus avalanche). Inputs to HVC are shown in blue, and outputs of HVC are shown in red. Based on Akutagawa & Konishi (*J Comp Neurol*, 2010).

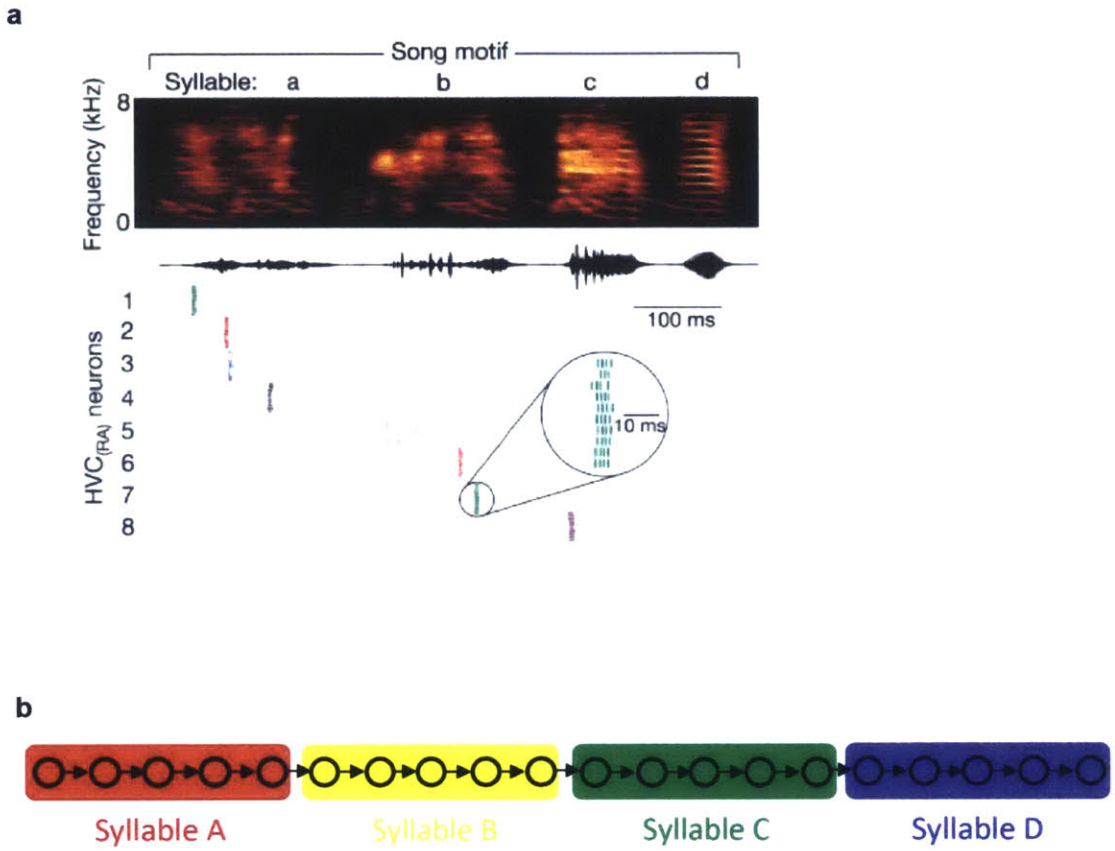

a
Song motif
Syllable: a b c d
Frequency (kHz)
8
0
100 ms
HVC(RA) neurons
1
2
3
4
5
6
7
8
10 ms
b
Syllable A
Syllable B
Syllable C
Syllable D

Figure 1.3

Figure 1.3: Activity pattern of HVC projection neuron in adult birds.

a, Activity of HVC_{RA} neurons in adult birds aligned to a motif. Top: Spectrogram of the four-syllable motif with microphone signal on the bottom. Bottom: Raster plot of 8 HVC_{RA} projection neurons sorted according to the latency of activation. Activity during 10 renditions of the same motif is shown for each neuron. Notice how each HVC_{RA} neuron exhibit a sparse burst during a particular time in the motif and different neurons are active at different times thus forming a sequence as a population (figure from Hahnloser et al., *Nature*, 2002). **b,** Schematic diagram of the hypothesized chain-like organization in HVC of adults. Each circle indicates a small population of HVC projection neurons, and arrows between them indicate functional connections between these neurons. Each syllable of adult song is represented by a distinct sequence in HVC.

Chapter 2

Growth of a rhythmic neural sequence underlies the transition from subsong to protosyllable stage

In the previous chapter, I have summarized the literature pointing to the importance of HVC in song learning. Despite the importance of HVC, single-unit activity in this area has not been previously characterized in juvenile birds. This chapter begins with an overview of the developmental change in the activity pattern of HVC neurons. This is followed by a detailed description of the changes that happen during the first major developmental milestone: emergence of protosyllables from subsong. This chapter ends by proposing that the growth of a rhythmic neural sequence in HVC is important for this transition.

Results

We recorded from populations of HVC projection neurons in birds at the subsong, plastic song, and adult song stages (n=1,149 neurons, 35 birds). Subsong was defined on the basis of a characteristic exponential distribution of song syllable durations (Aronov et al., 2011; Veit et al., 2011). Plastic song was further divided into an early protosyllable stage on the basis of a significant peak in the syllable duration distribution (Aronov et al., 2011; Veit et al., 2011), a later stage based on the acoustic differentiation of protosyllables into multiple syllable types (Tchernichovski et al., 2001; Liu et al., 2004), and a final 'motif' stage based on the appearance of a reliable sequence of syllables (Zann, 1996) (Figure 1.1, 2.2a; see Methods).

Syllable-locked activity in HVC

At all stages of vocal development, HVC projection neurons generated brief bursts of spikes during singing (Figure 2.1, 2.2b, c). However, we found that many of the

characteristics of HVC activity previously described in adult birds, such as precise temporal alignment to song, emerged gradually during vocal development. For example, in the subsong stage (n=12 birds), prior to the emergence of protosyllables, roughly half of neurons generated bursts that did not appear temporally locked to syllable onsets (Figure 2.3a, b, 2.5a), while the other half produced bursts that tended to occur at a particular latency relative to subsong syllable onsets (Figure 2.3c, d, 2.4a-d, 2.5a; 19/39 neurons exhibited syllable locking). However, in the protosyllable stage (n=13 birds), the majority of neurons generated bursts locked to protosyllables (Figure 2.1b, 2.4e, f, 2.5a; 104/135 neurons). This fraction continued to increase with development until, as previously described in adult birds (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Prather et al., 2008; Long et al., 2010; Fujimoto et al., 2011; Amador et al., 2013; Hamaguchi et al., 2014; Vallentin and Long, 2015), virtually every projection neuron generated bursts precisely locked to syllables (Figure 2.1c, 2.4g, h, 2.5a). This developmental increase in the syllable-locked activity was observed for both projection neuron types, HVC_{RA} and HVC_X (Figure 2.5b).

Development of rhythmic activity in HVC

Song development is characterized by a gradual change in song rhythm (Tchernichovski et al., 2001; Saar and Mitra, 2008; Glaze and Troyer, 2013). The change corresponds to an increase in the period between repetitions of the same sound, and is attributable to the addition of new song syllables (Tchernichovski et al., 2001). We wondered what changes in the activity of HVC neurons is associated with changes in song rhythm. In the subsong stage, few projection neurons exhibited rhythmic activity

(Figure 2.1a, 2.5a, 2.6b; 3/39 neurons). However, in the protosyllable stage, roughly half of the projection neurons generated rhythmic bursts (5-10 Hz), corresponding to fast rhythmic repetition of protosyllables (Tchernichovski et al., 2001; Liu et al., 2004; Aronov et al., 2011; Veit et al., 2011) (Figure 2.1b, 2.5a, 2.6c; 70/135 neurons; period 169 ± 6.4 ms, mean ± s.e.m.). Rhythmic bursts were typically locked to rhythmic protosyllables, but were also observed during less rhythmic portions of the song, particularly early in the protosyllable stage (Figure 2.11). On average, the period of the HVC burst rhythm gradually increased during the emergence of new syllable types and the formation of the song motif (Figure 2.5d). At this later stage, the majority of neurons were active once per motif, at a period corresponding to the motif repetition (Figure 2.1c, 2.5a, d, 2.6d; 297/377 neurons were rhythmic; period 456 ± 11.5 ms, mean ± s.e.m.). This developmental increase in the fraction of rhythmic neurons and the period was observed for both HVC_{RA} and HVC_X (Figure 2.5c, e).

Bout-related activity in HVC

A significant fraction of neurons (285/1,118 neurons) generated bursts related to song bouts—defined as epochs of continuous singing bounded by periods of silence (see Methods). These neurons generated brief bursts of spikes immediately prior to the onset of singing (137/285 neuron; Figure 2.7a-d, o) or immediately after the offset of singing (98/285 neuron; Figure 2.7e-h, p). Some of these neurons were active exclusively at bout onsets or at bout offsets (Figure 2.7a-h); other neurons were active at both (50/285 neurons; Figure 2.7i-k) and many neurons with bout-related activity also had syllable-locked activity (Figure 2.7*l*-n).

Activity of putative HVC interneurons during subsong

HVC contains both projection neurons and putative interneurons. What is the activity pattern of putative interneurons during subsong? To address this question, we recorded 15 putative interneurons (HVCi) during subsong. These neurons were evoked from antidromic stimulation in RA and/or Area X, but did not exhibit spike collision (Hahnloser et al., 2006). Outside singing, HVCi neurons exhibited tonic firing, and during subsong, the firing rate increased and the activity was modulated in a syllable-related manner (Figure 2.8a). To examine the relation between these modulation in neural activity and syllable onsets and offsets, histograms of neural activity aligned to syllable onsets or offsets were calculated (Figure 2.8b, c). Syllable-onset aligned histograms exhibited a prominent peak prior to subsong syllable onsets (Figure 2.8b), and syllable-offset aligned histograms showed a prominent dip prior to subsong syllable offsets (Figure 2.8c). This syllable-related activity was homogeneous across all HVCi neurons recorded during subsong. As a population, HVCi neurons exhibited a peak in activity 12.3 ± 5.7 ms (mean ± s.e.m.) prior to syllable onsets (Figure 2.8d, f), and a dip in activity 21.0 ± 1.6 ms (mean ± s.e.m.) prior to syllable offsets (Figure 2.8e, g). Notably, this syllable-related modulation occurred for syllables of all durations (Figure 2.8b, c).

Growth of neural sequences underlying rhythmic protosyllables

Rhythmic protosequences

Protosyllables are one of the earliest learned components of juvenile song (Tchernichovski et al., 2001; Liu et al., 2004; Aronov et al., 2011; Veit et al., 2011).

Lesions of HVC abolish the production of protosyllables (Aronov et al., 2011; Veit et al., 2011), suggesting that the observed rhythmic bursting of HVC neurons may play a role in the emergence of these early vocal elements. How is this rhythmic activity coordinated across the population of HVC projection neurons? Multiple recordings in the same bird revealed that different neurons were active at different times with respect to protosyllable onsets (Figure 2.9a, b). For individual birds in which more than 10 projection neurons were recorded in this stage (n=3 birds), the population of neurons provided a complete coverage of spiking activity throughout the duration of the protosyllable and the intervening gap (Figure 2.9d-f; >90% coverage; see Chapter 4). These findings suggest that protosyllables are generated by a rhythmic protosequence—a repeating motor program comprised of a continuous sequence of bursts in HVC.

Probabilistic neural activity during protosyllable stage

In contrast to the highly reliable bursting of HVC projection neurons in adult birds (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Long et al., 2010; Amador et al., 2013), we found that projection neurons in the protosyllable stage participated probabilistically in song syllables (Figure 2.9a, b). On average, projection neuron bursts occurred on 53% of protosyllables, although some neurons participated over 90% of the protosyllables (Figure 2.10a). Simultaneous paired recordings (Figure 2.9b) revealed that different projection neurons participated on relatively independent sets of protosyllables (median correlation r^2=0.072; median mutual information=0.056 bits; n=11 pairs; Figure 2.10b, c). These findings suggest that individual projection neurons participate probabilistically and largely independently in an ongoing rhythmic protosequence within

HVC. In addition, some neurons exhibited significant correlation between spike count and the acoustic features (Figure 2.10d-h), suggesting that variability in spiking activity of HVC projection neurons are weakly correlated with variability in the acoustic structure. Note that the degree of correlation between HVC activity and acoustic structure we observed is similar to that observed in RA (Sober et al., 2008).

Growth of a protosequence

We next examined the developmental emergence of this rhythmic protosequence. We focused on the distribution of burst latencies, which underwent striking changes through song development (Figure 2.4i-k). In the subsong stage, bursts have a significantly earlier distribution of latencies from syllable onset compared to the relatively broad coverage of burst latencies in the protosyllable stage (Figure 2.4i, j; P=0.02), with a larger fraction of bursts occurring before syllable onsets in the subsong stage (63%) compared to the protosyllable stage (43%). Even in the subsong stage, however, different neurons recorded in the same bird were precisely locked to syllable onsets at a range of latencies (Figure 2.4a-d). These findings suggest the existence, in the subsong stage, of transient sequential activity, initiated just prior to syllable onset but decaying within a few tens of milliseconds of syllable onset. This sequential activity appears to grow in the protosyllable stage to form longer sequences that can persist for more than a hundred milliseconds, throughout the duration of the protosyllable (Figure 2.4i-k, 2.9c-e).

Various forms of coupling between song and rhythmic bursts in HVC

Are rhythmic bursts in HVC always associated with rhythmic repetitions of protosyllables? While we found this to be true in many cases (Figure 2.1b, 2.9a, b, 2.11a, e), we have also observed other examples in which HVC bursts were not always associated with protosyllables. For example, there were cases where HVC exhibited continuous rhythmic bursting even when the song was not obviously rhythmic (Figure 2.11b-d). However, even in these cases, neurons were significantly locked to syllables (Figure 2.11f-h). In some cases, HVC bursts were tightly locked to a particular acoustic feature (Figure 2.11c), or a rhythmic modulation in the song amplitude within a long syllable (Figure 2.11d). These results demonstrate that rhythmic activity in HVC is coupled with rhythmic structure in the song, and that the exact nature of this coupling varies.

Discussion

Origin of syllable-locked activity in HVC during subsong

Lesion studies have shown that HVC is not necessary for subsong, but is necessary for protosyllables (Aronov et al., 2008; Aronov et al., 2011; Veit et al., 2011). Therefore, we initially hypothesized that HVC would not be active during subsong but becomes active during protosyllables. To our surprise, HVC exhibited singing-related activity even during subsong. Approximately half of the HVC projection neurons exhibited syllable-locking (Figure 2.5a) and all HVCi neurons exhibited syllable-locking activity (Figure 2.8f). Why does HVC exhibit syllable-locked activity without being necessary for subsong?

One could imagine two possibilities that explain this syllable locking. The first possibility is that, while HVC might not be necessary for subsong production (Aronov et al., 2008; Aronov et al., 2011), HVC could nonetheless be playing a premotor role during subsong. For example, bursts in HVC could be biasing the timing of subsong syllable onsets. This would explain the peak in activity prior to subsong syllable onsets.

The second possibility is that HVC receives an efference copy of subsong syllable onsets signals that originates in LMAN (Aronov et al., 2008). This efference copy signal could go through NIf or Uva, and activate HVC at subsong syllable onsets. Such an indirect pathway connecting LMAN and HVC has been characterized in adults (Hamaguchi and Mooney, 2012). In addition, Uva exhibits peak in activity prior to syllable onsets in adults (Danish, Aronov, and Fee, unpublished observations), so it is possible that Uva exhibits peak in activity prior to syllable onsets during subsong as well. The possibility that NIf or Uva carry efference copy signals can be tested by recording from these areas and lesioning HVC to eliminate its premotor role. If these areas indeed carry efference copy signals, syllable-related activity should remain even after HVC lesion.

Syllable-locked activity during subsong has also been observed in brain areas other than HVC and LMAN, such as Area X (Goldberg and Fee, 2012; Pidoux et al., 2015) and DLM (Goldberg and Fee, 2012). This raises the possibility that syllable-related activities in these downstream areas are being driven by syllable-related activity in HVC. Consistent with this hypothesis, recent study showed that syllable-related activity in Area X is abolished after HVC lesion (Pidoux et al., 2015).

Rhythmic sequence in HVC during protosyllables

During the protosyllable stage, HVC projection neurons exhibited rhythmic activity. While previous works have shown that HVC bursts can be rhythmic at the period of a motif (i.e. several hundred milliseconds) (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007), the period of the rhythmic bursts during protosyllables was much shorter (~100 ms; Figure 2.6c). Thus, over the course of development, there was an increase in the period of the rhythmic bursts (Figure 2.5d).

Moreover, as a population, different neurons had different latencies with respect to protosyllables, thus demonstrating sequential activity in HVC (Figure 2.9). These results demonstrate that HVC generates neural sequence even during the earliest stereotyped vocalizations of the juvenile bird. In the next two chapters, we will examine how birds use this protosequence to produce multiple syllable types.

We hypothesize that this rhythmic sequence in HVC shapes activity in downstream area RA (Olveczky et al., 2011) and imposes rhythmic structure on the subsong. This could be the neural mechanism that underlies the emergence of rhythmic and stereotyped vocalizations in juvenile birds. Depending on the relative influence of rhythmic activity in HVC and variable activity in LMAN, the coupling between rhythmic sequences in HVC and song output may vary (Figure 2.11). In some cases, rhythmic HVC activity was associated with the rhythmic repetition of protosyllables, consistent with HVC being dominant; in other cases, rhythmic HVC activity was associated with particular acoustic structure within the variable song, or a rhythmic modulation of the

song amplitude within a long syllable (Figure 2.11) consistent with a strong influence of LMAN on song.

The origin of rhythmic activity in HVC remains an open question. It is possible that external inputs to HVC are rhythmic, and that HVC forms sequences based on these rhythmic inputs. Such possibility will be explored in our computational model presented in Chapter 5. Alternatively, it is possible that the external inputs to HVC are not rhythmic, but HVC is able to assemble a sequence based on this non-rhythmic input. Once the protosequence is formed, it could repeat itself, thus generating rhythmic activity patterns. Computational models of chain growth in HVC have shown that HVC network could develop such a self-repeating loop during the self-organization process (Jun and Jin, 2007; Fiete et al., 2010).

Probabilistic activity during protosyllables

In contrast to the adult HVC activity, where bursts are reliably associated with syllables, we found that during the protosyllable stage, burst are probabilistic and could skip cycles (Figure 2.9a, b, 2.10a). This could either be due to factors intrinsic to HVC, such as weak connections between the neurons leading to probabilistic propagation of activity, or due to the influence of the other premotor area LMAN, which is mainly responsible for the variable vocalization in juveniles (Scharff and Nottebohm, 1991; Olveczky et al., 2005; Aronov et al., 2008; Andalman and Fee, 2009; Aronov et al., 2011). Indeed, previous studies have shown that activity in LMAN can influence HVC activity via a polysynaptic pathway (Hamaguchi and Mooney, 2012). The possibility that LMAN is introducing probabilistic activity in HVC can be experimentally tested by

recording HVC during protosyllables in combination with LMAN lesion (or inactivation). If LMAN is the source of the variability in HVC, LMAN lesion should eliminate the probabilistic firing of HVC neurons during protosyllables. A separate cortical area MMAN, which sends direct projections to HVC (Nottebohm et al., 1982; Bottjer et al., 1989; Foster and Bottjer, 2001), could also be involved in this process.

Bout-related activity in HVC

In addition to syllable-locked activity, we have observed bout-related activity in HVC (Figure 2.7). These neurons exhibited bursts prior to bout onsets or bursts after bout offsets. This type of activity related to singing has been observed in other brain areas. For example, a subset of LMAN neurons (Aronov et al., 2008) and putative MSNs in Area X (Goldberg and Fee, 2010) show activity related to bout onsets or offsets. Note, however, that bout-offset related activity in LMAN occurs *before* the end of the last syllable in the bout (Aronov et al., 2008). This is different from the bout-offset activity we observed in HVC, which occurs *after* the last syllable in the bout (Figure 2.7e-h, p). Given this difference, it remains to be seen whether bout-related activity in HVC is related to bout-related activity in LMAN and Area X.

Behavioral studies have shown that bout onsets and offsets, or bout edges, are important places in the song, and new syllables could initially appear preferentially at these bout edges (Lipkind and Tchernichovski, 2011; Lipkind et al., 2013). We propose a specific role for the bout-onset neurons with regards to bout-onset syllable differentiation in Chapter 4.

Bout-related activity has not been previously reported in HVC of adult birds. However, it is more difficult to clearly define the onsets of bouts in adults since adult song bouts are usually preceded by introductory notes. Note that there are neurons associated with introductory notes in HVC (Kozhevnikov and Fee, 2007; Rajan and Doupe, 2013). Whether bout-onset neuron we observed in juvenile birds are related to neurons that exhibit introductory note-related activity in adults is an open question.

Interestingly, neural activity related to initiation or termination of an action sequence has been observed in brain areas of other animals such as prefrontal cortex of primates during a sequential saccade task (Fujii and Graybiel, 2003), and basal ganglia of mice during a sequence of lever press (Jin and Costa, 2010). These results suggest an interesting possibility that bout-related signal we observed in HVC, which includes the corticostriatal neurons (HVC_X neurons), might be important for learning a motor sequence.

Methods

Animals

We used juvenile and young adult male zebra finches (*Taeniopygia guttata*) 44-112 days post hatch (dph) singing undirected song (n=32 birds). Animals were not divided into experimental groups; thus, randomization and blinding were not necessary. No statistical methods were used to predetermine sample size. Birds were obtained from the Massachusetts Institute of Technology zebra finch breeding facility (Cambridge, Massachusetts). The care and experimental manipulation of the animals were carried out in accordance with guidelines of the National Institutes of Health and were reviewed and approved by the Massachusetts Institute of Technology Committee on Animal Care.

All the juvenile birds were raised by their parents in individual breeding cages until 38 ± 5.2 dph (mean ± s.d.) when they were removed and were singly housed in custom-made sound isolation chambers (maintained on a 12:12 hour day-night schedule). In a subset of the birds (Bird 1, 2, 4, additional tutoring was carried out after removal from the breeding cages to facilitate song imitation. This was done by playback of the tutor song through a speaker (20 bouts per day). Additional tutoring was done for 12 days for Bird 1, 7 days for Bird 2, and 18 days for Bird 4. Bird identification key: Bird 1, to3965; Bird 2, to3779; Bird 3, to3017; Bird 4, to5640; Bird 5, to3396; Bird 6, to2309; Bird 7, to3412; Bird 8, to3567; Bird 9, to2462; Bird 10, to2331; Bird 11, to2427; Bird 12, to3352; Bird 13 to3563; Bird 14, to2819.

To compare the activity of HVC projection neurons in juvenile birds with that of adult birds, we also included neurons recorded in adults (>120 dph, n=3 birds) which

included a reanalysis of previously published HVC recordings performed in adult male zebra finches singing directed song (Kozhevnikov and Fee, 2007).

Song recordings

Songs were recorded with Sound Analysis Pro (Tchernichovski et al., 2000) or a custom-written MATLAB software (A. Andalman), which was configured to ensure triggering of recordings on all quiet vocalizations of juvenile birds (Aronov et al., 2008). The vertical axis range for all spectrograms is 500-8000 Hz.

Classification of song stages

We classified each day of juvenile singing into four song stages: subsong stage, protosyllable stage, multi-syllable stage, and motif stage (Figure 2.2a). Subsong stage (48 ± 4 dph, median ± inter-quartile range) is defined as having a syllable duration distribution well-fit by an exponential distribution (Aronov et al., 2011; Veit et al., 2011), with an upper limit for the Lilliefors goodness-of-fit statistic of 6. Following the subsong stage, birds enter the protosyllable stage (58 ± 10 dph, median ± i.q.r.) characterized by the presence of syllables with consistent timing reflected in a peak in the distribution of syllable durations (Tchernichovski et al., 2001; Liu et al., 2004; Aronov et al., 2011; Veit et al., 2011). The onset of the protosyllable stage was defined here as the first day in which the syllable duration distribution deviated from an exponential distribution (Lilliefors goodness-of-fit statistic greater than 6). Following the protosyllable stage, birds transition to multi-syllable stage (62 ± 12 dph, median ± i.q.r.) in which song contains multiple distinct syllable types (Tchernichovski et al., 2001; Liu et al., 2004).

These distinct syllable types were identifiable as multiple clusters in a scatter plot of syllable features (Tchernichovski et al., 2004) (e.g. Figure 3.2b). The motif stage (73 ± 21 dph, median ± i.q.r.) was defined as the song stage in which the bird produced a sequence of syllables in a fixed order (Zann, 1996). Finally, songs recorded in birds older than 120 dph were assigned as adult stage. A slightly older cutoff than the typical definition of adulthood in zebra finches (~90 dph) (Immelmann, 1969) was used, because some of our birds in the 90-120 dph continued to undergo some small developmental changes, as has been reported (Zann, 1996).

Syllable segmentation and bout extraction

Syllable segmentation of the juvenile song was done based on the song power in a spectral band between 1-4 kHz, as described previously (Aronov et al., 2008; Aronov et al., 2011; Veit et al., 2011). In a few cases, cutoff frequencies of the band-pass filters were adjusted to avoid the inclusion of high-frequency inspiratory sounds (Goller and Daley, 2001; Veit et al., 2011). Introductory notes were removed manually to avoid including HVC neurons that are rhythmically active during these notes (Rajan and Doupe, 2013). Song bouts were defined as a sequence of syllables separated by gaps less than 300 ms (Veit et al., 2011). Bout onset was defined as the onset of the first syllable in the bout, and bout offset was defined as the offset of the last syllable in the bout.

Syllable classification and labeling

Protosyllables were defined by their characteristic durations as has been described previously (Aronov et al., 2011; Veit et al., 2011). In short, to identify the protosyllables,

we first subtracted the best-fit exponential distribution (using 200-400 ms) from the syllable duration distribution, and fitted a Gaussian distribution to this residual; protosyllables were defined as syllables having durations within two standard deviations from the mean of this Gaussian distribution. We labeled protosyllables using the Greek letter 'α' in all our birds for consistency.

To label the emerging syllables in the juvenile song, we used the Greek letters β, γ, δ, and ε. In contrast, to label the syllables in the adult motif, we used the capital letters of the Latin alphabet A, B, C, etc. For birds in which the song learning trajectory was tracked developmentally, we labeled the syllables such that the correspondence between the juvenile syllables and adult syllables is straightforward: for example, α becomes A, β becomes B, γ becomes C, δ becomes D, and ε becomes E. Note that this labeling scheme leads to a slightly unconventional labeling of adult song in the sense that a motif can have letters in a reverse order (e.g. CBA in Figure 4.1a), or a motif might not have a syllable A (e.g. EDCB in Figure 4.5a).

Syllable labeling was done manually by visual inspection of the song spectrogram; this was done blind with respect to the neural activity. The existence of multiple distinct syllable types were confirmed by calculating the syllable duration and acoustic features commonly used to analyze birdsong syllables (Tchernichovski et al., 2000; Mandelblat-Cerf and Fee, 2014), and visualizing the clusters for each syllables in a two-dimensional space (Tchernichovski et al., 2004) (Figure 3.2b, Figure 4.7b, 4.8d). In some cases, syllable order was used as an additional indicator of syllable identity (e.g. Figure 4.5a, 70 dph; Figure 4.7a, 51 dph; Figure 4.8a, 59 dph).

In Bird 1, syllables β and γ were labeled manually by the visual inspection of the song spectrogram (Figure 3.2a). Since characterizing shared neurons and specific neurons depend on the reliable labeling of syllables, we took a conservative approach and only labeled syllables that were clearly identifiable and did not label the syllables that were ambiguous (fraction of syllables labeled as β or γ during 62-66 dph: 70 ± 5.5%, mean ± s.d.). We then estimated the error rate of our labeling procedure by plotting the labeled syllables (n=200 syllables per type on each day) in a two dimensional space of syllable duration and mean pitch goodness (Figure 3.2b), and obtained a decision boundary using linear discriminant analysis. We used mismatch between manual labeling and feature-based labeling to estimate the error rate for syllable β and γ. The error rate during the first five days of syllable differentiation (62-66 dph), when the labeling was most difficult, was only 1.1 % on average (range: 0.25-3.0%).

For the second round of differentiation in Bird 1, syllable order was used to assist in the labeling of syllables in early stages when syllables 'B' and 'D' were not easily distinguishable based on acoustic differences. Because these syllables underwent bout-onset differentiation, the first β after bout onset was labeled 'D'; later renditions of β in the bout were labeled 'B' (Figure 4.5a).

In Bird 2, several emerging syllables could be easily distinguished based on syllable durations (Figure 4.1d). Specifically, syllables whose durations were 110-160 ms, and 180-250 ms were defined as α and β, respectively. Syllables that were 10-75 ms in duration were labeled γ if they were followed by a β, and labeled ε otherwise.

Chronic neural recordings

Single-unit recordings of HVC projection neurons during singing were carried out using a motorized microdrive described previously (Fee and Leonardo, 2001; Okubo et al., 2014). Single-units were confirmed by the existence of the refractory period in the inter-spike interval (ISI) distribution (Figure 2.2b). Neurons that were active only during distance calls and not during singing (Kozhevnikov and Fee, 2007) were excluded from the analysis. In addition, neurons recorded for less than 5 seconds of singing were excluded since the short recording duration did not allow us to reliably quantify the activity pattern of these neurons.

Antidromic identification of HVC projection neurons was carried out with a bipolar stimulating electrode implanted in RA and Area X (single pulse of 200 μs every 1 second; current amplitude: 50-500 μA) (Hahnloser et al., 2002; Fee et al., 2004; Hahnloser et al., 2006; Kozhevnikov and Fee, 2007; Okubo et al., 2014). A subset of antidromically-identified projection neurons was further validated with collision testing (Hahnloser et al., 2002; Fee et al., 2004; Hahnloser et al., 2006; Kozhevnikov and Fee, 2007; Okubo et al., 2014). A subset of single units were identified as putative projection neurons based on sparse bursting, but could not be antidromically identified because they did not respond to antidromic stimulation or were lost before antidromic identification could be carried out (211/1,149 neurons). These neurons were included in the data set as unidentified HVC projection neurons (HVC_p).

Analysis of neural activity

Spikes were sorted offline using a custom MATLAB software (D. Aronov), and single-unit activities were represented as instantaneous firing rates, *R(t)*, defined at each time point as the inverse of the enclosing inter-spike interval as follows:

$$R(t) = \frac{1}{t_{i+1} - t_i}, \text{ for } t_i < t \leq t_{i+1}$$

where t_i is the time of the i-th spike.

Definition of bursts

HVC projection neurons exhibited bursts of action potentials during singing (Figure 2.1). Inter-spike interval (ISI) distribution of HVC projection neurons during singing exhibited two peaks with an inter-peak minimum near 30 ms; ISIs shorter than 30 ms corresponds to ISIs within bursts, and ISIs longer than 30 ms corresponds to ISIs between bursts (Figure 2.2b). Thus, we defined a 'burst' as a continuous group of spikes separated by intervals of 30 ms or less. Thus, by definition, bursts are separated from other spikes by intervals greater than 30 ms. Note that single spikes separated by more than 30 ms from both the preceding spike and the following spikes were also counted as a burst (Figure 2.6a). This definition of bursts using ISIs, in contrast to the definition of bursts based on setting a threshold on the instantaneous firing rate, allowed us to capture the rhythmic activity of HVC projection neurons that were composed of single spikes, as was frequently observed in juvenile birds. Burst time was defined as the center of mass of all the spikes within the burst. Burst width was defined as the interval between the first and the last spike in a burst (Figure 2.2c, top). Firing rate during burst was defined as a reciprocal of the mean inter-spike interval in a burst (Figure 2.2c, bottom). For the

calculation of burst width and firing rate during bursts, bursts composed of a single spike were excluded.

Syllable-related neural activity

To analyze the temporal relation between neural activity and song syllables, we aligned the spike times to syllable onsets and constructed a rate histogram (1 ms bins, smoothed over 20 bins). To test the significance of peak in the histogram, surrogate histograms were created by adding different amount of random time shifts to the spikes times (uniform distribution over the entire histogram which was between ±0.5 s from syllable onsets) on each trial (Goldberg and Fee, 2012). The peak of this surrogate histogram was recorded, and this shuffling procedure was repeated 1,000 times; P-values were obtained by analyzing the frequency with which the peaks of surrogate data were larger than that of the real data, and $P<0.05$ was considered significant (Goldberg and Fee, 2012). To visualize the population activity associated with protosyllables, we constructed a population raster plot by choosing 20 most active protosyllable renditions for each neuron, and by plotting different neurons in different colors (Figure 2.9d-f). For all the other population raster plots associated with identified syllables, 20 random renditions were chosen for display.

Bout-related neural activity

A subset of HVC projection neurons exhibited bout-related activity: bursting before bout onsets and/or after bout offsets (Figure 2.7). To quantify the pre-bout activity, we generated histograms aligned to bout onsets (Figure 2.7b, d, j) and found a peak in the

histogram in a 300 ms window prior to bout onset. We considered a neuron to be exhibiting 'pre-bout activity' if the size of this peak was significant ($P<0.05$) compared to peaks obtained from the surrogate histograms (identical to the procedure described above in *Syllable-related neural activity*). To eliminate the possibility of including syllable-related activity as bout-related activity, we did not consider a neuron to be exhibiting pre-bout activity if the neuron showed a peak in the bout-onset aligned histogram and a peak at a similar latency (less than 25 ms apart) for the syllable-onset aligned histogram. We considered a neuron to be exhibiting 'post-bout activity' if there was a significant peak in the bout-offset aligned histogram (Figure 2.7f, h, k, n) in a 300 ms window after bout-offset.

Quantification of the rhythmic neural activity

To quantify the rhythmic neural activity of HVC projection neurons, we used four different methods: inter-bursts interval, spike-train autocorrelation, spectrum of the spike train, and cepstrum of the spike train (Figure 2.6b-d). Only spikes that were produced during singing (i.e. between the onset of the first syllable and the offset of the last syllable in the bout) were used for the calculation of these measures. (1) *Inter-burst interval.* Intervals between burst times were calculated and the peak between 80-1000 ms was found. (2) *Spike-train autocorrelation*. To quantify the second-order statistics of the firing pattern of HVC neurons, spike-train autocorrelation, expressed as conditional firing rate (Rieke, 1997), was calculated (Figure 2.6b-d), and the peak between 80-1000 ms was found. The width of the center peak indicates the width of bursts, and multiple side lobes with regular intervals indicate rhythmic bursting (Figure 2.6c, d). (3) *Spectrum of the*

spike train. Rhythmicity of the single-unit activity was also quantified in the frequency domain using the multi-taper spectral analysis of spike trains treated as point processes (Jarvis and Mitra, 2001). We used the Chronux software to calculate spectrum for the spike trains (Mitra and Bokil, 2008; Bokil et al., 2010). First, bouts of singing were segmented into analysis windows of 1.5 second long, and then spectrum for each window was calculated using the multi-taper spectral analysis with time-bandwidth product NW = 3/2 and the number of tapers K=2. To obtain the mean spectrum for a given neuron, spectra calculated from all the analysis windows were averaged. Finally, we found the peak in the mean spectrum within the range 2-15 Hz. (4) *Cepstrum of the spike train*. HVC projection neurons often exhibited rhythmic bursts with precise inter-burst intervals (Figure 2.1b, c, Figure 2.6c, d). Thus, the spectrum of the spike train tended to have multiple peaks at the multiples of the fundamental frequency (Figure 2.6c, d). To represent these burst trains that have regular intervals in a more compact way, we calculated the cepstrum (a technique commonly used in speech processing to extract the period of glottal pulses) of the spike train defined as the inverse Fourier transform of the log spectrum (Oppenheim and Schafer, 2004), and found the peak in the cepstrum between 80-1000 ms.

To assess the significance of the peaks in these four measures, we compared the distribution of peak amplitude obtained from the real data with that of the surrogate data obtained by shuffling the bursts times. For this shuffling procedure, we first identified all the bursts during a bout of singing as described above (Figure 2.6a). We then randomly placed bursts sequentially in an interval that has the same duration as the song bout; when spikes from two bursts were closer than 30 ms, we repeated the random placement until

they were further apart by more than 30 ms (Figure 2.6a). Note that this randomization procedure only shuffles the burst times and preserves both the number of bursts and the ISIs within bursts. Then, all four metrics listed above were calculated by applying the same method to this surrogate spike trains. This shuffling was repeated (1,000 times for the IBI and auto-correlation, 100 times for the spectrum and cepstrum) and the P-values of the peak were calculated by analyzing the frequency in which the peaks from the surrogate spike trains were larger than the peak obtained from real data. A neuron was considered to be exhibiting 'rhythmic' bursting, if it had significant peaks in at least two of the four metrics. The period of the rhythm was defined as the location of the largest peak of spike-train autocorrelation between 80-1000 ms.

Quantification of the probabilistic neural activity during the protosyllable stage (Figure 2.10a)

Although many HVC projection neurons recorded in the juvenile bird exhibited rhythmic bursts, these bursts did not occur reliably on every cycle of the rhythm, but instead participated probabilistically (Figure 2.9a, b). To quantify the degree of participation, we first extracted the protosyllables based on syllable duration (see *Syllable classification and labeling* above) and examined the fraction of protosyllables in which at least one spike occurred (time-window between 30 ms prior to protosyllable onset to 10 ms after protosyllable offset). The fraction of protosyllables in which the neuron the neuron was active was obtained for all the HVC projection neurons recorded during the protosyllable stage that showed a significant rhythmic bursting (Figure 2.10a).

Analysis of simultaneously recorded pairs of neurons (Figure 2.10b, c)

To test whether probabilistic bursting of neurons in the protosyllable stage is coordinated across many neurons, we analyzed the correlation between pairs of simultaneously recorded neurons (Figure 2.10b). This analysis was restricted to pairs of neurons that were rhythmically bursting (n=11 pairs, 3 birds). Bursting activity of each neuron was converted to a binary string corresponding to its participation in each protosyllable (for the definition of protosyllables, see *Syllable classification and labeling* above). The activity of a neuron was assigned a '1' for a protosyllable if the neuron exhibited activity in a time-window between 30 ms prior to protosyllable onset to 10 ms after protosyllable offset. Only activity during protosyllables was analyzed to avoid including the highly variable subsong syllables which are likely generated outside HVC (Aronov et al., 2008; Aronov et al., 2011). For simultaneously recorded pairs of neurons, this procedure resulted in two binary strings corresponding to the protosyllable-related activity of each neuron. We then calculated the coefficient of determination r^2 by taking the square of the Pearson's correlation coefficient r between the two binary strings calculated for each neuron in the pair. The distribution of coefficient of determination is shown in Figure 2.10b.

We also carried out a mutual information analysis to quantify whether the activity of one neuron was predictive of the set of protosyllables for which the other neuron was active. Using the same binary representation described above, we calculated the joint probability distribution describing the four possible states of activity (neither neuron spikes, neuron A spikes, neuron B spikes, both neurons spike). The mutual information

was computed from this joint distribution, and the result in bits for each neuron pair is plotted in Figure 2.10c.

Statistics

Results are expressed as the mean ± s.d. or s.e.m. as indicated. For χ^2 tests, if the contingency table included a cell that has an expected frequency less than 5, Fisher's exact test was used (Kanji, 2006). All tests were two-sided, and $P<0.05$ was considered significant. Bonferroni correction was used to account for multiple comparisons.

Figure 2.2(a) To quantify the relation between song stage and age, we calculated Spearman's rank correlation coefficient ρ and the P-value under the null hypothesis that $\rho=0$. *(c)* We computed the statistical significance of developmental changes in burst width (top) and firing rate during bursts (bottom) by using Kruskal-Wallis test followed by a post-hoc pairwise test to compare each stage with the adult stage.

Figure 2.4(i-k) Wilcoxon rank-sum test was used to test whether the median of the syllable-onset aligned latency distribution was different between subsong and protosyllable stages.

Figure 2.5(a) The statistical significance of developmental changes in the fraction of HVC neurons that were syllable-aligned was assessed in two different ways: 1) Each stage was compared with the adult stage using the χ^2 test followed by a post-hoc pairwise test. 2) To quantify the developmental trend in the fraction of syllable-locked neurons, we calculated Pearson's correlation coefficient r between the binary value for each neuron (0, unlocked; 1, locked) and song stage (subsong: 1, protosyllable: 2, multi-syllables: 3, motif: 4, adult: 5). The P-value was calculated under the null hypothesis that $r=0$. The

significance of the developmental trend for rhythmic bursting was calculated similarly. Similar results were obtained for correlation between these metrics and the age at which each neuron was recorded, rather than song stage.

Figure 2.5*(d)* Statistical significance of developmental changes in the period of the HVC rhythm was also assessed in two different ways: 1) Each song stage was compared with the adult stage using the Kruskal-Wallis test followed by a post-hoc pairwise test. 2) To quantify the developmental trend in the period of the HVC rhythm, we calculated Pearson's correlation coefficient r between burst period and song stage. Similar results were obtained for correlation between burst period and the age at which each neuron was recorded.

Figure 2.5(b, c, e) To test whether fraction of syllable-locked neurons *(panel b)*, fraction of rhythmic neurons *(panel c)*, and period of HVC rhythm *(panel e)* significantly differed between HVC_{RA} and HVC_X, we used χ^2 test for all the pairwise comparisons with Bonferroni correction for multiple comparisons.

Data sharing

Activity of HVC projection neurons in juvenile birds presented in this thesis will be posted to Collaborative Research in Computational Neuroscience (CRCNS).
https://crcns.org/

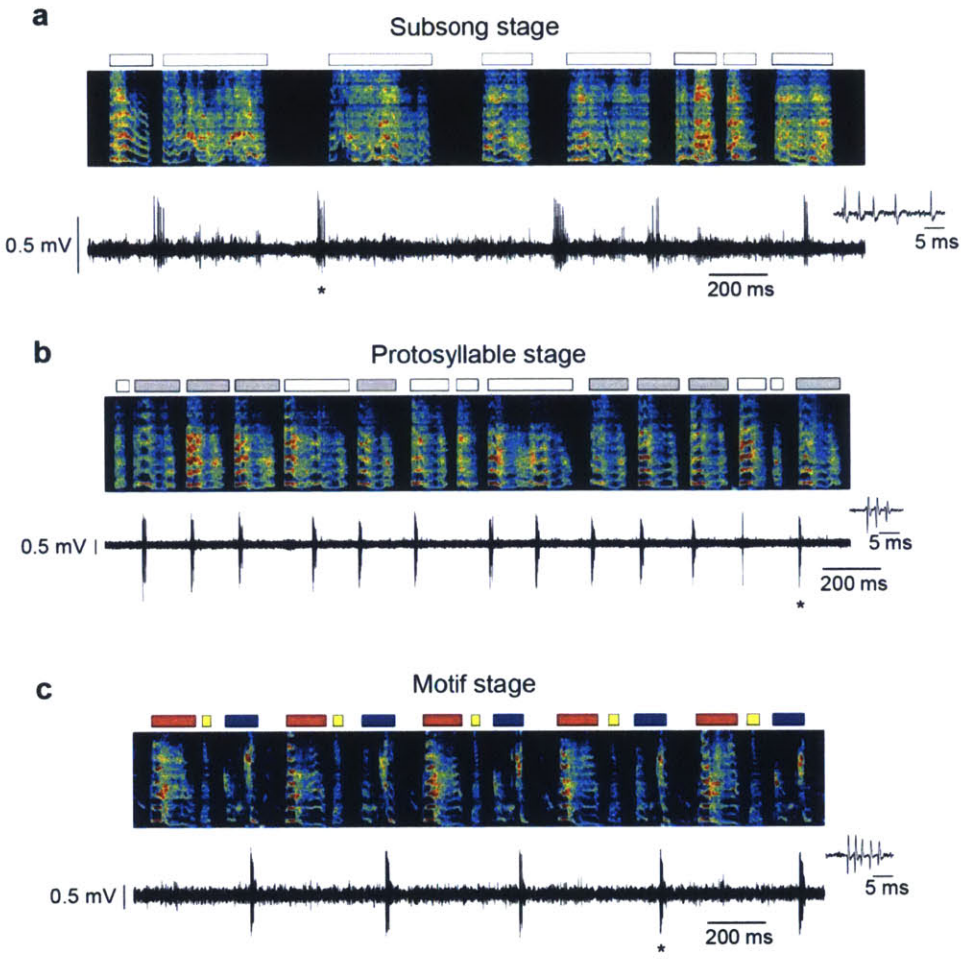

a
Subsong stage
0.5 mV
5 ms
200 ms
b
Protosyllable stage
0.5 mV
5 ms
200 ms
c
Motif stage
0.5 mV
5 ms
200 ms

Figure 2.1

Figure 2.1: Examples of HVC projection neuron activity in juvenile birds at different song stages.

a, Example neuron recorded in the subsong stage, prior to the formation of protosyllables (HVC_{RA}; 51 dph; Bird 7). Top: song spectrogram with syllables indicated above. Bottom: extracellular voltage trace. **b,** Example neuron recorded in the protosyllable stage (HVC_{RA}; 62 dph; Bird 2) showing rhythmic bursting locked to syllables. Protosyllables are indicated (gray boxes). **c,** Example neuron recorded after the formation of a motif (HVC_{RA}; 68 dph; Bird 6) showing bursts once per song motif. Different syllable types are indicated by boxes of different colors. Inset shows zoom of bursts indicated by asterisk. Spectrogram vertical axis range 500-8000 Hz.

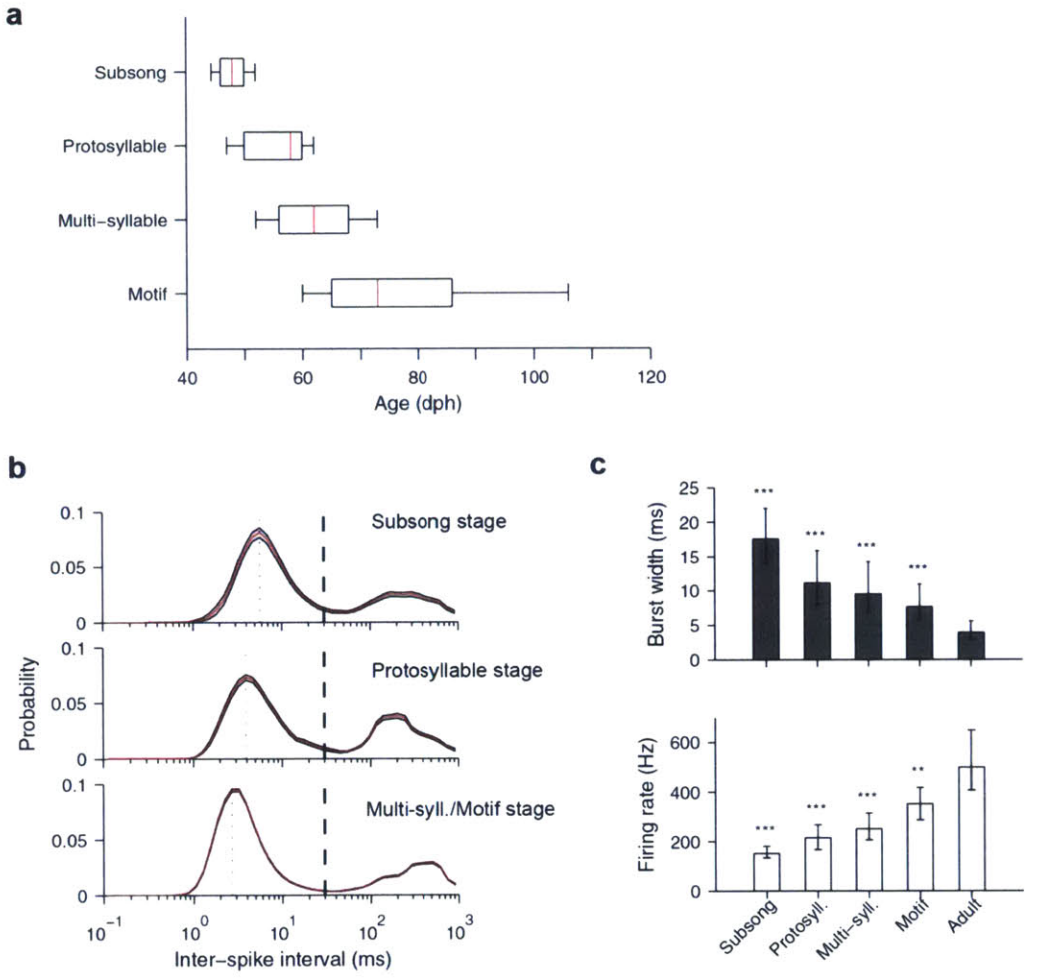

a
Subsong
Protosyllable
Multi-syllable
Motif
40
60
80
100
120
Age (dph)
b
Subsong stage
Protosyllable stage
Multi-syll./Motif stage
Probability
0.1
0.05
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Inter-spike interval (ms)
c
Burst width (ms)
25
20
15
10
5
0
Firing rate (Hz)
600
400
200
0
***
**
Subsong
Protosyll.
Multi-syll.
Motif
Adult

Figure 2.2

Figure 2.2: Quantification of HVC bursting at different song stages.

a, Age range of birds at which songs were classified into different developmental stages (Spearman's rank correlation between age and stage=0.61; red line indicates the median, box indicates the 25-75 percentile, and whiskers indicate 10-90 percentile; n=39, 135, 565, 377 neurons at each stage). **b,** Interspike-interval (ISI) distributions (mean ± s.e.m.) of HVC projection neurons that exhibited spiking during singing, at three stages of vocal development (n=38, 130, 921 neurons). ISI distributions computed with logarithmic binning show a bimodal structure; the peak around 3-5 ms indicates inter-spike intervals within bursts, and a broader peak around 100-400 ms indicates intervals between bursts. Dashed line indicates the threshold used for defining a burst (ISI=30 ms), and dotted line indicates the peak. Note the refractory period below 1 ms. **c,** Median burst width (top) and median firing rate during bursts (bottom) as a function of developmental stage (error bars indicate median ± quartiles; n=39, 135, 565, 377, 32 neurons; **P<0.01, ***P<0.001 post-hoc comparison with adult stage).

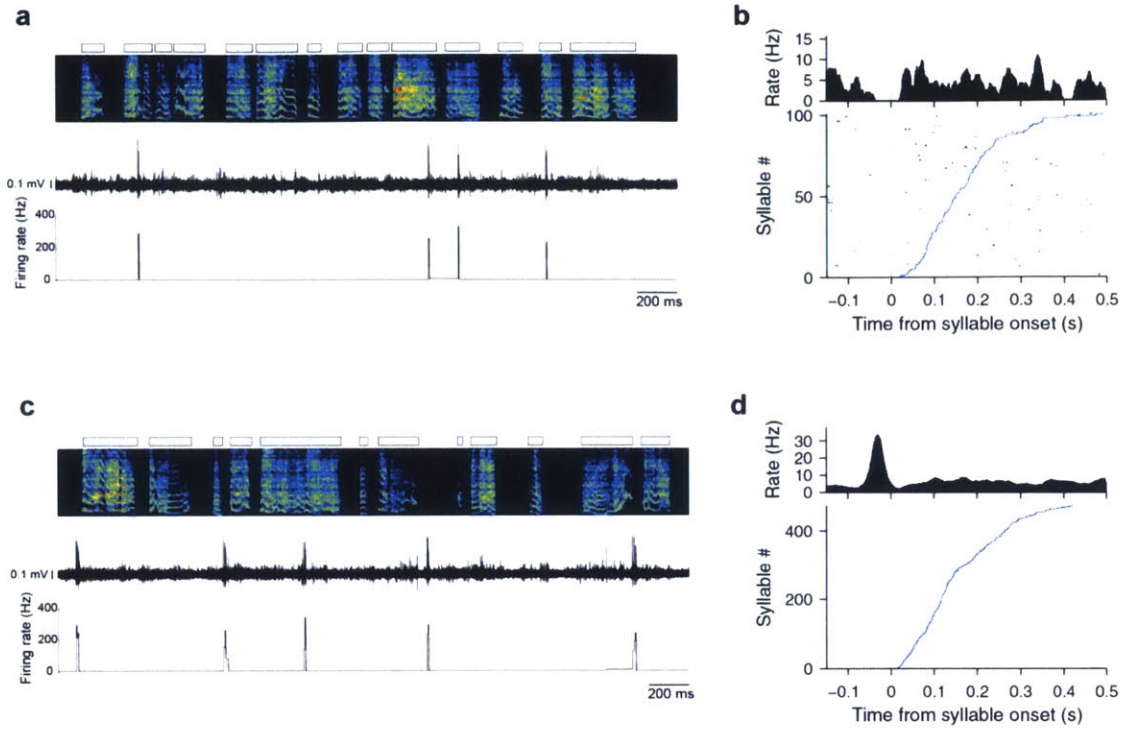

a
0.1 mV
Firing rate (Hz)
400
200
0
200 ms
b
Rate (Hz)
15
10
5
0
Syllable #
100
50
0
−0.1
0
0.1
0.2
0.3
0.4
0.5
Time from syllable onset (s)
c
0.1 mV
Firing rate (Hz)
400
200
0
200 ms
d
Rate (Hz)
30
20
10
0
Syllable #
400
200
0
−0.1
0
0.1
0.2
0.3
0.4
0.5
Time from syllable onset (s)

Figure 2.3

Figure 2.3: Activity of HVC projection neurons during subsong.

a, Neuron that exhibits bursts that are not aligned to subsong syllables (HVC_{RA}, Bird 7, 50 dph). **b,** Syllable-onset aligned raster and histogram for the neuron shown in panel a. Note the lack of clear peaks in the histogram. **c,** Neuron that preferentially bursts prior to subsong syllable onsets (HVC_{RA}, Bird 7, 51 dph). **d,** Syllable-onset aligned raster and histogram for the neuron shown in panel c. Note the clear peak prior to syllable onsets for syllables of all duration.

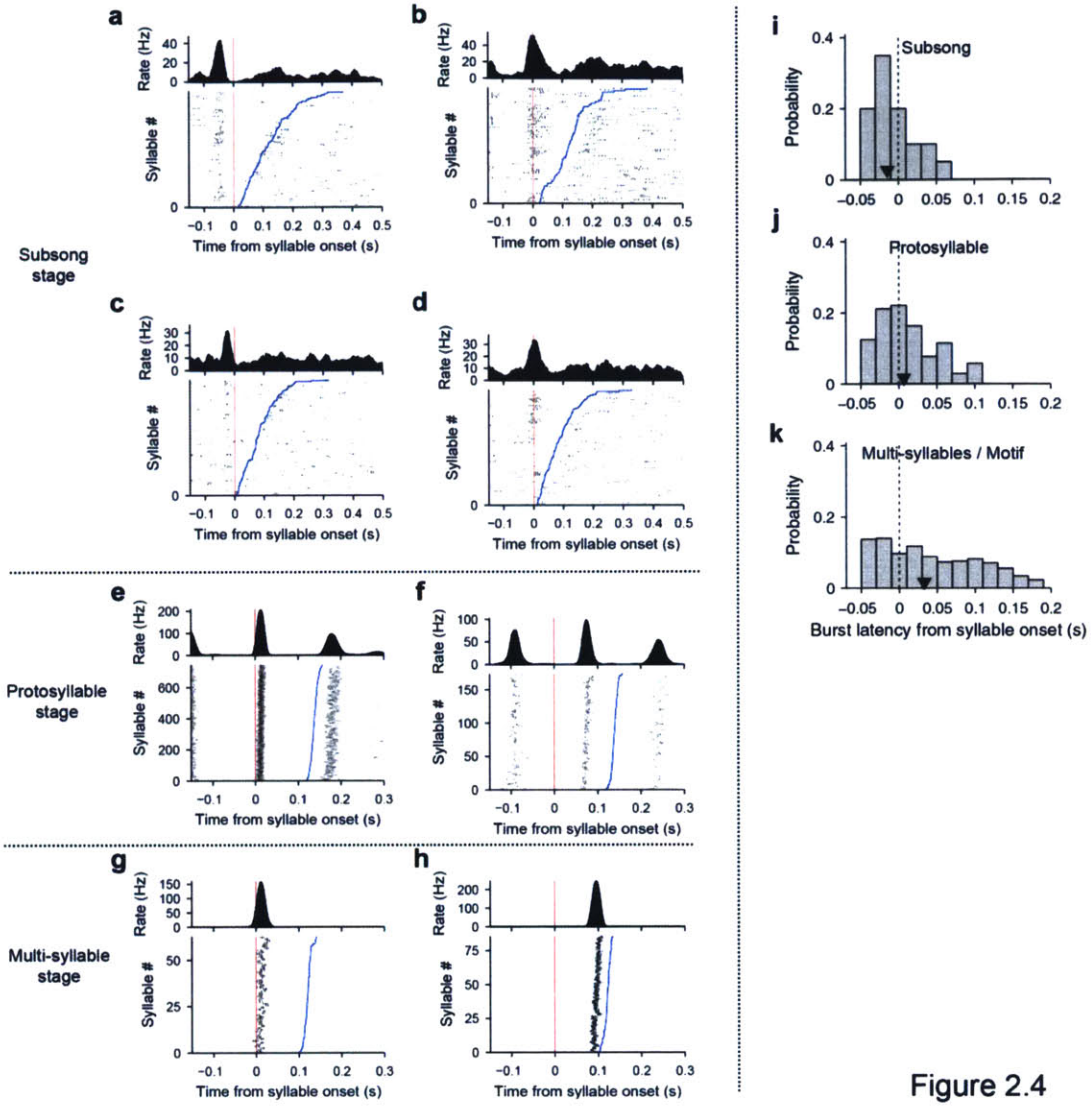

a
b
c
d
e
f
g
h
i
j
k
Rate (Hz)
Syllable #
Time from syllable onset (s)
Subsong stage
Protosyllable stage
Multi-syllable stage
Probability
Subsong
Protosyllable
Multi-syllables / Motif
Burst latency from syllable onset (s)

Figure 2.4

Figure 2.4: Latency distribution of syllable-locked bursts.

a-b, Two projection neurons recorded in a subsong bird (both HVC_X; 47 and 48 dph, respectively; Bird 9) Note different latencies of bursting. **c-d,** Two projection neurons recorded in a different subsong bird (both HVC_X; 47 and 44 dph, respectively; Bird 10). **e-f,** Syllable-onset-aligned raster plot and histogram of spiking activity of two neurons recorded during the protosyllable stage (Bird 2) that showed strong locking to protosyllable: **e,** for the same neuron as in Figure 2.1b (HVC_{RA}; 62 dph); **f,** for another neuron (HVC_{RA} ; 65 dph) locked at a time point near the middle of the protosyllable. **g-h,** Same as above, but for two neurons recorded in the motif stage (Bird 6): **g,** HVC_X neuron; 61 dph); **h,** Same neuron as in Figure 2.1c (HVC_{RA} ; 68 dph) showing strong locking to a time point late in the song syllable. **i-k,** Distribution of burst latencies relative to syllable onset at three different stages of vocal development: subsong (i), protosyllable (j), and multi-syllable/motif stages (k). Black triangles indicate the median burst times. n=19, 104, 813 neurons, respectively.

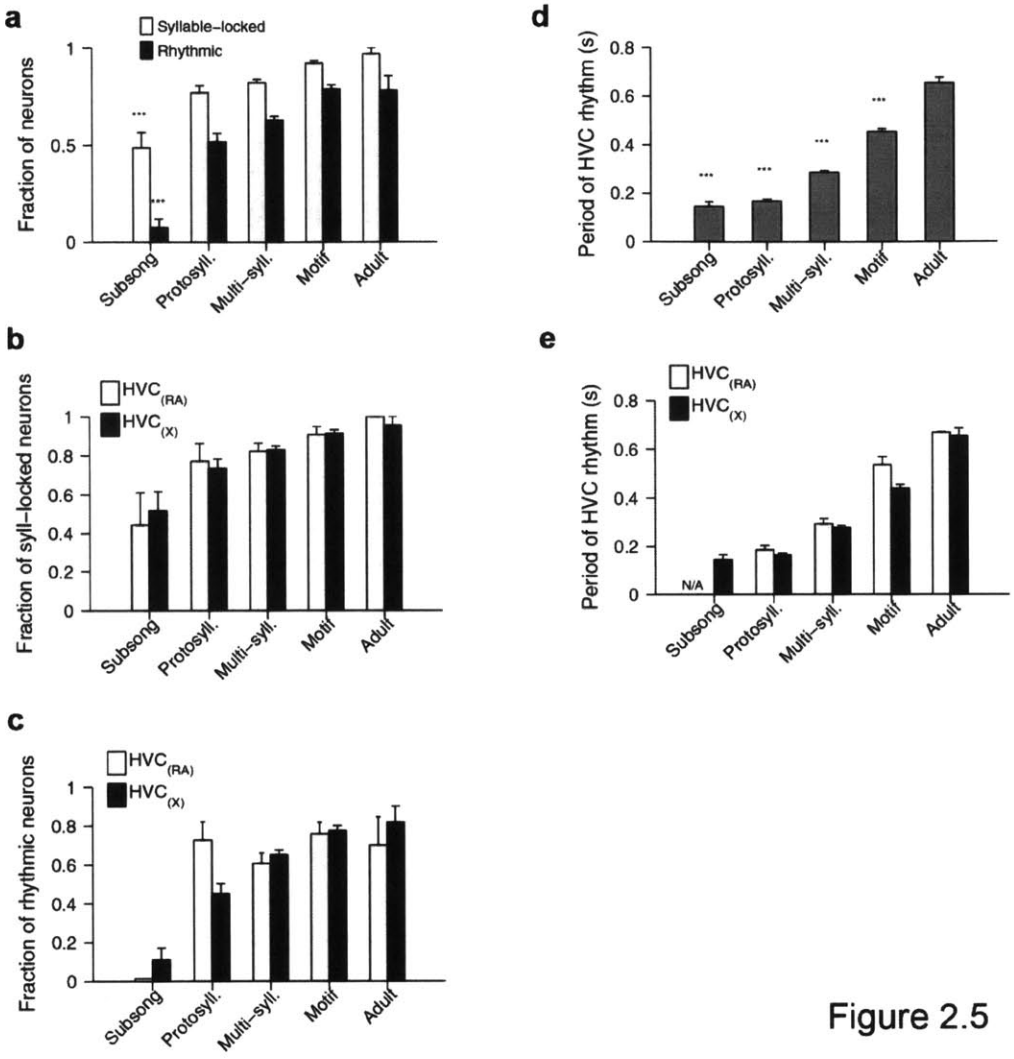

a
Syllable-locked
Rhythmic
Fraction of neurons
1
0.5
0
***
***
Subsong
Protosyll.
Multi-syll.
Motif
Adult
d
Period of HVC rhythm (s)
0.8
0.6
0.4
0.2
0
***
***
***
***
Subsong
Protosyll.
Multi-syll.
Motif
Adult
b
HVC(RA)
HVC(X)
Fraction of syll-locked neurons
1
0.8
0.6
0.4
0.2
0
Subsong
Protosyll.
Multi-syll.
Motif
Adult
e
HVC(RA)
HVC(X)
Period of HVC rhythm (s)
0.8
0.6
0.4
0.2
0
N/A
Subsong
Protosyll.
Multi-syll.
Motif
Adult
c
HVC(RA)
HVC(X)
Fraction of rhythmic neurons
1
0.8
0.6
0.4
0.2
0
Subsong
Protosyll.
Multi-syll.
Motif
Adult

Figure 2.5

Figure 2.5: Developmental progression of syllable-locking and rhythmic activity in HVC projection neurons in juvenile birds.

a, Fraction of neurons temporally locked to syllables (light gray), and fraction of neurons that exhibit rhythmic bursts (black) at each stage of song learning. Error bars indicate mean ± s.e.m (n=39, 135, 565, 377, 32 neurons, respectively). **b,** Fraction of neurons temporally locked to syllables analyzed separately for HVC_{RA} and HVC_X neurons. Error bars indicate mean ± s.e.m. (HVC_{RA}: 9, 22, 83, 53, 10 neurons analyzed at each stage, respectively; HVC_X: 27, 91, 376, 244, 22 neurons analyzed at each stage, respectively). **c**, Fraction of neurons that exhibited rhythmic bursts analyzed separately for HVC_{RA} and HVC_X neurons (HVC_{RA}: 9, 22, 84, 53, 10 neurons, respectively; HVC_X: 27, 91, 376, 244, 22 neurons, respectively). **d,** Mean period of the HVC rhythmicity as a function of song stage. Error bars indicate mean ± s.e.m. (n=3, 70, 356, 297, 25 neurons, respectively). ***P<0.001, post-hoc comparison with the adult stage. **e,** Mean period of HVC rhythmicity as a function of song stage analyzed separately for HVC_{RA} and HVC_X neurons (HVC_{RA}: 0, 16, 50, 40, 7 neurons, respectively; HVC_X: 3, 41, 245, 189, 18 neurons, respectively). Of the 14 comparisons between HVC_{RA} and HVC_X shown in e-g, only the period of HVC rhythm (panel g) during the motif stage showed significantly difference between the cell types (P<0.05 with Bonferroni correction for multiple comparisons).

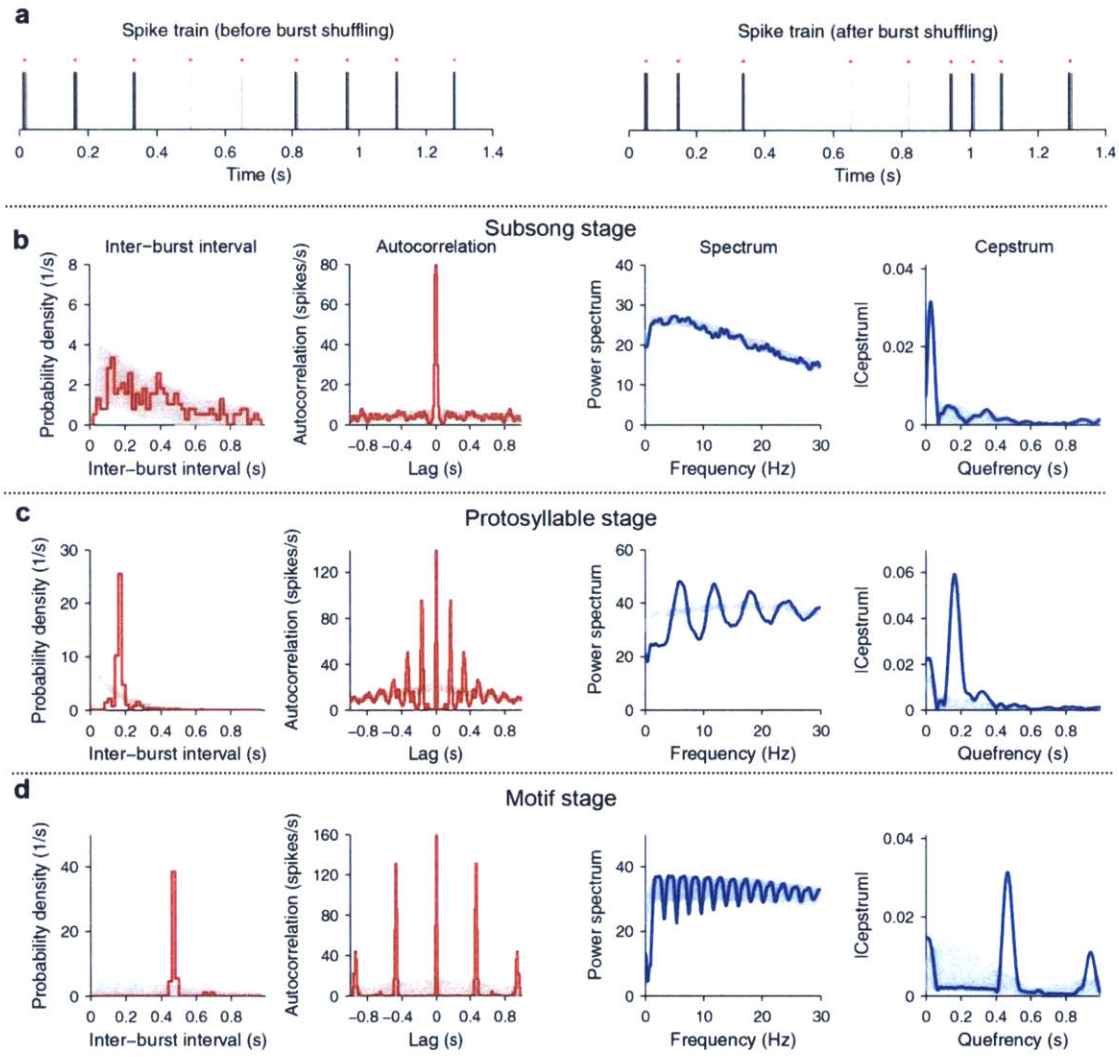

a
Spike train (before burst shuffling)
Spike train (after burst shuffling)
Time (s)
b
Subsong stage
Inter-burst interval
Autocorrelation
Spectrum
Cepstrum
Probability density (1/s)
Autocorrelation (spikes/s)
Power spectrum
|Cepstrum|
Inter-burst interval (s)
Lag (s)
Frequency (Hz)
Quefrency (s)
c
Protosyllable stage
d
Motif stage

Figure 2.6

Figure 2.6: Quantification of HVC projection neuron rhythmicity.

a, Illustration of burst shuffling procedure used to generate surrogate spike trains for assessing the significance of the rhythmicity. A spike train during a single bout of singing before (left) and after (right) burst shuffling. Detected bursts are indicated by red asterisks (see Methods for the definition of bursts). This procedure only shuffles the inter-burst interval and preserves both the inter-spike intervals within bursts and the number of bursts. **b-d,** Four different measures for quantifying rhythmicity: Inter-burst interval distribution, spike-train autocorrelation expressed as a conditional rate (Rieke, 1997), spectrum of the spike train, absolute valuc of the cepstrum of the spike train (see Methods). Gray shading indicates the 2.5-97.5% range for the surrogate data set with burst shuffling. **b,** Analysis of the neuron shown in Figure 2.1a recorded during subsong. **c,** Analysis of the neuron shown in Figure 2.1b recorded during the protosyllable stage. Note that the multiple periodic peaks in the spectrum are captured as a single peak in the cepstrum. **d,** Analysis of the neuron shown in Figure 2.1c recorded during the motif stage. The slow rhythm (~2 Hz) corresponds to the rhythmic repetition of a motif during song bouts.

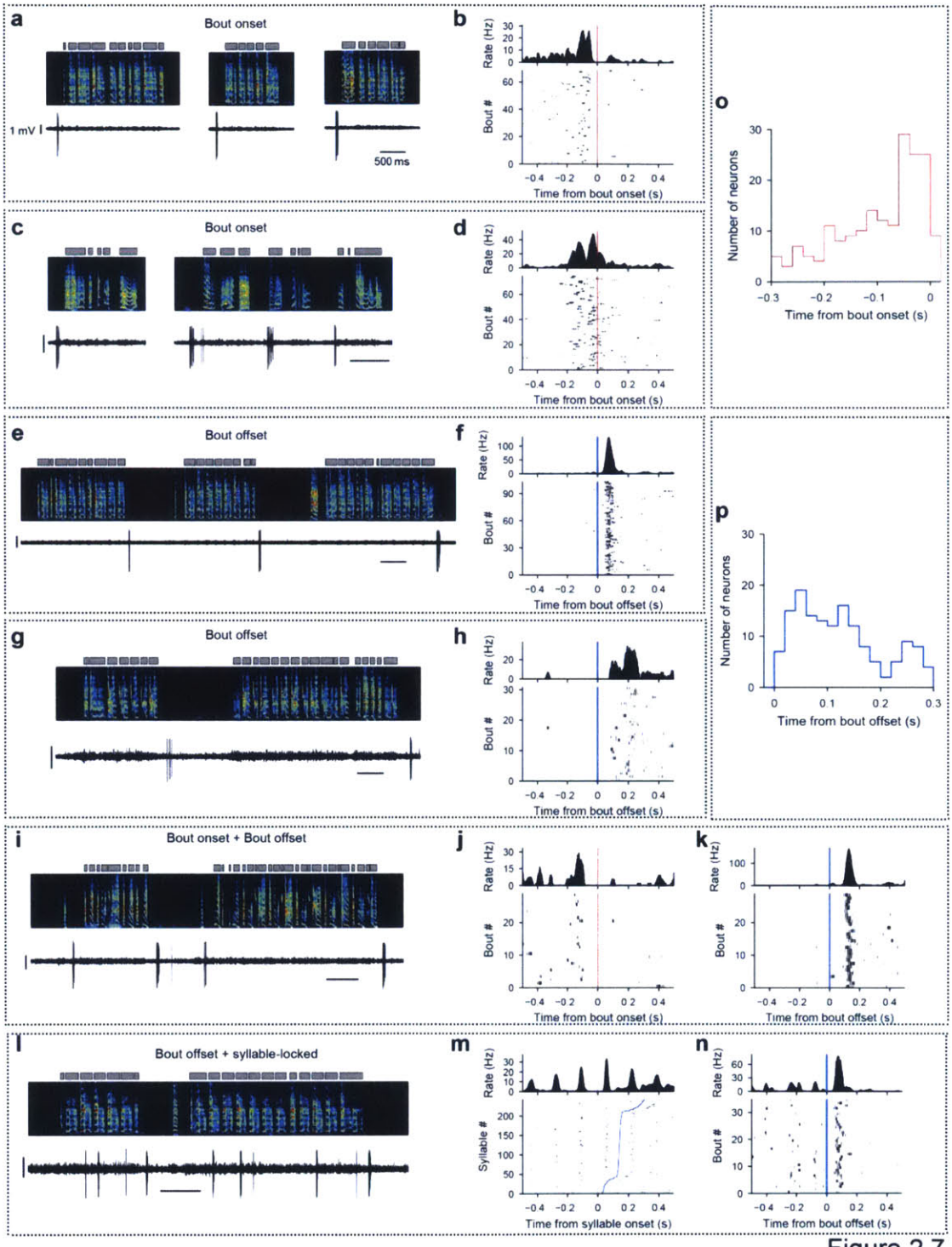

a
Bout onset
1 mV
500 ms
b
Rate (Hz)
Bout #
Time from bout onset (s)
c
Bout onset
d
Rate (Hz)
Bout #
Time from bout onset (s)
o
Number of neurons
Time from bout onset (s)
e
Bout offset
f
Rate (Hz)
Bout #
Time from bout offset (s)
g
Bout offset
h
Rate (Hz)
Bout #
Time from bout offset (s)
p
Number of neurons
Time from bout offset (s)
i
Bout onset + Bout offset
j
Rate (Hz)
Bout #
Time from bout onset (s)
k
Rate (Hz)
Bout #
Time from bout offset (s)
l
Bout offset + syllable-locked
m
Rate (Hz)
Syllable #
Time from syllable onset (s)
n
Rate (Hz)
Bout #
Time from bout offset (s)

Figure 2.7

Figure 2.7: Bout-related activity of HVC projection neurons.

a, Neuron bursting exclusively at bout onset (HVC_X; 61 dph; Bird 2). **b,** Bout-onset aligned histogram and raster plot for the neuron shown in (a). **c,** Another bout-onset neuron (HVC_X; 44 dph; Bird 11). **d,** Bout-onset aligned histogram and raster plot for the neuron shown in (c). **e,** Neuron bursting exclusively at bout offset (HVC_{RA}; 65 dph; Bird 2). **f,** Bout-offset aligned histogram and raster plot for the neuron shown in (e). **g,** Bout-offset neuron (HVC_X; 61 dph; Bird 1). **h,** Bout-offset aligned histogram and raster plot for the neuron shown in (g). **i,** Neuron active before bout onsets and also after bout offsets (HVC_X; 62 dph; Bird 6). **j-k,** Bout-onset (j) or bout-offset (k) aligned histogram and raster plot for the neuron shown in (i). **l,** Neuron exhibiting bout-offset activity, as well as syllable-locking (HVC_X; 63 dph; Bird 2). **m-n,** Syllable-onset (m) or bout-offset (n) aligned histogram and raster plot for the neuron shown in (l). **o,** Distribution of the peak latencies aligned to bout onsets for all the neurons that exhibited significant bout-onset activity (n=187 neurons). **p,** Distribution of the peak latencies aligned to bout offsets for all neurons that exhibited significant bout-offset activity (n=149 neurons). Scale bars: 1 mV, 500 ms.

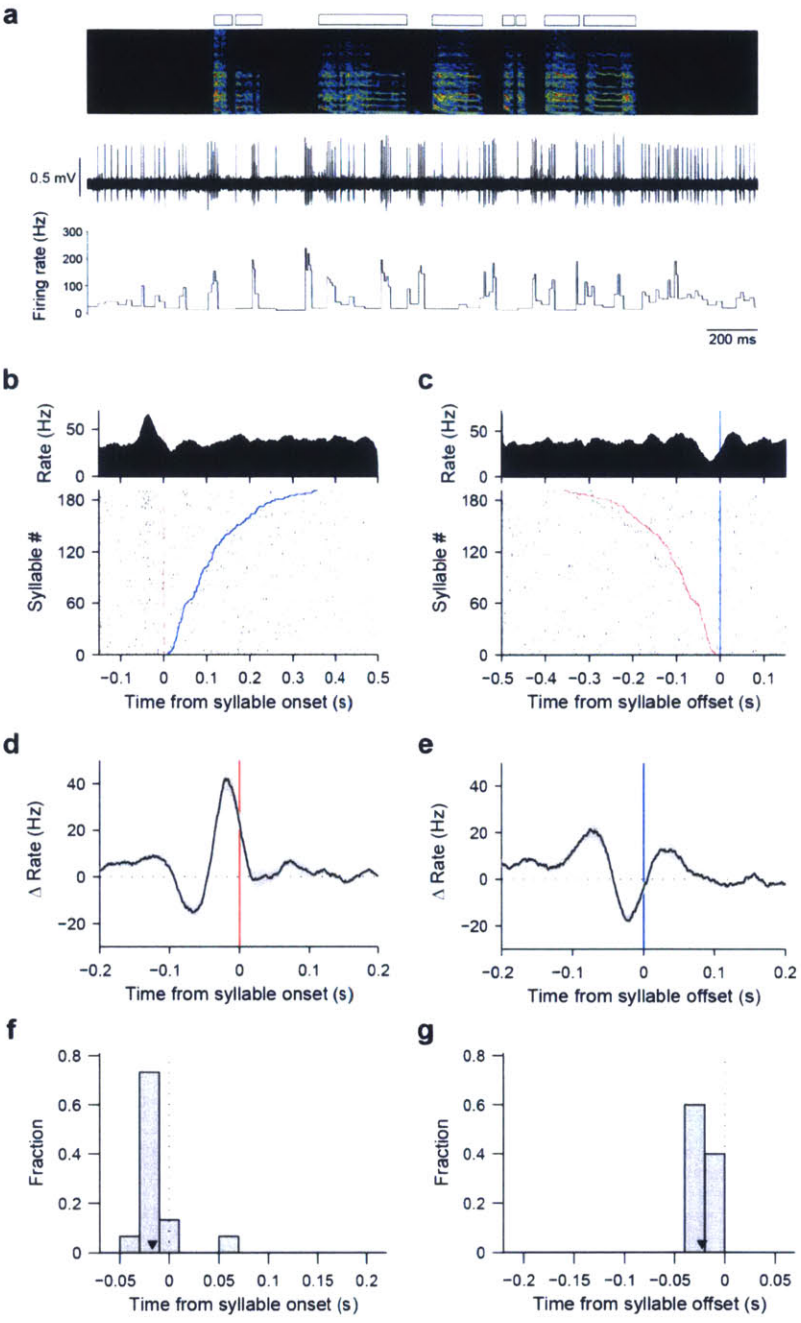

a
0.5 mV
Firing rate (Hz)
300
200
100
0
200 ms
b
Rate (Hz)
50
0
Syllable #
180
120
60
0
−0.1 0 0.1 0.2 0.3 0.4 0.5
Time from syllable onset (s)
c
Rate (Hz)
50
0
Syllable #
180
120
60
0
−0.5 −0.4 −0.3 −0.2 −0.1 0 0.1
Time from syllable offset (s)
d
Δ Rate (Hz)
40
20
0
−20
−0.2 −0.1 0 0.1 0.2
Time from syllable onset (s)
e
Δ Rate (Hz)
40
20
0
−20
−0.2 −0.1 0 0.1 0.2
Time from syllable offset (s)
f
Fraction
0.8
0.6
0.4
0.2
0
−0.05 0 0.05 0.1 0.15 0.2
Time from syllable onset (s)
g
Fraction
0.8
0.6
0.4
0.2
0
−0.2 −0.15 −0.1 −0.05 0 0.05
Time from syllable offset (s)

Figure 2.8

Figure 2.8: Activity of putative HVC interneurons during subsong.

a, Example of a putative HVC interneuron during subsong (47 dph; Bird 9). Top: song spectrogram with syllables indicated above. Middle: extracellular voltage trace. Bottom: instantaneous firing rate. **b,** Syllable-onset aligned raster plot and histogram for the neuron shown in panel a. Red and blue line indicate syllable onsets and offsets, respectively. Note the peak in activity prior to syllable onsets. **c,** Syllable-offset aligned raster plot and histogram for the neuron shown in panel a. Red and blue line indicate syllable onsets and offsets, respectively. Note the dip in activity prior to syllable offsets. **d-e,** Syllable onset (panel d) and syllable offset (panel e) aligned histogram averaged over all putative HVC interneurons recorded during subsong (n=15 neurons). Black line and gray shade indicate mean ± s.e.m. **f,** Distribution of peak latencies of syllable-onset aligned histograms shown for neurons that exhibited significant locking to syllable onsets (n=15 neurons out of 15 neurons recorded during subsong). **g,** Distribution of dip latencies of syllable-offset aligned histograms shown for neurons that exhibited significant locking to syllable offsets (n=10 neurons out of 15 neurons recorded during subsong). Black triangle indicates median latency.

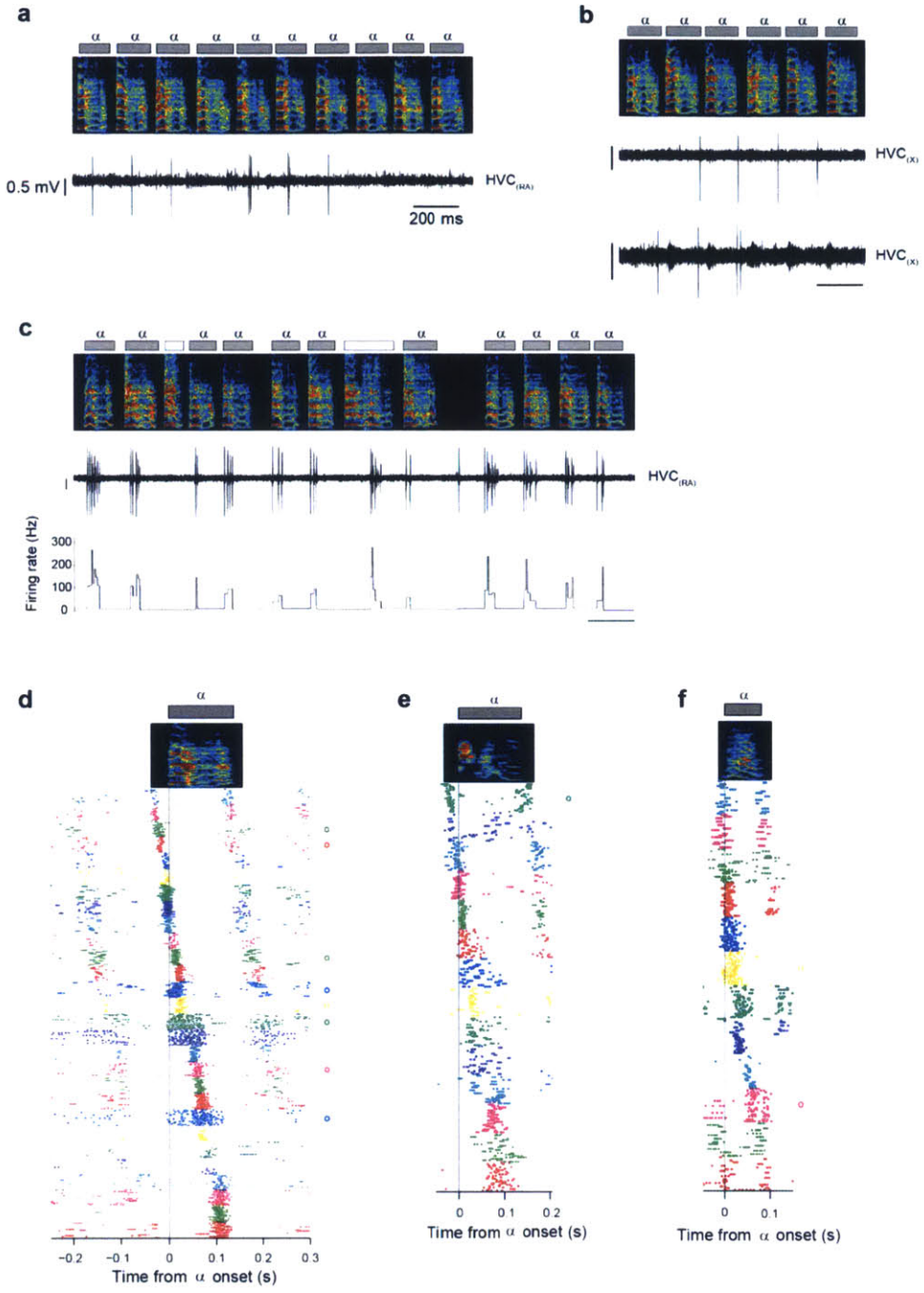

a
α
0.5 mV
HVC(RA)
200 ms
b
HVC(X)
HVC(X)
c
HVC(RA)
Firing rate (Hz)
300
200
100
0
d
−0.2 −0.1 0 0.1 0.2 0.3
Time from α onset (s)
e
0 0.1 0.2
Time from α onset (s)
f
0 0.1
Time from α onset (s)

Figure 2.9

Figure 2.9: Rhythmic sequences in HVC during the protosyllable stage.

a, Example neuron recorded during the protosyllable stage (HVC_{RA}; 62 dph; Bird 2). **b,** Example of a simultaneous recording of two HVC_X neurons from the same bird (64 dph; Bird 2). **c,** Example neuron recorded during the protosyllable stage that exhibited wide bursts (HVC_{RA}; 60 dph; Bird 2). This neuron is shown as the sixth HVC_{RA} neuron (indicated by a dark green circle on the right) in panel c. Note that as a population, bursts were wider in juvenile birds compared to adult birds (Figure 2.2c, top). **d,** Raster plot of 28 HVC projection neurons that were significantly locked to protosyllable onsets (57-64 dph; Bird 2). Neural activity is aligned to protosyllable onsets (vertical solid line at 0 s), and syllables are time warped to the mean offset time of the protosyllable (dotted line). HVC_{RA} neurons are indicated by circles on the right of the rasters. **e-f,** Additional examples of protosequences from two other birds: **e,** 14 projection neurons that were locked to protosyllables (56-59 dph; Bird 1); **f,** 12 projection neurons that were locked to protosyllables (48-49 dph; Bird 10).

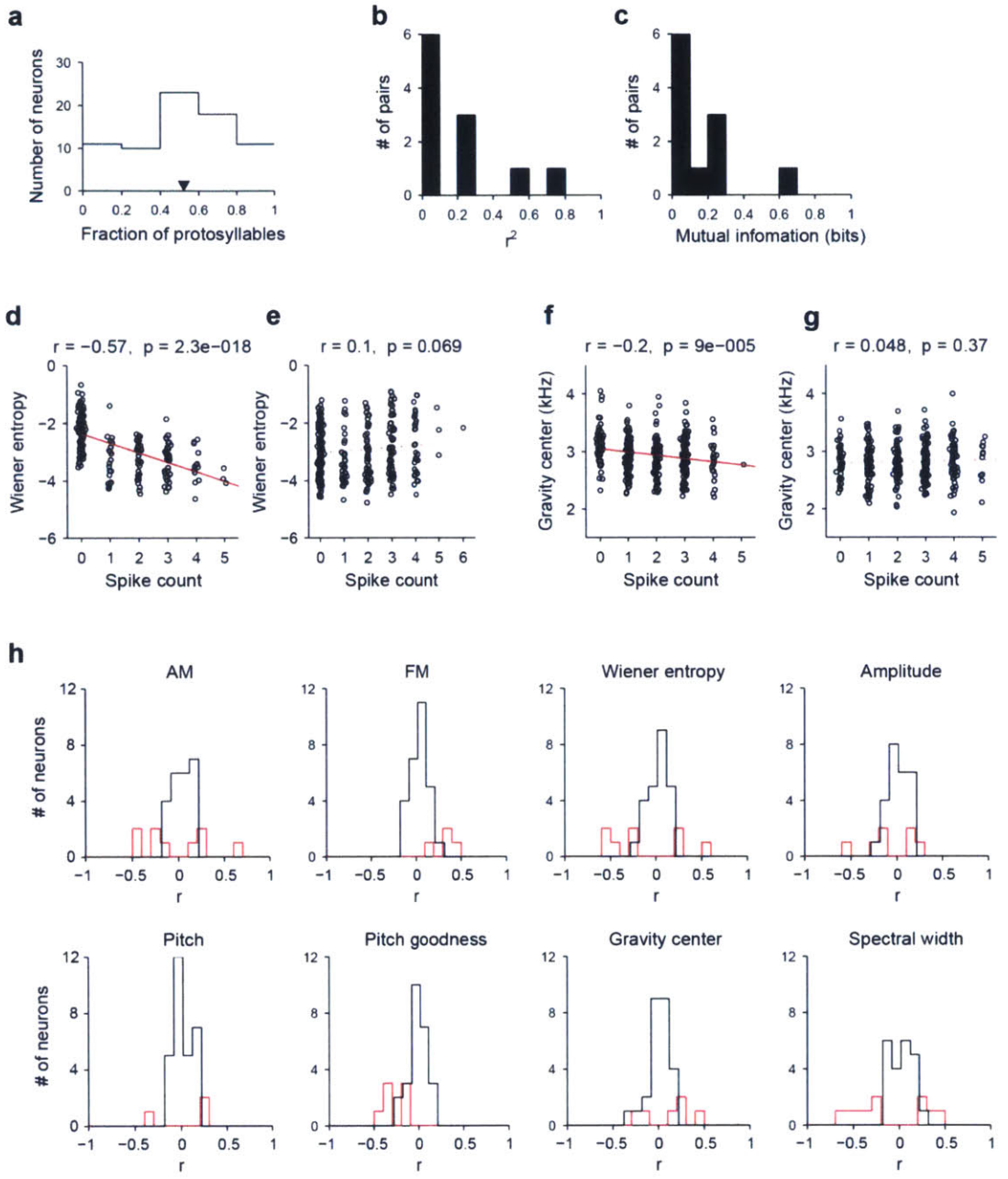

a
Number of neurons
Fraction of protosyllables
b
# of pairs
r²
c
# of pairs
Mutual infomation (bits)
d
r = −0.57, p = 2.3e−018
Wiener entropy
Spike count
e
r = 0.1, p = 0.069
Wiener entropy
Spike count
f
r = −0.2, p = 9e−005
Gravity center (kHz)
Spike count
g
r = 0.048, p = 0.37
Gravity center (kHz)
Spike count
h
AM
FM
Wiener entropy
Amplitude
Pitch
Pitch goodness
Gravity center
Spectral width
# of neurons
r

Figure 2.10

Figure 2.10: Analysis of the probabilistic neural activity during protosyllables.

a, Fraction of protosyllables on which spiking occurs, computed for rhythmic neurons recorded in the protosyllable stage (n=70 neurons). **b,** Histogram of the coefficient of determination r^2 across simultaneously recorded pairs of neurons in the protosyllable stage (n=11 pairs). **c,** Histogram of mutual information across simultaneously recorded pairs of neurons in the protosyllable stage (n=11 pairs). **d,** Example of a neuron with a significant correlation between spike count and Weiner entropy (HVC_p, 59 dph, Bird 1). **e,** Example of a neuron with a lack of correlation (HVC_X, 56 dph, Bird 1). **f,** Example of a neuron with a significant correlation between spike count and spectral center of gravity (HVC_X, 63 dph, Bird 2). **g,** Another example of a lack of correlation (HVC_p, 61 dph, Bird 2). **h,** Distribution of correlation coefficients between spiking activity and eight different acoustic features for 32 neurons recorded in the protosyllable stage. Neurons with statistically significant correlation are shown in red ($P<0.05$ with Bonferroni correction, 8 comparisons).

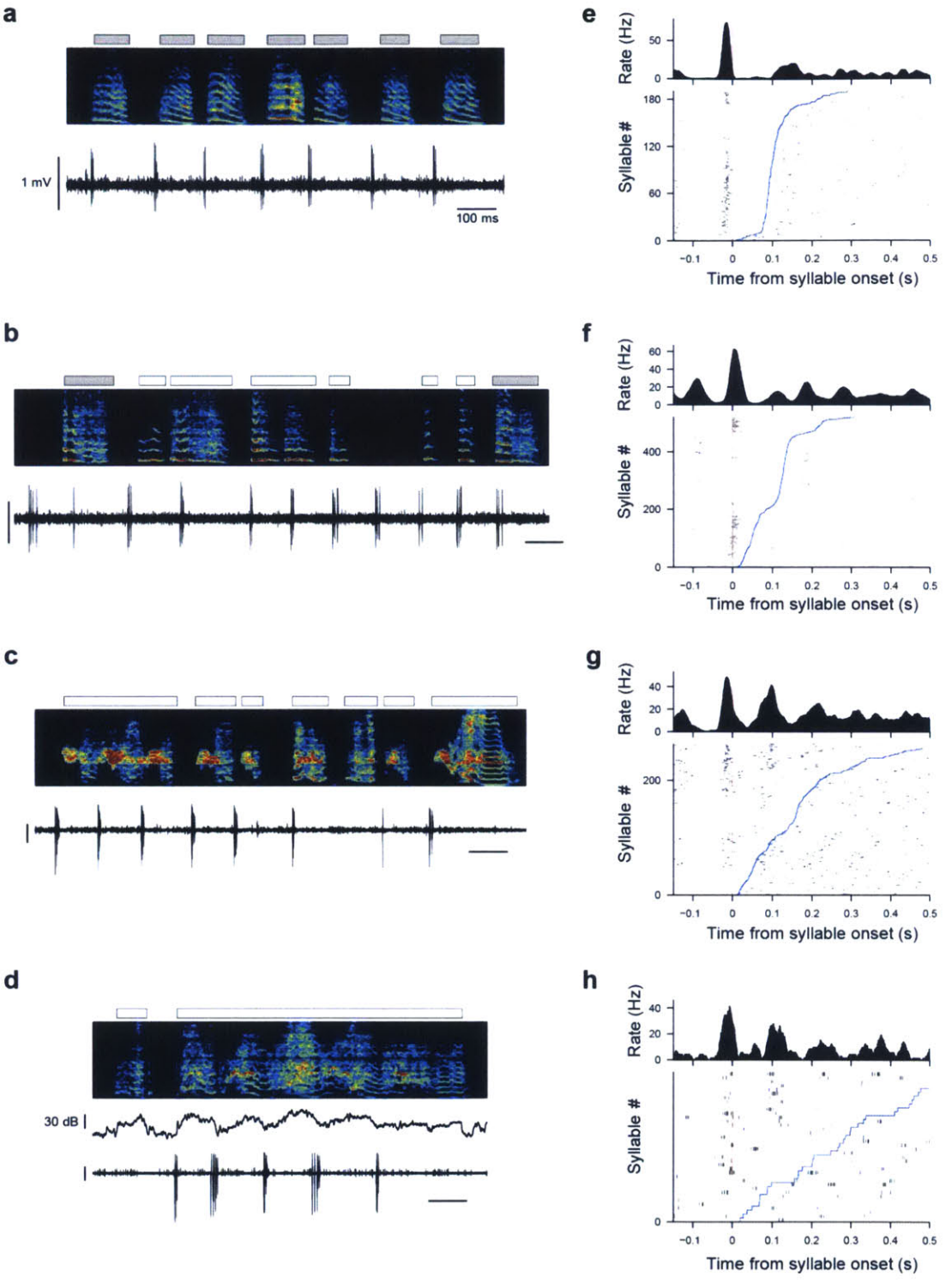

a
1 mV
100 ms
b
c
d
30 dB
e
Rate (Hz)
Syllable #
Time from syllable onset (s)
f
Rate (Hz)
Syllable #
Time from syllable onset (s)
g
Rate (Hz)
Syllable #
Time from syllable onset (s)
h
Rate (Hz)
Syllable #
Time from syllable onset (s)

Figure 2.11

Figure 2.11: Various forms of coupling between rhythmic HVC bursts and juvenile song.

a, Example neuron tightly-coupled to the rhythmic protosyllable (HVC_X; 51 dph; Bird 14). **b,** Example neuron rhythmically bursting throughout the song motif even during long gaps (HVC_{RA}; 57 dph; Bird 2). **c,** Example neuron coupled to the sub-syllabic acoustic structure (HVC_{RA}; 57 dph; Bird 12). **d,** Example of rhythmic bursts correlated with modulations in the song amplitude during a long syllable (HVC_X; 53 dph; Bird 12). **e-h,** Syllable-onset aligned rasters and histograms for neurons shown in panels a-d, respectively. Note that in panels f-h, there is a secondary peak around 0.1 s after syllable onset which corresponds to rhythmic bursting within long syllables.

Chapter 3

Splitting of neural sequence during syllable formation

In the previous chapter, I described how growth of a rhythmic neural sequence in HVC underlies the transition from subsong to protosyllables. How, then, do multiple distinct syllables of adult song arise, and how do distinct sequences in HVC, each corresponding to a different syllable, emerge during vocal learning?

One could imagine two hypotheses that can explain how a new sequence corresponding to a new syllable type emerge (Figure 3.1). First hypothesis, which we call '*de novo* sequence formation,' assumes that a new sequence is assembled every time the bird develops a new syllable type, perhaps using the same growth mechanism utilized in building the protosequence (Figure 3.1a). This hypothesis predicts that there will be little overlap in the neurons coding for two emerging syllables A and B.

Second hypothesis assumes that new sequences emerge by a gradual splitting of a single protosequence (Figure 3.1b). In this view, we predict that the neural sequences underlying newly emerging syllable types would be largely overlapped initially, with many neurons shared across the emerging syllables. Splitting would be associated with an increasing fraction of neurons selective for a specific emerging syllable type.

This chapter examines the predictions of these models in detail, and provides evidence that sequence splitting underlies the emergence of new syllable types. The concept of 'shared neuron' is introduced and its properties are described. We believe that the proposed mechanism is a fundamental process underlying syllable formation. Different learning strategies in which this process could be used to form new syllable types will be discussed in the next chapter.

Results

Alternating syllable differentiation

To test these hypotheses on neural mechanism underlying syllable differentiation, we recorded from a total of 769 HVC projection neurons in 6 juvenile birds while they acquired new syllable types. As a first example, we will describe changes in the HVC population activity in a bird that developed two distinct syllable types over the course of several days (Bird 1; Figure 3.2). Prior to 60 dph, this bird generated protosyllables (labeled α, see Methods) with substantial variability in acoustic features (Figure 3.2a, b). By 60 dph, however, two acoustically distinct syllables were apparent in the song spectrogram (Figure 3.2a). By 62 dph, these two syllables underwent further differential refinement (differentiation), as previously described (Tchernichovski et al., 2001). This resulted in two distinct clusters in the space of acoustic and temporal features, allowing reliable identification of two new syllable types labeled β and γ (Figure 3.2b; estimated classification error < 3%, see Methods).

Specific neurons

We wondered how HVC activity is related to the emergence of these two new syllable types. In total, we recorded 375 projection neurons in Bird 1 throughout vocal development over a nearly two month period (56-112 dph). During the protosyllable stage (56-59 dph), 14 out of 16 projection neurons generated a rhythmic sequence of bursts locked to protosyllables (Figure 3.2c, i). After the emergence of two syllable types, β and γ (62-72 dph), many neurons were selectively active during only β or γ, but not both (Figure 3.2d, e; β-specific: 41/105 neurons; γ-specific: 42/105 neurons; see

Methods). The bursts of these syllable-specific neurons exhibited a wide range of latencies, with spiking activity of neurons in each group spanning the entire duration of each syllable (Figure 3.2j).

Shared neurons

Notably, after the emergence of the two new syllable types (62-72 dph), we also observed a substantial population of neurons that were significantly active during both β and γ ('shared' neurons; Figure 3.2f, g, 22/105 neurons). These shared neurons burst rhythmically at the same period as the neurons recorded during the protosyllable stage (Figure 3.4a, b, f, g). Shared neuron bursts occurred even on renditions of β and γ that were readily distinguishable (Figure 3.2f, g), indicating that the appearance of shared neurons was not an artifact of syllable misclassification (see Methods). Further support for this view comes from simultaneous recordings revealing the co-occurrence, in different neurons, of shared and specific firing patterns (Figure 3.2g, 3.3a, b).

Similar latency of shared neuron bursts

A striking feature of shared neurons was revealed by examining the timing of their bursts. Across the population of neurons, bursts occurred at a wide range of latencies within syllables β and γ. Yet, crucially, the bursts of a shared neuron during β occurred at a similar latency as the bursts of that neuron during γ (Figure 3.5). Thus, as a population, shared neurons generated the same continuous sequence during both β and γ.

Developmental decrease in the fraction of shared neurons

To examine how the population coding in HVC evolves during the further refinement of song syllables, we also recorded the activity of HVC projection neurons in Bird 1 later in development (81-112 dph). We found that the fraction of shared neurons decreased compared to the earlier recordings (Figure 3.2k; 10/100 neurons; P=0.03). Thus, the refinement of β and γ into the adult syllables B and C coincides with a decrease in the fraction of shared neurons, producing a gradual splitting of these representations into distinct, largely non-overlapping neural sequences.

Broad syllable selectivity of shared neurons

How does the initially shared representation of emerging syllables transition to a later non-overlapping representation? We found that the population of shared neurons exhibited a broad range of selectivity for emerging syllables—some were equally active for both syllable types (Figure 3.6a, c) while others showed a significant bias, with higher activity in one syllable than the other (Figure 3.6b, d). These observations suggest that individual shared neurons can exist in a state intermediate between 'specific' and 'shared' (Figure 3.6i, j).

Mechanism underlying the increase in the period of HVC rhythm

We wondered if the emergence of bias and specificity in the activity of neurons toward one emerging syllable type is related to the gradual increase in the period of neural activity through development (Figure 2.5d). In Bird 1, for example, the tendency to alternate between syllables β and γ means that specific neurons had an inter-burst interval that was twice as long as that observed in the earlier protosyllable stage (Figure

3.2c-j, 3.4a-e). Note also that biased shared neurons exhibited a period twice that of unbiased shared neurons as defined by the peak of the spike-train autocorrelation (Figure 3.4g-j). Thus, the increase in the period of neural activity through skipping or alternating cycles of an underlying rhythm appears to be the basis of the increase in song period during vocal learning (Tchernichovski et al., 2001).

Acoustic differences associated with shared neuron bursts

One of the distinguishing features of the emergence of new syllable types is an apparent differentiation of the acoustic structure within the emerging syllables. However, it is possible that shared neurons may only be active at times within emerging syllables at which no acoustic differentiation has yet occurred—that is, at times when the emerging syllable types are acoustically identical. To test this possibility, we analyzed the trajectories of acoustic features of emerging syllable types around the times of shared neuron bursts (see Methods). For the majority of shared neurons (47/76 neurons), there was a statistically significant difference between the acoustic features in two shared syllables, measured in a brief premotor window after the burst time (10-40 ms; Figure 3.7, 3.8). Such differences in acoustic output were observed for both HVC_{RA} and HVC_X neurons (11/15 neurons and 36/61 neurons, respectively; Figure 3.7k-n). These findings suggest that the acoustic structure of emerging syllable types can be significantly different during shared neural sequences, and that both projection neuron types participate in such differentiating shared sequences.

Summary across birds

Although our key findings were described above for Bird 1, a similar pattern of HVC coding by shared and specific neurons was seen in a total of 6 birds for which recordings were made during the emergence of multiple syllable types (Birds 1-6; 185 shared neurons; 496 specific neurons; across 8 syllable pairs analyzed). Across three birds in which neurons were also recorded in later song stages, there was a significant decrease in the fraction of shared neurons during syllable development (n=5 syllable pairs; early-late comparison, $P=3\times10^{-6}$; Birds 1, 2, 4). As reported for Bird 1, a wide range of bias in neural activity of shared neurons was observed in all birds for which more than 3 shared neurons were recorded (Birds 1-5; mean bias: 0.28 ± 0.22 for shared neurons; 0.96 ± 0.11 for specific neurons). Neurons exhibiting an increased burst period by skipping cycles of an underlying rhythm were observed in 4 of the 6 birds (Birds 1, 3, 4, 6; e.g. Figure 4.7f-h, 4.8f, h).

Shared neurons were found in similar proportion across both HVC_{RA} and HVC_X neurons (19% and 28%, respectively; P=0.08; averaged over all developmental stages) and shared neurons of both cell types exhibited the property that bursts have similar latencies during the shared syllables (Figure 3.5). For both neuron types, we observed shared neurons that burst at times where there was a significant acoustic difference between the shared syllables (Figure 3.7, 3.8; 11/15 neurons and 36/61 neurons for HVC_{RA} and HVC_X types, respectively). These findings suggest that both projection neuron types participate in shared neural sequences, and that these shared sequences occur during acoustically distinguishable parts of the emerging syllables.

Discussion

Shared and specific neurons

This chapter addressed the question of how the neural activity in HVC changes during the emergence of two syllable types. We found that during the emergence of two syllable types, there are shared neurons, consistent with the prediction of the sequence splitting model and not the *de novo* sequence formation model (Figure 3.1). Since these shared neurons are a central part of the sequence splitting, their properties are discussed below in detail.

Identification of shared and specific neurons

Determining whether a given neuron is shared or specific depends on being able to identify two distinct syllable types reliably. Is it possible that the existence of shared neurons is an artifact due to the difficulty in classifying syllable types reliably in a variable juvenile song? Several lines of argument suggest that this is not the case.

First, two syllable types β and γ were clearly separable as shown by two distinct clusters in the two-dimensional space of acoustic features (Figure 3.2b). Moreover, we took a conservative approach by only labeled syllable that were clearly identifiable; for example, while songs on 60 dph hints at the existence of two emerging syllable types (Figure 3.2a), given the difficulty in labeling the syllables, we did not include neurons recorded on this day, but instead waited until 62 dph when the reliable labeling became possible (Figure 3.2a). Note that the problem of relying on acoustic to classify shared and specific neurons can be avoided if simultaneous recordings of tens of HVC projection neurons become possible in the future (as discussed in Chapter 5).

Second, quantification of acoustic features associated with shared neuron bursts showed that two syllables are statistically significantly different (Figure 3.7, 3.8). Moreover, because this analysis was done on narrow times around the time of shared neuron bursts, it shows that shared neurons exist even for parts of syllables that differ in acoustics.

Third, early in differentiation, there were two groups of neurons that had periods that differed by a factor of two (Figure 3.4). Calculating the period of the rhythmic bursts in HVC does not rely on syllable labeling. Thus, this result independently shows the existence of two groups of neurons with different properties.

Together, we believe that shared and specific neurons are not an artifact of misclassification.

Similar latency of shared neurons

Most of the shared neurons exhibited similar latencies for the syllables that they were associated with (Figure 3.5). This demonstrates that these two emerging syllable types not only share neurons but also share a neural sequence (Figure 3.2j). This shared sequence is one of the key features that give strong support to the sequence splitting model (Figure 3.1). While these neurons were not simultaneously recorded (Figure 3.2j), if enough neurons are recorded at the same time, we expect that the relative timing of bursts between simultaneously recorded shared neurons will be similar during the two syllables types that they are associated with.

Developmental decrease in the fraction of shared neurons

The sequence splitting model predicts that the fraction of shared neurons decrease as the splitting proceeds (Figure 3.1b). This is what we observed (Figure 3.2j, k). The significance of these results are considered separately for two projection neuron types, HVC_{RA} and HVC_X.

In adult birds, nearly all HVC_{RA} exhibit at most single burst during the motif, indicating that there are few shared HVC_{RA} neurons (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Long et al., 2010; Vallentin and Long, 2015). Thus, the fact that we observed many shared HVC_{RA} neurons early in syllable differentiation (Figure 3.7, 3.8) suggests that this is a special feature of juvenile birds

As for HVC_X, it is known that there are neurons that burst multiple times in the motif in adults (Kozhevnikov and Fee, 2007; Vallentin and Long, 2015). In fact, previous analysis has found that when an HVC_X neuron exhibits multiple bursts in the motif, these bursts tended to precede similar patterns of syllables and gaps, but not similar spectral patterns (Kozhevnikov and Fee, 2007). We speculate that some of these HVC_X neurons that burst multiple times might be a vestige of shared neurons.

Finally, it is important to note that our experimental design compares the fraction of shared neurons during early versus late in differentiation within a single bird; this approach eliminates additional variability that might exist in the fraction of shared neurons across different birds, and allows us to attribute the decrease in the fraction of shared neurons to development.

Developmental progression from shared neuron to specific neuron

Our data from the population of HVC projection neurons shows that fraction of shared neurons decrease during syllable differentiation. What does this result imply about the activity of the individual neurons? We speculate that individual neurons are initially shared between two syllables with equal levels of activity ('unbiased' shared neuron). Then, activity associated with one syllable will gradually decrease ('biased' shared neurons), until this weaker activity falls below threshold for spiking (specific neuron). Consistent with this hypothesis, we observed cases where individual neurons switched between shared and specific activity within several tens of minutes of recording (Figure 3.2h, 3.3c-e)[3]. Moreover, shared neurons had different degrees of bias toward one of the emerging syllables (Figure 3.6), suggesting that at any given point in time, different neurons exhibit different degrees of transition from shared to specific. Testing this idea requires long-term tracking of HVC activity over multiple days (as discussed in Chapter 5).

Increase in period of rhythmicity during development

Behavioral studies have shown that 'period,' measured as a median duration between two repetitions of sound, increases during the course of vocal development (Tchernichovski et al., 2001). Similarly, we have observed a developmental increase in the period of HVC rhythm quantified by the peak in spike-train autocorrelation (Figure 2.5d). Here, we outline a potential mechanism of how this can happen. During the protosyllable stage, most neurons are rhythmically bursting at the protosyllable rhythm

[3] Note that switching from specific to shared activity in a single neuron might be due to a depolarization of the cell membrane caused by the extracellular electrode. However, we observed neurons switching from shared to specific (Figure 3.2h, 3.3d) suggesting that not all of our observations of switching can be explained by the change in cell health.

with a period T (typically 100-200 ms). As two syllable types emerge, shared neurons with no bias exhibit the same period T. As soon as shared neurons start having bias in the activity level, the spike-train autocorrelation will tend to have a larger peak at $2T$ than at T (compare Figure 3.4g and h), while the peak in inter-burst interval remains to be at T. As neurons stop being active on one of the syllable types and thus start to skip cycles of the rhythm, both the peak in inter-burst interval (Figure 3.4c-e) and the peak in spike-train autocorrelation (Figure 3.4i, j) become $2T$. This mechanism is similar to the 'period doubling' often observed in nonlinear dynamical system (Strogatz, 1994). Note that while 'skipping of cycles' as a mechanism for increase in period may be general, period doubling is likely to be special for cases where newly emerging syllable types alternate, as in Bird 1 (Figure 3.2).

Methods

Analysis of neural activity

Shared and specific neurons

To examine whether a given HVC projection neuron was active during multiple syllable types ('shared' neuron) or was active only during a specific syllable type ('specific' neuron), we first constructed a syllable-onset aligned histogram (1 ms bin smoothed over 20 bins) for each syllable type. Spike times were linearly time warped (Leonardo and Fee, 2005) to the mean duration of that syllable to reduce the trial-to-trial variability in the spike timing associated with the variation in the syllable duration. Next, the significance of the peak in the histogram was determined by comparing it with the peak size obtained from the shuffled histogram using the same method described above (*Syllable-related neural activity)*

We defined 'shared' and 'specific' neurons in the context of a particular syllable differentiation process (e.g. β and γ from Bird 1 in Figure 3.2; α and β from Bird 2 in Figure 4.2; B and D from Bird 1 in Figure 4.5). 'Specific' neurons were defined as neurons that had a significant peak in the syllable-aligned histogram for only one syllable type, whereas 'shared' neurons were defined as neurons that had significant peaks for both syllable types (time window 30 ms prior to syllable onset to 10 ms post syllable offset). We visually inspected the syllable-aligned histograms to confirm that the same burst is not detected twice (i.e. being associated with an offset of one syllable and an onset of the next syllable). We took a conservative approach and only considered a neuron to be shared if the peak was significant for both syllable types. However, some

specific neurons had weak activity for other syllables that did not reach significance (e.g. Figure 4.2d).

Note that we defined shared and specific neurons in the context of a particular pair of syllables undergoing differentiation. For example, in a bird that exhibited hierarchical differentiation (Bird 1), we saw examples of neurons that were B-specific when considering B-C differentiation but shared when considering B-D differentiation. Thus, when considering all the syllables in the motif, our definition of shared and specific neuron based on syllable pairs will underestimate the fraction of shared neurons and overestimate the fraction of specific neurons.

Quantification of the similarity of latencies in shared neurons (Figure 3.5)

To test whether shared neurons were active at similar latencies for multiple syllable types, we first calculated the latency of the peak in the syllable onset (or offset)-aligned histograms. We then plotted the latency of the peak for one syllable against that of another syllable (Figure 3.5). When a shared neuron was active for three or more syllables, two syllables associated with two highest firing rates were chosen. To quantify whether shared neurons were active at similar latencies for two syllable types, we calculated the Pearson's correlation coefficient r between two latencies.

Quantification of the activity level difference in shared neurons (Figure 3.6)

To quantify the difference in the activity level for multiple syllable types in the shared neurons, we calculated the 'bias' defined as follows

$$Bias = 1 - \frac{\min(r_1, r_2)}{\max(r_1, r_2)}$$

where r_i is the peak firing rate in the syllable-aligned histogram for syllable i. Bias of 0 indicates equal activity level for both syllable types, whereas bias of 1 indicates exclusive activity for only one of the syllable types (Figure 3.6j).

Analysis of acoustic features associated with bursts of shared neurons (Figure 3.7, 3.8)

During song development, shared neurons are defined by the fact that they burst within more than one emerging syllable type. We wondered if the bursts of shared neurons were associated with different acoustic signals at the time of the shared bursts. (An alternative possibility is that shared neurons burst only at times within the emerging syllable types when the acoustic signals are identical.) An example of a neuron analyzed here is shown in Figure 3.7a (from the same data shown in Figure 3.2f). This neuron bursts just after the onset of both syllables β and γ. We analyzed the acoustic differences in a 0-50 ms analysis window after the burst time, but were most interested in acoustic differences in a narrower premotor window (10-40 ms), as this corresponds to the premotor latency for which one expects HVC neurons to exert an effect on vocal output (Fee et al., 2004; Ashmore et al., 2005; Ali et al., 2013).

For each neuron analyzed, all syllables in which the neuron generated a burst were identified. The analysis was carried out for every syllable rendition on which the neuron burst, and was restricted to only those syllables. Syllables had previously been labeled by type (i.e. β and γ). We first directly visualized the spectral differences between the two syllable types using a sparse contour representation (Lim et al., 2012; Markowitz

et al., 2013), which is suitable for constructing an 'average' spectrogram. The analysis was carried out on the sound signal extracted from a 50 ms window after each burst. In many cases, this spectral representation revealed consistent differences between the different syllable types in this analysis window (Figure 3.7b, c).

One complication is that some of the shared neurons burst prior to syllable onsets or immediately before syllable offsets such that the 10-40 ms window after the bursts were obscured by silent gaps (9 of 24 HVC_{RA} neurons and 59 of 120 HVC_X neurons were obscured). These neurons were excluded from analysis of acoustic difference.

We further quantified differences in the acoustic signals by extracting time varying acoustic and spectral features in a window 0-50 ms after burst time (see subsection *Definition of bursts*). We used 8 acoustic features previously established to analyze birdsongs (Wiener entropy, spectral center of gravity, spectral width, pitch, pitch goodness, sound amplitude, amplitude modulation, frequency modulation)(Tchernichovski et al., 2000; Mandelblat-Cerf and Fee, 2014). The 8-dimensional vector of features was calculated in 1 ms steps over the 50 ms analysis window (Figure 3.7d, e).

Because each syllable was labeled, we could determine if the feature trajectories were significantly different for syllables labeled β and those labeled γ, and to make this determination at every time step in the analysis window (Figure 3.7d; s.e.m. indicated by shaded region around mean trajectory). Rather than quantify the difference in these trajectories one feature at a time, we used Fisher's discriminant analysis (Duda et al., 2001) to project the 8-dimensional acoustic feature vector onto a single dimension that gives maximum separability between the two syllable types. The projected direction is

determined independently at each time point, and the feature vectors of all syllable renditions are projected, at each time point, to yield a distribution of projected samples. For most neurons, the different syllable types produce visibly different distributions of projected samples (Figure 3.7f) indicating distinct acoustic structure. The separability of the distributions (in one dimension) of projected samples for different syllable types was quantified using the d-prime metric (d'), corresponding to the distance between the means of the distributions, normalized by the pooled variance (Markowitz et al., 2013)

$$d' = \frac{\mu_A - \mu_B}{\sqrt{\frac{1}{2}(\sigma_A^2 + \sigma_B^2)}}$$

Because the features evolve in time, this analysis is carried out independently at each 1 ms step in the 50 ms analysis window, and the d' was plotted as a function of time (Figure 3.7g). Statistical significance of the d' trajectory was assessed by randomizing the syllable labels and rerunning the d' analysis on shuffled datasets (N=1,000 shuffles). For each randomization, the peak value of d' in 10-40 ms premotor window was recorded; significance threshold was as set as the 95 percentile of the distribution of these peak values. A shared neuron was determined to have significant acoustic difference between the shared syllables *only* if the d' trajectory *remained above this significance threshold for the entire premotor window of 10-40 ms after the burst.* Note that, in the simulated data, none of the 1,000 surrogate runs generated a d' trajectory that met this stringent criterion.

Statistics

Figure 3.2(j,k) To test whether the fraction shared neurons differed between early and late stages of syllable differentiation, we used the χ^2 test on a 2 × 2 contingency table (shared/specific, early/late). *Significance across all birds:* To calculate whether the fraction of shared neurons differed between early and late stages of syllable differentiation over all birds (n=5 syllable pairs in 3 birds), we used the Cochran-Mantel-Haenszel test for repeated tests of independence (McDonald, 2014).

Figure 3.5 To calculated the relation between latencies of bursts associated with shared neurons, we calculated the Pearson's correlation coefficient *r* together with the P-value under the null hypothesis that $r=0$.

Figure 3.7(m, n) To test whether mean d' metric were different between HVC_{RA} and HVC_X, we used Wilcoxon rank-sum test. Only neurons with d' trajectories that were significant (continuously from 10-40 ms) were included in this comparison.

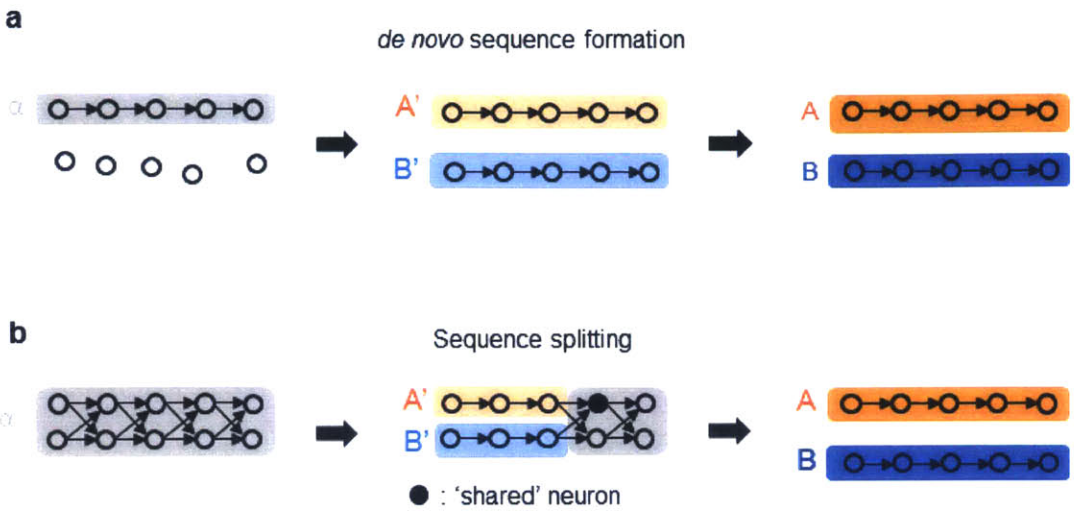

a
de novo sequence formation
A'
B'
A
B
b
Sequence splitting
A'
B'
A
B
● : 'shared' neuron

Figure 3.1

Figure 3.1: Two hypotheses on the neural mechanisms of the emergence of new syllable types.

Individual circles indicate HVC projection neurons, and arrows between them indicate connections between them. Gray box indicates protosyllable α, and orange and blue boxes indicate two emerging syllable types A' and B' which are refined to become distinct syllable types A and B. **a,** According to the *de novo* sequence formation model, sequences are assembled one at a time whenever new syllable types emerge. First, there is a protosequence coding for the protosyllable α (left) and this becomes the sequence that codes for A', while a new sequence forms to code for the emerging syllable type B' (middle). This hypothesis predicts that there is little overlap between the sequences coding for A and B throughout this process. **b,** The second hypothesis, called sequence splitting, hypothesizes that a protosequence (left) splits into two to give rise to two sequences that codes for A or B (right). Notice that during the splitting process, cross connections between the A' sequence and B' sequence are gradually eliminated (middle). This hypothesis predicts that there are 'shared' neurons active for both of the emerging syllable types A' and B' (e.g. neuron shown in black). The fraction of shared neurons decreases as the splitting proceeds, and all neurons become specific in the end (right). Note that in the case depicted here, the splitting is occurring from the beginning of the sequence for illustrative purpose. However, there can be other ways in which splitting could happen, such as nearly simultaneous elimination of cross connections between sequences A' and B' throughout the syllable.

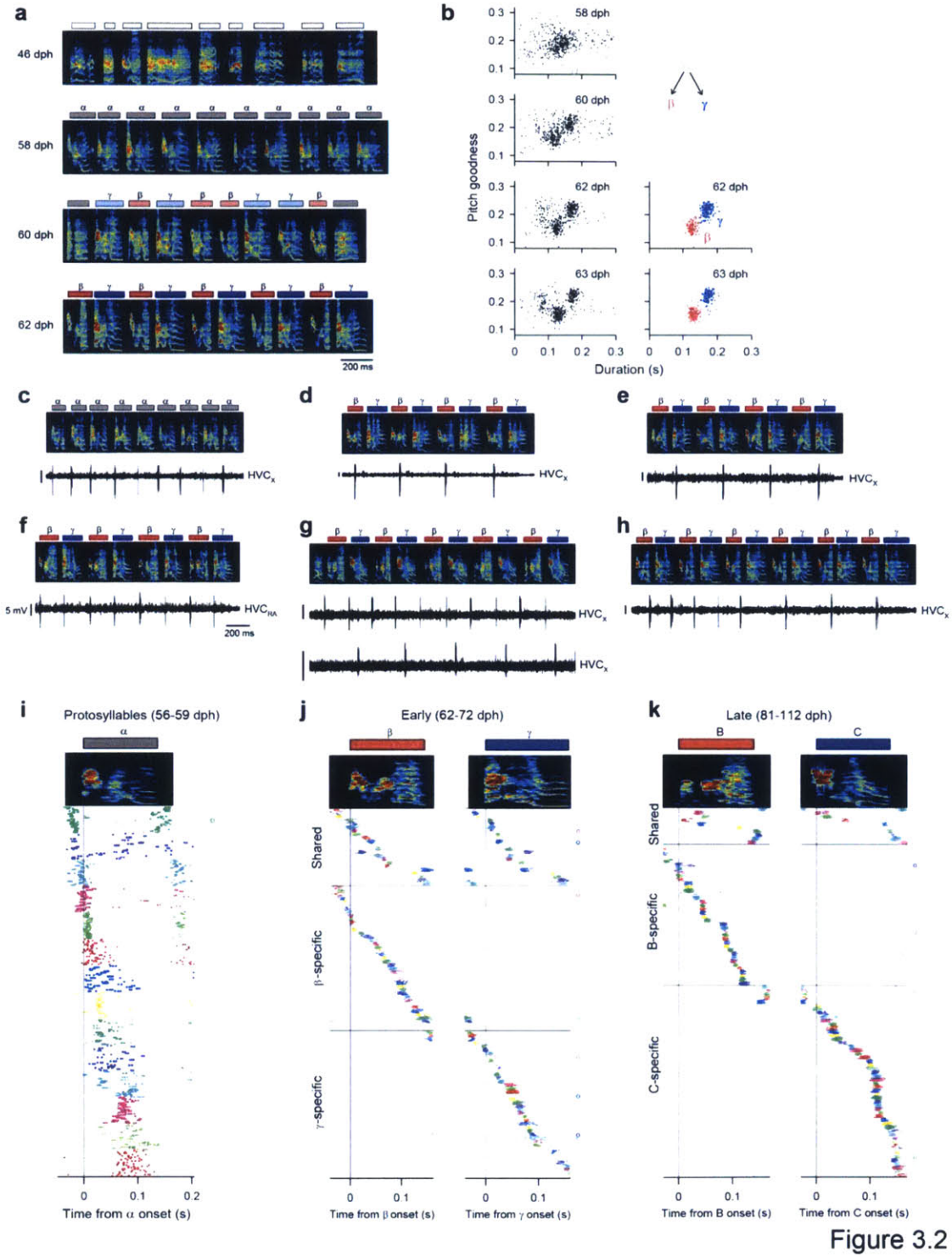

a
46 dph
58 dph
60 dph
62 dph
200 ms
b
58 dph
60 dph
62 dph
63 dph
Pitch goodness
Duration (s)
c
d
e
HVC_x
f
5 mV
HVC_RA
200 ms
g
h
i
Protosyllables (56-59 dph)
Time from α onset (s)
j
Early (62-72 dph)
Shared
β-specific
γ-specific
Time from β onset (s)
Time from γ onset (s)
k
Late (81-112 dph)
B
C
Shared
B-specific
C-specific
Time from B onset (s)
Time from C onset (s)

Figure 3.2

Figure 3.2: Shared and specific sequences during alternating differentiation.

All data from Bird 1. **a,** Song examples during the emergence of syllables β and γ. Panels: i, subsong stage (46 dph); ii, rhythmic repetition of protosyllable α (58 dph); iii, rhythmic repetition of variants of the protosyllable (β and γ ; 60 dph); iv, further acoustic differentiation of β and γ (62 dph). **b,** Scatter plot of syllable duration versus mean pitch goodness (each dot is one syllable rendition; n=400 syllables per day; unclassified syllables gray). **c,** Neuron recorded during protosyllable stage (56 dph). **d,** β-specific neuron (64 dph). **e,** γ-specific neuron (65 dph). **f,** Shared neuron active during both β and γ (68 dph). **g,** Simultaneously-recorded pair: shared neuron (top) and γ-specific neuron (bottom; 71 dph). **h,** Neuron switching between shared and specific spiking (63 dph). **i,** Population raster of 14 neurons aligned to protosyllable onsets. **j,** Raster of 105 projection neurons early in syllable differentiation showing shared (top) and specific (bottom) sequences. **k,** Same as panel **j** but for 100 neurons recorded after differentiation of β and γ into adult syllables B and C. Scale bars: (c-h) 0.5 mV, 200 ms.

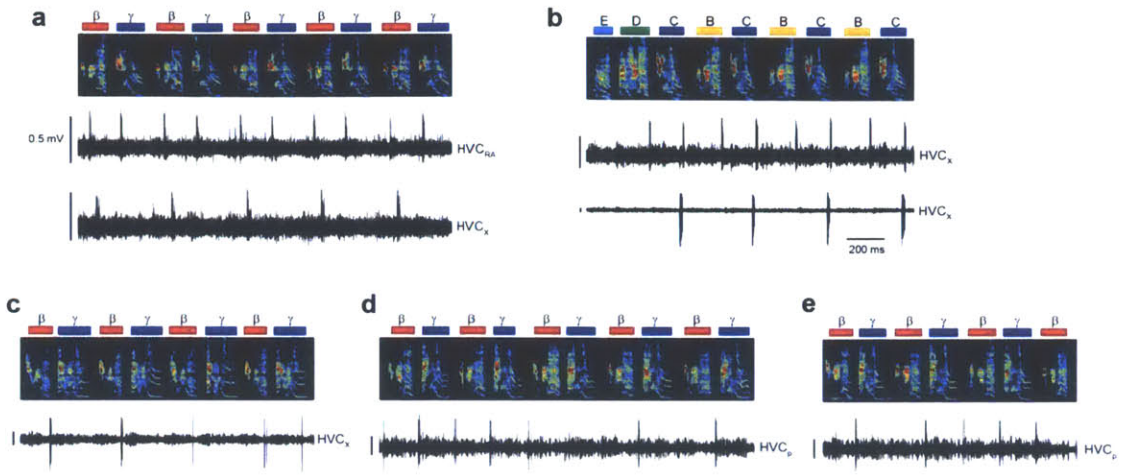

a
β γ β γ β γ β γ β γ
0.5 mV
HVC$_{RA}$
HVC$_{X}$
b
E D C B C B C B C
HVC$_{X}$
HVC$_{X}$
200 ms
c
β γ β γ β γ β γ
HVC$_{X}$
d
β γ β γ β γ β γ β γ
HVC$_{P}$
e
β γ β γ β γ β
HVC$_{P}$

Figure 3.3

Figure 3.3: Increase in the period of HVC rhythmicity during differentiation.

All data are from Bird 1. **a,** Paired recording of a shared neuron and a β-specific neuron (69 dph; neuron type indicated at right of each voltage trace). **b,** Paired recording of a shared neuron and a C-specific neuron (110 dph). **c,** Same neuron as in Figure 3.2h showing switch from being β-specific in the first part to being shared in the second part. **d,** Neuron switching from being shared in the first half to being γ-specific in the second half (68 dph). **e,** Same neuron as in (d), but going from being γ-specific in the first half to being shared in the second half. Scale bars: 0.5 mV, 200 ms.

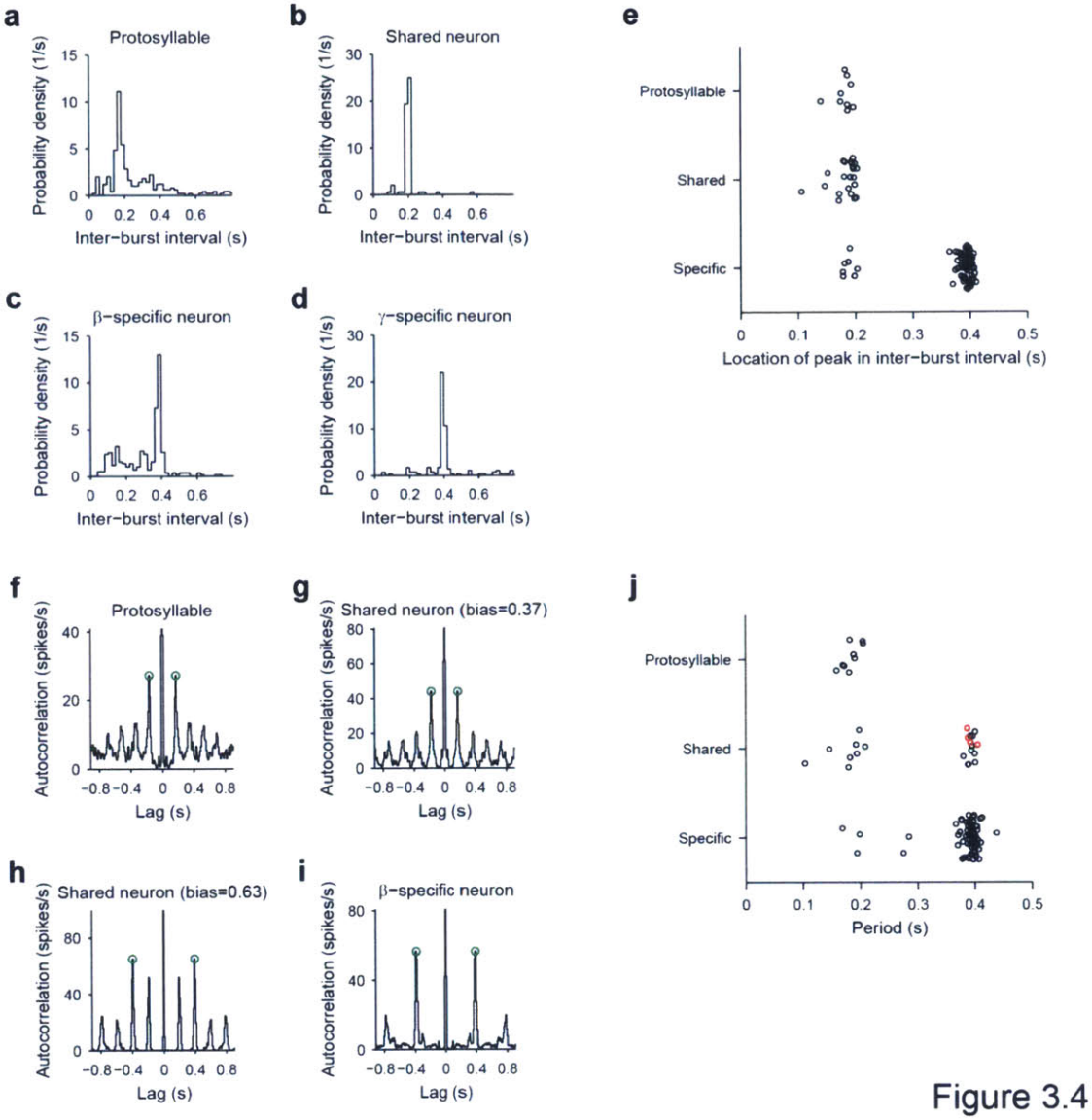

a
Protosyllable
Probability density (1/s)
Inter−burst interval (s)
b
Shared neuron
Probability density (1/s)
Inter−burst interval (s)
c
β−specific neuron
Probability density (1/s)
Inter−burst interval (s)
d
γ−specific neuron
Probability density (1/s)
Inter−burst interval (s)
e
Protosyllable
Shared
Specific
Location of peak in inter−burst interval (s)
f
Protosyllable
Autocorrelation (spikes/s)
Lag (s)
g
Shared neuron (bias=0.37)
Autocorrelation (spikes/s)
Lag (s)
h
Shared neuron (bias=0.63)
Autocorrelation (spikes/s)
Lag (s)
i
β−specific neuron
Autocorrelation (spikes/s)
Lag (s)
j
Protosyllable
Shared
Specific
Period (s)

Figure 3.4

Figure 3.4: Increase in the period of HVC rhythmicity during differentiation.

a-d, Inter-burst interval (IBI) distributions for shared and specific neurons. **a,** for the neuron in Figure 3.2c. recorded during protosyllable stage. **b,** for the shared neuron shown in the top panel of Figure 3.2g. **c,** for the β-specific neuron shown in Figure 3.2d. **d,** for the γ-specific neuron shown in Figure 3.2e. **e,** Population summary of the peak in IBI for the neurons recorded during the protosyllable stage (n=9), shared neurons (n=22), and specific neurons (n=83) that were recorded between 62-72 dph. Note that shared neurons had the same most-probable IBI as neurons recorded during the protosyllable stage, and that most specific neurons had a most-probable IBI twice as long. **f-i,** Examples of spike-train autocorrelations for neurons showing increasing periodicity (red dots indicate peak in autocorrelation function). **f,** Neuron recorded in protosyllable stage. **g,** Shared neuron with low bias (bias = 0.37). **h,** Shared neuron with high bias (bias = 0.63). **i,** Specific neuron. **j,** Population summary of the period of neural activity, as determined from the peak in the spike train autocorrelation. Same set of neurons shown in (e). Note that shared neurons with a high bias (panel a) exhibit a period twice that of neurons with low bias (panel g), and twice that of neurons recorded during the protosyllable stage (panel f).

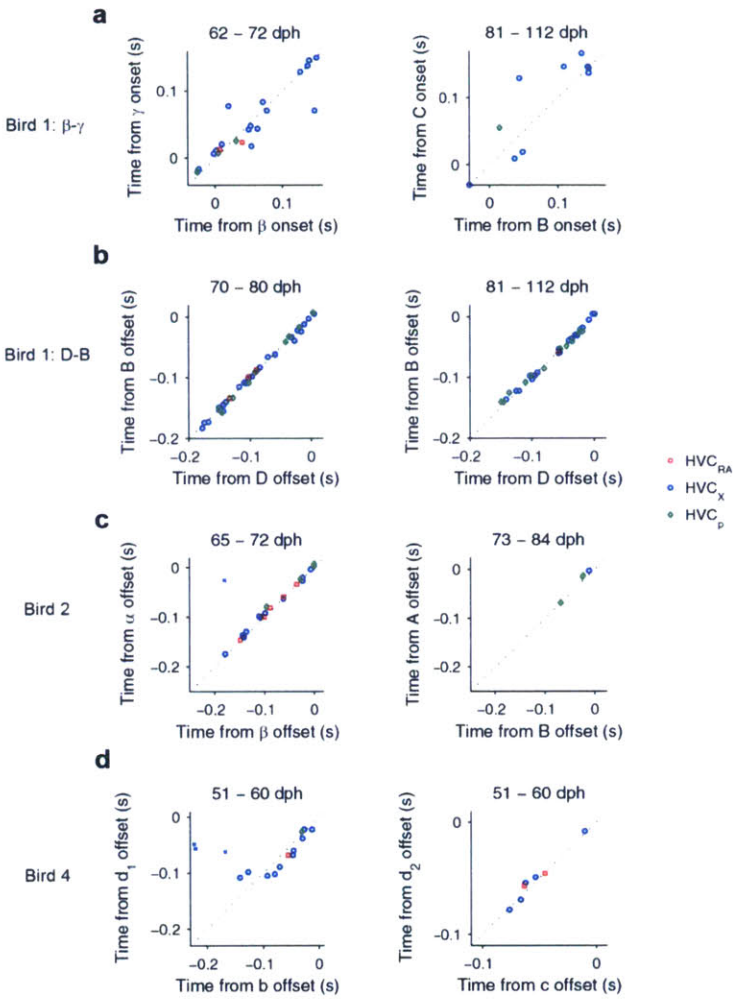

a
Bird 1: β-γ
62 – 72 dph
81 – 112 dph
Time from γ onset (s)
Time from C onset (s)
Time from β onset (s)
Time from B onset (s)
b
Bird 1: D-B
70 – 80 dph
81 – 112 dph
Time from B offset (s)
Time from D offset (s)
c
Bird 2
65 – 72 dph
73 – 84 dph
Time from α offset (s)
Time from A offset (s)
Time from β offset (s)
Time from B offset (s)
d
Bird 4
51 – 60 dph
51 – 60 dph
Time from d1 offset (s)
Time from d2 offset (s)
Time from b offset (s)
Time from c offset (s)
HVC_RA
HVC_X
HVC_P

Figure 3.5

Figure 3.5: Analysis of latency of shared neurons.

a-d, Latencies of shared neuron bursts, color-coded by cell type: HVC_{RA} (red), HVC_X (blue), and HVC_p (green). **a,** Neurons in Bird 1 shared between syllables β and γ (Figure 3.2) during the early (top panel; r=0.91, P<0.001) and late (bottom panel; r=0.87, P=0.005) stages of syllable differentiation. **b,** Neurons in Bird 1 shared between syllables D and B (Figure 4.5) during the early (top panel; r>0.99, P<0.001) and late (bottom panel; r>0.99, P<0.001) stages of syllable differentiation. **c,** Neurons in Bird 2 shared between syllables β and α (Figure 4.2) during the early (top panel; r>0.99, P<0.001) and late (bottom panel; r>0.99, P<0.001) stages of syllable differentiation. A shared neuron that had two peaks during the syllable α is shown with an 'x' symbol; this was not included in the calculation of correlation. **d,** Neurons in Bird 4 shared between syllables 'b' and 'd_1' (Figure 4.8*l*) during the early stage of syllable differentiation (top panel; r=0.89, P<0.001). Also shown (bottom panel) are neurons in Bird 4 shared between syllables 'c' and 'd_2' (Figure 4.8n) during early stage of syllable differentiation (r=0.98, P<0.001).

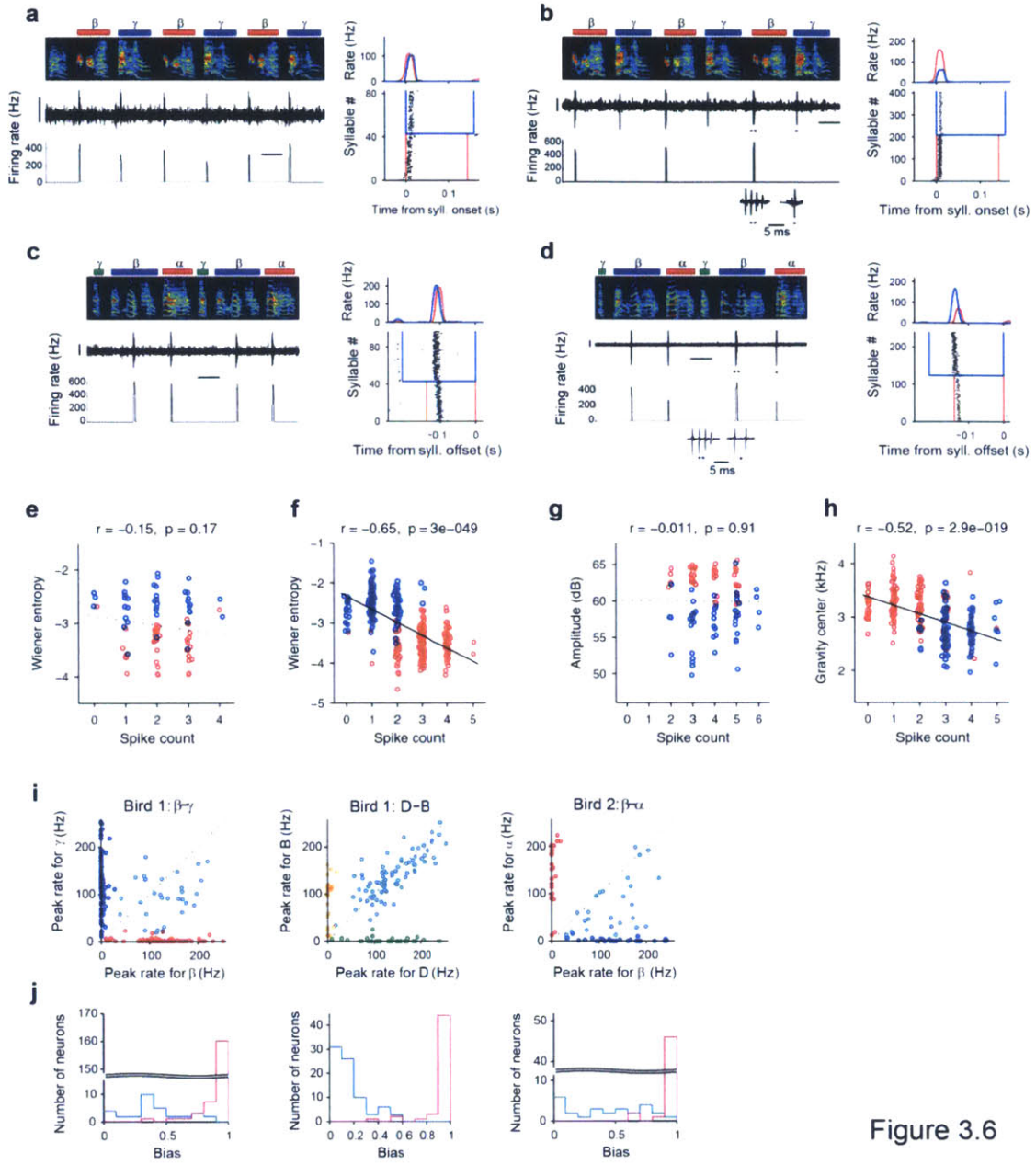

a
b
c
d
e
f
g
h
i
j
Firing rate (Hz)
Rate (Hz)
Syllable #
Time from syll. onset (s)
Time from syll. offset (s)
5 ms
r = −0.15, p = 0.17
r = −0.65, p = 3e−049
r = −0.011, p = 0.91
r = −0.52, p = 2.9e−019
Wiener entropy
Amplitude (dB)
Gravity center (kHz)
Spike count
Bird 1: β–γ
Bird 1: D–B
Bird 2: β–α
Peak rate for β (Hz)
Peak rate for D (Hz)
Peak rate for B (Hz)
Peak rate for γ (Hz)
Peak rate for α (Hz)
Number of neurons
Bias

Figure 3.6

Figure 3.6: Analysis of syllable selectivity (bias) of shared neurons.

a, Shared neuron (left) and syllable-onset-aligned raster plot and histogram (right) showing similar peak firing rates for syllables β and γ (bias = 0.07; HVC_p; 68 dph, Bird 1). **b,** Shared neuron (left) and syllable-onset-aligned raster plot and histogram (right) showing higher peak firing rate for syllable β than γ (bias = 0.63; HVC_{RA}; 68 dph, Bird 1). **c,** Shared neuron with a similar peak firing rate (bias = 0.06; HVC_X; 69 dph, Bird 2). **d,** Shared neuron with a higher peak firing rate for syllable β than α (bias = 0.55; HVC_X; 68 dph, Bird 2). **e,** Lack of correlation between spike count and Wiener entropy for an unbiased shared neuron (same neuron as panel a). **f,** Significant correlation between spike count and Wiener entropy in a biased shared neuron (same neuron as panel b). **g,** Lack of correlation in an unbiased shared neuron (same neuron as panel c). **h,** Significant correlation in a biased shared neuron (same neuron as panel d). Note that the existence of biased activity and acoustic differences between syllables automatically introduces a correlation between spike count and acoustic features. **i,** Comparison of the peak firing rate during two different syllables, quantified by the height of the peak in the syllable-aligned spike histogram. Each individual dot is a neuron; shared neurons are colored in cyan (unbiased neurons are near the diagonal); specific neurons are colored according to the syllable with which it is associated, and appear near the axes. **j,** Distribution of the biases for shared neurons (cyan) and specific neurons (magenta). This 'bias' ranged from 0, representing equal activity, to 1, representing activity exclusive to either one of the syllables (see Methods). Specific neurons exhibited a bias tightly clustered around one (0.97 ± 0.087, mean ± s.d.). In contrast, shared neurons exhibited a broad range of bias (0.27 ± 0.24).

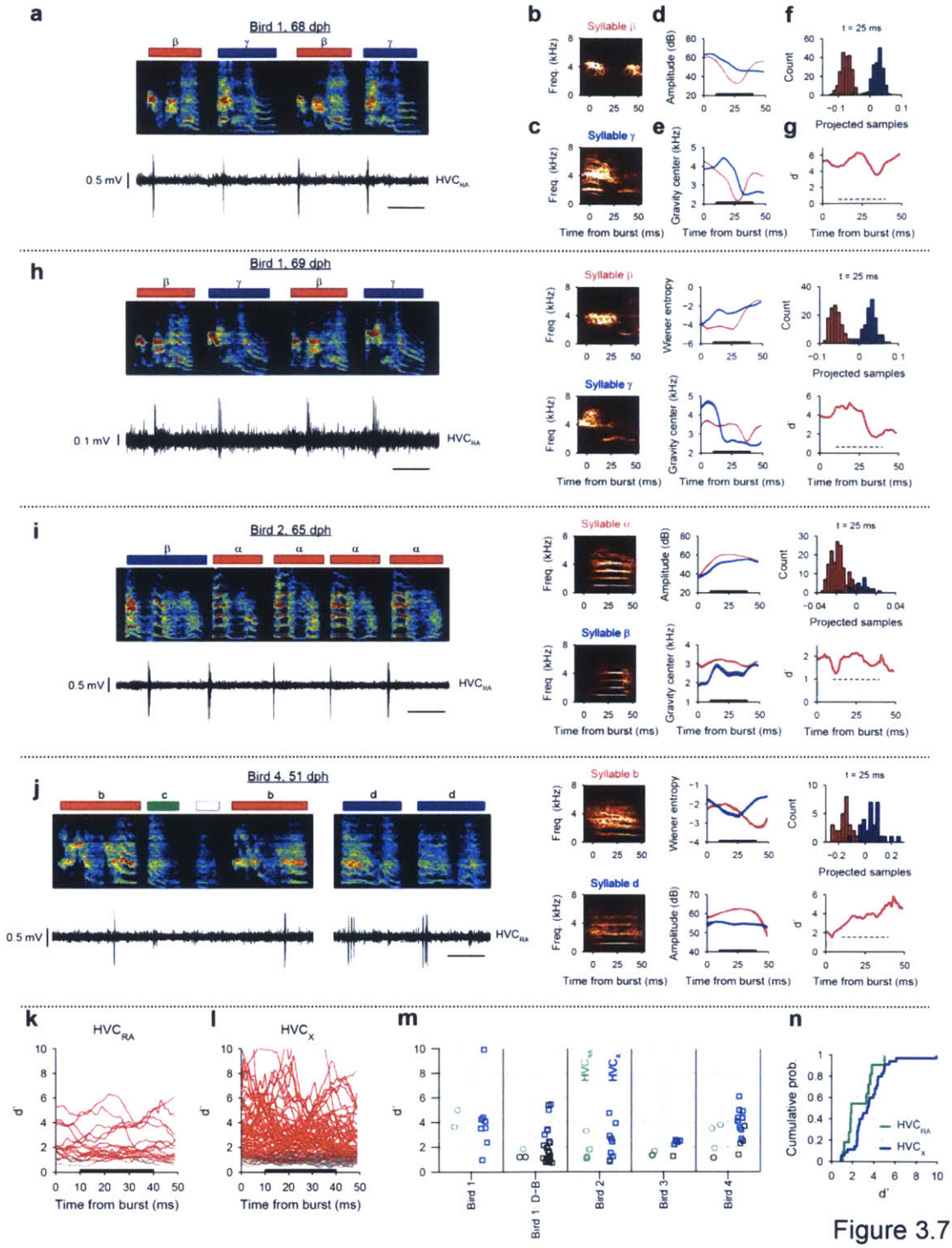

a
Bird 1, 68 dph
β
γ
β
γ
0.5 mV
HVC_RA
b
Syllable β
Freq (kHz)
c
Syllable γ
Freq (kHz)
Time from burst (ms)
d
Amplitude (dB)
e
Gravity center (kHz)
Time from burst (ms)
f
t = 25 ms
Count
Projected samples
g
d'
Time from burst (ms)
h
Bird 1, 69 dph
β
γ
β
γ
0.1 mV
HVC_RA
Syllable β
Freq (kHz)
Wiener entropy
t = 25 ms
Count
Projected samples
Syllable γ
Freq (kHz)
Gravity center (kHz)
d'
Time from burst (ms)
i
Bird 2, 65 dph
β
α
α
α
α
0.5 mV
HVC_RA
Syllable α
Freq (kHz)
Amplitude (dB)
t = 25 ms
Count
Projected samples
Syllable β
Freq (kHz)
Gravity center (kHz)
d'
Time from burst (ms)
j
Bird 4, 51 dph
b
c
b
d
d
0.5 mV
HVC_RA
Syllable b
Freq (kHz)
Wiener entropy
t = 25 ms
Count
Projected samples
Syllable d
Freq (kHz)
Amplitude (dB)
d'
Time from burst (ms)
k
HVC_RA
d'
Time from burst (ms)
l
HVC_X
d'
Time from burst (ms)
m
d'
Bird 1
Bird 1 D-B
Bird 2
Bird 3
Bird 4
n
Cumulative prob
HVC_RA
HVC_X
d'

Figure 3.7

Figure 3.7: Analysis of the acoustic differences associated with shared neurons.

a, HVC_{RA} neuron shared between syllables β and γ (same neuron as Figure 3.2f). **b-c,** Average spectrogram (sparse contour representation) computed for β and γ, centered on a 50 ms window immediately after the burst. **d-e,** Song amplitude (panel d) and spectral center of gravity (also known as the mean frequency; panel e) as a function of time from burst for β (red) and γ (blue). Lines show average across all renditions on which the neuron was active. Shading around lines show s.e.m. (sometimes too small to be visible). **f,** Distribution of projected samples for β (red) and γ (blue) at t=25 ms. **g,** d-prime analysis of separability of projected samples for syllables β and γ. 95% confidence interval of the shuffled d' trajectories are shown (gray band at bottom). Also shown (dashed horizontal line) is the 95 percentile of the distribution of peak values of d' in the surrogate data. **h-j,** Acoustic analysis for three additional HVC_{RA} neurons. Panels are analogous to **a-g**. **k,** Plot of d' trajectories for all shared HVC_{RA} neurons. Significant d' values are shown in red. Non-significant values are shown as gray line. **l,** Same as (k) but for shared HVC_X neurons. **m,** Population summary of mean d' (averaged over 10-40 ms after burst time). Each symbol represents a different shared neuron and each column indicates a different bird. Analysis is shown separately for each neuron type: for each bird, HVC_{RA} neurons are shown at left (green circles) and HVC_X neurons are shown at right (blue squares). Neurons with no significant acoustic differences are indicated with black symbols. **n,** Cumulative distribution of mean d' for shared HVC_{RA} (green; n=11 neurons) and shared HVC_X neurons (blue; n=36 neurons). Only significant d' metric are included in the cumulative. No significant difference in the distribution of d' metric was observed between HVC_{RA} and HVC_X neurons (P=0.1).

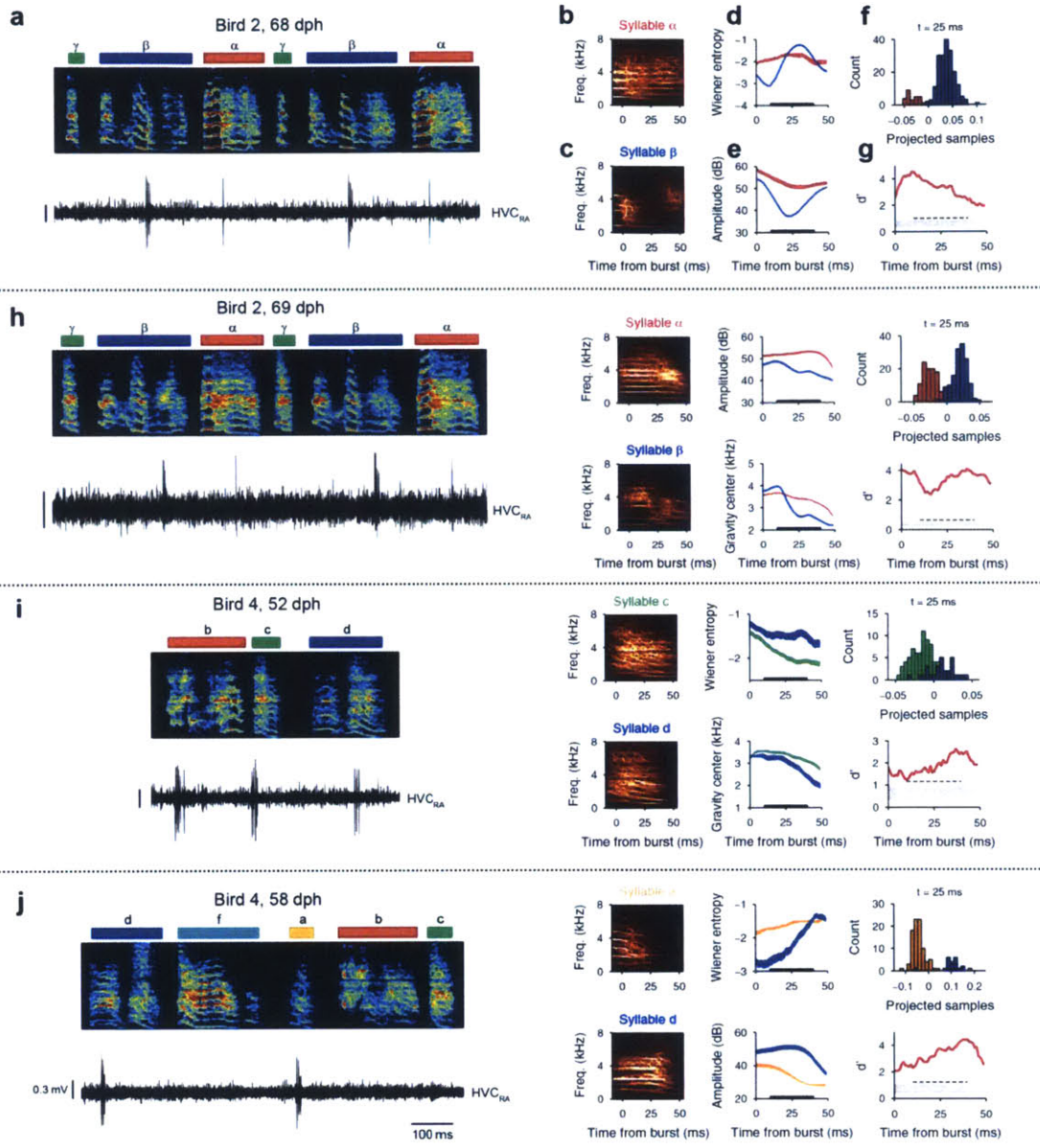

a
Bird 2, 68 dph
γ β α γ β α
HVC$_{RA}$
b
Syllable α
Freq. (kHz)
c
Syllable β
Freq. (kHz)
Time from burst (ms)
d
Wiener entropy
e
Amplitude (dB)
Time from burst (ms)
f
t = 25 ms
Count
Projected samples
g
d'
Time from burst (ms)
h
Bird 2, 69 dph
γ β α γ β α
HVC$_{RA}$
Syllable α
Freq. (kHz)
Amplitude (dB)
Count
Projected samples
Syllable β
Freq. (kHz)
Gravity center (kHz)
d'
Time from burst (ms)
i
Bird 4, 52 dph
b c d
HVC$_{RA}$
Syllable c
Freq. (kHz)
Wiener entropy
Count
Projected samples
Syllable d
Freq. (kHz)
Gravity center (kHz)
d'
Time from burst (ms)
j
Bird 4, 58 dph
d f a b c
0.3 mV
HVC$_{RA}$
100 ms
Freq. (kHz)
Wiener entropy
Count
Projected samples
Syllable d
Freq. (kHz)
Amplitude (dB)
d'
Time from burst (ms)

Figure 3.8

Figure 3.8: Additional examples of the acoustic differences associated with shared HVC_{RA} neurons.

Additional 4 examples that compellingly demonstrate acoustic differences associated with bursts of shared HVC_{RA} neurons (same format as Figure 3.7). **a-g,** Example recording of an HVC_{RA} neuron active during both syllables α and β (Bird 2), along with analysis of acoustic differences in the 50 ms window following the burst time. **h,** Example of a shared HVC_{RA} neuron from the same bird (Bird 2). Note that this neuron participated in the shared sequence shown in Figure 4.2g. **i,** A shared HVC_{RA} neuron in Bird 4. This neuron was active on syllables 'b', 'c', 'd', but we focus here on the acoustic differences between syllables 'c', and 'd' since this neuron was part of the shared sequence for those two syllables (Figure 4.8n). **j,** An HVC_{RA} neuron shared between syllables 'a' and 'd' (Bird 4). Note that in all of these four example neurons, d' trajectory is above the significance threshold (see Methods) for the entire 10-40 ms premotor latency.

Chapter 4

Sequence splitting underlying different learning strategies

Analysis of the coverage and clustering of HVC bursts was done in collaboration with Galen F. Lynch (Department of Brain and Cognitive Sciences, MIT).

In Chapter 3, I have shown that sequence splitting in HVC underlies the formation of new syllables during alternating syllable differentiation, i.e. the development of two new syllables by alternating increasingly different variants of the protosyllable. This is reminiscent of 'sound differentiation in situ' (Tchernichovski et al., 2001) combined with the 'serial repetition' learning strategy (Liu et al., 2004). However, previous research showed that juvenile birds can use several different learning strategies including syllable differentiation at bout edges (Lipkind and Tchernichovski, 2011; Lipkind et al., 2013) and learning the entire motif simultaneously (Liu et al., 2004). What is the neural mechanism underlying these different strategies of song learning? Is splitting of sequence also involved or is there an alternative neural mechanism?

In this chapter, I examined these questions by recording neural activity in HVC during these different learning strategies. As a result, I found that key features of sequence splitting model, such as shared sequence and its decrease over development, was also found during these different strategies. This suggests that there is a common neural mechanism underlying diverse learning strategies employed by a juvenile zebra finch.

Results

Behavioral analysis of bout-onset differentiation

Behavioral studies have shown that in some birds, new syllables preferentially occur at bout onset or offset (Lipkind and Tchernichovski, 2011; Lipkind et al., 2013). Does the formation of new syllables at bout onset also involve the splitting of a neural sequence as described in Chapter 3? Data related to this question were obtained in three

birds in which sufficient numbers of neurons were recorded during bout-onset syllable formation. One bird in which projection neurons were recorded throughout song development is illustrative (Bird 2; 57-84 dph). During learning, this bird acquired a three-syllable motif ('CBA'; Figure 4.1a; see Methods). Tracking the syllable structure through development revealed that syllables A and B derived from a common protosyllable, and that syllable B emerged specifically at bout onsets (Figure 4.1b).

The emergence of syllable B followed a distinctive pattern. As in other birds, a rhythmically repeated protosyllable (labeled α) emerged early on, by 60 dph (Figure 4.1a). By 63 dph, each string of repeated protosyllables was consistently preceded by a brief sound (ε) at bout onsets (Figure 4.1a, c). By day 65 dph, the silent gap between ε and the first α in the bout was eliminated, giving rise to a new syllable β (Figure 4.1a-c). Because β was formed by the concatenation of ε and α, the acoustic features of α and the latter part of β were initially highly overlapping (Figure 4.1e, f). Over the next 6 days (65-70 dph), α and β diverged acoustically (differentiated) to become two distinct syllables (Figure 4.1e, f). The changes were sufficiently continuous that α and β could be tracked into adulthood and identified with adult syllables A and B (Figure 4.1e, f).

Splitting of a neural sequence during bout-onset differentiation

To examine the neural mechanisms underlying the emergence of the new syllable β at bout onsets, we analyzed the firing patterns of 125 HVC projections neurons recorded in this bird (Bird 2; Figure 4.2). Early in song development (57-64 dph), prior to the concatenation of ε with α, we recorded 35 projection neurons. The majority of the neurons were rhythmic (n=28) and participated in the protosyllable sequence (Figure

2.9a-c); another population of neurons (n=4) were active at bout onsets, bursting immediately before the brief element ε (Figure 4.2a). Later, after the emergence of β (by the concatenation of ε and α), we recorded 51 projection neurons over a one-week period (65-72 dph), of which 43 were active during α or β. Among these, 21 were active during either α or β, but not both (α-specific: 7; β-specific: 14; Figure 4.2g). We also observed a population of shared neurons active during both α and β (22 neurons; Figure 4.2b, c, g). Notably, the shared neurons burst at nearly the same times with respect to syllable offsets during α and β (Figure 3.5b; correlation of burst times>0.99), and thus, as a population, produced nearly identical sequences during these two syllables (Figure 4.2g). These findings show that α and β initially share a highly overlapping neural representation, and support the view that β is initially formed as a concatenation of the motor program for α with that of the bout-onset element ε.

What are the neural representations of adult syllables A and B, and how do they emerge from the initially overlapping representations of α and β? To address this question, we also recorded 39 projection neurons later in song development (73-84 dph), 32 of which were active during A or B. Compared to the earlier stage (65-72 dph), we observed a much larger fraction of syllable-specific neurons (Figure 4.2h; α-specific: 14; β-specific: 14), and a correspondingly smaller fraction of shared neurons (4/32 shared neurons; $P=5\times10^{-4}$). These results suggest that the emergence of syllables at bout onset may also be implemented by the gradual splitting of a protosequence into increasingly non-overlapping 'daughter' sequences. Additional examples of shared neural sequence during bout-onset differentiation were observed in two other birds (Figure 4.3, Bird 5; Figure 4.5, Bird 1).

Insertion of an additional vocal element at the end of a protosyllable

The example just described included adding a vocal element ε at the beginning of a protosyllable. We have also observed a case where an additional vocal element was added at the end of a protosyllable (Figure 4.4; Bird 12). In this bird, after subsong, the song started having rhythmic repetitions within long syllables. We call this repeating unit protosyllable α, even though for most of the time, these protosyllables were not separated by gaps (Figure 4.4a). Protosyllable α was associated with rhythmic bursts in HVC (Figure 4.4c). Around the same time, a vocal element that consists of a harmonic stack, which we call β, was inserted within the rhythmic repetition of α (Figure 4.4a, c). When β initially appeared, it appeared within syllables (e.g. Figure 4.4c, first bout), but after a few days, the end of β was reliably associated with a gap (Figure 4.4a, 61 dph). Simultaneous recording of HVC projection neurons at this stage revealed rhythmic bursting of a neuron associated with α and a single burst of a different neuron associated with β (Figure 4.4c). At around 70 dph, the bird started repeating two repetitions of α reliably; we refer to the first repetition α as α_1 and second repetition of α as α_2, and these were followed by a β. This three-part syllable α_1-α_2-β differentiated into an adult syllable A (Figure 4.4b). Consistent with the hypothesis that α_1 and α_2 are two vocal elements that differentiated from α, there were shared neurons between α_1 and α_2 (Figure 4.4d, h). There were also neurons specific for α_1 or α_2 (Figure 4.4e, f, h), as well as neurons specific for β (Figure 4.4g, h). Thus, this is an interesting case of constructing a long adult syllable by adding a vocal element at the end of a repetition of protosyllables.

Sequence splitting during hierarchical differentiation

Example #1 (Figure 4.5; Bird 1)

So far, we have examined cases where a single protosyllable differentiates into two new syllable types. Can the sequence splitting be used iteratively to generate more sylalble types? We will begin by examining one bird (Bird 1; Figure 4.5) in which bout-onset differentiation occurred after the earlier differentiation of the protosyllable into daughter syllables β and γ (as shown in Figure 3.2). While syllable γ gradually transformed into adult syllable C, syllable β differentiated again to form two adult syllables, B and D (Figure 4.5a-d). At the neural level, syllables B and D shared a highly overlapping neural sequence, consistent with the idea that these two syllables originated from the same precursor sequence (Figure 4.5f, g). These observations further suggest that new syllables may emerge through a hierarchical process—that is, by the splitting of sequences that are themselves the daughter sequences of an earlier splitting process.

Example #2 (Figure 4.6; Bird 6)

Another example of hiearachical differentiation was observed in a bird that developed a six-syllable motif from subsong (Figure 4.6a). Tracking of the song over 20 days suggested that a protosyllable α separated into β and γ, and that one of these syllables γ further split into C and D, thus demonstraing hierarchical differentiation (Figure 4.6b). More specifically, the bird started repeating protosyllable α on 50 dph, and we hypothesize that the bird started inserting inspirations in two places. First, inspirations that were inserted between 'α's, which we call ε, started to vocalize; this vocal element ε had most of the acoustic power concentrated in high-frequency band (4-6 kHz), and is

consistent with the hypothesis that ε is produced by an inspiratory pressure pulse (Goller and Daley, 2001; Veit et al., 2011). Second, inspirations were sometimes inserted within protosyllable α leading to the separation of α into β and γ (Figure 4.6a 52 dph, d). Next, the acoustics of syllables α, β, γ, and ε were refined without changing the syllable order, demonstrating sound differentiation 'in situ' (Tchernichovski et al., 2001). Finally, γ differentiated into syllables C and D (Figure 4.6b, d), while α became A of adults. In summary, we speculate that adult sylablle A, C, and D all originated from the protosyllable α (Figure 4.6b). Indeed, acoustic trajectory for syllables A, C, and D were initially similar, but rapidly diverged over the course of 4 days (Figure 4.6c).

Few neurons that were recorded in this bird were consistent with the hypothesis presented above; there were neurons that were shared between C and D (Figure 4.6e, h), C and A (Figure 4.6f, h), and D and A (Figure 4.6g, h). Thus, we speculate that A, C, and D all differentiated from the protosylalble α thorough a hierarchical differentiation (Figure 4.6b).

Sequence splitting during motif strategy

Example #1 (Figure 4.7; Bird 3)

Yet another strategy for the acquisition of new syllables is the 'motif strategy', in which multiple syllables emerge nearly simultaneously to form an entire sequence of syllables (Liu et al., 2004). Does this strategy also involve the splitting of neural sequences? We were able to examine this question in two birds in which syllable types emerged from a rhythmic 'protosyllable' modulation in song amplitude to form most or all of a motif.

In Bird 3, at 45 dph, the song began to acquire rhythmic modulation in song amplitude around 9 Hz (Figure 4.7a, c), and recordings at this stage (45-46 dph) revealed projection neurons that burst rhythmically and were locked to the song rhythm (Figure 4.7e; 2/7 neurons). Over the next five days (47-51 dph), this bird acquired a reliable pattern of 4-5 acoustically distinct elements ('syllables'), each generated in a different cycle of the 9 Hz rhythm (Figure 4.7a, b; see Methods for segmentation and labeling procedure). During this period (47-51 dph), the acoustic structure in each syllable was gradually refined, resulting in an excellent match to the tutor song even at this early age (Figure 4.7a).

Electrophysiological recordings reveal that all syllables of the early song motif share a partially overlapping motor program. Of the 13 HVC projection neurons recorded in this stage (48-51 dph), 11 were significantly locked to syllables: two were specific for a single syllable type, and nine were shared, generating bursts during two or more syllable types (Figure 4.7f-h). Furthermore, each syllable type in the emerging motif had at least one neuron in this data set that was shared with another syllable. For every pair of syllables in which shared neurons were active, bursts occurred at a similar phase of the rhythm (Figure 4.7i, j). These findings suggest that each syllable type is generated by a neural sequence partially overlapping with that of the other syllables—consistent with the view that all of these syllables emerged from the splitting of a common prototype neural sequence.

Example #2 (Figure 4.8; Bird 4)

We obtained another example of a sequence splitting during motif strategy in a different bird (Figure 4.8; Bird 4). Based on an analysis of acoustic signals and neural recordings, we have formulated a hypothesis for how the song of this bird developed, which we will describe step-by-step, from the formation of the protosyllable to the emergence of the complete motif. In this bird, we hypothesize that the fundamental protosyllable element corresponds to the prominent 10 Hz peak in the rhythm spectrum and the 70 ms peak in the duration distribution (Figure 4.8b). This view is further supported by the presence of neurons in the protosyllable stage that generate rhythmic bursts at 10 Hz (Figure 4.8e, f; 11/18 neurons were rhythmic, 5/11 rhythmic neurons exhibited periodicity at 10 Hz). An analysis of burst latencies relative to protosyllable rhythm revealed a sequence of bursts (Figure 4.8k; computed using phase segmentation at 10 Hz; see Methods).

In this bird, the rhythmic protosyllables differentiated nearly simultaneously, at an early age (52 dph, Figure 4.8a), into a complete sequence of distinct syllables that subsequently formed the adult song, suggesting this bird employed a 'motif strategy.' One complication of this simple view is that there may have been an early partial splitting of the short protosyllable α into two 'daughter' protosyllables α_1 and α_2, which alternated to produce the elements of the final motif. Two lines of evidence based on neural activity support this view: many neurons recorded at early stages (<50 dph) exhibited a prominent 5 Hz periodicity in their rhythmic bursting. (Figure 4.8f, h; 6/11 rhythmic neurons), rather than the expected 10 Hz period (Figure 4.8f, top). This observation led us to consider the possibility that the 100 ms neural sequence corresponding to the dominant 10 Hz protosyllable rhythm underwent a partial splitting during the protosyllable stage—

similar to the alternating differentiation described for Bird 1 (Figure 3.2; Figure 4.2). This would result in two distinct alternating protosyllable sequences α_1 and α_2 (Figure 4.8c). Such splitting would effectively double the period of the protosyllable rhythm, and could account for the 'doubled' protosyllables and the 5 Hz peak in the rhythm spectrum (Figure 4.8b).

The existence of short and doubled protosyllables led us to hypothesize that the short syllables of the adult motif ('a', 'c', and 'e') arose from the short protosyllables, while long adult syllables ('b' and 'd', and possible 'f') arose from doubled protosyllables (Figure 4.8c). Early syllable 'e' is later dropped by the juvenile, although it appears in the tutor song.

Furthermore, the analysis of shared sequences (Figure 4.8*l*-o) revealed a predominance of shared neurons between syllable elements in alternating cycles of the underlying 10 Hz rhythm. For example, shared neurons were observed between syllables 'a', 'b_2' and 'd_1' (Figure 4.8i for neuron shared between 'a' and 'd_1'; Figure 4.8*l* for neurons shared between 'b_2' and 'd_1'). Shared neurons were also observed between syllables 'b_1', 'c', and 'd_2' (Figure 4.8h for neuron shared between 'b_1', 'c', and 'd_2'; panel n for neurons shared between 'c' and 'd_2'). In contrast, many fewer shared neurons were observed between neighboring cycles of the underlying rhythm, although examples of this can be found (Figure 4.8j).

Repetition of two syllable types (Figure 4.9)

In the alternating differentiation discussed in Chapter 3, two emerging syllables were produced in alternation (e.g. β-γ-β-γ). However, we also observed that each of the

two emerging syllables can be repeated such as β-β-β-β and γ-γ-γ-γ (Figure 4.9; Bird 13). In this bird, the song was rhythmic at 7 Hz (Figure 4.9a, b), and over the course of 3 days (46-49 dph), two distinct syllable types β and γ appeared (Figure 4.9c) and both of these syllables were repeated multiple times (Figure 4.9a). Correspondingly, there were neurons that were shared between β and γ (3/19 neurons; Figure 4.9d, g), neurons that were specific for β (12/19 neurons; Figure 4.9e, g), and neurons that were specific for γ (4/19 neurons; Figure 4.9f, g). These results are consistent with the possibility that rhythmic neural sequence in HVC coding for β-β-β-β and γ-γ-γ-γ differentiated from a common rhythmic sequence at an earlier stage.

Coverage of syllables and motifs by HVC bursts

Finally, we analyzed our dataset to test two models of HVC coding. As described in Chapter 1, there are currently two views on how HVC encodes syllables. One model of HVC function ('continuous-time model') hypothesizes that HVC bursts drive premotor activity at each moment in time (Hahnloser et al., 2002; Fee et al., 2004; Leonardo and Fee, 2005), and control the temporal precision of learning (Fiete et al., 2004; Charlesworth et al., 2011; Fee and Goldberg, 2011; Ravbar et al., 2012). This model would imply a continuous, though not necessarily uniform, coverage of HVC bursts throughout song. Another model ('gesture model') hypothesizes that HVC bursts only occur at transition points between gestures, or gesture trajectory extrema (GTEs) (Amador et al., 2013).

While explicitly testing the prediction of the gesture model requires calculating GTEs, which could be quite difficult given the variability of juvenile songs, it was

possible to test the prediction of the continuous-time model: complete coverage of syllables by projection neuron bursts. We tested this prediction by analyzing the coverage and clustering of bursts of both syllable-aligned activity and motif-aligned activity.

Coverage of syllables by HVC bursts (Figure 4.10)

An analysis of HVC bursts in juvenile birds revealed nearly continuous coverage of bursts in most syllables (>90% coverage for 17 of 20 syllables; n=4 birds; analysis restricted to syllables with more than 10 associated bursts; includes protosyllable and multi-syllable stages; Figure 4.10a). In three syllables with coverage less than 90%, the lower coverage was consistent with undersampling due to a smaller number of recorded neurons in these birds (Figure 4.10b; P>0.05; see Methods).

We also tested the uniformity of HVC bursts by using three different measures: burst-density, burst-density variance, and Ripley's L metric (Figure 4.10c-*l*; see Methods). As a result, only one out of 17 syllables showed deviation from uniform distribution (Figure 4.10*l*). Note that even this syllable had a near-complete coverage (Figure 4.10b). Overall, given the very large number of HVC neurons in each hemisphere (Walton et al., 2012) ($>10^4$), our measurements are consistent with a continuous representation of time throughout song syllables.

Coverage of motifs by HVC bursts (Figure 4.11)

We also examined how HVC projection neuron bursts are aligned to motifs. Gesture model predicts that there will be gaps in the motif where there are no HVC bursts. On the other hand, continuous-time model predicts that HVC bursts provide

nearly complete coverage of the song. Data from previous studies (Kozhevnikov and Fee, 2007; Amador et al., 2013) were not large enough to discriminate between these two models.

Here we recorded more than 50 bursts from individual birds (Figure 4.11). Direct visualization of the location of HVC projection neuron burst revealed that these bursts are distributed throughout the motif. Detailed quantifications of these results which include the analysis of putative HVC interneurons are presented elsewhere (Lynch, Okubo, Hanuschkin, Hahnloser, and Fee; under review). In sum, analyses based on syllables and motifs are consistent the continuous-time model of HVC coding.

Discussion

This chapter examined whether sequence splitting presented in Chapter 3 also underlies other strategies of syllable formation. Summary of the findings and possible interpretations are presented for each of these strategies.

Bout-onset differentiation

Previous behavioral studies have described that new syllables can appear at bout onsets or bout offsets, a phenomenon also termed the 'edge effect' (Lipkind and Tchernichovski, 2011; Lipkind et al., 2013). We have observed the formation of syllables at bout onset in our data as well (Figure 4.1). In addition, we found that vocal elements initially produced at bout onsets can either become a syllable on its own (Figure 4.5a) or can merge with the first protosyllable to give rise to a longer syllable (Figure 4.1). Continuous tracking of syllables until adulthood allowed us to demonstrate that the

syllable that appeared at bout onset was eventually incorporated in to the adult motif (Figure 4.1f), thus ruling out the possibility that this syllable only appeared transiently during song development. Moreover, birds that exhibited bout-onset differentiation imitated their tutor song well (e.g. Figure 4.1a) suggesting that bout-onset differentiation is a functional mechanism for learning new syllable types during song imitation.

Neural recordings revealed that sequence splitting also underlies bout-onset differentiation. In fact, decrease in the fraction of shared neurons was more pronounced in the example shown in Figure 4.2g, h than the alternating differentiation (Figure 3.2j, k). This is because merging of short-bout onset clement and the protosyllable created a new syllable type that had a longer duration, allowing us to label the two syllable types purely based on duration (Figure 4.1d); this lead to identification of two distinct syllable types early on when their acoustic structures were still highly similar (Figure 4.1e, f)[4]. Correspondingly, neural sequences producing these two syllables were initially highly similar (Figure 4.2g). However, later in development, the sequence became increasing different as the acoustic structures of these syllables diverged (Figure 4.2h). These results suggest the possibility that neural sequences that code for bout-onset elements can facilitate the splitting of neural sequence coding for protosyllables. This would give rise to two distinct sequences coding for different syllable types. One particular way in which this might occur is presented in Chapter 5.

Insertion of an additional element at the end of a protosequence

Bird 12, shown in in Figure 4.4, demonstrates two phenomena that show how

[4] Note that in alternating differentiation presented in Chapter 3, the syllable types were similar in duration and had to be classified based on acoustic differences. Thus, it is possible that by the time two syllable types can be reliably labeled, there are fewer shared neurons.

long syllables can be learned. The first phenomenon is that two back-to-back repetitions of a protosyllable without a gap, or 'doubling', can form a basis for forming long syllables in adults. The fact that these two repetitions of protosyllables α_1 and α_2 were acoustically similar (Figure 4.4a, d-h), together with the existence of neurons shared between them (Figure 4.4d, h), suggest that both α_1 and α_2 differentiated from an earlier protosyllable (α) (Figure 4.4b). This is similar to a previously reported case where back-to-back repetition of a protosyllable differentiated into distinct adult syllable types (Tchernichovski et al., 2001).

The second phenomenon that this bird illustrates is that a vocal element can be inserted and merged at the end of a protosyllable to generate a new syllable type. Initially, vocal element β was inserted within syllables (e.g. Figure 4.4c, first bout). Over multiple days, β became increasingly followed by a long gap (Figure 4.4a, 71 dph). Thus, β became the last vocal element in the bout. Whether this is similar to the emergence of syllables at bout offsets (Lipkind et al., 2013) is an open question.

Hierarchical differentiation

Splitting of neural sequence in its basic form, as presented in Chapter 3, can produce two sequences from a protosequence, and thus two distinct syllable types form a single protosyllable. However, adult zebra finch songs are typically composed of 3-7 syllable types. What mechanisms might allow birds to generate more than two syllable types? Here, we elucidated the neural mechanisms underlying several strategies for the emergence of multiple syllable types.

Hierarchical differentiation is a process in which a daughter sequence from first

sequence splitting further undergoes a second round of splitting (Figure 4.5, 4.6). This allows the bird to use the sequence splitting recursively giving rise to many distinct syllable types. Future research might able to bring this hierarchical differentiation under tighter experimental control by using a sequential tutor paradigm where the juvenile is first exposed to a tutor with a simple song, and then is exposed to a second tutor with a more complex song (Ravbar et al., 2012; Lipkind et al., 2013).

Note that the example presented in Figure 4.5 might be a case where the splitting of the neural sequence in HVC was incomplete. Consistent with this idea, in the later recordings, fraction of shared neurons between syllables D and B remained large (Figure 4.5g), compared to the smaller fraction of shared neurons syllables B and C of the same bird (Figure 3.2k). However, syllables B and C differentiated from parent syllable α much earlier in development (~60 dph, Figure 3.2b) than D and B differentiated from β (~80 dph, Figure 4.5c). One might speculate that the splitting of D and B may have failed to reach completion before the bird reached adulthood, possibly preventing further splitting.

Neural mechanisms underlying 'motif strategy'

Earlier behavioral study found that juvenile birds could use multiple strategies to learn the same song such as 'serial repetition' strategy and 'motif strategy' (Liu et al., 2004). We have also observed both strategies in our data, although the motif strategy was less common. This bias could be due to the fact that our birds were singly housed, whereas the original study that identified the motif strategy used a social setting (Liu et al., 2004).

Here we found that basic features described in Chapter 3 were also found in two birds that employed the motif strategy. These features include rhythmic neural sequence early on, and shared neurons with similar latencies (Figure 4.7i, j, 4.8*l*, n). Thus, it seems that there is a common neural mechanism between the motif strategy and other strategies. It would be interesting to determine the underlying neural mechanisms that give rise to these differences in learning strategy in different birds.

Emergence of two rhythmic repetitions of syllables (Figure 4.9)

We also observed an example where the bird developed two distinct syllable types β and γ, and repeated each of these syllables (Figure 4.9a) instead of alternating between them. Neural recording in this bird showed that there are shared and specific sequences underlying the song (Figure 4.9g), consistent with the sequence splitting. The smaller fraction of shared neurons in this bird, and the highly distinct acoustic structures between β and γ, suggest that perhaps this splitting happened at an early age. Note that the song of this bird consists of repetition of syllables (Figure 4.9a) and is not typical of zebra finches[5]. Thus, it remains to be seen whether this strategy is normally used in zebra finch song learning.

Time vs. gesture code in HVC

Our findings bear on several recent models of song representation in HVC. One model hypothesized that HVC bursts provide a continuous coverage of the song and act as a temporal bases for motor control (Hahnloser et al., 2002; Fee et al., 2004; Leonardo

[5] This song might be more similar to that of a Bengalese finch which contains repetitions of the same syllable type.

and Fee, 2005) and learning (Fiete et al., 2004; Charlesworth et al., 2011; Fee and Goldberg, 2011; Ravbar et al., 2012). Our analysis of the coverage of syllables by HVC bursts show that HVC provides nearly complete coverage in most of the cases (Figure 4.10b).

Another model of HVC coding has emphasized the finding that bursts may occur more often at particular times in the song, related to 'gestures' in the vocal control parameters (Amador et al., 2013). Our finding that bursts are more concentrated around syllable onsets early in vocal development (Figure 2.4i, j) suggests that HVC may generate protosyllables as primitive gestures that serve as a scaffold on which later song syllables develop (Tchernichovski et al., 2001). During development, HVC activity appears to evolve such that, as a population, bursts occur more uniformly throughout song syllables (Figure 2.4i-k), while the activity of individual neurons becomes sparser and more precise. At the same time, vocal gestures become more complex and precise as syllables develop into their adult forms. Thus, our data are consistent with the view that gestures are continuously controlled on a moment-by-moment basis by the sequence activity in HVC (Fee et al., 2004; Leonardo and Fee, 2005).

Methods

Syllable segmentation based on the song rhythmicity ('phase segmentation')

For Bird 3 ('motif strategy'), it was difficult to segment syllables consistently using previous methods based on setting a threshold on the sound amplitude (Aronov et al., 2008; Aronov et al., 2011; Veit et al., 2011). To overcome this limitation, we segmented syllables based on the phase of the rhythmicity in the song ('phase segmentation'). The peak of the song rhythm, defined as the spectrum of the sound amplitude during singing (Saar and Mitra, 2008), exhibited a peak around 9 Hz (Figure 4.7c). To estimate the instantaneous phase of this rhythm, we first band-pass filtered the sound amplitude between 6-12 Hz (Figure 4.7c, d). The band-pass filtered signal was then processed using the Hilbert transform (MATLAB command *hilbert*) to compute the instantaneous amplitude and phase (Figure 4.7d). Next, we set a threshold on this instantaneous amplitude to find the rhythmic part of the song. Finally, within this rhythmic part, song was segmented by detecting threshold crossings of the instantaneous phase (Figure 4.7d). Phase segments that contain no sounds or calls were manually removed. Similarly, phase segmentation (with band-pass filter 7-13 Hz) was used to segment the song during the protosyllable stage for Bird 4 (Figure 4.8a, e, f). Note that this method is best suited for segmenting songs that are rhythmic, but in which syllable boundaries are not strongly rhythmic. This is typical of birds employing the 'motif strategy' (Liu et al., 2004).

Analysis of neural activity

Quantification of the similar phases in shared neurons (Figure 4.7i, j)

For the bird whose song was segmented based on the phase of the rhythm (Bird 4, Figure 4.7), we asked whether shared neurons were active at similar phases. To quantify the phase of the neural activity, we first detected the burst times during singing, and for each burst, we assigned an instantaneous phase extracted from the song using the Hilbert transform (see the section on phase segmentation above). Then, mean phase of all the bursts produced during a particular syllable type was calculated (φ_i, **where** $i = 1,2,\ldots,5$ **indicates syllables**). Finally, top two syllable types that had the largest number of bursts were chosen and the difference between the mean phases for these two syllables ($|\Delta\varphi| = |\varphi_m - \varphi_n|$, **where** m **and** n **are indices of syllables**) was obtained (Figure 4.7i). We tested the significance of this value by comparing the value of $|\Delta\varphi|$ against that obtained from the shuffled data where the pairing of phases were randomized across all shared neurons (Figure 4.7j; 1,000 shuffles). P-values were obtained by analyzing the frequency with which $|\Delta\varphi|$ of surrogate data was smaller than that of the real data, and $P<0.05$ was considered significant.

Analysis of coverage by the HVC projection neuron bursts (Figure 4.10a, b)

We wondered whether projection neuron bursts effectively span the entire duration of juvenile song syllables, or whether bursts are highly localized to specific times, leaving other times in the syllable unrepresented (Amador et al., 2013). It is clear from the syllable aligned raster plots that some syllables were completely covered by bursts (Figure 3.2k, syllable 'C'), while other syllables showed some gaps in the burst

coverage (Figure 4.2h, syllable 'A'). To further quantify this aspect of the HVC representation during singing, we analyzed the fraction of time within the syllables of juvenile birds that were 'covered' by the recorded projection neurons bursts ('covered fraction').

We first determined the region of the song syllable covered by each HVC projection neuron burst. We generated a syllable-onset or offset locked histogram of spike times recorded from a single neuron over every rendition of the song syllable. Initial identification of candidate burst events was determined by smoothing the histogram (9 ms sliding square window, 1 ms steps), and setting a threshold to define a window in which to analyze burst spikes (2 Hz for protosyllable stage birds; 10 Hz threshold for older juveniles). To eliminate low-probability spike events, we only considered bursts for which spiking activity (at least one spike) occurred in the candidate burst window on at least 25% of the renditions for that syllable. Bursts were included only if they occurred between 30 ms prior to syllable onset and 10 ms after syllable offset.

For candidate bursts that met these criteria, all spikes occurring in the burst window were considered as contributing to that burst. Based on earlier measurements of postsynaptic currents and potentials of HVC and RA neurons (Garst-Orozco et al., 2014), each HVC spike in the burst window was conservatively assumed to exert a postsynaptic effect lasting no more than 5 ms. Thus, each spike in the dataset was replaced with a 5 ms postsynaptic square pulse (beginning at the spike time). We considered a region of the syllable to be 'covered' by this burst if at least three of these post-synaptic pulses overlapped at that time within the burst, across renditions of the syllable. This procedure

yielded a small 'patch' of time covered by the burst. The patches associated with each different neuron were combined with a logical 'OR' operation to determine the total coverage time of the syllable (again in a window from 30 ms prior to syllable onset to 10 ms after syllable offset). The covered time was divided by the duration of the syllable window to determine the covered fraction. Only syllables that had more than 10 neurons bursting within the syllable window were analyzed. This criterion excluded syllables from Bird 4 (shown in Figure 4.7), from which relatively few neurons were recorded.

While most syllables had nearly complete burst coverage (>90%), one syllable had coverage of only 73%, which could potentially be due to the relatively smaller number of neurons recorded in this bird. Thus, we asked whether the measured coverage is consistent with sparse sampling of the recorded bursts from a large number of uniformly placed bursts. To simulate this, we calculated the covered fraction for 1,000 surrogate datasets in which the 'covered patches' for each burst were randomly shuffled within the syllable. A random offset was added to the time of each patch, and a circular shift was used, allowing the patches to wrap around the edges of the syllable window. The distribution of covered fractions was determined over all shuffled surrogate datasets, and the 2.5-97.5 percentiles (95% confidence interval) of this distribution were determined (Figure 4.10b).

Analysis of the distribution of HVC projection neuron bursts (Figure 4.10c-l)

We wondered whether, as a population, the bursts of HVC projection neurons were distributed randomly over the duration of individual syllables, or if they exhibited a statistically significant tendency to cluster at a few times in the syllable. We addressed

this question by computing a metric of burst clustering in the observed data and comparing this metric with that expected if the bursts were randomly shuffled in time.

To do this, we first determined a list of times at which bursts occurred in each syllable across the population of neurons in the dataset (Figure 4.10c, f, i; bottom panels). The burst time for each neuron was extracted from a histogram of spikes aligned to syllable onsets or offsets, as described in Methods subsection entitled *Syllable-related neural activity*. The vast majority of neurons generated at most only a single burst in any one syllable. Further analysis was carried out only for syllables within which 10 or more neurons showed a significant burst (Figure 4.10b).

We computed the metric of burst clustering in two different ways. First we examined fluctuations in burst density in different time bins within each syllable (burst-density variance). Time-varying estimates of burst density for each syllable were computed by convolving the histogram of bursts (calculated in 1 ms bins) with a Gaussian (with s.d. of 4 ms; normalized to an area of one; Figure 4.10c, f, i; top panels). For some syllables, the burst density exhibited large peaks (Figure 4.10c), while for others it appeared smoother (Figure 4.10i). To determine if any of these peaks were larger than expected for random placement of burst times, we computed the distribution of peak values for 10,000 surrogate datasets in which burst times were randomly shuffled across the duration of the syllable. Peaks in burst density of observed syllables were considered significant if they exceeded P=0.05 of the surrogate data (Bonferroni correction for multiple comparisons).

We also assessed burst clustering by computing the burst-density variance, for each syllable, as the variance of values of burst density across the syllable. To determine

if the observed burst-density variance is significantly higher than expected for a random placement of burst times, we computed the burst-density variance for 10,000 surrogate datasets. Burst clustering was considered significant if the observed burst-density variance exceeded P=0.05 of the distribution of values for the randomized data (Bonferroni correction for multiple comparisons).

We also quantified burst clustering using a bin-free statistical measure of clustering called Ripley's-L (Ripley, 1981; Dixon, 2013). In brief, this is determined by counting the number of bursts within an integration time window τ of any other burst, and subtracting the count expected if the bursts were uniformly distributed. The result is typically expressed in units of time by normalizing by the average rate of the process. The resulting value is positive if bursts tend to be clustered on a time scale τ (Figure 4.10e), and negative if bursts tend to be dispersed on that timescale (Figure 4.10k). Some hypotheses of HVC function have emphasized the possibility that bursts may be clustered on a 5 ms timescale (Amador et al., 2013), thus we report the values of the Ripley's-L metric at τ=5 ms (Figure 4.10*l*). Significance was determined by comparing observed values of Ripley's-L to the distribution of values found in surrogate datasets with randomized burst times. The choice of τ did not change the number of syllables for which significant clustering was found (i.e. similar results were obtained at τ=10 ms).

Motif-aligned activity of HVC projection neuron bursts (Figure 4.11)

We visualize the population activity of HVC projection neuron bursts during motif production by plotting the smoothed firing rate of each bursts. To eliminate jitter in spike timing due to small variations in syllable durations, we used piecewise linear time-

warping (Leonardo and Fee, 2005) using syllable onsets and offsets as a reference. To obtain a smoothed firing rate for each burst, we first constructed a peri-event time histogram for each neuron by counting the number of spikes in 1 ms bins over all renditions of the song motif. This histogram was then smoothed by convolving it with a square kernel (width: 9 ms, step size: 1 ms). Candidate burst events were identified as periods when the smoothed firing rate exceeded a threshold of 10 Hz, and only spikes falling within the window between threshold crossings were considered. Candidate burst events were excluded if spikes occurred on 50% or fewer of the recorded motif renditions. Groups of spike meeting these criteria were classified as bursts. To visualize the location of bursts, we plotted motif-aligned smoothed firing rate of each bursts using a heat map (Figure 4.11). Note that multiple bursts from a single neuron are separated into different rows for visualization.

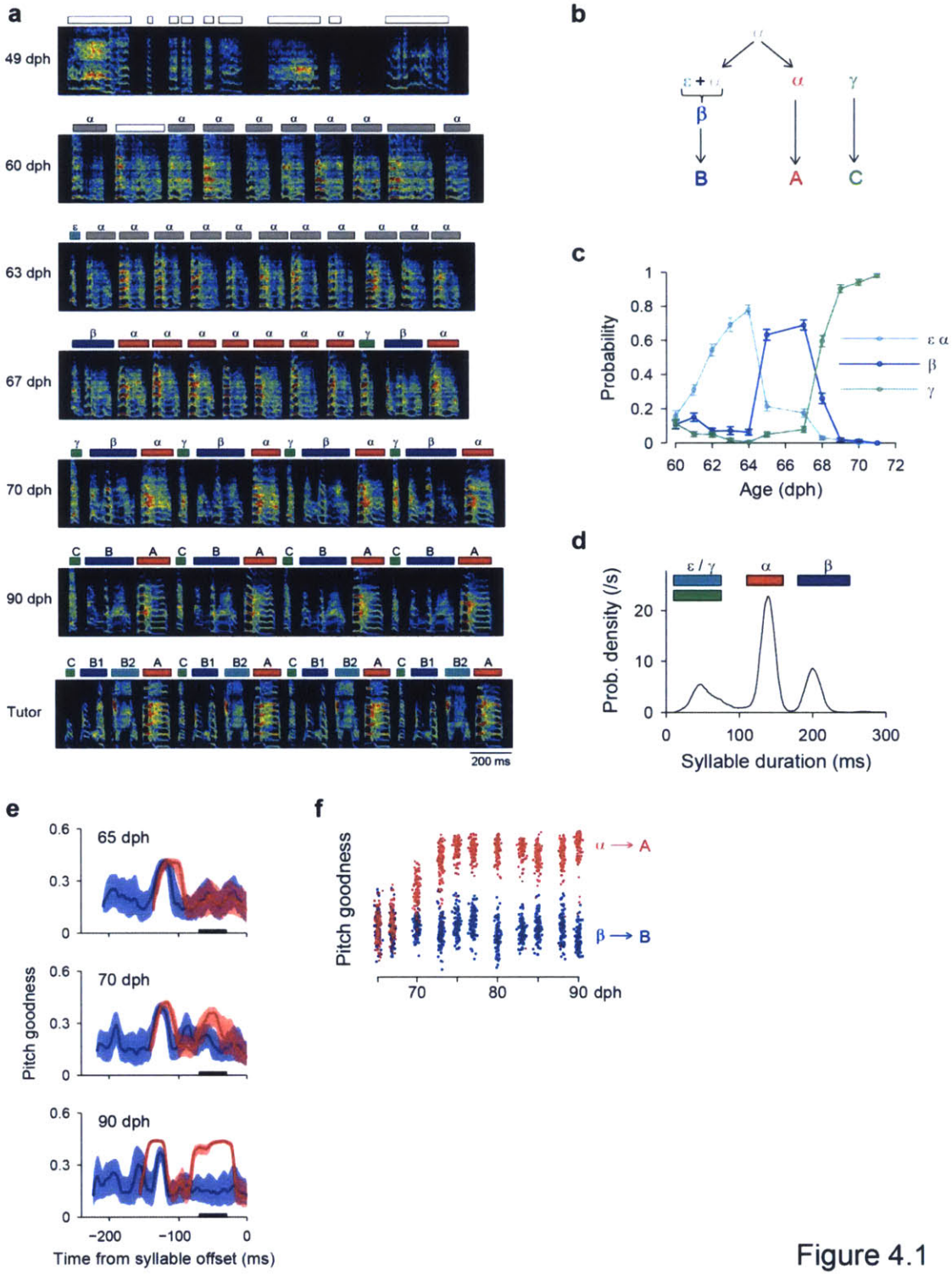

a
49 dph
60 dph
63 dph
67 dph
70 dph
90 dph
Tutor
200 ms
b
ε + α
α
γ
β
B
A
C
c
Probability
Age (dph)
ε α
β
γ
d
ε / γ
α
β
Prob. density (/s)
Syllable duration (ms)
e
65 dph
70 dph
90 dph
Pitch goodness
Time from syllable offset (ms)
f
Pitch goodness
α → A
β → B
dph

Figure 4.1

Figure 4.1: Song analysis of bout-onset differentiation.

All song and neural data are from Bird 2. **a,** Progression of song examples during the emergence of syllable β. Panels: i, subsong (49 dph); ii, emergence of protosyllable α from subsong (60 dph); iii, appearance of bout-onset element ε (63 dph); iv, fusion of bout-onset element with first protosyllable to form new syllable β (67 dph); v-vi, gradual differential acoustic refinement of β and α, and crystallization of β and α into a song motif with another bout-onset element γ to form the adult motif CBA (70, 90 dph); vii, tutor song for comparison. **b,** Schematic of the syllable formation. **c,** Developmental time course of the occurrence probability of different syllable types at bout onsets. Error bars indicate mean ± s.e.m. **d,** Syllable duration distribution showing three non-overlapping peaks (67 dph). Colored bars indicated syllable duration ranges used for syllable labeling. This separation of durations allowed automatic determination of syllable identity. **e,** Pitch goodness trajectory of syllables α (red) and β (blue) at three stages of vocal development, showing that these syllables initially share common acoustic structure (median ± quartiles; n=100 syllables of each type per day). Black bar indicates region used to compute data in (f). **f,** Scatter plot of mean pitch goodness showing differential refinement of syllables α (red) and β (blue) through development (n=100 syllables of each type per day).

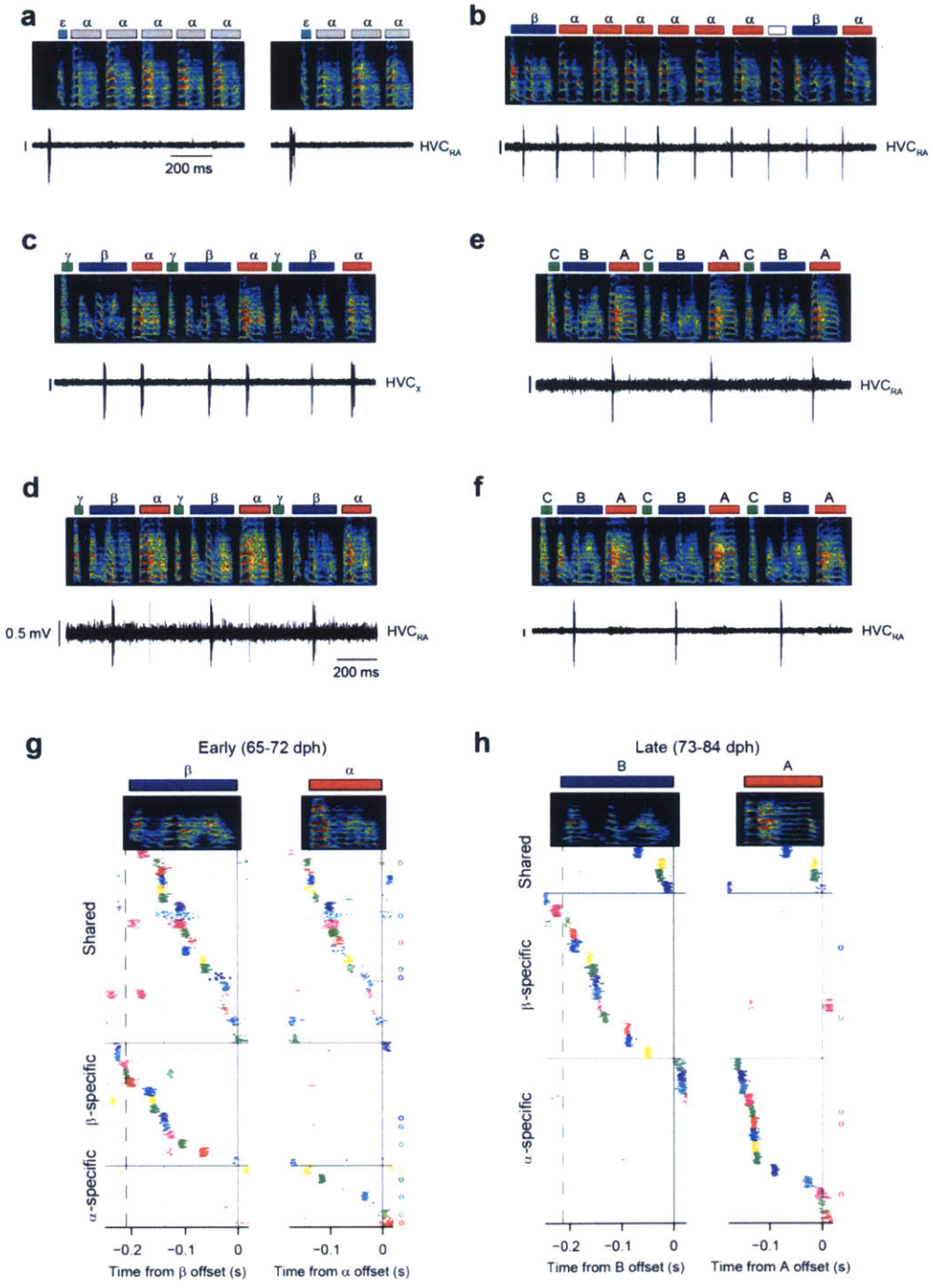

a
ε α α α α α
ε α α α
200 ms
HVC$_{RA}$
b
β α α α α α α β α
HVC$_{RA}$
c
γ β α γ β α γ β α
HVC$_{X}$
e
C B A C B A C B A
HVC$_{RA}$
d
γ β α γ β α γ β α
0.5 mV
HVC$_{RA}$
200 ms
f
C B A C B A C B A
HVC$_{RA}$
g
Early (65-72 dph)
β
α
Shared
β-specific
α-specific
−0.2 −0.1 0
Time from β offset (s)
−0.1 0
Time from α offset (s)
h
Late (73-84 dph)
B
A
Shared
β-specific
α-specific
−0.2 −0.1 0
Time from B offset (s)
−0.1 0
Time from A offset (s)

Figure 4.2

Figure 4.2: Shared and specific sequences during bout-onset differentiation.

All song and neural data are from Bird 2. **a,** Bout-onset neuron reliably active before bout-onset element ε (64 dph). **b,** Rhythmic bursts of a shared neuron active during both α and β (HVC_{RA}; 65 dph). **c,** Shared neuron active during both syllables α and β (HVC_X; 70 dph). **d,** Neuron active during both syllables α and β (HVC_{RA}; 69 dph). Activity during α was weak making this neuron β-specific according to our strict criteria. **e,** A–specific neuron (HVC_{RA}; 80 dph). **f,** B-specific neuron (HVC_{RA}; 73 dph). **g,** Population raster plot of 43 HVC projection neurons recorded early in the emergence of syllables β and α (65-71 dph). Shared neurons (top), β-specific neurons (middle), and α-specific neurons (bottom). **h,** Raster plot of 32 neurons recorded later in song learning (73-84 dph). Note the decrease in the fraction of shared neurons at this later stage. Dashed line indicates the mean onset time for β, and dotted line indicates the mean onset time for α. Scale bars (a-f): 0.5 mV, 200 ms.

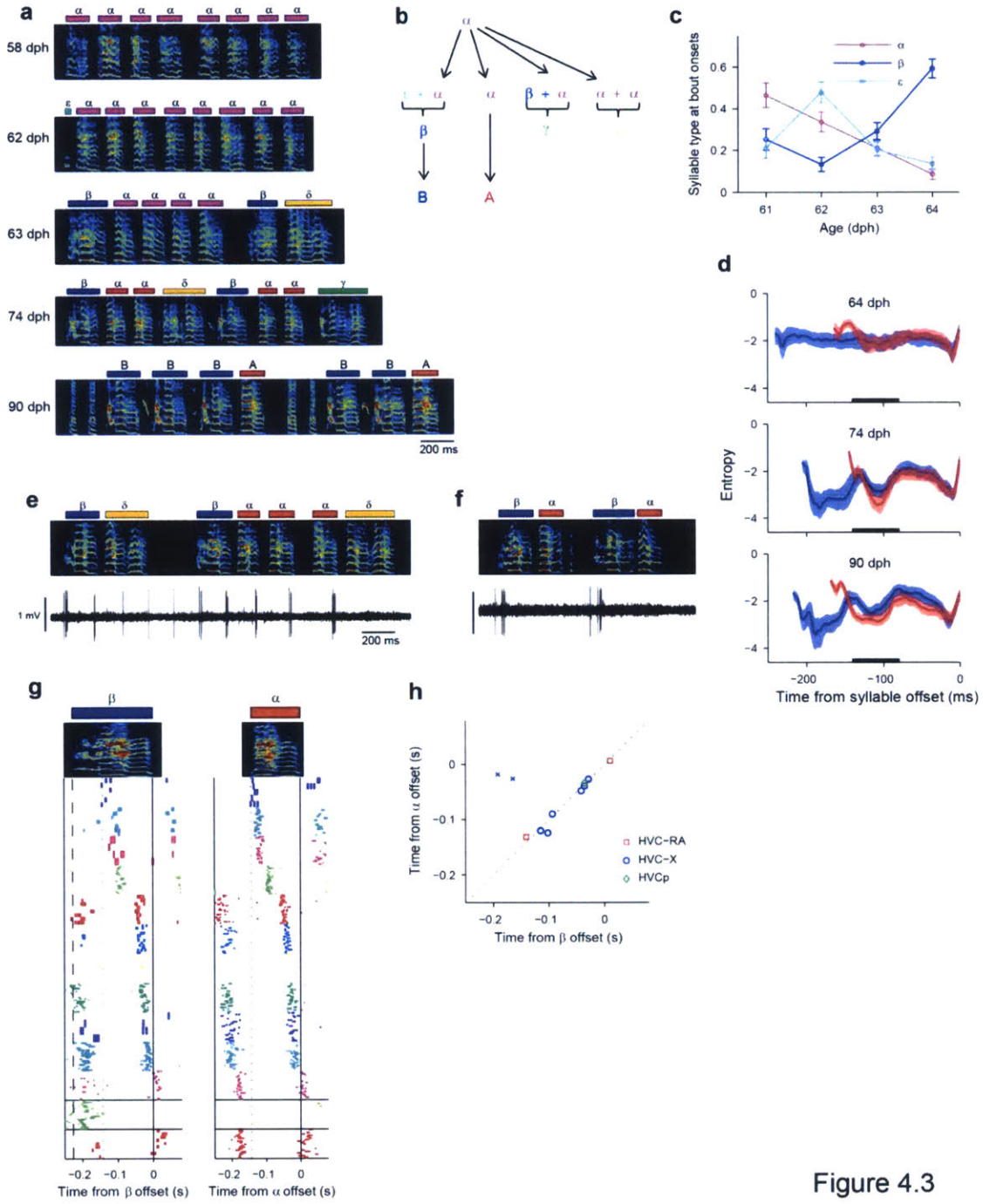

a
58 dph
62 dph
63 dph
74 dph
90 dph
200 ms
b
B
A
c
Syllable type at bout onsets
Age (dph)
d
64 dph
74 dph
90 dph
Entropy
Time from syllable offset (ms)
e
1 mV
200 ms
f
g
Time from β offset (s)
Time from α offset (s)
h
Time from α offset (s)
Time from β offset (s)
HVC-RA
HVC-X
HVCp

Figure 4.3

Figure 4.3: Another example of bout-onset differentiation.

All song and neural data are from Bird 5. **a,** Progression of song examples during the emergence of syllable β. Panels: i, rhythmic repetition of protosyllable α (58 dph); ii, appearance of bout-onset element ε (62 dph); iii, fusion of bout-onset element with first protosyllable to form new syllable β (63 dph); iv, gradual differential acoustic refinement of β and α (74 dph). γ (fusion between β and α), and δ (fusion between two α's) appeared during development but was not retained in the adult song; v, crystallization of β and α into adult syllables B and A, respectively (90 dph). **b,** Schematic of the syllable differentiation. **c,** Developmental time course of the occurrence of different syllable types at bout onsets. Error bars indicate s.e.m. **d,** Entropy trajectory of syllables α (red) and β (blue) at three stages of vocal development (median ± quartiles; n=80 syllables of each type per day). **e,** Rhythmic burst of a neuron shared between β and α (HVC_X, 63 dph). **f,** β-specific neuron (HVC_X, 63 dph). **g,** Population raster of 13 projection neurons aligned to β and α (63-66 dph). Dashed line indicates the mean onset time for β, and dotted line indicates the mean onset time for α. Shared neurons (top), a β-specific neuron (middle), an α-specific neuron (bottom) showing that shared neurons generated the same sequence between syllables β and α. Notice that there are many neurons shared between the first half of β (that originated from ε) and the second half of β (that originated from α), raising the possibility that ε and α might have shared a common origin. **h,** Latencies of the peak firing rate for all the shared neurons (n=11 neurons; correlation coefficient r = 0.98; P=0.001). Two shared neurons that had two peaks during the syllable β are shown in x and were removed from the calculation. However, even in these neurons, the smaller peak had a similar latency as that during α. Scale bars (e, f): 1 mV, 200 ms.

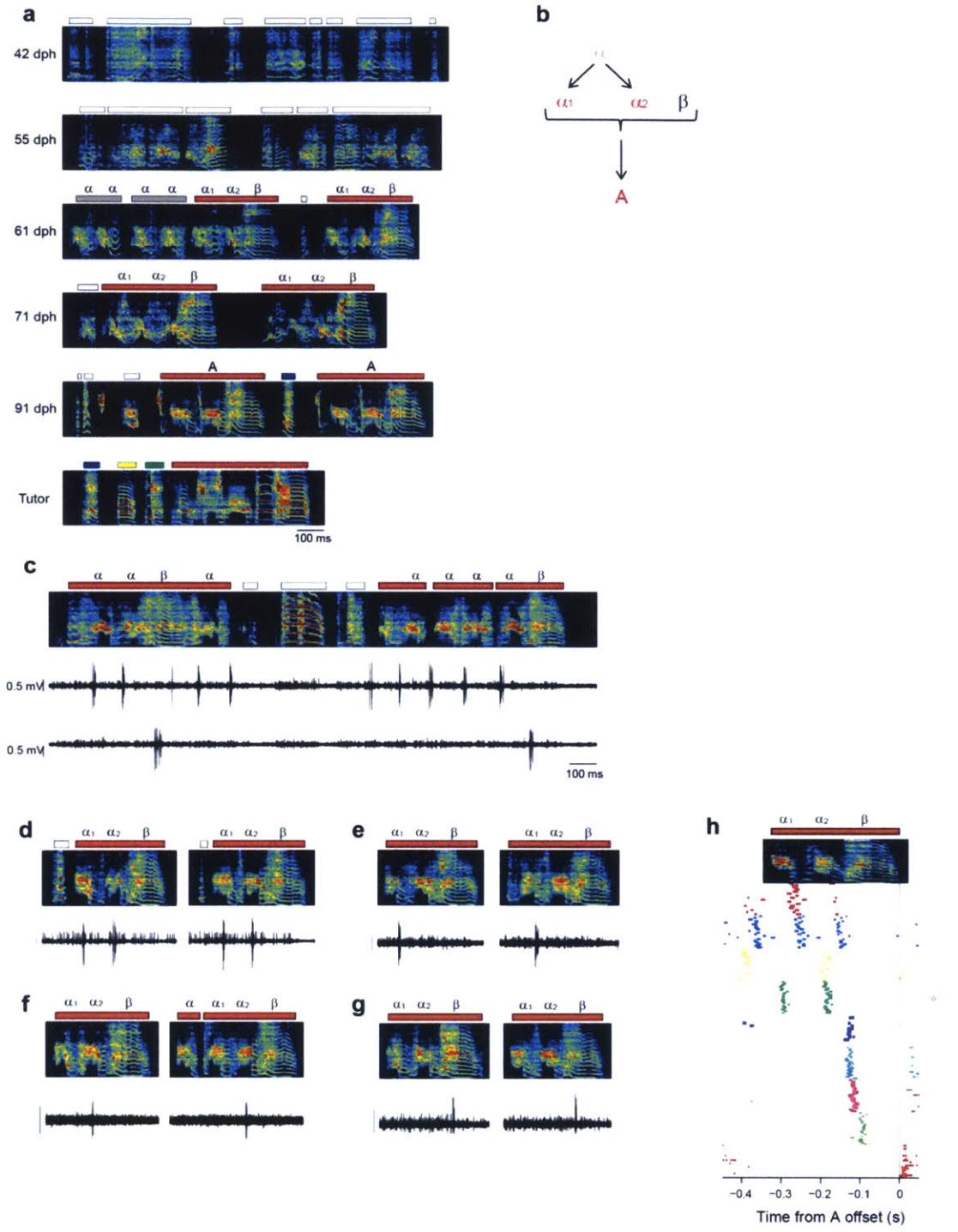

a
42 dph
55 dph
61 dph
71 dph
91 dph
Tutor
100 ms
b
α1
α2
β
A
c
0.5 mV
0.5 mV
100 ms
d
e
f
g
h
−0.4
−0.3
−0.2
−0.1
0
Time from A offset (s)

Figure 4.4

Figure 4.4: Insertion of a vocal element at the end of a protosyllable.

All song and neural data from Bird 12. **a,** Progression of song examples during the emergence of a long syllable A from protosyllable α and from an inserted syllable β. Panels: i, subsong (42 dph). ii, At 55 dph, the song of this bird began to acquire rhythmic 'protosyllable' modulations in song. iii, Harmonic stack syllable β were inserted in the song (61 dph). iv: Reliable repetition of α_1, α_2, and β to form a long syllable A (71 dph). v: Refinement of the acoustics of A (91 dph). vi: Tutor song which also has a long syllable that ends with harmonic stack. **b,** Schematic of syllable formation. Protosyllable α differentiated into two different variants α_1 and α_2. These, together with an inserted vocal element β, formed a long syllable A of adult song. **c,** Simultaneous recording. Top: Neuron that is rhythmically active during protosyllable α. Bottom: Neuron active for the inserted element β (HVC_{RA} and HVC_p, respectively; 57 dph). **d,** Shared neuron active during both α_1 and α_2 (HVC_{RA}; 70 dph). **e,** Neuron specific for α_1 (HVC_p; 70 dph). **f,** Neuron specific for α_2 (HVC_X; 69 dph). **g,** Neuron specific for β (HVC_{RA}; 68 dph). **h,** Population raster of 9 projection neurons aligned to syllable A (49-55 dph). 20 most active trials are shown for each neuron. Scale bars (c-g): 0.5 mV, 100 ms.

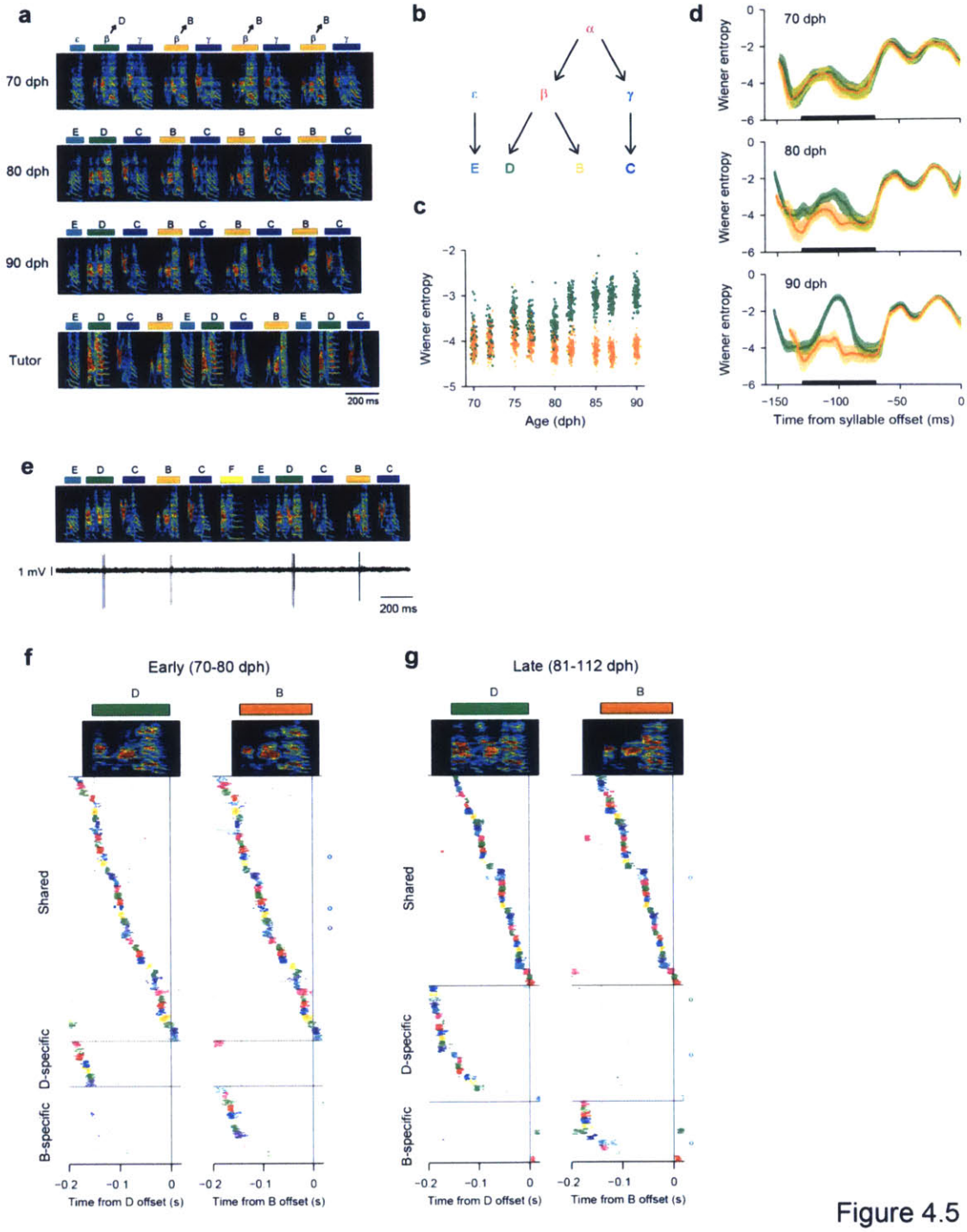

a
70 dph
80 dph
90 dph
Tutor
200 ms
b
α
ε
β
γ
E
D
B
C
c
Wiener entropy
Age (dph)
d
70 dph
80 dph
90 dph
Wiener entropy
Time from syllable offset (ms)
e
1 mV
200 ms
f
Early (70-80 dph)
Shared
D-specific
B-specific
Time from D offset (s)
Time from B offset (s)
g
Late (81-112 dph)
Shared
D-specific
B-specific
Time from D offset (s)
Time from B offset (s)

Figure 4.5

Figure 4.5: Hierarchical differentiation of syllables.

All song and neural data from Bird 1. **a,** Progression of song examples during the emergence of two syllables B and D from a common precursor syllable β, which had undergone earlier differentiated from a protosyllable α (Bird 1; same bird as Figure 3.2). Panels: i (70 dph), After the initial differentiation of the protosyllable into β and γ (at ~62 dph), the bird produced a rhythmic alternation of these two syllables, and the alternating sequence was reliably preceded at bout onsets by a short vocal element ε (ε-β-γ-β-γ-β-γ…). Note that the first repetition of β in each bout (labeled D) is acoustically identical to later repetitions (labeled B); panel ii (80 dph), the first repetition of β in the bout (syllable D) undergoes differential acoustic refinement compared to later repetitions (syllable B); iii, syllable B, C and D, together with bout-onset element ε, crystallize into adult motif EDCB (90 dph), that approximately matches the tutor motif (panel iv). **b,** Schematic of the progression of syllable emergence. **c,** Scatter plot of the mean Wiener entropy showing differential acoustic refinement of syllables B (orange) and D (green) through development (n=100 syllables of each type per day). **d,** Wiener entropy trajectory of syllables B and D at three stages of vocal development, showing that these syllables initially share common acoustic structure. Black bar indicates region used to compute data in c (median ± quartiles; n=100 syllables of each type per day). **e,** Rare example of a shared HVC_{RA} neuron in the adult bird (106 dph). **f,** Population raster plot of 60 HVC projection neurons recorded between 70-80 dph that were significantly locked to syllable D or B. **g,** Same as in **f,** but for 70 neurons recorded later in song learning (81-112 dph).

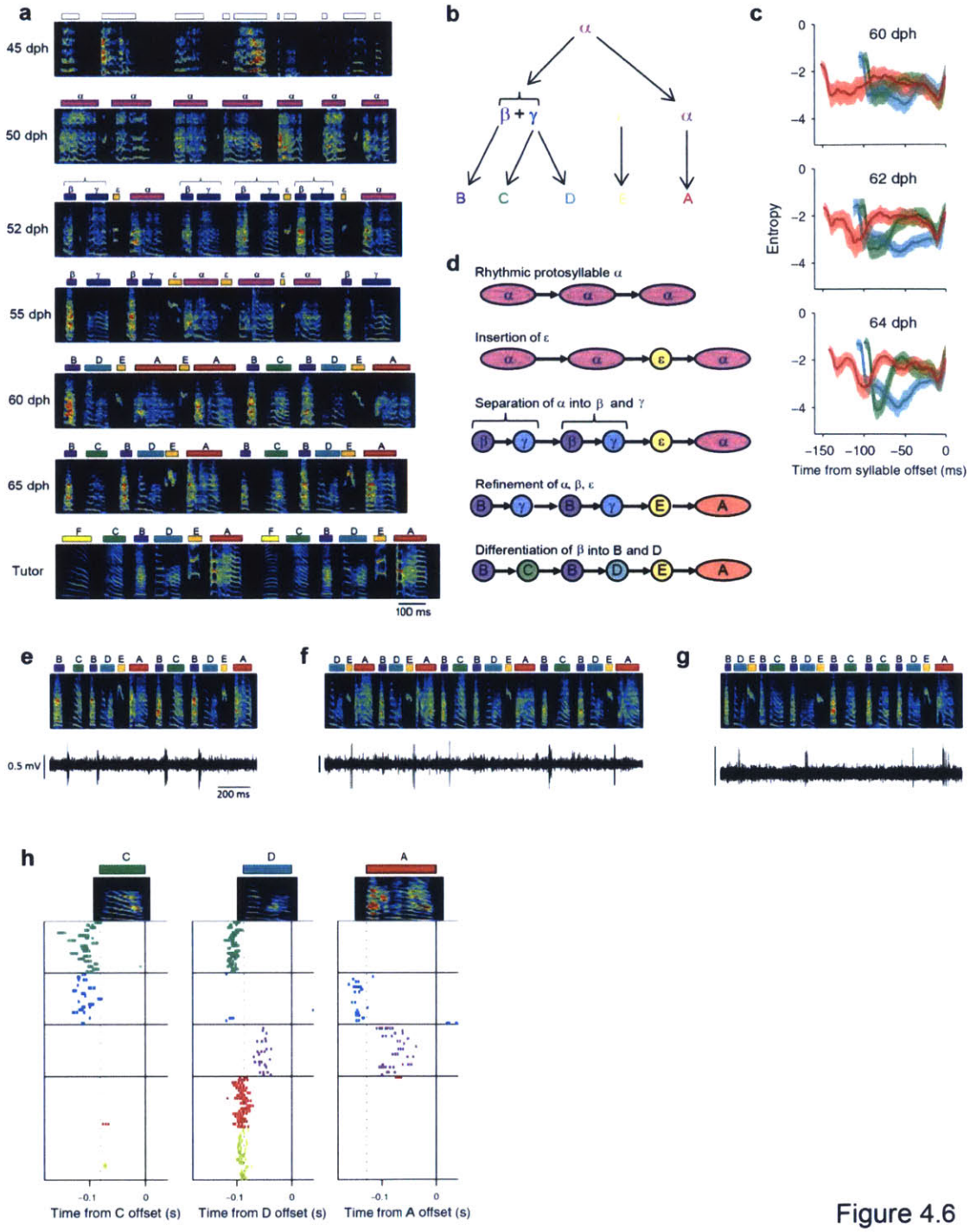

a
45 dph
50 dph
52 dph
55 dph
60 dph
65 dph
Tutor
100 ms
b
α
β + γ
α
B
C
D
E
A
c
60 dph
62 dph
64 dph
Entropy
0
−2
−4
−150 −100 −50 0
Time from syllable offset (ms)
d
Rhythmic protosyllable α
Insertion of ε
Separation of α into β and γ
Refinement of α, β, ε
Differentiation of β into B and D
e
0.5 mV
200 ms
f
g
h
C
D
A
-0.1 0
Time from C offset (s)
Time from D offset (s)
Time from A offset (s)

Figure 4.6

Figure 4.6: Another example of hierarchical differentiation.

All song and neural data from Bird 6. **a,** Progression of song examples during the emergence of a 6-sylalble motif with hierarchical differentiation. Panels: i subsong (45 dph); ii (50 dph), rhythmic repetition of the protosyllable α; iii (52 dph), insertion of putative inspiratory syllable ε, and separation of α into β and γ; iv (55 dph), acoustic differentiation of α, β and γ; v (60 dph), hierarchical differentiation of γ into C (followed by a B) and D (followed by an E); vi (65 dph), formation of a six-syllable motif 'BCBDEA'; vii, tutor song for comparison. **b,** Schematic of the progression of syllable emergence. **c,** Wiener entropy trajectory of syllables A (red), C (green) and D (cyan) at three stages of vocal development, showing that the acoustic structures of these syllables gradually diverged (median ± quartiles; n=50 syllables of each type per day). **d,** A hypothesis describing the emergence of a six-syllable motif. Rhythmic repetition of protosyllable α turned into a motif 'BCBDEA' through the insertion of putative inspiratory events, and through separation of α into β and γ, followed by a second differentiation of γ into C and D (panel b). **e,** Shared neuron between C and D (HVC_X; 64 dph). **f,** Shared neuron between C and A (HVC_p; 62 dph). **g,** Shared neuron between D and A (HVC_X; 60 dph). **h,** Population raster plot of 5 HVC projection neurons recorded between 60-64 dph that were significantly locked to syllable C, D, or A. From top to bottom: C-D shared neuron (panel e), C-A shared neuron (panel f), D-A shared neuron (panel g), and D specific neuron.

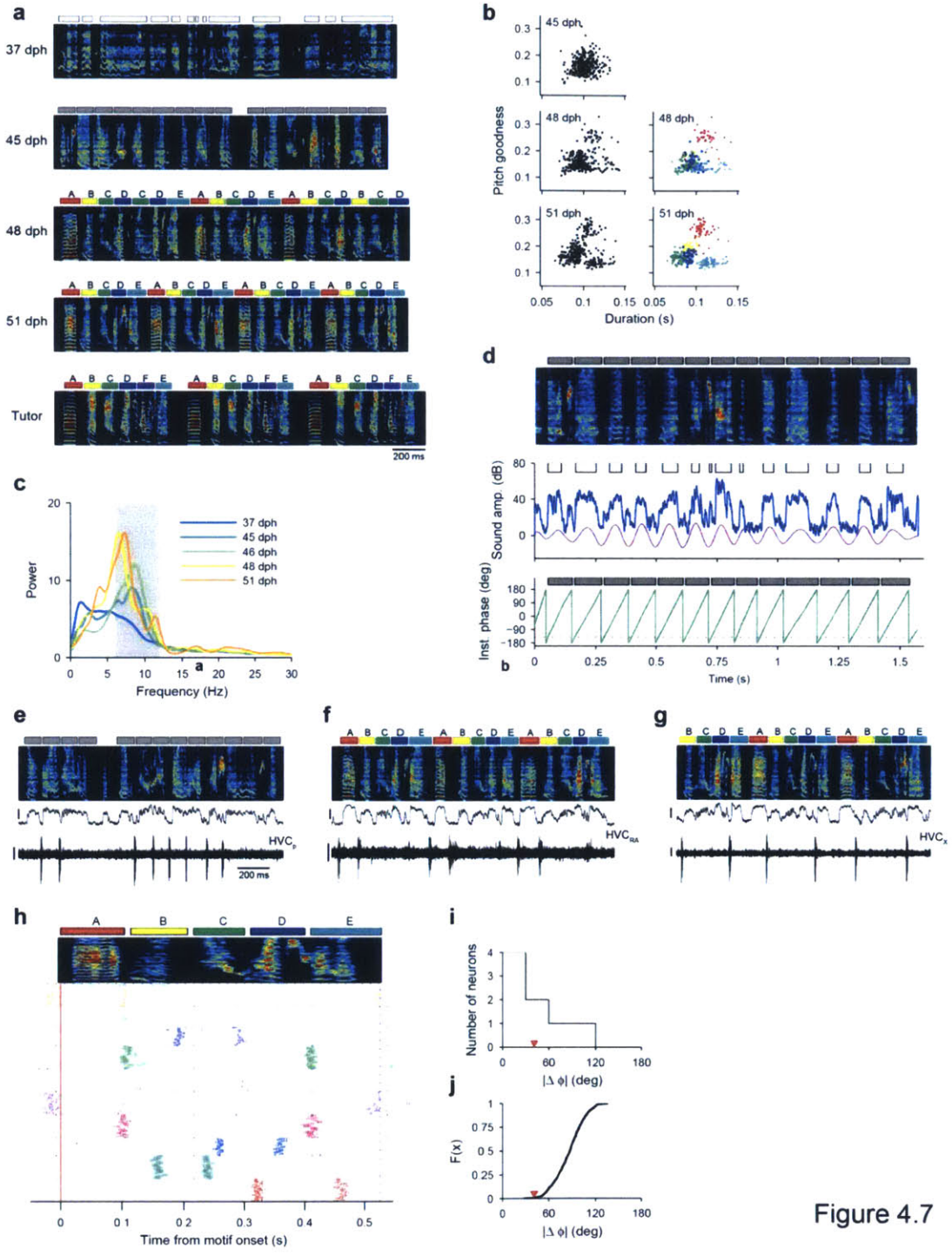

a
37 dph
45 dph
48 dph
51 dph
Tutor
200 ms
b
45 dph
48 dph
48 dph
51 dph
51 dph
Pitch goodness
Duration (s)
c
37 dph
45 dph
46 dph
48 dph
51 dph
Power
Frequency (Hz)
d
Sound amp (dB)
Inst. phase (deg)
Time (s)
e
HVC
200 ms
f
HVC
g
HVC
h
A
B
C
D
E
Time from motif onset (s)
i
Number of neurons
|Δ ϕ| (deg)
j
F(x)
|Δ ϕ| (deg)

Figure 4.7

Figure 4.7: Simultaneous formation of multiple syllable types into an entire motif.

All song and neural data from Bird 3. **a,** Song examples during motif strategy. Panels: i, subsong (37 dph); ii, At 45 dph, the song of this bird began to acquire rhythmic 'protosyllable' modulation in song amplitude around 9 Hz, iii, Over the next five days (47-51 dph), this bird acquired a reliable pattern of 4-5 acoustically distinct elements ('syllables'); iv, acoustic structure in each syllable was gradually refined, resulting in an excellent match to the tutor song (panel v) even at this early age (51 dph). **b,** Scatter plot of syllable duration and pitch goodness of each syllable (n=300 syllables in total per day). Syllables are color-coded for 48 and 51 dph (right). **c,** Song rhythmicity quantified as spectrum of the sound amplitude. Gray shade: pass band for the band-pass filtering used for the phase segmentation in (d). **d,** Phase segmentation (see Methods). Top: song spectrogram with phase segments in gray boxes. Middle: sound amplitude (blue) and band-pass filtered sound amplitude (magenta). Syllable segmentation based on sound amplitude is shown in white. Bottom: instantaneous phase (green) of the band-pass filtered sound amplitude. Phase segments (gray boxes) are obtained by detecting threshold crossing (black dotted line) of the instantaneous phase. **e,** Rhythmic neuron (45 dph). **f,** Neuron shared between A and B (48 dph). **g,** Neuron shared between B and E (49 dph). **h,** Population rasters aligned to the five-syllable motif for neurons that were significantly locked to syllables (n=10 neurons). **i,** Histogram of phase difference between the two syllables for the shared neurons (n=8 neurons; phase difference: 41 ± 33.9 deg, mean ± s.d.). **j,** Cumulative distribution of the phase difference after randomizing the pairing. Red dotted line: threshold for significance (P=0.05). Red triangle: location of the actual data (P=0.013). Scale bars: (e-g) 30 dB, 0.3 mV, 200 ms.

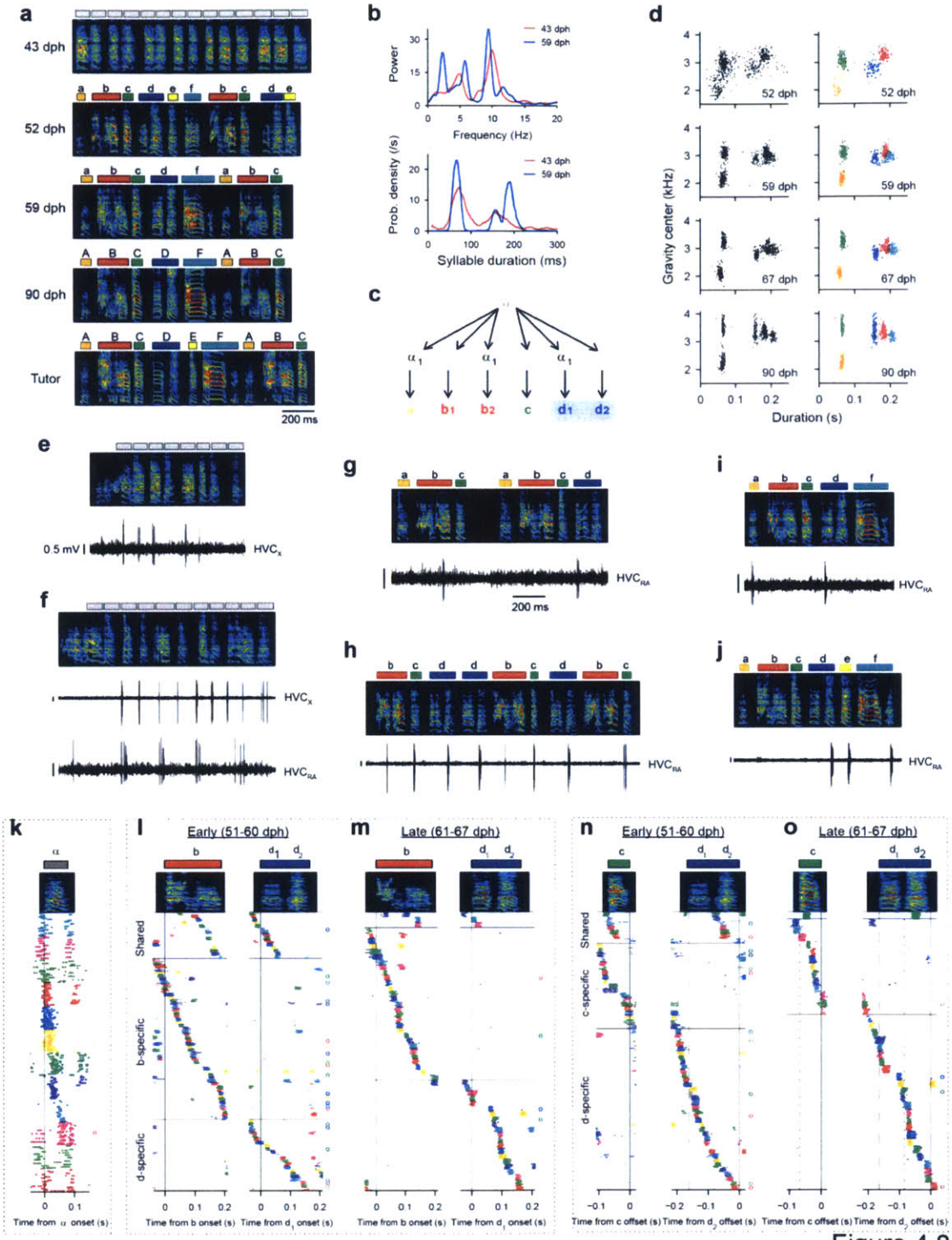

a
43 dph
52 dph
59 dph
90 dph
Tutor
200 ms
b
43 dph
59 dph
Power
Frequency (Hz)
Prob. density (/s)
Syllable duration (ms)
c
d
Gravity center (kHz)
Duration (s)
e
0.5 mV
HVC$_X$
f
HVC$_X$
HVC$_{RA}$
g
HVC$_{RA}$
200 ms
h
HVC$_{RA}$
i
HVC$_{RA}$
j
HVC$_{RA}$
k
Time from α onset (s)
l
Early (51-60 dph)
Shared
b-specific
d-specific
Time from b onset (s)
Time from d$_1$ onset (s)
m
Late (61-67 dph)
Time from b onset (s)
Time from d$_1$ onset (s)
n
Early (51-60 dph)
Shared
c-specific
d-specific
Time from c offset (s)
Time from d$_2$ offset (s)
o
Late (61-67 dph)
Time from c offset (s)
Time from d$_2$ offset (s)

Figure 4.8

Figure 4.8: Another example of shared burst sequences during the emergence of a motif. All song and neural data from Bird 4. The majority of syllables emerged nearly simultaneously in a relatively fixed order, consistent with a 'motif strategy.' **a,** Sample song spectrograms from four different ages during song development. Tutor song shown at the bottom. Phase segments are shown above the spectrogram for song on 43 dph. **b,** Top: Song rhythm spectrum calculated in the protosyllable stage (43 dph) and after motif formation (59 dph). Note the pronounced peaks at 5 Hz and 10 Hz in both stages. Bottom: Syllable duration distribution in the protosyllable stage (43 dph) and after motif formation (59 dph) showing two peaks. At 43 dph, the peak at 70 ms indicates short protosyllables corresponding to one cycle of the 10 Hz rhythm, and the peak at 140 ms indicates longer syllables formed by two protosyllables fused across two cycles of the 10 Hz rhythm (doubled protosyllables). **c,** Hypothesized mechanism of motif construction, based on the examination of acoustic structure and analysis of neural burst sequences (see below). **d,** Scatter plots of syllable duration versus mean spectral center of gravity at four stages of vocal development. Each dot represents a single syllable (n=500 syllables per day). Also shown are scatter plots with syllable labels indicated in color (right). **e-j,** Example recordings of HVC projection neurons during singing. **e,** Neuron bursting at 10 Hz rhythm (48 dph). Phase segments are shown. **f,** (Top) Neuron bursting at the 10 Hz protosyllable rhythm (49 dph). (Bottom) Simultaneous recording of a neuron bursting on alternate cycles of 10 Hz rhythm. Phase segments are shown. **g,** Shared neuron bursting on second half of syllable 'b' (labeled b_2) and first half of syllable 'd' (labeled d_1) (51 dph). **h,** Shared neuron bursting rhythmically on first half of 'b' (b_1), syllable 'c' and second half of 'd' (d_2) (51 dph). **i,** Shared neuron bursting on 'a' and first half of 'd' (d_1)

(58 dph). **j,** Shared neuron bursting on second half of 'd' (d_2), 'e' and last part of 'f' (57 dph). **k-o,** Spike raster plots showing bursts of HVC projection neurons aligned to syllable onsets or offsets. **k,** Raster plot of 12 HVC projection neurons that were significantly locked to protosyllable onsets (48-49 dph). Neural activity is aligned to protosyllable onsets (vertical solid line at 0 s), and syllables are time warped to the mean offset time of the protosyllable (dotted line). **l,** Raster plot showing neurons active in either syllable 'b' or syllable 'd'. This panel shows neurons recorded earlier (51-60 dph). **m,** Similar to panel *l*, but showing neurons recorded later in development (61-67 dph). **n,** Spike rasters showing neurons active during syllables 'c' and 'd'. This panel shows neurons recorded earlier (51-60 dph). **o,** Similar to panel m, but showing neurons recorded later in development (61-67 dph).

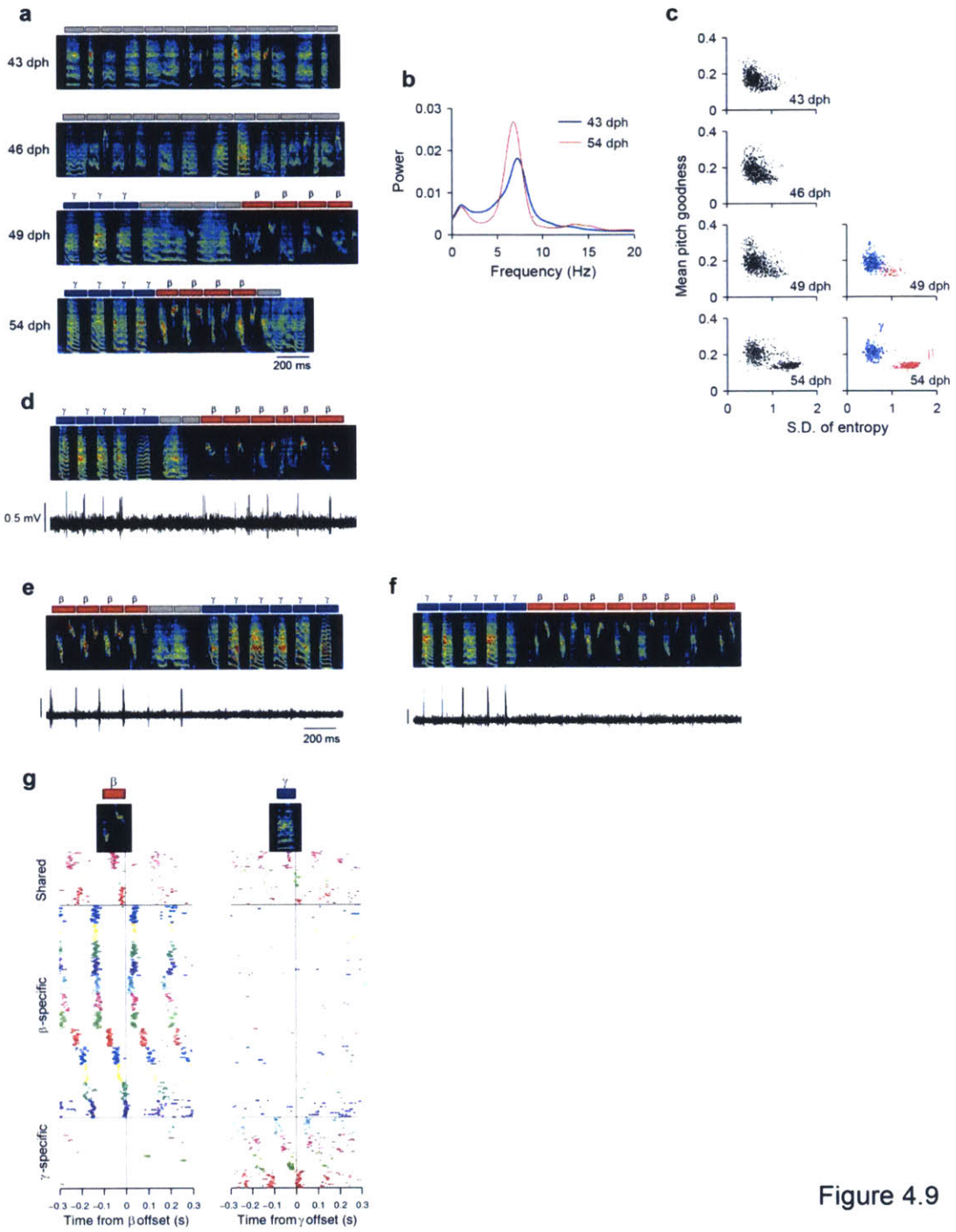

a
43 dph
46 dph
49 dph
54 dph
200 ms
b
Power
Frequency (Hz)
43 dph
54 dph
c
Mean pitch goodness
S.D. of entropy
43 dph
46 dph
49 dph
54 dph
d
0.5 mV
e
200 ms
f
g
Shared
β-specific
γ-specific
Time from β offset (s)
Time from γ offset (s)

Figure 4.9

Figure 4.9: Repetition of two syllable types.

All song and neural data from Bird 13. **a,** Sample song spectrograms and phase segments from four different ages during song development. Phase segments are shown. On 49 dph, two distinct rhythmic repetitions of syllables β-β-β and γ-γ-γ start to appear, and these become more acoustically distinct on 54 dph. **b,** Song rhythm spectrum calculated at 43 and 54 dph. Song rhythmicity exhibited a peak at 7 Hz, and band-pass filter centered at 7 Hz was used for the phase segmentation. **c,** Scatter plots of standard deviation of the entropy versus mean pitch goodness of the phase segments at four stages of vocal development. Each dot represents a single syllable (n=500 syllables per day). Also shown are scatter plots with syllable labels indicated in color (right). **d,** Shared neuron between β and γ (HVC_X; 49 dph). **e,** Neuron specific for β (HVC_X; 55 dph). **f,** Neuron specific for γ (HVC_X; 52 dph). **g,** Spike raster plots showing bursts of HVC projection neurons aligned to syllables β or γ.

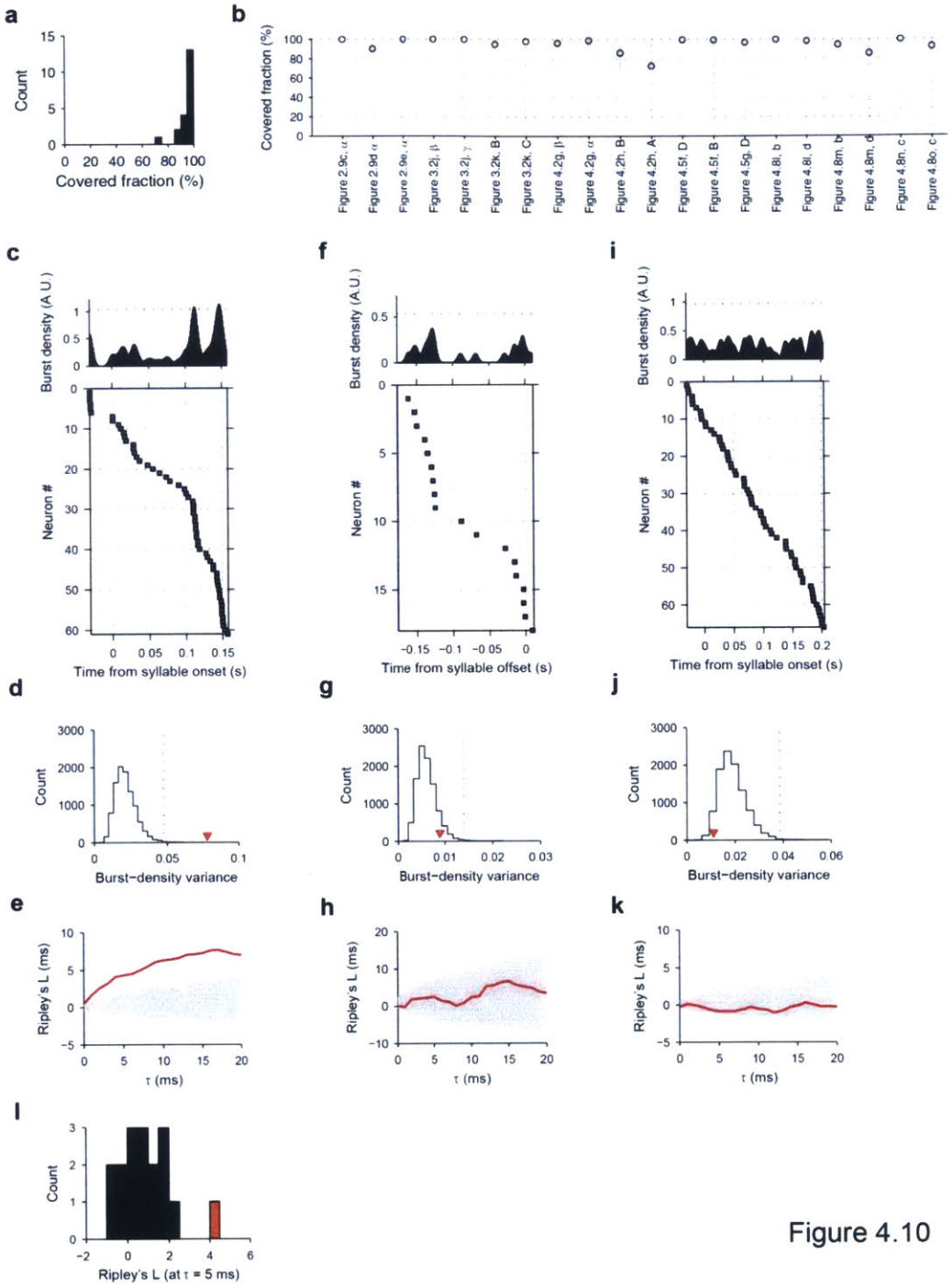

a
Count
Covered fraction (%)
b
Covered fraction (%)
Figure 2.9c, α
Figure 2.9d α
Figure 2.9e, α
Figure 3.2j, β
Figure 3.2j, γ
Figure 3.2k, B
Figure 3.2k, C
Figure 4.2g, β
Figure 4.2g, α
Figure 4.2h, B
Figure 4.2h, A
Figure 4.5f, D
Figure 4.5f, B
Figure 4.5g, D
Figure 4.8i, b
Figure 4.8i, d
Figure 4.8m, b
Figure 4.8m, d
Figure 4.8n, c
Figure 4.8o, c
c
Burst density (A.U.)
Neuron #
Time from syllable onset (s)
f
Burst density (A.U.)
Neuron #
Time from syllable offset (s)
i
Burst density (A.U.)
Neuron #
Time from syllable onset (s)
d
Count
Burst-density variance
g
Count
Burst-density variance
j
Count
Burst-density variance
e
Ripley's L (ms)
τ (ms)
h
Ripley's L (ms)
τ (ms)
k
Ripley's L (ms)
τ (ms)
l
Count
Ripley's L (at τ = 5 ms)

Figure 4.10

Figure 4.10: Coverage and clustering analysis of HVC projection neuron bursts.

a, Summary histogram of the covered fraction for all analyzed syllables (n=20). Note that 17/20 syllables had a covered fraction higher than 90%. **b,** We analyzed the fraction of time within each syllable that is continuously covered by the projection neuron bursts active during that syllable (covered fraction). Results are shown for 20 syllables for which more than 10 neurons exhibited bursts associated with that syllable. Shown is the covered fraction (open circles) for each syllable for which raster plots are shown in the figures. Also shown is the 95% confidence interval (vertical gray bars) for the covered fraction found in surrogate datasets in which the bursts of that syllable were randomly shuffled in time throughout the syllable (see Methods). Note that the observed covered fraction is within the 95% confidence interval for all syllables. **c-l,** Analysis of the uniformity of burst distribution was carried out for syllables in the intermediate and late stages of differentiation (n=17 syllables). **c-e,** Example of burst uniformity analysis for one syllable (Figure 3.k, syllable 'C') that exhibited significant temporal clustering of bursts. Burst times of all 62 individual neurons recorded for this syllable are shown in the bottom panel (square symbols). Shown at top is a smoothed (Gaussian with s.d. of 4 ms) histogram of burst density. Red dotted line indicates the threshold for statistical significance ($P<0.05$ with Bonferroni correction, 17 comparisons). **d,** Distribution of burst-density variance (black trace) for the shuffled data (N=10,000 runs). Also shown is burst-density variance for the actual data (red triangle). Red dotted line indicates the threshold for statistical significance ($P<0.05$ with Bonferroni correction, 17 comparisons). **e,** Ripley's L metric of clustering (red trace; see Methods) as a function of the size of the integration window τ. Positive values indicate clustering; negative values indicate

dispersion. Gray band indicates the 95% confidence interval for the shuffled data. Note the highly significant clustering of bursts, even though this syllable has a covered fraction of 100%. **f-h,** Same as in panels c-e but for the syllable 'A' shown in Figure 4.2h. Note that the apparent clustering of bursts around -0.125 s is not significant. **i-k,** Same as in panels c-e but for the syllable 'b' shown in Figure 4.8*l*. Notice that in this last example (panel i), the bursts seem to tile uniformly throughout the syllable, which leads to a slightly negative values in the Ripley's L metric (panel k). Also note that such a uniform distributions of burst times is relatively unlikely in a randomly distributed set of bursts, as measured by burst-density variance (P=0.045; panel j). **l,** Summary histogram for the Ripley's L metric for all 17 syllables (measured at a 5 ms integration window). Only one syllable (panel e) showed a statistically significant deviation from the randomly shuffled data as quantified by Ripley's L (red data point). The same result holds for analysis of peak burst density and burst-density variance.

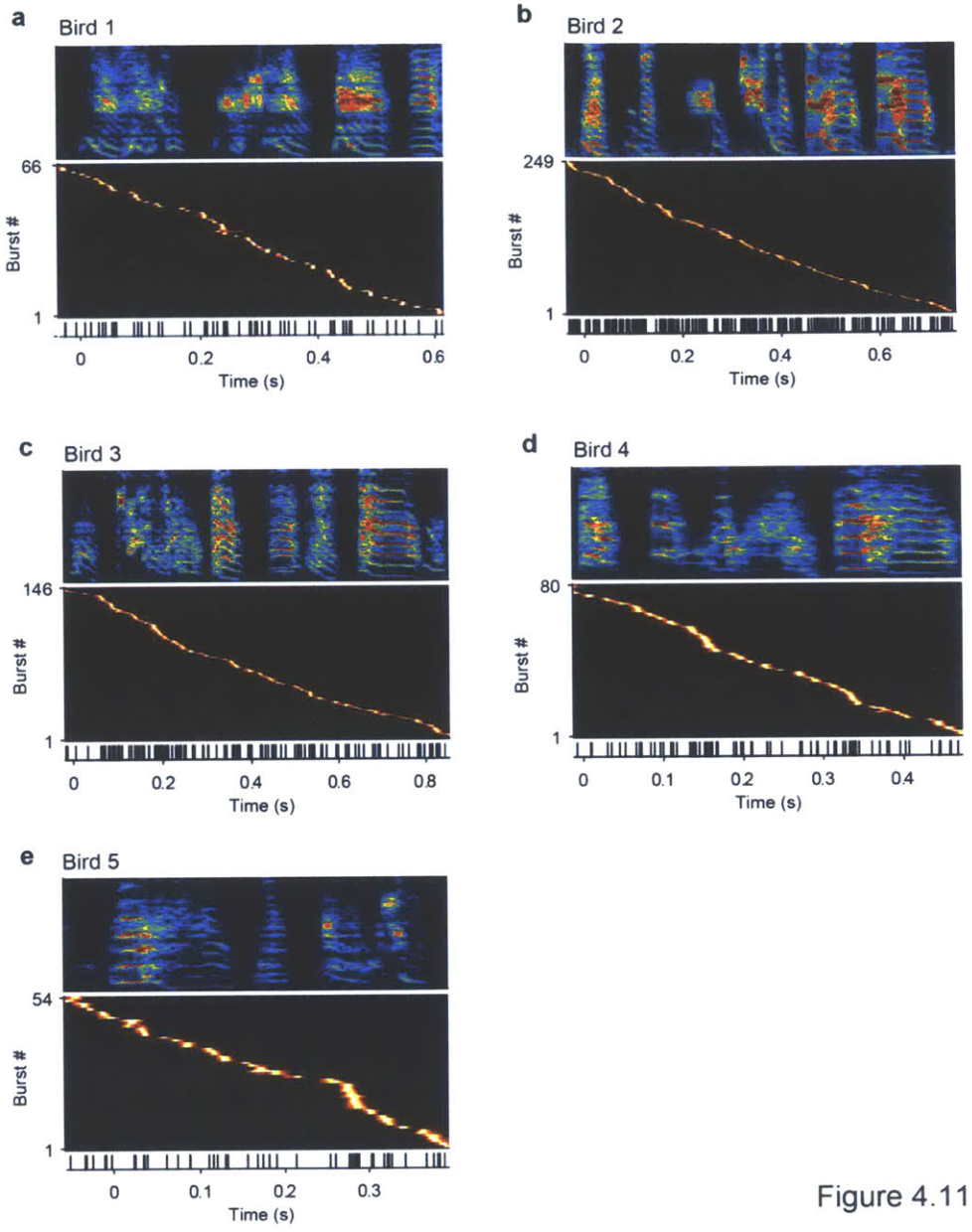

a
Bird 1
66
Burst #
1
0
0.2
0.4
0.6
Time (s)
b
Bird 2
249
Burst #
1
0
0.2
0.4
0.6
Time (s)
c
Bird 3
146
Burst #
1
0
0.2
0.4
0.6
0.8
Time (s)
d
Bird 4
80
Burst #
1
0
0.1
0.2
0.3
0.4
Time (s)
e
Bird 5
54
Burst #
1
0
0.1
0.2
0.3
Time (s)

Figure 4.11

Figure 4.11: Populations of HVC projection neuron bursts aligned to motifs.

a, Motif-aligned activity of HVC projection neurons. Top: Song spectrogram of the motif. Middle: Heat map of smoothed firing rate (normalized by the peak for each bursts.) sorted by the latency of bursts. Note that multiple bursts from a single neuron are treated separately, and appear at different rows. Bottom: Black tick marks indicate the center of the gravity of each burst. Data from previously published HVC recordings in an adult bird (Kozhevnikov and Fee, 2007). **b-e,** Same as panel a, but for 4 different birds that had a stable motif.

Chapter 5

Discussion

The simulation of a network model presented in this chapter was developed in collaboration with Emily L. Mackevicius (Department of Brain and Cognitive Sciences, MIT) and Hannah L. Payne (Department of Neurobiology, Stanford University).

In the previous chapters, I have characterized the activity patterns of HVC at different song stages. In this chapter, I will synthesize these results into a single model of the developmental change in HVC activity during song learning. First, I will give an overview of the model in words, and then I will describe simulation results of a neural network model. Next, I will point out the limitations of the model, and list future experiments that could test the model. Finally, I will discuss the broad implications of the findings.

A model on the developmental change in HVC activity

A schematic of our model of the developmental progression in HVC is depicted in Figure 5.1. This model assumes that the first step in vocal learning is a growth of a neural sequence ('protosequence'), which repeats itself rhythmically and generates protosyllables (Figure 5.1a). Protosequence is a 'wide' sequence in a sense that many neurons are active at a given time. Next, this wide sequence splits to give rise to two or more daughter sequences that corresponds to multiple emerging syllable types (Figure 5.1b). During splitting, there are neurons that are active for both of these syllables at similar latencies ('shared neurons'). As the splitting proceeds, fraction of shared neurons decrease, and more neurons become specific for a particular syllable type. Finally, this splitting can occur multiple times in a hierarchical manner, giving rise to more syllable types (Figure 5.1c). Detailed mechanistic descriptions on each of these steps are depicted below.

Mechanistic model of sequence splitting in HVC

Here, we describe a numerical simulation of a neural network model that implements the concepts presented above (details can be found in Appendix B). This model is based on the idea that sequential bursting results from the propagation of activity through a synaptically-connected chain of neurons within HVC (Buonomano, 2005; Li and Greenside, 2006; Jin et al., 2007; Jun and Jin, 2007; Gibb et al., 2009; Fiete et al., 2010; Long et al., 2010), and thus retains the essential features of the continuous-time model (Fee et al., 2004; Leonardo and Fee, 2005). Yet it also captures the excess burst density observed at protosyllable onsets, as found in the gesture model (Amador et al., 2013).

Recent modeling studies have shown that a combination of two synaptic plasticity rules—spike-timing dependent plasticity (STDP) and heterosynaptic competition—can transform a randomly connected network into a feedforward chain that generates sparse sequential activity similar to that observed in HVC (Jun and Jin, 2007; Fiete et al., 2010). We hypothesize that the same mechanisms can lead to the formation of a single chain that generates a rhythmic protosyllable, as well as the splitting of this chain into multiple daughter chains for different syllable types. To test this hypothesis, we constructed a simple network of binary units representing HVC projection neurons (Fiete et al., 2010). The network undergoes three stages of development corresponding to the subsong, protosyllable, and differentiation stages.

Network at initial conditions: Subsong

The model neurons are initially connected with random excitatory weights, a state that in our model captures HVC activity in the subsong stage. Motivated by the narrow distribution of burst latencies at syllable onsets in subsong (Figure 2.4i), we hypothesize that a subset of HVC neurons receives an external input at syllable onsets and serve as a seed from which chains grow during later learning stages (Buonomano, 2005; Jun and Jin, 2007). To visualize network activity before learning (the subsong stage in our model; Figure 5.2a), we activated seed neurons at random intervals corresponding to the highly variable timing of syllable onsets in subsong birds (Aronov et al., 2008; Aronov et al., 2011). Activation of the seed neurons produced a transiently propagating sequence of network activity that decayed rapidly (within tens of milliseconds; Figure 5.2a). Thus, only a subset of HVC neurons were reliably locked to syllable onsets, and the rapid decay ensured that the latencies of these neurons were tightly clustered at syllable onsets (Figure 5.2e), as observed in our recordings (Figure 2.4i).

Growing a protosyllable chain

In the next stage, the network is trained to produce a single protosyllable by activating seed neurons rhythmically (100 ms period). The connections are modified according to STDP and heterosynaptic long-term depression (hLTD) (Jun and Jin, 2007; Fiete et al., 2010). As a result, connections were strengthened among the population of neurons that were sequentially activated after syllable onsets. This strengthening, together with the incorporation of additional neurons, resulted in the formation of a robust feedforward synaptically-connected chain that supported stable propagation of activity (Figure 5.2b). This chain was 'wide' in the sense that many neurons were coactive at each time in the sequence. At this stage, all neurons in the network were locked to song

syllables (Figure 5.2b). Individual neurons were rhythmically active, each at a distinct latency in the rhythm, and the distribution of latencies was broader than in subsong (Figure 5.2e; compare Figure 2.4i-k) Together, these features resemble the activity of HVC projection neurons observed in singing birds at the protosyllable stage (Figure 2.9d-f).

Splitting the protosyllable chain

We next hypothesized that this single chain could be induced to split into two chains by altering the pattern of seed neuron activity. The splitting would proceed through the same competitive synaptic interactions that formed the initial protosyllable chain. We chose to divide the seed neurons into two groups activated on alternate cycles of the rhythm. Local inhibition (Kosche et al., 2015) and synaptic competition were also increased (see Appendix B). This resulted in a gradual splitting of the wide chain into two narrower daughter chains (Figure 5.2c, d). During the splitting process, we observed neurons specific to each of the emerging syllable types, as well as shared neurons that were active at the same latencies in both syllable types (Figure 5.2c). Just as we observed in our data, the distribution of burst latencies in the model continued to broaden during learning (Figure 5.2e), and the fraction of shared neurons decreased over the course of learning. As we also observed in our birds (Figure 2.5d), the average period of rhythmic bursting in model neurons increased during chain splitting as neurons 'specific' for each emerging syllable began to participate only on alternate cycles of the protosyllable rhythm (Figure 5.2d, Figure 5.3d). These observations suggest that a previously proposed synaptic mechanism of chain formation, with a few simple modifications, can capture

numerous features in our neural data associated with the formation of protosyllables and the emergence of multiple syllable types during vocal development.

In this model, we envision that the formation of daughter chains in HVC is translated into the emergence of new syllable types as follows: During the splitting process, as two distinct sequences of specific neurons develop, their downstream projections can be independently modified such that each of the emerging chains of specific neurons can drive a distinct pattern of downstream motor commands, allowing distinct acoustic structure in the emerging syllable types (Fiete et al., 2004). Such differential acoustic refinement is consistent with previous behavioral observation that the altered acoustic structure of new syllables emerges in place, without moving or reordering sound components ('sound differentiation in situ') (Tchernichovski et al., 2001).

This model naturally explains the apparent 'decoupling' of shared projection neuron bursts from acoustic structure in the vocal output—i.e., the fact that the bursts of shared neurons become associated with two distinct acoustic outputs during the differentiation of two syllable types (Figure 3.7, 3.8). Specifically, during syllable differentiation, a shared neuron participates with different ensembles of neurons during each of the emerging sequences (Figure 5.2c), and these different ensembles can drive different vocal outputs.

Other strategies for syllable formation

Our model can reproduce other strategies by which birds learn new syllable types, such as the simultaneous emergence of a three-syllable motif ('motif strategy') by dividing the seed neurons into three subpopulations (Figure 5.3a-d).

Our model was also able to simulate syllable differentiation at bout onsets (Lipkind and Tchernichovski, 2011; Lipkind et al., 2013). Inspired by our observation of bout-onset neurons (Figure 2.7, 4.2a), we implemented bout-onset differentiation in the model by including a population of seed neurons activated at bout onsets, in addition to the seed neurons rhythmically activated at protosyllable onsets (Figure 5.3e; see Methods). This caused the protosyllable chain to split in such a way that one daughter chain was reliably activated at bout onsets and the other daughter chain followed rhythmically (Figure 5.3f-h).

Experimental test of the model

Long-term tracking of HVC activity

Our model predicts that individual neurons initially start out being shared between multiple syllable types and gradually become specific for a particular syllable. To conclusively test this transition from shared to specific activity within a given neuron, it is necessary to develop a novel technique that allows stable long-term recordings over several days, since the current method only allows recording from HVC projection neurons for an hour at most (Appendix A).

One approach is to use electrodes that can stably record single-unit activity over multiple days. For example, a recent study demonstrated the feasibility of obtaining long-term recordings of HVC activity using array of carbon fiber electrodes (Guitchounts et al., 2013). While the number of projection neurons recorded per animal was small in this study, it provides a good starting point for future technological development.

A complementary approach is to perform a long-term imaging of neural activity using genetically-encoded calcium indicators (Komiyama et al., 2010; Ziv et al., 2013; Masamizu et al., 2014; Peters et al., 2014). Indeed, a recent work succeeded in imaging HVC activity during singing using calcium indicators in adult birds (Markowitz et al., 2015). It will thus be interesting to modify this technique to image HVC activity from juvenile birds.

Simultaneous recording of HVC population activity in juvenile birds

In the current study, identification of shared and specific neurons relied on labeling syllables based on their acoustic features (e.g. Figure 3.2b) or syllable duration (e.g. Figure 4.1d). A potential issue with this approach is that by the time one can identify distinct syllable types reliably, the number of shared neurons might already be low in number. To overcome this problem, it is desirable to identified shared and specific neurons based on the population activity of simultaneously recorded neurons instead of relying on acoustics of syllables. For example, if the population activity during singing exhibits two 'trajectories' in high-dimensional space corresponding to the two emerging syllables, one could identify whether a given neuron is shared or specific using neural activity alone. In the current study, we had at most 5 electrodes within HVC. With this design, simultaneous recordings of a pair of HVC projection neuron were fairly common, but it was difficult to obtain recordings from more than three projection neurons at once (Appendix A). Thus, it will be useful to develop new technologies to record from many HVC projection neurons simultaneously either by using electrode arrays or using imaging mentioned above.

Synaptic pruning during development

What is the anatomical basis of sequence splitting? Is the sequence splitting associated with pruning of the cross connections between HVC projection neurons (Figure 5.1)? Recent technological advances might make it possible to address this question. One way to observe the pruning process is to use time-lapse *in vivo* imaging of fluorescent proteins, as has been applied to synaptic elimination in neuromuscular junction (Lichtman and Smith, 2008). In other words, one could observe fluorescently labeled axon terminals of a single HVC_{RA} neuron and test whether they undergo pruning. An alternative approach is to perform a connectomics analysis of the HVC microcircuit using the electron microscope (Seung, 2009) in both juveniles and adults and to test whether HVC_{RA} neurons are more interconnected in juveniles than in adults.

Synaptic plasticity rule in HVC

Our model assumes that spike-timing dependent plasticity in HVC is important for sequence formation and subsequent sequence splitting. However, no study has so far reported a synaptic plasticity rule between HVC projection neurons. To test whether STDP occurs among HVC neurons, paired recordings of projection neurons in slice could be performed. However, given a relatively infrequent connections between HVC_{RA} neurons (Mooney and Prather, 2005; Kosche et al., 2015), one might need to develop a high-throughput approach to test such connections and their plasticity.

Changes in inputs to HVC

In our model, we hypothesized a particular developmental change in the inputs to seed neurons in HVC: early on, seed neurons receive rhythmic inputs; during syllable differentiation, seed neurons are divided into two groups and receive alternating inputs. While this is only one way in which sequence splitting can occur, it will be of interest to test whether this hypothesized change occur in inputs to HVC during development. This could be tested by recording from neurons in brain areas that project to HVC (e.g. NIf, Uva, Av, and MMAN; Figure 1.2) and observing how their activity changes during song development.

Limitations of the model

Features that are not captured by our model

While our model captures the majority of the experimental observations, there are features that are not explained by the model. First, we observed that during the protosyllable stage, individual HVC projection neurons exhibit rhythmic bursting, and that this rhythmic bursting is probabilistic and is often associated with skipping cycles (Figure 2.9a, b, 2.10a). While differences in acoustic features of syllables could explain part of this variability (Figure 2.10d-h), it is likely that there is an additional source of variability such as probabilistic propagation of bursts in HVC. In our model of HVC network, we assumed that during the protosyllable stage, there is an all-to-all connections from one layer to the next (Figure 5.1a). However, it is possible that connections in the real HVC network are sparser and the propagation is more variable. Alternative source of variability could be LMAN, another premotor area that is important for generating

variable vocalizations in juveniles (Scharff and Nottebohm, 1991; Olveczky et al., 2005; Aronov et al., 2008; Hamaguchi and Mooney, 2012) (see Discussion in Chapter 2).

Second feature that is not captured by our model is the transitions between different syllable types. In other words, our network model explains how neural sequences associated with each syllable arises during development, but does not explain how one sequence becomes reliably activated after another sequence to give rise to a stereotyped syllable order. For example, we have observed both alternating differentiation (β-γ-β-γ; Chapter 3, Figure 3.2a) and repeating syllables (β-β-β, γ-γ-γ; Figure 4.9a). Syllable order could be encoded as connections within HVC network or it could depend on how external inputs activate different sequences in HVC. These are interesting questions for future research.

Alternative interpretation of our results

Our model hypothesizes that a single sequence splits into multiple sequences (Figure 5.1). However, it is also possible that there are two or more distinct sequences even from the beginning, and that these sequences are co-activated and run simultaneously early in development. During the emergence of new syllable types, these sequences are activated more independently, and in adults, each sequence corresponding to a distinct adult syllable type is activate one at a time in a fixed order. According to this model, plasticity might be mostly happening at the inputs to HVC as opposed to within HVC. Although this is a different mechanism than the one proposed in Figure 5.1, it could still be called a 'splitting of sequence' in a broad sense since multiple sequences are initially co-activated and run simultaneously, but over time become independent.

It is also possible that the developmental changes in the HVC network are mostly due to change in functional connections as opposed to anatomical connections. For example, increased inhibition between the different sequences might be sufficient to explain our results without assuming synaptic pruning. Consistent with the idea that inhibition also plays a role, when HVC_{RA} neurons in adults were recorded with local application of gabazine, a $GABA_A$ antagonist, these HVC_{RA} neurons exhibited more bursts per motif, suggesting that inhibition within HVC is important for ensuring sparse bursting of HVC_{RA} in adults (Kosche et al., 2015). Again, change in functional connectivity between the HVC_{RA} neurons via increase in inhibition can still be interpreted as a 'splitting of neural sequence' in a broad sense.

Alternative mechanisms of syllable formation

While our observations support the view that emerging syllable types can be formed by the splitting of a protosyllable sequence, we cannot rule out the possibility that at least some syllable types arise from the *de novo* formation of a new sequence. For example, in several birds, we observed a short vocal element at bout onsets that did not appear to differentiate acoustically from a prototype syllable (thus not bout-onset differentiation), and was retained in the adult song as a distinct syllable type (for example, 'E' in Bird 1, Figure 4.5a; 'C' in Bird 2, Figure 4.1a). Note that it is difficult to conclusively demonstrate *de novo* formation using neural recordings; lack of shared neurons might mean that the recordings were not performed early enough when there were shared neurons, or that the sampling was insufficient to reveal their existence.

Possible advantages of sequence splitting

Regardless of the specific underlying neural mechanisms, the process of splitting a prototype neural sequence may be advantageous in allowing learned components of a prototype motor program to be reused in each of the daughter motor programs. For example, one of the earliest aspects of vocal learning is the coordination between singing and breathing (Veit et al., 2011), specifically, the alternation between vocalized expiration and non-vocalized inspiration typical of adult song (Goller and Cooper, 2004). The protosequence in HVC would allow the bird to learn the appropriate coordination of respiratory and vocal musculature, presumably through synaptic plasticity in the downstream connections of HVC neurons. Duplication of the protosequence through splitting would result in two 'functional' daughter sequences, each already capable of proper vocal/respiratory coordination. Thus, each daughter sequence would be well suited, through further refinement of the downstream projections, to serve as a substrate for a new syllable type.

Relation to other studies

Development of rhythmic motor patterns in invertebrates

Neural networks that generate rhythmic patterns have been extensively studied in invertebrates. For example, studies on stomatogastric ganglion (STG) in crustaceans have succeeded in explaining how rhythmic motor output can be generated by biophysics of individual neurons and synaptic interactions among them (Marder and Bucher, 2001; Nusbaum and Beenhakker, 2002; Marder and Bucher, 2007). Interestingly, while the adult STG circuit is capable of generating at least two distinct rhythms (pyloric rhythm

and gastric mill rhythm), the STG of embryos only generate a single rhythm (Casasnovas and Meyrand, 1995; Rehm et al., 2008). Since the number of neurons in the circuit remains constant over development, researchers have argued that adult circuit capable of generating multiple rhythms differentiates from a single embryonic circuit that generates a single rhythm (Casasnovas and Meyrand, 1995). Later studies showed that this developmental change in rhythm production is partly due to a change in neuromodulation since embryonic circuit was able to produce adult-like pattern when appropriate neuromodulators were applied (Le Feuvre et al., 1999; Rehm et al., 2008). It is interesting that a differentiation of neural network to produce multiple motor patterns is observed during development of an innate behavior of invertebrates, and not just in vocal learning of songbirds.

Learning-related change from shared to specific activity in rat hippocampus

Changes in neural activity from shared to specific pattern during learning have been observed in the CA1 region of the rat hippocampus. For example, in one study, naïve rats were introduced to two environments, circular and square boxes (Lever et al., 2002b). Initially, patterns of neural activity in circular and square environments were highly similar, with many neurons exhibiting firing in similar locations ('homotopic' fields; similar to 'shared neurons with similar latency' in our case). Activity patterns associated with these two environments began to diverge from each other over several days during the repeated exposure (Lever et al., 2002b). Some cells became only active in one environment ('monotopic'; similar to 'specific neurons') or fired in two different locations in the two environment ('heterotopic') (Lever et al., 2002a; Lever et al., 2002b).

While the mechanism underlying sequence generation might differ between rodent hippocampus (Mehta et al., 2002) and songbird HVC (Long et al., 2010), these results suggests that change in neural activity from shared to specific activity pattern is commonly observed during learning.

Speech acquisition in humans

Previous studies have proposed that vocalizations in human infants develop from 'reduplicated' babbling (such as 'ba ba ba') to 'variegated' babbling (such as 'ba da ga') (Oller, 1980; Stark, 1980), but the experimental support for this proposal has been elusive (Lipkind et al., 2013). A recent study reexamined this issue, and found that the occurrence of reduplicated babbling did not decrease over development as a whole, but when the same data was aligned to the appearance of each syllable type, there was a clear decrease in the frequency of reduplicated babbling (Lipkind et al., 2013). This result suggests that when a new syllable first appears, it tends to be reduplicated, and over time, it becomes variegated. This could either be due to differentiation of new syllable types from a single prototype syllable, or due to learning new transitions between existing syllable types (Lipkind et al., 2013). While this is not a direct test of our model, it is nonetheless reminiscent of the alternating differentiation shown in Chapter 3. In sum, even though the model presented in Figure 5.1 might not hold in its simplest form in humans, given the similarity in behavior and the brain organization of birds and humans (see Chapter 1), it is possible that a mechanism similar to sequence splitting is involved in speech acquisition in humans as well. Further test of these ideas will require unbiased

and comprehensive recording of infant vocalizations[6] and non-invasive methods to record neural activity during vocalizations in infants.

Rhythmic stereotypies in humans

The work presented here might also be related to the development of motor behaviors in human infants other than speech acquisition. For example, Thelen and others have characterized the development of rhythmic repetitious movements common in human infants during the first year. These movements are called 'rhythmic stereotypies' and include kicking, rocking, banging, and rubbing (Thelen, 1981, 1995). Researchers first compared kicking movements when the infants were lying on their back and walking that appeared later in development, and have argued that 'rhythmic stereotypies bridge the gap between uncoordinated and well-coordinated movement' and that 'rhythmic movements are the essential subroutines for the assembly of coordinated movements'(Thelen, 1981). However, there is an important difference between Thelen's view and ours. The former hypothesizes that the rhythmic stereotypies occur due to central pattern generators in the spinal cord while the cortex slowly matures (Thelen, 1981), whereas our work suggests that juvenile rhythmic patterns are generated by the same premotor cortical area that is used for later adult behavior.

Duplication and divergence in biology

The proposed neural mechanism of splitting neural sequence for generating multiple syllable types in songbirds is part of 'duplication and divergence' often observed

[6]An example of a technology that allow collection of such data is the Human Speechome Project initiated by Deb Roy at MIT.

in biology at many levels of analysis. For example, at the smallest level, it has been proposed that gene duplication followed by divergence through independent mutations give rise to the evolution of novel gene functions (Ohno, 1970; Taylor and Raes, 2004). A classic example of such event is the evolution of multiple types of rhodopsin gene for color vision (Nathans et al., 1986).

A similar mechanism can give rise to new cell types in neural circuits. For example, using molecular fingerprinting, some have argued that three types of cells in the vertebrate retina—rods, cones, and bipolar cells—have evolved from a common ciliary photoreceptor precursors (Arendt, 2008)[7]. In this example, photoreceptors (rods and cones) and bipolar cells have 'segregation of functions' resulting in specialized descendants: photoreceptors are specialized for visual transduction, and bipolar cells are specialized for communications with the retinal ganglion cells. On the other hand, rods and cones have 'divergence of functions' since they are both light sensitive, but are specialized at detecting different wavelengths (Arendt, 2008).

Finally, for the acquisition of complex behaviors, the duplication of neural sequences by splitting, followed by independent differentiation through learning, may provide a mechanism for constructing complex motor programs. This could happen during learning within an individual animal as demonstrated in this thesis, but could also occur at slower time scales over the course of evolution. For example, voltage-sensitive dye imaging of neurons in a segmental ganglion of leeches revealed that there are many neurons active during both swimming and crawling ('multifunctional interneurons'; similar to 'shared neurons' in our terminology) while there are also neurons that are

[7] Several lines of evidence support this hypothesis: common use of ribbon synapses, existence of a cilium called Landolt's club in bipolar cells, and similar morphology at an early differentiation stage.

specific for swimming or crawling ('dedicated interneurons'; similar to 'specific neurons') (Briggman et al., 2005; Briggman and Kristan, 2006). Interestingly, because all species of leech can crawl but only half of the species swim (Sawyer, 1986), it has been suggested that neural circuit for swimming might have evolved by modifying the network for crawling (Briggman and Kristan, 2006, 2008).

Conclusions

Previous studies have shown that sequences of sparse bursts in the premotor area HVC of songbirds generate stereotyped vocalizations in adult birds. Since this discovery, HVC has become an important brain area to test general ideas on sequence generations in neural circuits. To address how these sequences arise during development, I recorded populations of neurons in juvenile birds during song learning. As a result, I identified a novel phenomenon, growth and subsequent splitting of a rhythmic neural sequence, as a neural mechanism for forming new syllable types. Based on these results, I proposed a concrete mechanistic model on how sequence splitting could be implemented in neural circuits within HVC. I hope future studies will continue to bridge the gap between behavioral observations of song learning, an ethologically-relevant and rich behavior, and the underlying neural mechanisms. I also wish that juvenile HVC becomes an important testbed for evaluating general ideas on the development of neural sequences during learning.

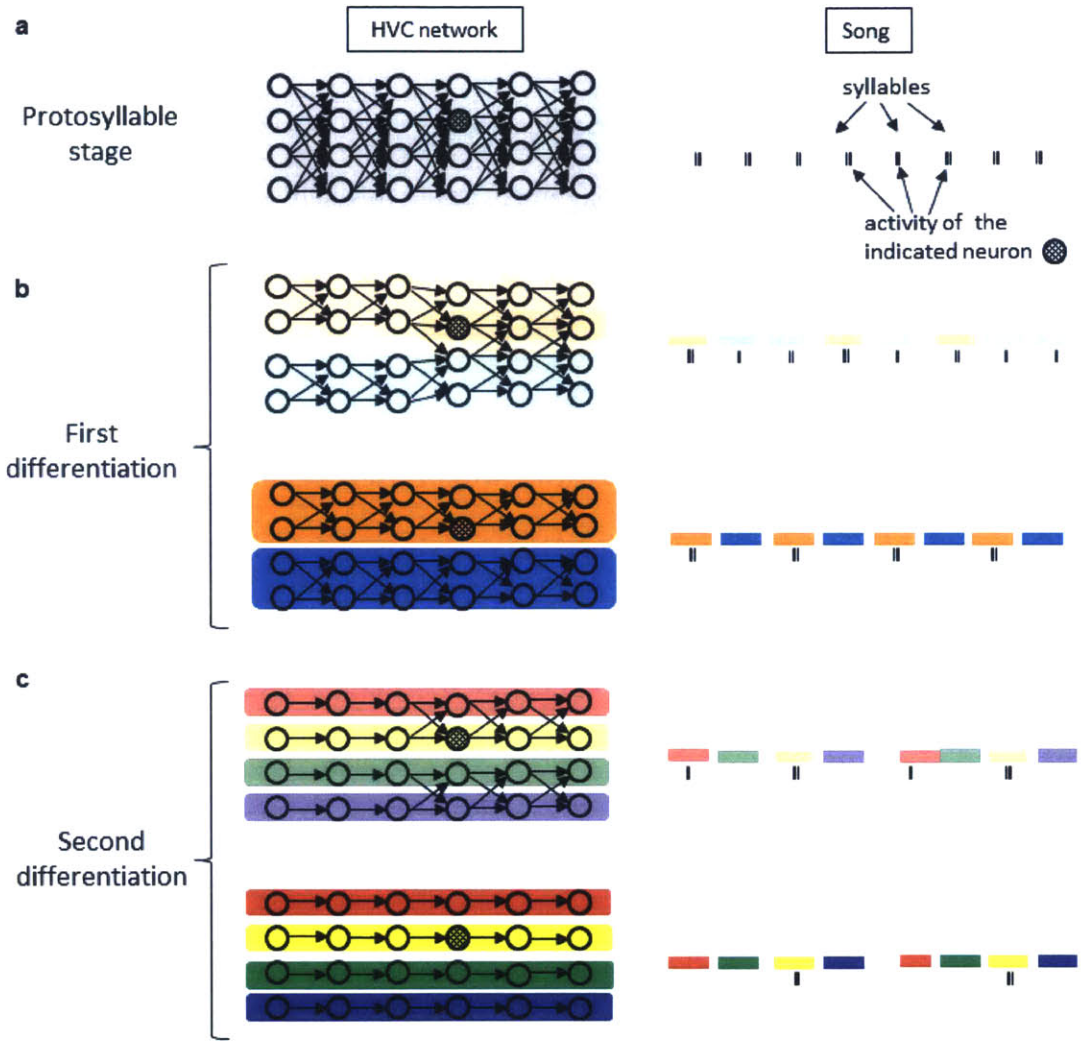
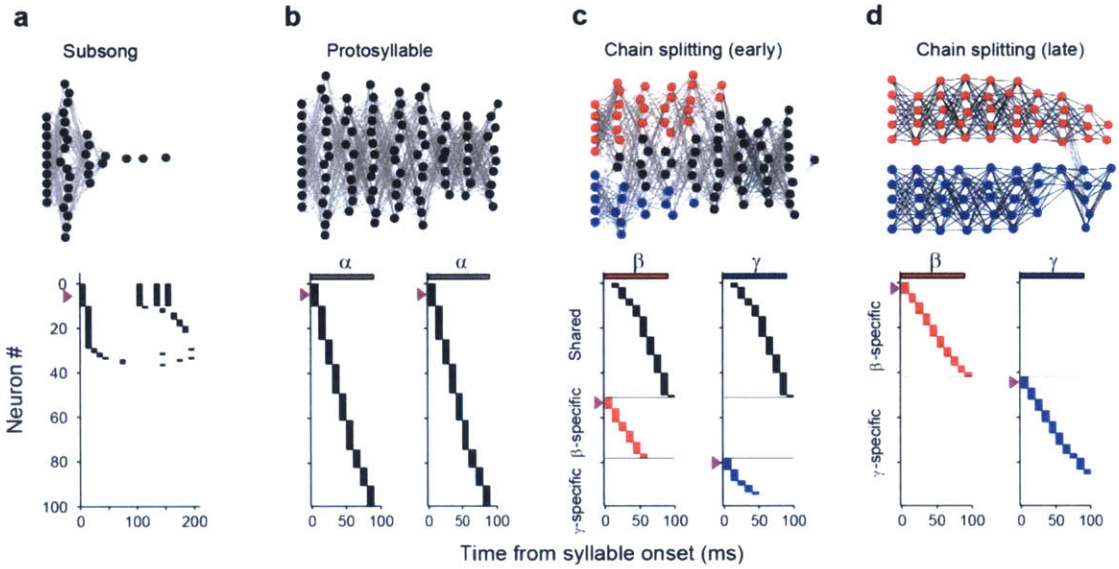

a
HVC network
Song
Protosyllable stage
syllables
activity of the indicated neuron
b
First differentiation
c
Second differentiation

Figure 5.1

Figure 5.1: Schematic of the developmental change in HVC network.

HVC network (left) and the corresponding song that it produces (right) are shown for different developmental stages. Individual circles indicate HVC projection neurons, and arrows indicate connections between them. Spiking activity of a particular neuron in the network (black hatch) is shown on the right. **a,** Protosyllable is encoded by a 'wide' sequence in HVC with many neurons active at a given time. Each neuron is rhythmically active during protosyllables. **b,** Splitting of a protosequence gives rise to two daughter sequences coding for orange and blue syllables. The indicated neuron is initially active in the middle of both of these syllables (shared neuron), but becomes specific for the orange syllable once the splitting is complete. **c,** Second differentiation splits two sequences into four. In the end, each neuron in the network is active at a particular time in a particular syllable thus exhibiting the sparse coding of syllables characteristic of adult HVC.


a
Subsong
b
Protosyllable
c
Chain splitting (early)
d
Chain splitting (late)
Neuron #
α
α
β
γ
β
γ
Shared
β-specific
γ-specific
β-specific
γ-specific
Time from syllable onset (ms)

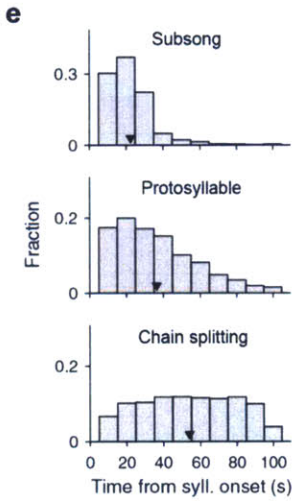

e
Subsong
Protosyllable
Chain splitting
Fraction
Time from syll. onset (s)

Figure 5.2

Figure 5.2: A model of sequence formation and splitting in HVC.

a-d, Network structure and activity for four stages in the emergence of new syllable types: subsong, protosyllable, during chain splitting, after chain splitting. Top: network diagrams showing model HVC projection neurons and the connections between them (darker lines indicate stronger connections). Neurons are sorted by relative latency from syllable onsets (see Methods). Bottom: population activity plot for each syllable type, showing activity of all participating neurons, sorted by relative latency within the categories of shared (on top) and specific (below). Seed neurons receiving external inputs are colored magenta in the subsong and protosyllable stages, and red and blue when they receive alternating inputs during the splitting process. Horizontal color bars on the top indicate emerging syllable types. **a,** Subsong stage: activation of seed neurons produces a rapidly-decaying burst of sequential activity in a subset of neurons. **b,** Protosyllable stage: rhythmic activation of seed neurons induces the formation of a 'wide' protosyllable chain. **c,** During chain splitting: alternating activation of subsets of seed neurons drives subsets of neurons to become 'specific' (red and blue neurons in the network diagram); shared neurons shown in black. **d,** After chain splitting: competition has eliminated connections between neurons belonging to different daughter chains. **e,** Distribution of burst latencies at three different stages in model sequence development corresponding to subsong, protosyllable, and multi-syllable stage.

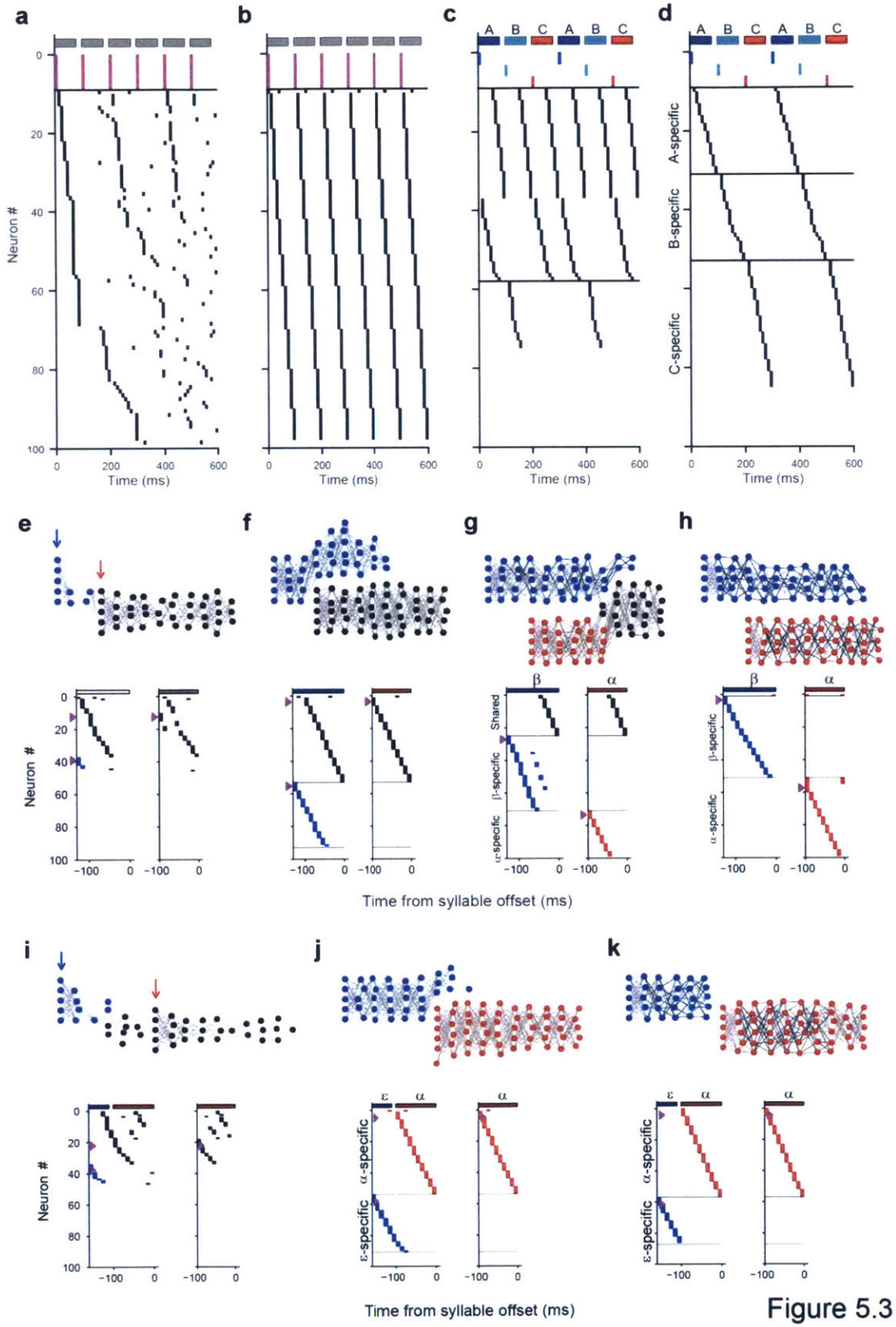

a
b
c
d
A B C A B C
Neuron #
Time (ms)
A-specific
B-specific
C-specific
e
f
g
h
Shared
β-specific
α-specific
Time from syllable offset (ms)
i
j
k
ε-specific
α-specific
Time from syllable offset (ms)

Figure 5.3

Figure 5.3: Other strategies for syllable formation.

a-d, Model of simultaneous formation of multiple syllable types into an entire motif ('motif strategy'). **a-b,** Protosyllable seed neurons (magenta) were activated rhythmically to form a protosequence. **c,** Seed neurons were then divided into three sequentially activated subgroups, resulting in the rapid splitting of the protosequence into three daughter sequences. In intermediate stages, individual neurons exhibited varying degrees of specificity and sharedness for the emerging syllable types. **d,** After learning, neurons were active sequentially throughout the entire 'motif,' but individual neurons were active only during one of the resulting syllables, forming three distinct non-overlapping sequences. **e-h,** Bout-onset differentiation results from activation of bout-onset seed neurons (blue arrow) followed by rhythmic activation of protosyllable seed neurons (red arrow). Network diagrams show protosyllable formation (panels e, f) and splitting of chains specific for bout-onset syllable β and specific for later repetitions of the protosyllable α (panels c, d; blue and red, respectively; shared neurons: black). **i-k,** Network diagrams and raster plots showing an example of the formation of a new syllable chain at bout onset. In the network diagrams, neurons specific for each emerging syllable type are colored blue and red respectively. Bout-onset seed neurons and protosyllable seed neurons are shown by blue and red arrows, respectively. The three panels represent the early protosyllable stage (panel i), the late protosyllable stage (panel j), and the final stage (panel k). The protosyllable seed neurons were driven strongly throughout the learning process, so they did not become outcompeted on some cycles of the protosyllable chain, and a separate bout-onset chain remained in the final stage.

Appendix A

In vivo recording of single-unit activity during singing in zebra finches

This chapter was previously published as:

Okubo T.S., Mackevicius E.L., and Fee M.S. (2014) *In vivo* recording of single-unit activity during singing in zebra finches. *Cold Spring Harbor Protocols* (12): 1273-83.

Abstract

Zebra finches have proven to be an important organism for investigating the neural mechanisms underlying vocal production and learning. Previous anatomical and gene expression studies have identified an interconnected set of brain areas in the zebra finch that are important for singing. To further advance our understanding of how these various brain areas act together to enable the animal to learn and produce a highly-stereotyped song, it is necessary to record activity of individual neurons during singing. Here, we describe a protocol for recording single-unit activity in freely moving zebra finches using a miniature motorized microdrive. This protocol includes procedures for both the microdrive implant surgery and the electrophysiological recordings. We also describe methods for identifying neurons based on the projection pattern using antidromic stimulation and the spike collision test. Finally, common pitfalls, troubleshooting tips, and possible combinations with other methods are discussed.

Materials

Reagents

- Analgesic (e.g. Buprenorphine, 0.05 mg/kg, 0.75% in saline)
- Antibiotic (e.g. Baytril, 5 mg/kg, 0.1% in saline)
- Betadine solution (10% in distilled water)
- Dental acrylic (e.g. Flow-It Accelerated Light Cure, Pentron Clinical)
- General anesthetic (e.g. Isofluorane)
- Liquid soldering flux for stainless steel (e.g. Stay-Clean, Harris Products Group)
- Local anesthetic (e.g. Bupivacaine, 0.2 mg/kg, 0.05 mg/mL in saline)
- Phosphate-buffered saline
- Silicon elastomer (e.g. Kwik-Cast, World Precision Instruments)
- Tissue adhesive (e.g. Vetbond, 3M)

Equipment

Electrophysiological recording setup

The schematic of the setup is shown in Figure A1a.

- Amplifier for neural signals with a gain of 100 (e.g. operational amplifier TL084, Texas Instruments)
- Audio monitor (e.g. AM10, Grass Technologies)
- Bird cage (approx. 23 cm in diameter, 20 cm in height)
- Cable (e.g. male connector with wires A7876-001, female connector A7877-001, Omnetics Connector) with surface mount field-effect transistors (FETs, e.g.

2SK3796-3-TL-E, ON Semiconductor) in a source-follower configuration on the microdrive side of the cable (Figure A2c, d).

- Collision test circuit: an electrical circuit that triggers the pulse generator upon detecting a threshold crossing event due to an occurrence of a spike. Operational amplifier (e.g. TL084, Texas Instruments), inverter chip (e.g. SN74LS04N, Texas Instruments), timer chip (e.g. LM555, Fairchild Semiconductor), an LED, resistors, and capacitors (Figure A3a).
- Computer for data acquisition
- Data acquisition system (e.g. card: PCI-6052E; connectors: BNC-2090A; cable: SH68-68-EP; all from National Instruments) connected to a computer
- Filters: high-pass filter (cutoff frequency of 300 Hz), and a low-pass filter (cutoff frequency of 15 kHz) for a sampling frequency of 40 kHz.
- Microphone (e.g. AT803, Audio-Technica) and an amplifier (e.g. MA3, Tucker-Davis Technologies) for song recordings
- Motor controller for the microdrive (e.g. Fee-modified MP-285, Sutter Instrument) Details can be found in Fee and Leonardo, 2001.
- Motorized commutator with a torque feedback motor controller (Fee and Leonardo, 2001)
- Oscilloscope (e.g. TDS2024C, Tektronix)
- Preamplifier: high-pass filter (cutoff frequency 2 Hz) followed by an instrumentation amplifier with a gain of 10 (e.g. INA2141U, Texas Instruments) (Figure A2a, b).
- Pulse generator (e.g. Master-8, AMPI)

- Soundproof chamber (e.g. ENV-018MD-EMS, Med Associates)
- Stimulus isolator (e.g. ISO-Flex, AMPI)
- Torque sensor: Hall sensor (e.g. HGT-2101, Lake Shore Cryotronics), a ball bearing with I.D. 0.125", O.D. 0.25" (e.g. BBSRIF-418XXX, Quality Bearings & Components), a hollow tubing with an O.D. 0.25", a bent piece of metal, a disc magnet, and a female connector with wires (e.g. A8109-001, Omnetics Connector) (Figure A2a, b)

Surgical setup

- Amplifier for the neural signal (e.g. Model 1800, A-M Systems)
- Audio monitor (e.g. AM10, Grass Technologies)
- Bipolar stimulating electrode (e.g. 2 mil stainless steel microwire # 304, California Fine Wire; 2 pin socket nano connector A9069-001 for the electrode, 2 pin male nano connector A9577-001 for the holder, Omnetics Connector; 5 mil PFA-coated silver wire #786000 A-M systems; 2" long wooden stick). (Figure A3b)
- Dental burr (e.g. round carbide burr FG #1/4, Midwest)
- Dental drill handpiece (e.g. Tradition, Midwest)
- Dura pick (e.g. holder RS-6060, dissection needle RS-6067, Roboz Surgical Instrument)
- Glass capillary tube (e.g. 3-000-203-G/X, Drummond Scientific) for measuring the head angle
- Heating blanket (e.g. Homeothermic blanket system, Harvard Apparatus)

- Hypodermic needle (30G, e.g. 305106, Becton, Dickinson)
- Isofluorane vaporizer (e.g. Isotec 5, Datex Ohmeda)
- Light source (e.g. KL 1500 LCD, Schott)
- Miniature motorized microdrive (Fee and Leonardo, 2001; Otchy and Olveczky, 2012) (Figure A1b)
- Oscilloscope (e.g. TBS1062, Tektronix)
- Pulse generator (e.g. Master-8, AMPI)
- Recording electrode mounted on a microdrive for single-unit recording (e.g. 10 megohm quartz/platinum-tungsten fiber electrodes, Thomas Recording)
- Recording electrode used during surgery to observe spontaneous activity in RA and multiunit "hash" from antidromic stimulation (e.g. Carbostar-1, Kation Scientific)
- Rotary encoder (e.g. encoder: E4-300-118; cable: CA-FC5-SH-MIC4; display: ED3; all from US Digital) for measuring the head angle
- Soldering iron (e.g. MT1500, Weller)
- Stainless steel insect pins (e.g. 26002-10, Fine Science Tools)
- Stereoscope (e.g. MZ9.5, Leica) on a boom stand
- Stereotaxic device (e.g. Digital Stereotaxic Instrument with Fine Drive, Leica): a stereotaxic device designed for mice can be used by replacing the tooth bar with a beak clamp.
- Stimulus isolator (e.g. ISO-Flex, AMPI)
- Surgical instruments (scalpel, forceps, etc.)
- Wire cutter (e.g. 570E, Copper Hand Tools)

Methods

Here, we describe a detailed protocol for recording single-unit activity during singing in zebra finches. As an example, we describe procedures for recording from projection neurons in HVC (used as a proper name) by placing bipolar stimulating electrodes in the robust nucleus of the arcopallium (RA) and Area X (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007) (Figure A4a), the two main projection targets of HVC. However, the procedure is general and can be modified to record from other brain areas with minor modifications. Experiments were carried out in accordance with guidelines of the US National Institutes of Health and were reviewed and approved by the Massachusetts Institute of Technology Committee on Animal Care.

Microdrive implant surgery

1. First, prepare a bipolar stimulating electrode and its holder. Construct an electrode holder by attaching a 2 pin male nano connector onto a wooden stick (Figure A3b). Then, construct a bipolar stimulating electrode by inserting a 2 pin socket nano connector into the electrode holder and soldering the stripped ends of a silver wire (8 cm in length) and a stainless steel microwire onto each of the two pins of the connector (Figure A3b). The stainless steel microwires are implanted in the brain and the silver wires are used to connect the stimulating electrode with the microdrive. Straighten the stainless steel microwires, make them parallel spaced by 500 μm, and cut them to an appropriate length (3 mm for RA, 4 mm for Area X). Cover the pins of the connector and the solder joints using silicon elastomer. Finally, test the stimulating electrodes by connecting the silver wires of

the stimulating electrode to the stimulus isolator and placing the tip of the electrode in saline solution and passing a small current (~10 μA). Bubbles should come out only from the tip of the electrode that is connected to the negative voltage.

Use of special liquid soldering flux is necessary to solder the stainless steel microwires onto the connector pins. It is important to handle the stainless steel microwires with plastic forceps since sharp metal forceps can damage the insulation. This problem can be detected as bubbles coming out from the damaged site during the bubble test.

2. Prepare a microdrive (Figure A1b) for implant (Fee and Leonardo, 2001; Otchy and Olveczky, 2012) and adjust the length of the electrodes so that they will sit at an appropriate depth when implanted (Figure A1c). Fill the polyimide guide tubes with mineral oil to prevent fluid from entering during the implantation.

   The electrodes should be implanted to a depth of a few hundred micrometers above the brain region of interest. At this depth, the polyimide guide tube should be touching the surface of the brain without penetrating it. However, implanting the electrodes above HVC can be difficult since HVC is only a few hundred micrometers below the surface of the brain. In this case, electrodes can be implanted into HVC and retracted after the implant surgery so that the electrodes sit just above HVC until the bird recovers from the surgery.

3. Inject the bird with analgesic and antibiotics 20 minutes prior to the surgery.
4. Put the bird under anesthesia (e.g. Isofluorane, initial induction 3 % for 3 min, maintain at 1-2 %).

5. Pluck the feathers on the bird's head.
6. Put the bird on the stereotaxic device, and tighten the ear bars and the beak bar.
7. Clean the skin on the head with betadine and apply local anesthetic.
8. Make an incision in the skin along the midline with a scalpel, and retract the skin.

   Juvenile skulls are very soft. Avoid damaging it during this procedure.
9. Set the head angle to zero degrees.

   This is done by placing a glass capillary in the anterior midline groove just behind the beak. A zero degree head angle corresponds to this glass capillary being horizontal (check with a reference line on the stereotax).
10. Rotate the head forward 20 degrees using a glass capillary and a protractor, or using a rotary-encoded ear bar.
11. Using a dental drill with a round burr, make several pairs of small holes in the skull for the insect pins and one for the ground wire. Drill slowly so as not to damage the dura or the brain underneath.

    Insect pins act as anchors for the implant. This is not necessary in the adult since the dental acrylic can flow between the two layers of the skull. Make sure the holes are positioned so that the insect pins do not impede access to HVC, RA, or Area X. In addition, anterior sinus just behind the beak can be filled with dental acrylic and also be used as an anchor point.
12. Carefully thread a pin underneath the skull through each pair of holes. Put a small drop of silicon elastomer to cover both holes (Figure A1c).

    Insect pins should go through the holes on the skull tangentially without penetrating the dura.

13. In one of the holes, insert the ground wire a few millimeters under the skull so that it is positioned between the skull and the surface of the dura.

    Make sure that the ground wire is not floating in the air space between the inner and outer leaflets of the skull.

14. Identify the bifurcation of the mid-sagittal sinus (lambda), and use this as a reference point for stereotaxic coordinates.

    Juvenile skulls are thin and the mid-sagittal sinus can be seen through the overlying skull. For adults, it is necessary to thin the skull with a dental drill.

15. Implant a stimulating electrode in Area X:

    i. Mount the stimulating electrode on the stereotax using the stimulating electrode holder. Make sure the electrode is vertically straight by looking at it from the front and from the side to check the alignment with a reference line on the stereotax. To implant the electrode in Area X, move the midpoint of the bipolar stimulating electrode to 5.8 mm anterior, 1.5 mm lateral relative to lambda.

    ii. Make a craniotomy at the coordinates above.

        a) Drill through the outer leaflet of the skull.

        b) Thin the inner leaflet of the skull with a dental drill and remove it using fine forceps.

        Juvenile birds have a single-layered skull and thus step a) is unnecessary. Once a piece of skull is removed, it is important to prevent the surface of the brain and the

dura from drying. The brain can be kept moist by applying saline on the surface every few minutes.

c) Remove a small piece of dura using a dura pick.

iii. Gently lower the stimulating electrode and insert it into the brain 2.8 mm deep from the surface to target the center of Area X.

iv. Cover the exposed part of the craniotomy with silicon elastomer. Secure the stimulating electrode by using dental acrylic to connect the insect pins on the skull to the connector on the stimulating electrode. Then, carefully remove the electrode holder.

16. Implant a stimulating electrode in RA:

i. Rotate the head forward to a head angle of 90 degrees.

ii. Set the recording electrode in the manipulator and make sure it is vertically straight.

iii. Make a craniotomy at 1.8 mm posterior, 2.4 mm lateral for RA. Do not remove the dura at this stage.

iv. Using the recording electrode, search for the characteristic spontaneous activity (i.e. regular firing) of RA which starts around 1.5 mm beneath the surface. Make several penetrations 200 μm apart in the anterior-posterior axis and the medial-lateral axis to map out the extent of RA. Make sure to keep the brain surface moist throughout this process.

v. Find the center of RA and note its coordinates. Make a scratch on the skull in the vicinity of the craniotomy with a needle and note the coordinates of the center of this mark so that it could be used as a local reference point.

vi. Replace the recording electrode with a stimulating electrode, and by using the local reference, move the stimulating electrodes to the center of RA.

vii. Remove the dura and implant the stimulating electrode with one pole in the center of RA using the same procedure described for Area X.

17. Implant a microdrive in HVC:

i. Make the craniotomy above HVC at 45 degree head angle, 0.4 mm anterior, and 2.5 mm lateral. Do not remove the dura at this stage.

ii. Hook up the stimulating electrode (either RA or Area X) to the stimulus isolator and start stimulating every second (single pulse of 200 μs wide, 100 μA in amplitude) using the pulse generator.

iii. Insert the recording electrode in HVC, and search for the antidromic 'hash' (Figure A4b). Make several penetrations 200 μm apart and find the center of HVC and note its coordinates. Make a scratch on the skull in the vicinity of the craniotomy with a needle and note the coordinates of the center of this mark.

iv. Replace the recording electrode with a microdrive, and by using the local reference, move the microdrive electrodes to the center of HVC previously identified with the recording electrode.

v. Remove the dura and slowly lower the microdrive electrodes into the brain to the desired depth (Figure A1c).

Before lowering the electrodes, visually confirm that there is nothing in the way of electrodes and the virtual ground wire, and that the microdrive case will not bump into the stimulus electrodes. At the target

depth, the polyimide guide tubes should be touching the surface of the brain (Figure A1c). Filling the guide tubes with mineral oil prior to surgery should prevent any blood or fluid from entering the guide tubes, which can dry and prevent the electrodes from moving.

vi. Cover the exposed brain with silicon elastomer.

vii. Secure the microdrive in place by using dental acrylic to connect the microdrive case and the insect pins on the skull (Figure A1b).

viii. Solder the silver wires from the stimulating electrodes and the ground wire to the appropriate pins on the microdrive connector.

18. Close the incision anterior and posterior to the microdrive by pulling the skin up close to the dental acrylic and applying a small amount of tissue adhesive to the skin using a glass capillary.
19. Carefully remove the bird from the stereotaxic device and let the animal recover from the surgery.

Make sure that the height of the perch in the cage is appropriately set so that the microdrive will not get caught.

Single-unit recording during singing

1. Let the bird recover from the surgery for a few days. Wait until the bird starts singing.
2. Place the bird in the recording cage, and plug the cable into the microdrive (Figure A1a). Turn on the motorized commutator and make sure that it is working properly so that the cable does not become twisted over time.

3. Slowly lower the electrodes into HVC by moving the motors on the microdrive using the motor controller.

   Be careful not to activate the motor for more than one second at a time since this can overheat and damage the motor. Advance it with small steps, stopping for a few seconds in between each step.

4. Start stimulating RA or Area X (single pulse of 200 μs wide, 100 μA in amplitude) and observe the antidromic "hash" to confirm that the electrodes are in HVC (Figure A4b).

5. When the bird starts singing, isolate a single-unit by slowly moving the electrodes in 10-15 μm steps and waiting for a few minutes in between to check whether there is a single-unit on any of the electrodes. One can either search for a single-unit evoked by antidromic stimulation, or isolate a unit during singing and then check the response to antidromic stimulation to confirm its identity. Current required for the antidromic stimulation is typically within the range of 50-100 μA, but occasionally up to 300 μA.

   Note that local interneurons also respond to antidromic stimulation. Thus, other criteria, (e.g. latency variability of the evoked spike, spike collision test) will be necessary to distinguish projection neurons from local interneurons (Fuller and Schlag, 1976; Swadlow, 1998; Fee et al., 2004).

6. If the neuron fires spontaneously, one can perform a spike collision test. This is done by triggering antidromic stimulation off of a spontaneous spike with a latency of a few milliseconds. In this case, the antidromic spike should be blocked (Figure A4d).

We use a simple electrical circuit to detect threshold crossing events associated with occurrences of spikes (Figure A3a). The output of this circuit is used to trigger the pulse generator (e.g. Master-8) which then triggers the stimulus isolator (e.g. ISO-Flex). It is important to have a timer in the circuit (e.g. LM555 timer chip) that prevents multiple stimulations from being triggered for example by a burst of spikes. Without this mechanism, the brain will be stimulated for every spike in a burst which could lead to tissue damage.

7. If the single-unit is lost, advance the electrodes further to find a new unit.

   Holding time for a single-unit varies from a few minutes to a few hours.

8. Once the electrodes go through the entire depth of HVC, back up the electrodes above HVC.

9. Gently hold the bird and turn the lateral positioner on the microdrive so that the electrodes can penetrate into fresh tissue. This is typically done at the end of the day so that the bird has some time to recover from handling.

   Lateral positioner is typically rotated 1/8 of a turn, corresponding to roughly 20 μm.

10. Repeat these procedures over multiple days until the end of the experiment.

    Typically, one can perform one or two penetrations per day. To minimize the formation of gliosis in HVC, electrodes are retracted above HVC in between recording sessions.

11. At the end of the experiment, make an electrolytic lesion at the site of recording by passing a negative DC current (15 μA for 10-15 sec).

The size of the lesion varies depending on the brain area, and the current parameters need to be adjusted accordingly.

12. Use standard histological techniques to identify the electrolytic lesion and the electrode tracks to verify the recording site.

Troubleshooting

Lack of antidromic responses

A common problem encountered is that after the implant surgery, when the bird is connected to the cable, no antidromic response is observed (Step 4 of Single-unit recording during singing). This is likely due to the misplacement of the microdrive during surgery. To prevent this problem, it is important to implant the microdrive electrodes at precisely the same location as the site where the antidromic response was found using the recording electrode. To achieve this, we make a scratch on the surface of the skull in the vicinity of the craniotomy as a landmark (Step 17 (iii) of Microdrive implant surgery). This reduces the measurement error by providing a local reference. Another way to prevent this problem is to connect the microdrive to an amplifier during the implant process. This will ensure that the antidromic response is observed before the microdrive is secured to the skull.

Microphonics

During singing, vibration of the skull can be picked up by the electrodes and their connecting wires on the microdrive ("microphonics"). This can potentially interfere with neural recordings (Step 5 of Single-unit recording during singing). The occurrence of

microphonics can easily be detected by listening to the neural channel on the audio monitor since you will hear the song of the bird from the neural channel. To reduce the microphonics, it may help to secure the electrode connection wires to the microdrive case using a thin layer of acrylic. This will prevent these wires from picking up vibrations.

Movement artifacts

Another common problem is the presence of electrical artifacts associated with animal movement (Step 5 of Single-unit recording during singing). Differential recording between the recording electrode and the virtual ground reference electrode (a platinum-iridium wire located in the vicinity of the recording electrode; Figure A1b, c) helps remove the movement artifact (Fee and Leonardo, 2001). In addition, field-effect transistors (FETs) are used at the microdrive end of the cable (Figure A2c) in a source-follower configuration for each neural channel and the virtual ground signal (Figure A2d). These act as unity-gain buffer amplifiers and the signal follows the voltage of the electrodes. This also reduces the signal impedance seen by the cable and makes the signals more robust to noise and artifacts.

Discussion

Here, we have described a protocol for recording single-unit activity from antidromically-identified neurons in singing zebra finches. This protocol has several advantages. First, high-impedance electrodes can be used in a microdrive to obtain well-isolated single-units. Second, remote control of electrode position using the motorized microdrive allows neurons to be isolated without handling the bird. This is particularly advantageous when

working with juvenile birds since they may not sing for many hours after handling. Third, by using a lateral positioner to move the electrodes into a fresh tissue before each penetration, one can record from well-isolated neurons over the course of several weeks. One disadvantage is that it is difficult to record from a single-unit for more than a few hours. An alternative approach using an array of carbon fiber electrodes may improve long-term stability (Guitchounts et al., 2013).

The current setup in our lab allows an array of five high-impedance electrodes to be positioned by a single motor. Using this configuration, single-unit recordings from 3-8 projection neurons are typically obtained in a single penetration of HVC which corresponds to 30-40 recording sites through the 500 μm depth of HVC. A simultaneous recording of two neurons is obtained on approximately every other penetration (Figure A4c).

It is possible to combine single-unit recordings during singing with other manipulations. These include fiber tract transection (Olveczky et al., 2011), lesioning (Aronov et al., 2008; Goldberg and Fee, 2012; Kojima et al., 2013), and inactivation (Olveczky et al., 2011). Paired recordings from the cell body of a neuron and a calyceal axon terminal (Person and Perkel, 2007; Kojima and Doupe, 2009; Leblois et al., 2009) have been obtained during singing (Goldberg and Fee, 2012). Furthermore, this basic motorized microdrive design has been modified for intracellular recordings during singing (Long et al., 2010).

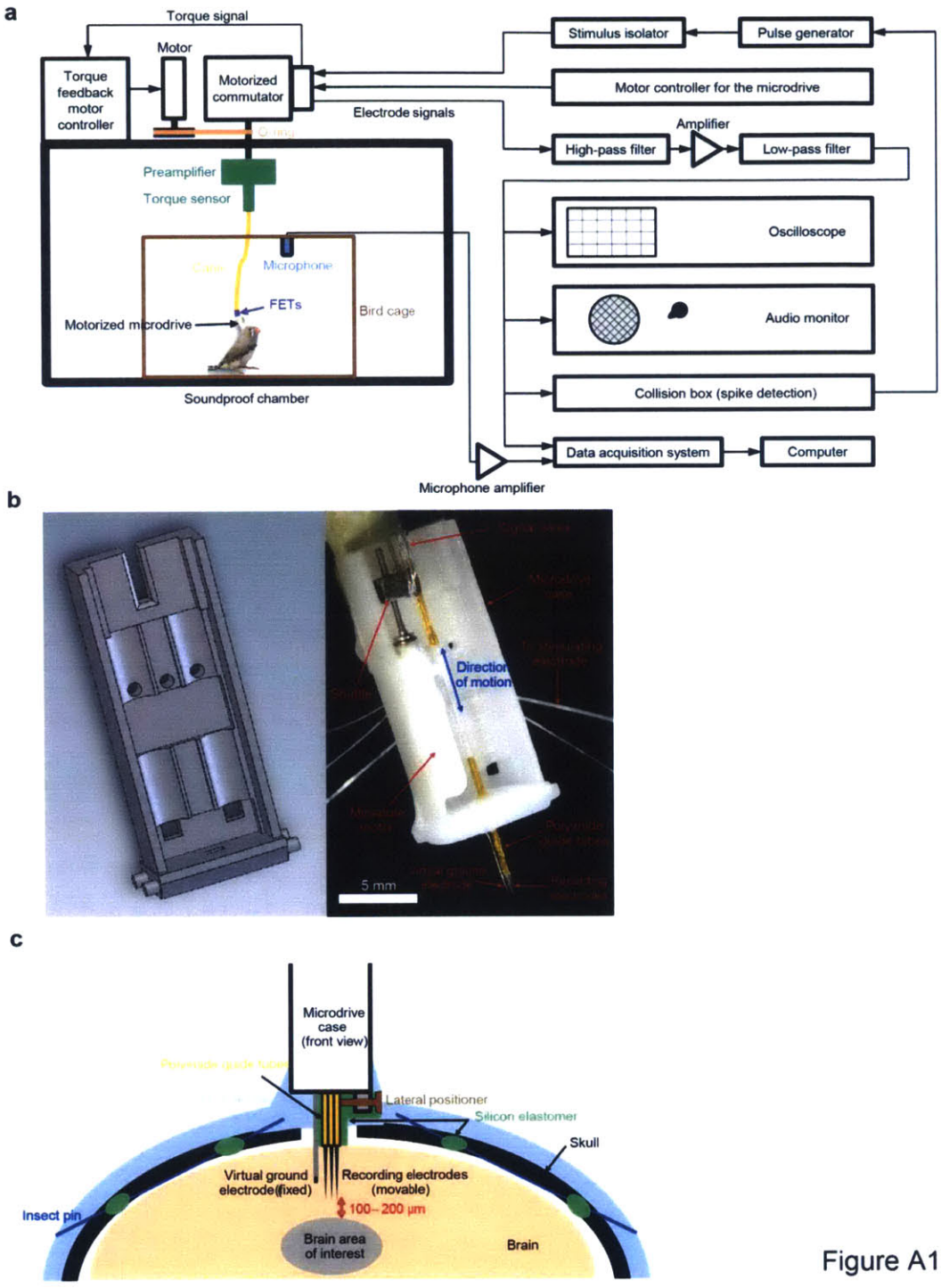

a
Torque signal
Motor
Torque feedback motor controller
Motorized commutator
Electrode signals
Stimulus isolator
Pulse generator
Motor controller for the microdrive
Amplifier
High-pass filter
Low-pass filter
Preamplifier
Torque sensor
Cable
Microphone
FETs
Motorized microdrive
Bird cage
Soundproof chamber
Oscilloscope
Audio monitor
Collision box (spike detection)
Data acquisition system
Computer
Microphone amplifier
b
Direction of motion
5 mm
c
Microdrive case (front view)
Polyimide guide tube
Lateral positioner
Silicon elastomer
Skull
Virtual ground electrode (fixed)
Recording electrodes (movable)
100–200 µm
Insect pin
Brain area of interest
Brain

Figure A1

Figure A1: Schematic of the recording setup and the microdrive implant.

a, Schematic of the electrophysiological recording setup. **b,** 3D CAD drawing (left) and a photograph of a motorized microdrive (right). Male Omnetics connector is located on the back. Lateral positioner is omitted for clarity but can be seen in panel c. Before the implant, the microdrive is protected by covering the open parts with a transparency film. **c,** Schematic showing positioning of the microdrive with respect to the skull and the brain. Insect pins go underneath the skull between the two holes in the skull and act as anchors. The lateral positioner provides a way to move the electrodes laterally so that they can enter fresh tissue during each penetration.

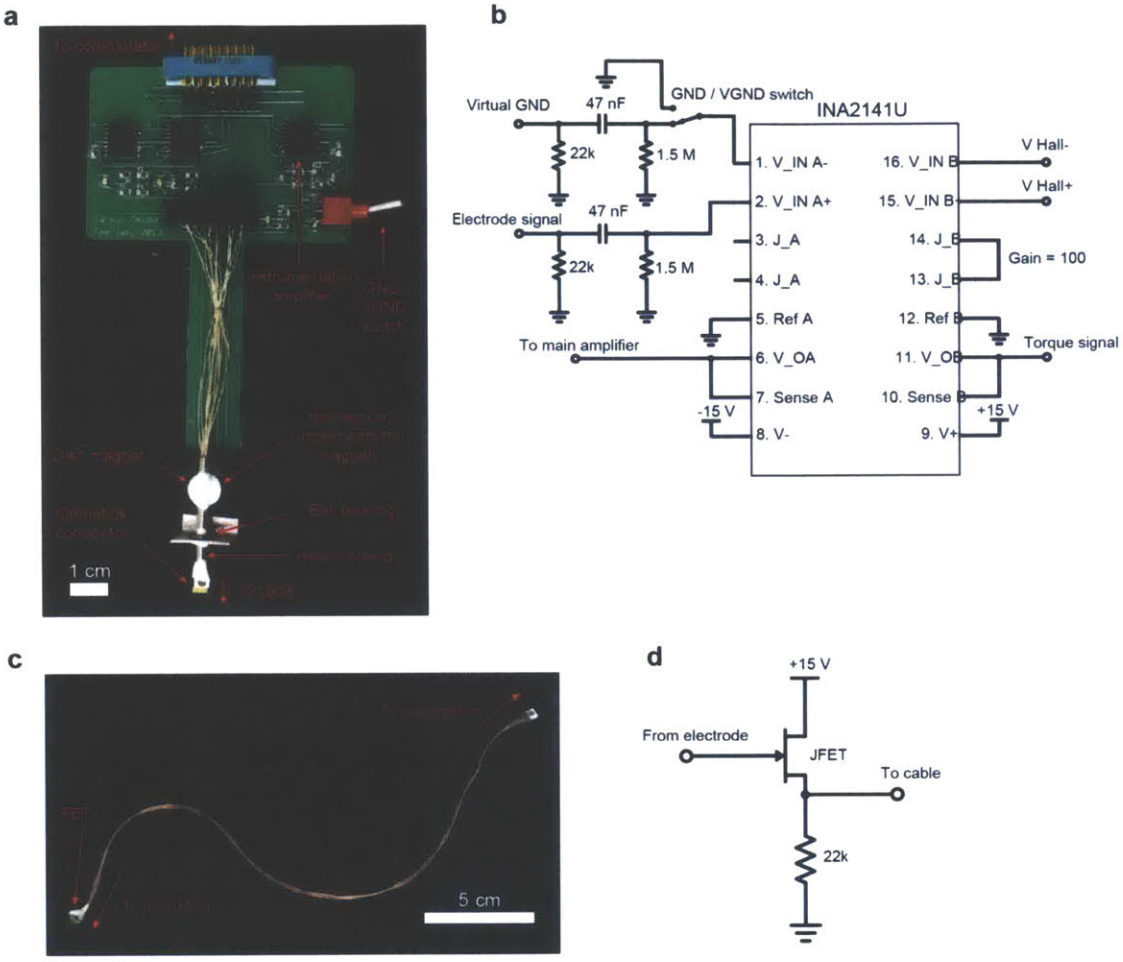

a
1 cm
b
Virtual GND
47 nF
GND / VGND switch
INA2141U
22k
1.5 M
Electrode signal
47 nF
22k
1.5 M
To main amplifier
-15 V
+15 V
1. V_IN A-
2. V_IN A+
3. J_A
4. J_A
5. Ref A
6. V_OA
7. Sense A
8. V-
16. V_IN B-
15. V_IN B+
14. J_B
13. J_B
12. Ref B
11. V_OB
10. Sense B
9. V+
V Hall-
V Hall+
Gain = 100
Torque signal
c
5 cm
d
+15 V
From electrode
JFET
To cable
22k

Figure A2

Figure A2: Preamplifier and cable.

a, Photograph of the preamplifier and the torque sensor. Rotation of the cable is detected by a Hall sensor placed underneath the disc magnet. **b,** Diagram of the preamplifier circuit. For simplicity, only one neural channel is drawn. 47 nF capacitor and 1.5 M ohm resistor act as a high-pass filter (cutoff frequency: 2.3 Hz) prior to the amplification. 22 k ohm resistor for the source follower (shown in Fig 2D) is also included. GND/VGND switch allows one to select whether GND or VGND is used for the differential amplification. Right half of INA2141U is used for amplifying the torque signal from the Hall sensor.

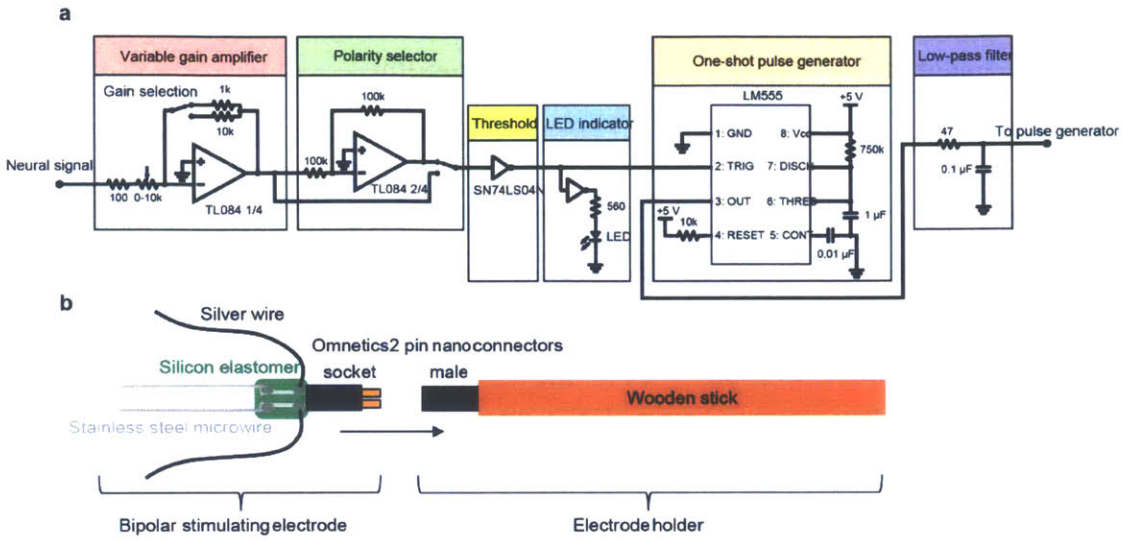

a
Variable gain amplifier
Gain selection
1k
10k
Neural signal
100
0-10k
TL084 1/4
Polarity selector
100k
100k
TL084 2/4
Threshold
SN74LS04
LED indicator
560
LED
One-shot pulse generator
LM555
+5 V
1: GND
2: TRIG
3: OUT
4: RESET
8: Vcc
7: DISC
6: THRES
5: CON
750k
1 µF
0.01 µF
+5 V
10k
Low-pass filter
47
0.1 µF
To pulse generator
b
Silver wire
Silicon elastomer
Omnetics2 pin nanoconnectors
socket
male
Wooden stick
Stainless steel microwire
Bipolar stimulating electrode
Electrode holder

Figure A3

Figure A3: Techniques for antidromic identification.

a, Diagram for a collision test circuit. Gain and polarity of the amplifier are manually adjusted online depending on the amplitude and polarity of the spike so that they can be detected by crossing the threshold of the inverter. Timer chip LM555 is used in the monostable mode to generate one-shot pulse to trigger the pulse generator (e.g. Master 8) which triggers the stimulus isolator (e.g. ISO-Flex). It also acts as a hold-off to prevent multiple stimulations faster than 1Hz. **b,** Schematic of the bipolar stimulating electrode and its holder.

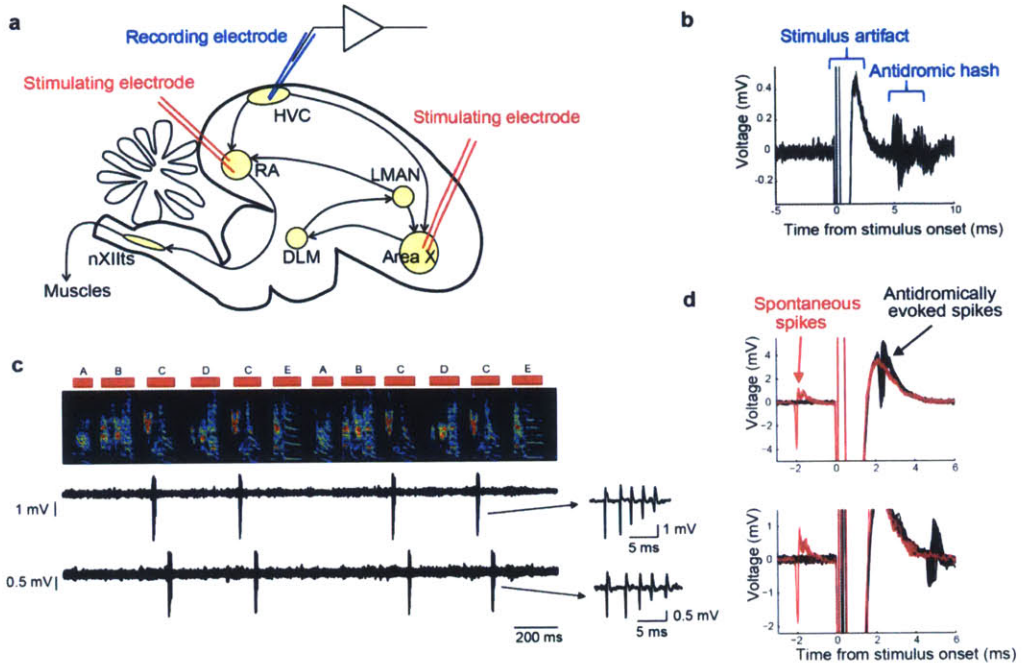

a
Recording electrode
Stimulating electrode
HVC
Stimulating electrode
RA
LMAN
nXIIts
DLM
Area X
Muscles
b
Stimulus artifact
Antidromic hash
Voltage (mV)
Time from stimulus onset (ms)
c
A B C D C E A B C D C E
1 mV
0.5 mV
200 ms
1 mV
5 ms
0.5 mV
5 ms
d
Spontaneous spikes
Antidromically evoked spikes
Voltage (mV)
Time from stimulus onset (ms)

Figure A4

Figure A4: Example recordings from HVC.

a, Schematic of the song system in the zebra finch indicating the locations of bipolar stimulating electrodes in RA and Area X, and a recording electrode in HVC. **b,** An example of a multiunit "hash" recorded in HVC following the electrical stimulation in RA (current amplitude 200 μA). Overlay of 20 trials aligned to the onset of the stimulation. **c,** An example of a simultaneous single-unit recording of two X-projecting HVC neurons during singing, recorded in an adult male zebra finch. (Top) Spectrogram of the song with syllable labels. (Bottom) Raw voltage traces from two different electrodes. (Right) Enlarged bursts showing individual action potentials. **d,** Collision tests for the neurons shown in panel c. Voltage traces are aligned to the onset of the electrical stimulation of Area X. Antidromically evoked spikes (black traces) are absent when the stimulation is triggered 2 milliseconds after spontaneous spikes (red traces).

Appendix B

Neural network model of sequence splitting

The simulation of a network model presented in this Appendix was developed in collaboration with Emily L. Mackevicius (Department of Brain and Cognitive Sciences, MIT) and Hannah L. Payne (Department of Neurobiology, Stanford University).

This appendix provides detailed methods on the neural network simulation presented in Chapter 5 (Figure 5.2, 5.3).

Methods

Binary neuron model

To illustrate a potential mechanism of chain splitting, we chose to implement the model as simply as possible. We modeled neurons as binary units and simulated their activity in discrete time steps (Fiete et al., 2010); at each time step (10 ms), the i-th neuron either bursts ($\boldsymbol{x_i = 1}$) or is silent ($\boldsymbol{x_i = 0}$).

Network architecture

A network of 100 binary neurons is recurrently connected in an all-to-all manner, with $\boldsymbol{W_{ij}}$ representing the synaptic strength from presynaptic neuron j to postsynaptic neuron i. Self-excitation is prevented by setting $\boldsymbol{W_{ii} = 0}$ for all i at all times (Fiete et al., 2010). Synaptic weights are initialized with random uniform distribution such that each neuron receives, on average, its maximum total input. During learning, the strength of each synapse is constrained to be within the interval $\boldsymbol{[0, w_{max}]}$, while the total incoming and outgoing weights of each neuron are both constrained by the "soft bound" $\boldsymbol{W_{max} = m * w_{max}}$ where m represents a target number of saturated synapses per neuron (Fiete et al., 2010) (see *Synaptic plasticity rule* section for details).

Network dynamics

The activity of each neuron in the network was determined in two steps; calculating the net feedforward input that comes from the previous time step, and determining whether that is enough to overcome the recurrent inhibition in the current time step.

First, the net feedforward input to the i-th neuron at time step t, $A_i^{net}(t)$, was calculated by summing the excitation, feedforward inhibition, neural adaptation, and external inputs:

$$A_i^{net}(t) = \left[A_i^E(t) - A^{Iff}(t) - A_i^{adapt}(t) + B_i(t) - \theta_i\right]_+$$

where $[z]_+$ indicates a rectification (equal to z if z>0 and 0 otherwise). $A_i^E(t) = \sum_j W_{ij}x_j(t-1)$ is the excitatory input from network activity on the previous time step. $A^{Iff}(t) = \beta\sum_j x_j\,(t-1)$ is a feedforward inhibitory input (Fiete et al., 2010), where β sets the strength of this feedforward inhibition. $A_i^{adapt}(t) = \alpha y_i$, is an adaptation term (Fiete et al., 2010) where α is the strength of adaptation, and y_i is a low-pass filtered record of recent activity in x_i with time constant $\tau_{adapt} = 40$ ms; that is $\tau_{adapt}\frac{dy_i}{dt} = -y_i + x_i$. $B_i(t)$ is the external input to neuron i at time t. For seed neurons, this term consists of training inputs (see section on *Seed neurons*). For non-seed neurons, it consists of random inputs with probability p_{in} = 0.01 in each time step and size $W_{max}/10$. Finally, θ_i is a threshold term used to reduce the excitability of seed neurons, making them less responsive to recurrent input than are other neurons in the network. For seed neurons $\theta_i = 10$ and for non-seed neurons $\theta_i = 0$. Including this term improves robustness of the training procedure by eliminating occasional situations in which seed neuron activity may be dominated by recurrent rather than external inputs. In these cases, external inputs may fail to exert proper control of network activity.

Second, we determined whether the i-th neuron will burst or not at time step t by examining whether the net feedforward input $A_i^{net}(t)$ exceeds the recurrent inhibition $A^{I\_rec}(t)$. We implemented recurrent inhibition by estimating the total activity of the network at time t:

$$A^{I\_rec}(t) = \gamma \sum_i A_i^{net}(t)$$

and feeding it back to all the neurons. Parameter γ sets the strength of the recurrent inhibition. We assume that this recurrent inhibition operates on a fast time scale(Kosche et al., 2015) (i.e. faster than the duration of a burst). Thus, the final output of the i-th neuron at time t becomes:

$$x_i(t) = \Theta[A_i^{net}(t) - A^{I\_rec}(t)]$$

where $\Theta[z]$ is the Heaviside step function (equal to 1 if $z > 0$and 0 otherwise). To induce splitting, γ was gradually stepped up to γ_{split} following a sigmoid with time constant τ_γ and inflection point t_0:

$$\gamma(t) = \frac{\gamma_{split}}{1 + e^{-(t-t_0)/\tau_\gamma}}$$

Seed neurons

A subset of neurons was designated as seed neurons, which received external training inputs used to shape network activity during learning (Buonomano, 2005; Jun and Jin, 2007). The external training inputs activate seed neurons at syllable onsets, reflecting the observed onset-related bursts of HVC neurons during the subsong stage (Figure 2.4i). The pattern of these inputs was adjusted in different stages of learning, and each strategy of syllable learning was implemented by different patterns of seed neuron training inputs.

Alternating differentiation (Figure 5.2a-e): Ten neurons were designated as seed neurons and received strong external input (W_{max}) to drive network activity. In the subsong stage, seed neurons were driven (by external inputs) synchronously and randomly with probability 0.1 in each time step corresponding to the random occurrence of syllable onsets in subsong (Aronov et al., 2008; Aronov et al., 2011). This was done only to visualize network activity; no learning was implemented at the subsong stage. During the protosyllable stage, seed neurons were driven synchronously and rhythmically with a period T = 100 ms. The protosyllable stage consisted of 500 iterations of 10 pulses each. To initiate chain splitting, the seed neurons were divided into two groups and each group was driven on alternate cycles. The splitting stage consisted of 2000 iterations of 5 pulses in each group of seed neurons (1 second total).

Motif strategy (Figure 5.3a-d): This was implemented in a similar manner as alternating differentiation, except that 9 seed neurons were used, and for the splitting stage, seed neurons were divided into 3 groups of 3 neurons, each driven on every third cycle.

Bout-onset differentiation (Figure 5.3e-h): Seed neurons were divided into two groups: 5 bout-onset seed neurons and 5 protosyllable seed neurons. At all learning stages, external inputs were organized into bouts consisting of four separate input pulses: Bout-onset seed neurons were driven at the beginning of each bout. Then, 30 ms later, protosyllable seed neurons were driven three times with an interval of T = 100 ms. In the protosyllable stage, inputs to all seed neurons were of strength W_{max}. In the splitting stage, the input to protosyllable seed neurons was decreased to $W_{max}/10$. This allowed

neurons in the bout-onset chain to suppress, through fast recurrent inhibition, the activity of protosyllable seed neurons during bout-onset syllables.

Each iteration of the simulation was 5 seconds long, consisting of 10 bouts, as described directly above, with random inter-bout intervals. The protosyllable stage consisted of 100 iterations, and the splitting stage consisted of 500 iterations.

Bout-onset syllable formation (Figure 10i-k): Input to seed neurons was set high ($2.5 * W_{max}$), and maintained at this high level throughout development. This prevented protosyllable seed neurons from being inhibited by neurons in the bout-onset chain. Furthermore, strong external input to the protosyllable seed neurons terminated activity in the bout-onset chain through fast recurrent inhibition, thus preventing further growth of the bout-onset chain, as occurs in bout-onset differentiation.

As in bout-onset differentiation, each iteration of the simulation was 5 seconds long, consisting of 10 bouts with random inter-bout intervals. The protosyllable stage consisted of 100 iterations, and the splitting stage consisted of 500 iterations.

Synaptic plasticity rule

As in previous models (Jun and Jin, 2007; Fiete et al., 2010), we hypothesized two plasticity rules in our model: Hebbian spike timing dependent plasticity (STDP) to drive sequence formation (Abbott and Blum, 1996; Dan and Poo, 2006), and heterosynaptic long term depression (hLTD) to introduce competition between synapses of a given neuron (Jun and Jin, 2007; Fiete et al., 2010). STDP is governed by the antisymmetric plasticity rule with a short temporal window (one burst duration):

$$\Delta_{ij}^{STDP}(t) = \eta \left[x_i(t)x_j(t-1) - x_i(t-1)x_j(t) \right]$$

where the constant η sets the learning rate. hLTD limits the total strength of weights for neuron *i*, and the summed weight limit rule for incoming weights is given by:

$$\Delta_{i*}^{hLTD}(t) = \eta \left[\sum_k \left(W_{ik}(t-1) + \Delta_{ik}^{STDP}(t) \right) - W_{max} \right]_+$$

and for outgoing weights from neuron j:

$$\Delta_{*j}^{hLTD}(t) = \eta \left[\sum_k \left(W_{kj}(t-1) + \Delta_{kj}^{STDP}(t) \right) - W_{max} \right]_+$$

At each time step, total change in synapse weight is given by the combination of STDP and hLTD:

$$\Delta W_{ij}(t) = \Delta_{ij}^{STDP}(t) - \varepsilon \Delta_{i*}^{hLTD}(t) - \varepsilon \Delta_{*j}^{hLTD}(t)$$

where ε sets the relative strength of hLTD.

Model parameters: subsong (Figure 5.2a)

In our implementation of the subsong stage, there was no learning. Subsong model parameters were: $\beta = 0.115, \alpha = 30, \eta = 0, \varepsilon = 0, \gamma = 0.01.$

Model parameters: alternating differentiation (Figure 5.2b-d)

After subsong, learning progressed in two stages: the protosyllable stage and the splitting stage. Parameters that remained constant over development were: $\beta = 0.115, \alpha = 30, \eta = 0.025, \varepsilon = 0.2$. To induce chain splitting, W_{max} was increased from 1 to 2, m was decreased from 10 to 5, and γ was increased from 0.01 to 0.18 following a sigmoid with time constant $\tau_\gamma = 200$ iterations and inflection point t_0= 500 iterations into the splitting stage. No change in parameters occurred prior to the chain-splitting stage.

Model parameters: motif strategy (Figure 5.3a-d)

Parameters that remained constant over development were: $\boldsymbol{\beta = 0.115,\ \alpha = 30,\ \eta = 0.025,\ \varepsilon = 0.2}$. To induce chain splitting, $\boldsymbol{w_{max}}$ was increased from 1 to 2, $\boldsymbol{m}$ was decreased from 9 to 3, and $\boldsymbol{\gamma}$ was increased from 0.01 to 0.18 following a sigmoid with time constant $\tau_\gamma = 200$ iterations and inflection point $t_0 = 500$ iterations into the splitting stage.

Model parameters: bout-onset differentiation (Figure 5.3e-h)

Parameters that remained constant over development were: $\boldsymbol{\beta = 0.13,\ \alpha = 30,\ \eta = 0\ .05,\ \varepsilon = 0.14}$. To induce chain splitting, $\boldsymbol{w_{max}}$ was increased from 1 to 2, $\boldsymbol{m}$ was decreased from 5 to 2.5, and $\boldsymbol{\gamma}$ was increased from 0.01 to 0.04 following a sigmoid with time constant $\tau_\gamma = 200$ iterations and inflection point $t_0 = 250$ iterations into the splitting stage.

Model parameters: formation of a new syllable at bout onset (Figure 5.3i-k)

Parameters that remained constant over development were: $\boldsymbol{\beta = 0.13,\ \alpha = 30,\ \eta = 0.05,\ \varepsilon = 0.15}$. To induce chain splitting, $\boldsymbol{w_{max}}$ was increased from 1 to 2, $\boldsymbol{m}$ was decreased from 5 to 2.5, and $\boldsymbol{\gamma}$ was increased from 0.01 to 0.05 following a sigmoid with time constant $\tau_\gamma = 200$ iterations and inflection point $t_0 = 250$ iterations into the splitting stage.

Shared and specific neurons

Neurons were classified as participating in a syllable type if the syllable onset-aligned histogram exhibited a peak that passed a threshold criterion. The criteria were chosen to include neurons where the histogram peak exceeded 90% of surrogate histogram peaks. Surrogate histograms were generated by placing one burst at a random latency in each syllable. (For example, in the protosyllable stage, the above criterion was found to be equivalent to having 5 bursts at the same latency in a bout of 10 protosyllables.) During the splitting phase, neurons were classified as shared if they participated in both syllable types, and specific if they participated in only one syllable type.

Visualizing network activity

We visualized network activity in two ways: network diagrams and raster plots of population activity (e.g. Figure 5.2a-d top and bottom panels, respectively). In both cases, we only included neurons that participated in at least one of the syllable types (see *Shared and specific neurons* above for participation criteria).

Network diagrams: Neurons are sorted along the x-axis based on their relative latencies. Neurons are sorted along the y-axis based on the relative strength of their synaptic input from specific neurons (or seed neurons) of each type (red or blue). Lines between neurons correspond to feedforward synaptic weights, and darker lines indicate stronger synaptic weights. For clarity of plotting, only the strongest six outgoing and strongest nine incoming weights are plotted for each neuron.

Population raster plots: Neurons are sorted from top to bottom according to their latency. Groups of seed neurons are indicated by magenta arrows. Shared neurons are

plotted at the top and specific neurons are plotted below. As for network diagrams, neurons that did not reliably participate in at least one syllable type were excluded.

Further details for Figure 5.2a-d: Panels show network and raster at four different stages. (a) subsong stage (before learning), (b) end of protosyllable stage (iteration 500), (c) early chain splitting stage (iteration 992), (d) late chain-splitting stage (iteration 2500),

Further details for Figure 5.3e-h: (e) early protosyllable stage (iteration 5), (f) late protosyllable stage (iteration 100), (g) early chain splitting stage (iteration 130), (h) late chain splitting stage (iteration 600).

Bibliography

Abbott LF, Blum KI (1996) Functional significance of long-term potentiation for sequence learning and prediction. Cereb Cortex 6:406-416.
Abe K, Matsui S, Watanabe D (2015) Transgenic songbirds with suppressed or enhanced activity of CREB transcription factor. Proc Natl Acad Sci U S A 112:7599-7604.
Abeles M (1991) Corticonics : neural circuits of the cerebral cortex. Cambridge ; New York: Cambridge University Press.
Agate RJ, Scott BB, Haripal B, Lois C, Nottebohm F (2009) Transgenic songbirds offer an opportunity to develop a genetic model for vocal learning. Proc Natl Acad Sci U S A 106:17963-17967.
Akutagawa E, Konishi M (2010) New brain pathways found in the vocal control system of a songbird. J Comp Neurol 518:3086-3100.
Aldridge JW, Berridge KC (1998) Coding of serial order by neostriatal neurons: a "natural action" approach to movement sequence. J Neurosci 18:2777-2787.
Ali F, Otchy TM, Pehlevan C, Fantana AL, Burak Y, Olveczky BP (2013) The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong. Neuron 80:494-506.
Amador A, Perl YS, Mindlin GB, Margoliash D (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. Nature 495:59-64.
Andalman AS, Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. Proc Natl Acad Sci U S A 106:12518-12523.
Andalman AS, Foerster JN, Fee MS (2011) Control of vocal and respiratory patterns in birdsong: dissection of forebrain and brainstem mechanisms using temperature. PLoS One 6:e25461.
Arendt D (2008) The evolution of cell types in animals: emerging principles from molecular studies. Nat Rev Genet 9:868-882.
Aronov D, Fee MS (2011) Analyzing the dynamics of brain circuits with temperature: design and implementation of a miniature thermoelectric device. J Neurosci Methods 197:32-47.
Aronov D, Andalman AS, Fee MS (2008) A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science 320:630-634.
Aronov D, Veit L, Goldberg JH, Fee MS (2011) Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. J Neurosci 31:16353-16368.
Ashmore RC, Wild JM, Schmidt MF (2005) Brainstem and forebrain contributions to the generation of learned motor behaviors for song. J Neurosci 25:8543-8554.
Bertram R, Daou A, Hyson RL, Johnson F, Wu W (2014) Two neural streams, one voice: pathways for theme and variation in the songbird brain. Neuroscience 277:806-817.
Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18:10464-10472.
Bokil H, Andrews P, Kulkarni JE, Mehta S, Mitra PP (2010) Chronux: a platform for analyzing neural signals. J Neurosci Methods 192:146-151.
Bottjer SW, Miesner EA, Arnold AP (1984) Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science 224:901-903.

Bottjer SW, Halsema KA, Brown SA, Miesner EA (1989) Axonal connections of a forebrain nucleus involved with vocal learning in zebra finches. J Comp Neurol 279:312-326.
Brainard MS, Doupe AJ (2013) Translating birdsong: songbirds as a model for basic and applied medical research. Annu Rev Neurosci 36:489-517.
Briggman KL, Kristan WB, Jr. (2006) Imaging dedicated and multifunctional neural circuits generating distinct behaviors. J Neurosci 26:10925-10933.
Briggman KL, Kristan WB (2008) Multifunctional pattern-generating circuits. Annu Rev Neurosci 31:271-294.
Briggman KL, Abarbanel HD, Kristan WB, Jr. (2005) Optical imaging of neuronal populations during decision-making. Science 307:896-901.
Buonomano DV (2005) A learning rule for the emergence of stable dynamics and timing in recurrent networks. J Neurophysiol 94:2275-2283.
Casasnovas B, Meyrand P (1995) Functional differentiation of adult neural circuits from a single embryonic network. J Neurosci 15:5703-5718.
Catchpole C, Slater PJB (2008) Bird song : biological themes and variations, 2nd Edition. Cambridge England ; New York: Cambridge University Press.
Charlesworth JD, Tumer EC, Warren TL, Brainard MS (2011) Learning the microstructure of successful behavior. Nat Neurosci 14:373-380.
Crandall SR, Aoki N, Nick TA (2007) Developmental modulation of the temporal relationship between brain and behavior. J Neurophysiol 97:806-816.
Dan Y, Poo MM (2006) Spike timing-dependent plasticity: from synapse to perception. Physiol Rev 86:1033-1048.
Daou A, Ross MT, Johnson F, Hyson RL, Bertram R (2013) Electrophysiological characterization and computational models of HVC neurons in the zebra finch. J Neurophysiol 110:1227-1245.
Denk W, Briggman KL, Helmstaedter M (2012) Structural neurobiology: missing link to a mechanistic understanding of neural computation. Nature reviews Neuroscience 13:351-358.
Desrochers TM, Jin DZ, Goodman ND, Graybiel AM (2010) Optimal habits can develop spontaneously through sensitivity to local cost. Proc Natl Acad Sci U S A 107:20512-20517.
Diesmann M, Gewaltig MO, Aertsen A (1999) Stable propagation of synchronous spiking in cortical neural networks. Nature 402:529-533.
Dixon PM (2013) Ripley's K Function. In: Encyclopedia of Environmetrics: John Wiley & Sons.
Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci 22:567-631.
Doupe AJ, Perkel DJ, Reiner A, Stern EA (2005) Birdbrains could teach basal ganglia research a new song. Trends Neurosci 28:353-363.
Doya K, Sejnowski TJ (1995) A Novel Reinforcement Model of Birdsong Vocalization Learning. In: Advances in Neural Information Processing Systems (Tesauro G, Touretzky DS, Leen TK, eds), pp 101-108. Cambridge: MIT Press.
Doya K, Sejnowski TJ (1998) A Computational Model of Birdsong Learning by Auditory Experience and Auditory Feedback. In: Central Auditory Processing and Neural Modeling (Poon PWF, Brugge JF, eds). New York: Springer.

Doya K, Sejnowski TJ (2000) A Computational Model of Avian Song Learning. In: The New Cognitive Neurosciences (Gazzaniga MS, ed). Cambridge: MIT Press.
Drew PJ, Abbott LF (2003) Model of song selectivity and sequence generation in area HVc of the songbird. J Neurophysiol 89:2697-2706.
Duda RO, Hart PE, Stork DG (2001) Pattern Classification, 2nd Edition. New York: Wiley.
Dutar P, Vu HM, Perkel DJ (1998) Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVc of the adult zebra finch. J Neurophysiol 80:1828-1838.
Farries MA, Perkel DJ (2002) A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. J Neurosci 22:3776-3787.
Fee MS, Leonardo A (2001) Miniature motorized microdrive and commutator system for chronic neural recording in small animals. J Neurosci Methods 112:83-94.
Fee MS, Scharff C (2010) The songbird as a model for the generation and learning of complex sequential behaviors. Ilar J 51:362-377.
Fee MS, Long MA (2011) New methods for localizing and manipulating neuronal dynamics in behaving animals. Current opinion in neurobiology 21:693-700.
Fee MS, Goldberg JH (2011) A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. Neuroscience 198:152-170.
Fee MS, Kozhevnikov AA, Hahnloser RH (2004) Neural mechanisms of vocal sequence generation in the songbird. Ann N Y Acad Sci 1016:153-170.
Feenders G, Liedvogel M, Rivas M, Zapka M, Horita H, Hara E, Wada K, Mouritsen H, Jarvis ED (2008) Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. PLoS One 3:e1768.
Fiete IR, Seung HS (2008) Neural network models of birdsong production, learning, and coding. In: Encyclopedia of Neuroscience (Squire LR, Spitzer N, Gage F, Albright T, Bloom F, eds). New York: Elsevier.
Fiete IR, Fee MS, Seung HS (2007) Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. J Neurophysiol 98:2038-2057.
Fiete IR, Hahnloser RH, Fee MS, Seung HS (2004) Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. J Neurophysiol 92:2274-2282.
Fiete IR, Senn W, Wang CZ, Hahnloser RH (2010) Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. Neuron 65:563-576.
Forssberg H (1999) Neural control of human motor development. Current opinion in neurobiology 9:676-682.
Foster EF, Bottjer SW (2001) Lesions of a telencephalic nucleus in male zebra finches: Influences on vocal behavior in juveniles and adults. J Neurobiol 46:142-165.
Fujii N, Graybiel AM (2003) Representation of action sequence boundaries by macaque prefrontal cortical neurons. Science 301:1246-1249.
Fujimoto H, Hasegawa T, Watanabe D (2011) Neural coding of syntactic structure in learned vocalizations in the songbird. J Neurosci 31:10023-10033.

Fuller JH, Schlag JD (1976) Determination of antidromic excitation by the collision test: problems of interpretation. Brain Res 112:283-298.
Garst-Orozco J, Babadi B, Olveczky BP (2014) A neural circuit mechanism for regulating vocal variability during song learning in zebra finches. Elife 4:e03697.
Gibb L, Gentner TQ, Abarbanel HD (2009) Inhibition and recurrent excitation in a computational model of sparse bursting in song nucleus HVC. J Neurophysiol 102:1748-1762.
Glaze CM, Troyer TW (2013) Development of temporal structure in zebra finch song. J Neurophysiol 109:1025-1035.
Goldberg JH, Fee MS (2010) Singing-related neural activity distinguishes four classes of putative striatal neurons in the songbird basal ganglia. J Neurophysiol 103:2002-2014.
Goldberg JH, Fee MS (2012) A cortical motor nucleus drives the basal ganglia-recipient thalamus in singing birds. Nat Neurosci 15:620-627.
Goldberg JH, Adler A, Bergman H, Fee MS (2010) Singing-related neural activity distinguishes two putative pallidal cell types in the songbird basal ganglia: comparison to the primate internal and external pallidal segments. J Neurosci 30:7088-7098.
Goldfield EC, Kay BA, Warren WH, Jr. (1993) Infant bouncing: the assembly and tuning of action systems. Child Dev 64:1128-1142.
Goldin MA, Alonso LM, Alliende JA, Goller F, Mindlin GB (2013) Temperature induced syllable breaking unveils nonlinearly interacting timescales in birdsong motor pathway. PLoS One 8:e67814.
Goller F, Daley MA (2001) Novel motor gestures for phonation during inspiration enhance the acoustic complexity of birdsong. Proc Biol Sci 268:2301-2305.
Goller F, Cooper BG (2004) Peripheral motor dynamics of song production in the zebra finch. Ann N Y Acad Sci 1016:130-152.
Graybiel AM (2008) Habits, rituals, and the evaluative brain. Annu Rev Neurosci 31:359-387.
Guitchounts G, Markowitz JE, Liberti WA, Gardner TJ (2013) A carbon-fiber electrode array for long-term neural recording. J Neural Eng 10:046016.
Hahnloser RH, Ganguli S (2013) Vocal Learning with Inverse Models. In: Principles of Neural Coding (Quiroga RQ, Panzeri S, eds): CRC Press.
Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature 419:65-70.
Hahnloser RH, Kozhevnikov AA, Fee MS (2006) Sleep-related neural activity in a premotor and a basal-ganglia pathway of the songbird. J Neurophysiol 96:794-812.
Hamaguchi K, Mooney R (2012) Recurrent interactions between the input and output of a songbird cortico-basal ganglia pathway are implicated in vocal sequence variability. J Neurosci 32:11671-11687.
Hamaguchi K, Tschida KA, Yoon I, Donald BR, Mooney R (2014) Auditory synapses to song premotor neurons are gated off during vocalization in zebra finches. Elife 3:e01833.

Hanuschkin A, Ganguli S, Hahnloser RH (2013) A Hebbian learning rule gives rise to mirror neurons and links them to control theoretic inverse models. Front Neural Circuits 7:106.
Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, Doya K (1999) Parallel neural networks for learning sequential procedures. Trends Neurosci 22:464-471.
Immelmann K (1969) Song development in the zebra finch and other estrildid finches. In: Bird Vocalizations (Hinde RA, ed), pp 61-74. Cambridge, UK: Cambridge University Press.
Isoda M, Tanji J (2002) Cellular activity in the supplementary eye field during sequential performance of multiple saccades. J Neurophysiol 88:3541-3545.
Jarvis ED (2004) Learned birdsong and the neurobiology of human language. Ann N Y Acad Sci 1016:749-777.
Jarvis ED, Yu J, Rivas MV, Horita H, Feenders G, Whitney O, Jarvis SC, Jarvis ER, Kubikova L, Puck AE, Siang-Bakshi C, Martin S, McElroy M, Hara E, Howard J, Pfenning A, Mouritsen H, Chen CC, Wada K (2013) Global view of the functional molecular organization of the avian cerebrum: mirror images and functional columns. J Comp Neurol 521:3614-3665.
Jarvis ED et al. (2005) Avian brains and a new understanding of vertebrate brain evolution. Nature reviews Neuroscience 6:151-159.
Jarvis MR, Mitra PP (2001) Sampling properties of the spectrum and coherency of sequences of action potentials. Neural Comput 13:717-749.
Jin DZ, Ramazanoglu FM, Seung HS (2007) Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. J Comput Neurosci 23:283-299.
Jin X, Costa RM (2010) Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature 466:457-462.
Jun JK, Jin DZ (2007) Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. PLoS One 2:e723.
Kanji GK (2006) 100 Statistical Tests, 3rd Edition. London ; Thousand Oaks, Calif.: Sage Publications.
Kawai R, Markman T, Poddar R, Ko R, Fantana AL, Dhawale AK, Kampff AR, Olveczky BP (2015) Motor cortex is required for learning but not for executing a motor skill. Neuron 86:800-812.
Kojima S, Doupe AJ (2009) Activity propagation in an avian basal ganglia-thalamocortical circuit essential for vocal learning. J Neurosci 29:4782-4793.
Kojima S, Kao MH, Doupe AJ (2013) Task-related "cortical" bursting depends critically on basal ganglia input and is linked to vocal plasticity. Proc Natl Acad Sci U S A 110:4756-4761.
Komiyama T, Sato TR, O'Connor DH, Zhang YX, Huber D, Hooks BM, Gabitto M, Svoboda K (2010) Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. Nature 464:1182-1186.
Konishi M (1985) Birdsong: from behavior to neuron. Annu Rev Neurosci 8:125-170.
Kosche G, Vallentin D, Long MA (2015) Interplay of Inhibition and Excitation Shapes a Premotor Neural Sequence. Journal of Neuroscience 35:1217-1227.

Kozhevnikov AA, Fee MS (2007) Singing-related activity of identified HVC neurons in the zebra finch. J Neurophysiol 97:4271-4283.
Kubota M, Taniguchi I (1998) Electrophysiological characteristics of classes of neuron in the HVc of the zebra finch. J Neurophysiol 80:914-923.
Le Feuvre Y, Fenelon VS, Meyrand P (1999) Central inputs mask multiple adult neural networks within a single embryonic network. Nature 402:660-664.
Leblois A, Bodor AL, Person AL, Perkel DJ (2009) Millisecond timescale disinhibition mediates fast information transmission through an avian basal ganglia loop. J Neurosci 29:15420-15433.
Leonardo A, Konishi M (1999) Decrystallization of adult birdsong by perturbation of auditory feedback. Nature 399:466-470.
Leonardo A, Fee MS (2005) Ensemble coding of vocal control in birdsong. J Neurosci 25:652-661.
Lever C, Burgess N, Cacucci F, Hartley T, O'Keefe J (2002a) What can the hippocampal representation of environmental geometry tell us about Hebbian learning? Biol Cybern 87:356-372.
Lever C, Wills T, Cacucci F, Burgess N, O'Keefe J (2002b) Long-term plasticity in hippocampal place-cell representation of environmental geometry. Nature 416:90-94.
Li M, Greenside H (2006) Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus HVC. Phys Rev E Stat Nonlin Soft Matter Phys 74:011918.
Lichtman JW, Smith SJ (2008) Seeing circuits assemble. Neuron 60:441-448.
Lichtman JW, Denk W (2011) The big and the small: challenges of imaging the brain's circuits. Science 334:618-623.
Lim Y, Shinn-Cunningham B, Gardner TJ (2012) Sparse Contour Representations of Sound. IEEE Signal Proc Let 19:684-687.
Lipkind D, Tchernichovski O (2011) Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. Proc Natl Acad Sci U S A 108 Suppl 3:15572-15579.
Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahasi M, Suzuki K, Feher O, Ravbar P, Okanoya K, Tchernichovski O (2013) Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. Nature 498:104-108.
Liu WC, Gardner TJ, Nottebohm F (2004) Juvenile zebra finches can use multiple strategies to learn the same song. Proc Natl Acad Sci U S A 101:18177-18182.
Long MA, Fee MS (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature 456:189-194.
Long MA, Jin DZ, Fee MS (2010) Support for a synaptic chain model of neuronal sequence generation. Nature 468:394-399.
Mandelblat-Cerf Y, Fee MS (2014) An automated procedure for evaluating song imitation. PLoS One 9:e96484.
Marder E, Bucher D (2001) Central pattern generators and the control of rhythmic movements. Curr Biol 11:R986-996.
Marder E, Bucher D (2007) Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs. Annu Rev Physiol 69:291-316.

Markowitz JE, Ivie E, Kligler L, Gardner TJ (2013) Long-range order in canary song. PLoS Comput Biol 9:e1003052.
Markowitz JE, Liberti WA, 3rd, Guitchounts G, Velho T, Lois C, Gardner TJ (2015) Mesoscopic patterns of neural activity support songbird cortical sequences. PLoS Biol 13:e1002158.
Markram H, Lubke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275:213-215.
Marler P (1970) Birdsong and speech development: could there be parallels? Am Sci 58:669-673.
Masamizu Y, Tanaka YR, Tanaka YH, Hira R, Ohkubo F, Kitamura K, Isomura Y, Okada T, Matsuzaki M (2014) Two distinct layer-specific dynamics of cortical ensembles during learning of a motor task. Nat Neurosci 17:987-994.
McDonald JH (2014) Handbook of Biological Statistics, 3rd Edition. Baltimore: Sparky House Publishing.
Mehta MR, Lee AK, Wilson MA (2002) Role of experience and oscillations in transforming a rate code into a temporal code. Nature 417:741-746.
Mello CV (2014) The Zebra Finch, Taeniopygia guttata: An Avian Model for Investigating the Neurobiological Basis of Vocal Learning. Cold Spring Harbor Protocols 2014:pdb emo084574.
Mitra P, Bokil H (2008) Observed Brain Dynamics: Oxford University Press.
Mooney R (2000) Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. J Neurosci 20:5420-5436.
Mooney R (2009) Neural mechanisms for learned birdsong. Learn Mem 16:655-669.
Mooney R, Prather JF (2005) The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. J Neurosci 25:1952-1964.
Nathans J, Thomas D, Hogness DS (1986) Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science 232:193-202.
Nottebohm F, Stokes TM, Leonard CM (1976) Central Control of Song in Canary, Serinus-Canarius. Journal of Comparative Neurology 165:457-486.
Nottebohm F, Kelley DB, Paton JA (1982) Connections of vocal control nuclei in the canary telencephalon. J Comp Neurol 207:344-357.
Nusbaum MP, Beenhakker MP (2002) A small-systems approach to motor pattern generation. Nature 417:343-350.
Ohno S (1970) Evolution by Gene Duplication: Springer-Verlag.
Okubo TS, Mackevicius EL, Fee MS (2014) In Vivo Recording of Single-Unit Activity during Singing in Zebra Finches. Cold Spring Harbor Protocols.
Oller DK (1980) The emergence of the sounds of speech in infancy. In: Child Phonology (Yeni-Komshian GH, Kavanagh JF, Ferguson CA, eds): Academic Press, Inc.
Olveczky BP, Andalman AS, Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol 3:e153.
Olveczky BP, Otchy TM, Goldberg JH, Aronov D, Fee MS (2011) Changes in the neural control of a complex motor sequence during learning. J Neurophysiol 106:386-397.
Oppenheim AV, Schafer RW (2004) From frequency to quefrency: A history of the Cepstrum. IEEE Signal Proc Mag 21:95-106.

Otchy TM, Olveczky BP (2012) Design and assembly of an ultra-light motorized microdrive for chronic neural recordings in small animals. J Vis Exp.
Person AL, Perkel DJ (2007) Pallidal neuron activity increases during sensory relay through thalamus in a songbird circuit essential for learning. J Neurosci 27:8687-8698.
Peters AJ, Chen SX, Komiyama T (2014) Emergence of reproducible spatiotemporal activity during motor learning. Nature 510:263-267.
Pfenning AR et al. (2014) Convergent transcriptional specializations in the brains of humans and song-learning birds. Science 346:1256846.
Pidoux M, Bollu T, Riccelli T, Goldberg JH (2015) Origins of basal ganglia output signals in singing juvenile birds. J Neurophysiol 113:843-855.
Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature 451:305-310.
Rajan R, Doupe AJ (2013) Behavioral and neural signatures of readiness to initiate a learned motor sequence. Curr Biol 23:87-93.
Ravbar P, Lipkind D, Parra LC, Tchernichovski O (2012) Vocal exploration is locally regulated during song learning. J Neurosci 32:3422-3432.
Rehm KJ, Deeg KE, Marder E (2008) Developmental regulation of neuromodulator function in the stomatogastric ganglion of the lobster, Homarus americanus. J Neurosci 28:9828-9839.
Reiner A, Yamamoto K, Karten HJ (2005) Organization and evolution of the avian forebrain. Anat Rec A Discov Mol Cell Evol Biol 287:1080-1102.
Rieke F (1997) Spikes : Exploring the Neural Code. Cambridge, Mass.: MIT Press.
Ripley BD (1981) Spatial Statistics. New York: Wiley.
Saar S, Mitra PP (2008) A technique for characterizing the development of rhythms in bird song. PLoS One 3:e1461.
Sawyer RT (1986) Leech biology and behaviour. Oxford: Oxford Science Publications.
Scharff C, Nottebohm F (1991) A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. J Neurosci 11:2896-2913.
Scharff C, Kirn JR, Grossman M, Macklis JD, Nottebohm F (2000) Targeted neuronal death affects neuronal replacement and vocal behavior in adult songbirds. Neuron 25:481-492.
Scott BB, Velho TA, Sim S, Lois C (2010) Applications of avian transgenesis. Ilar J 51:353-361.
Seeds AM, Ravbar P, Chung P, Hampel S, Midgley FM, Jr., Mensh BD, Simpson JH (2014) A suppression hierarchy among competing motor programs drives sequential grooming in Drosophila. Elife 3:e02951.
Seung HS (2009) Reading the book of memory: sparse sampling versus dense mapping of connectomes. Neuron 62:17-29.
Simpson HB, Vicario DS (1990) Brain pathways for learned and unlearned vocalizations differ in zebra finches. J Neurosci 10:1541-1556.
Sober SJ, Wohlgemuth MJ, Brainard MS (2008) Central contributions to acoustic variation in birdsong. J Neurosci 28:10370-10379.

Stark RE (1980) Stages of speech development in the first year of life. In: Child Phonology (Yeni-Komshian GH, Kavanagh JF, Ferguson CA, eds): Academic Press, Inc.
Strogatz SH (1994) Nonlinear dynamics and Chaos : with applications to physics, biology, chemistry, and engineering. Reading, Mass.: Addison-Wesley Pub.
Suthers RA, Zollinger SA (2004) Producing song: the vocal apparatus. Ann N Y Acad Sci 1016:109-129.
Swadlow HA (1998) Neocortical efferent neurons with very slowly conducting axons: strategies for reliable antidromic identification. J Neurosci Methods 79:131-141.
Tanji J (2001) Sequential organization of multiple movements: involvement of cortical motor areas. Annu Rev Neurosci 24:631-651.
Taylor JS, Raes J (2004) Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet 38:615-643.
Tchernichovski O, Mitra PP (2002) Towards quantification of vocal imitation in the zebra finch. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 188:867-878.
Tchernichovski O, Mitra PP, Lints T, Nottebohm F (2001) Dynamics of the vocal imitation process: how a zebra finch learns its song. Science 291:2564-2569.
Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP (2000) A procedure for an automated measurement of song similarity. Anim Behav 59:1167-1176.
Tchernichovski O, Lints TJ, Deregnaucourt S, Cimenser A, Mitra PP (2004) Studying the song development process: rationale and methods. Ann N Y Acad Sci 1016:348-363.
Thelen E (1981) Rhythmical Behavior in Infancy - an Ethological Perspective. Dev Psychol 17:237-257.
Thelen E (1995) Motor development. A new synthesis. Am Psychol 50:79-95.
Tinbergen N (1951) The study of instinct. Oxford Eng.: Clarendon Press.
Troyer TW, Doupe AJ (2000a) An associational model of birdsong sensorimotor learning II. Temporal hierarchies and the learning of song sequence. J Neurophysiol 84:1224-1239.
Troyer TW, Doupe AJ (2000b) An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. J Neurophysiol 84:1204-1223.
Troyer TW, Bottjer SW (2001) Birdsong: models and mechanisms. Current opinion in neurobiology 11:721-726.
Tumer EC, Brainard MS (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature 450:1240-1244.
Vallentin D, Long MA (2015) Motor origin of precise synaptic inputs onto forebrain neurons driving a skilled behavior. J Neurosci 35:299-307.
Veit L, Aronov D, Fee MS (2011) Learning to breathe and sing: development of respiratory-vocal coordination in young songbirds. J Neurophysiol 106:1747-1765.
Vu ET, Mazurek ME, Kuo YC (1994) Identification of a forebrain motor programming network for the learned song of zebra finches. J Neurosci 14:6924-6934.
Vu ET, Schmidt MF, Mazurek ME (1998) Interhemispheric coordination of premotor neural activity during singing in adult zebra finches. J Neurosci 18:9088-9098.

Walton C, Pariser E, Nottebohm F (2012) The zebra finch paradox: song is little changed, but number of neurons doubles. J Neurosci 32:761-774.
Wang CZ, Herbst JA, Keller GB, Hahnloser RH (2008) Rapid interhemispheric switching during vocal production in a songbird. PLoS Biol 6:e250.
Wild JM (2004) Functional neuroanatomy of the sensorimotor control of singing. Ann Ny Acad Sci 1016:438-462.
Wild JM, Williams MN, Howie GJ, Mooney R (2005) Calcium-binding proteins define interneurons in HVC of the zebra finch (Taeniopygia guttata). J Comp Neurol 483:76-90.
Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. Science 273:1871-1875.
Zann RA (1996) The Zebra Finch: A Synthesis of Field and Laboratory Studies. Oxford, UK: Oxford University Press.
Zeigler HP, Marler P (2008) Neuroscience of Birdsong. Cambridge England ; New York: Cambridge University Press.
Ziv Y, Burns LD, Cocker ED, Hamel EO, Ghosh KK, Kitch LJ, El Gamal A, Schnitzer MJ (2013) Long-term dynamics of CA1 hippocampal place codes. Nat Neurosci 16:264-266.