THE UNIVERSITY OF CHICAGO

AUDITORY FEEDBACK AND NEURONAL DYNAMICS IN THE SONGBIRD FOREBRAIN

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON NEUROBIOLOGY

BY

GRAHAM CAVIN FETTERMAN

CHICAGO, ILLINOIS

JUNE 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The work in this thesis would not have been possible without the assistance and support of a great many people.

Primacy of place must go to my advisor, Dan Margoliash. Dan has consistently pushed me to refine my reasoning when designing experiments and analyses – to think through in greater detail the meaning of a hypothesis or result before accepting it, and to never lose sight of either ethological relevance or evolutionary implications. What quality I can claim as a scientist today is to a significant degree due to this critical encouragement. This thesis also involved a number of challenges, from failing equipment to a global pandemic, and Dan has throughout projected a relentless but pragmatic optimism. No setback could daunt him. If I can retain a measure myself of that calm buoyancy in the face of reversal, I will consider it a real accomplishment. His deep commitment to the well-being of those in his charge, beyond their scientific accomplishments, is another excellent quality I hope to emulate. Dan, thank you.

Several other scientists were influential in guiding me towards this path. I would like to thank Melanie Fields, who first showed me the intricate beauty to be found throughout biology, and in the nervous system in particular. I came to understand how much of science involves cajoling temperamental equipment and diagnosing unexpected failures in a peerless laboratory course taught by Marvin Makinen. He showed me that a measurement obtained from a machine is meaningless if you don't understand how the machine works – as well as the pivotal role played in this process by coffee. David Biron welcomed me into his lab before it even existed, and when I had little experience conducting any kind of research. He encouraged me to learn everything I possibly could, from machining to plasmid design to computer vision, on the off chance it might come in handy. Nearly everything has. And it was David who insisted that if I were going to study

neuroscience at UChicago, I would be a fool to pass up the chance to rotate with some professor named “Margoliash.”

My thesis committee – Nicho Hatsopoulos, Murray Sherman, and Jason Maclean – has provided excellent feedback and guidance throughout this process. My work has benefitted a great deal from their advice, and even more from their probing questions.

I could not have asked for a better group of colleagues than the members of the lab, past and present. Arij Daou taught me how to implant accelerometers, and brought crucial mathematical and cell-biological perspectives to our discussions. Kyler Brown shepherded me through my first experiments in lab, including recording my first BOS-selective HVC neuron, and taught me a great deal about data analysis, effective software design, and surviving graduate school. Sofija Canavan and I suffered together through the trials of surgical refinement and implant design, and kept one another sane throughout. Nelson Medina’s enthusiasm for science and lunch was infectious, and made him an admirable officemate and whiteboard co-conspirator. Daniel Baleckaitis provided so much help it cannot be listed exhaustively – but it included animal care, organizational assistance, device manufacture, and particularly surgical training. His good humor and patience, as well, were a lodestone for the lab’s culture. I do not understand how any lab functions without him. All of these lab members, as well as Etienne Manderscheid, Tim Brawn, Peter Malonis, Daniel Lam, Andy Savoy, and Mia Paletta, provided good company and rousing scientific discussion. And I must not omit our small honorary postdocs: the (sadly, late) Professor Grey, and the incorrigible Dr Patches.

Finally, I would like to thank my friends and family. Friends old and new have helped me weather the stress and toil of thesis research. My parents, Marc and Nancy, have been, as always, my unflagging cheerleaders, at once pushing me to do better and cautioning me against burnout. And my wife Anna has been a pillar of support and patience. I cannot overstate her importance.

ABSTRACT

Birdsong is a useful model behavior for the study of the production of learned vocalizations. Its neural substrates are reasonably well mapped, but how the brain regions responsible for song generation accomplish this goal is still poorly understood. Much work has focused on the forebrain sensorimotor nucleus HVC, which is critical for song production and learning. Observations of extremely sparse bursts produced by projection neurons in HVC have been used to argue for a representation in this nucleus of time alone, with no relation to ongoing motor behavior.

Other prior work in HVC has produced compelling evidence for rapid changes (in less than 4 hours) in intrinsic properties of a particular class of neuron in HVC as a result of a behavioral manipulation, delayed auditory feedback (DAF). As almost all examples of behavioral manipulation of auditory feedback in the zebra finch (a common model organism in birdsong research) require experimental durations of days to weeks, the possibility of a behavioral correlate of these intrinsic property changes was enticing.

In this thesis, I examined both auditory feedback-related phenomena and HVC neuronal activity. First, I exposed birds to short periods of DAF and analyzed their singing behavior for changes, which were apparent nearly immediately. These changes were particularly related to the production of a preparatory behavior just prior to song onset. These results provide evidence of auditory feedback integration in zebra finch song at short latency, as well as for the use of auditory feedback to actively shape preparatory activity.

Second, I reanalyzed a rich dataset of HVC neuronal activity and found that basal ganglia-projecting neurons in HVC produce bursts which comprise sequences during singing that appear to act as copies of one another, allowing information about neuronal activity in the past to be projected forward into the future. These recurring sequences could constitute a mechanism by

which feedback, including auditory feedback, is integrated into ongoing HVC activity to produce a representation of song performance quality, a necessary precursor to correction of erroneous behavioral output.

Third, I presented to sleeping birds auditory stimuli resembling delayed auditory feedback (pseudo-DAF), and observed responses involving both suppressed firing as well as potentiated activity relative to a well-studied auditory stimulus, the bird's own song (BOS). These responses help to explain how DAF produces the changes in singing activity and neuronal intrinsic properties described above – abnormal neuronal responses interfere with ongoing HVC activity, producing error signals which ultimately result in behavioral change.

Altogether, these results argue for the importance of auditory feedback to ongoing song production in the zebra finch. They also suggest likely features of neuronal responses to vocal production errors, and they outline a potential mechanism by which errors (assessed via auditory feedback or otherwise) are recognized and propagated within and beyond HVC.

CHAPTER 1

INTRODUCTION

Learned vocalization

Vocal production in any vertebrate requires precise control over small muscles for extended periods of time, in tight coordination with respiratory activity. Moreover, these respiratory control systems are evolutionarily older (Brocklehurst et al, 2020), meaning that not only must vocal production be structured to allow respiration to occur, it must do so by interfacing with and activating networks that evolved under different constraints. These factors make it a particularly complex type of motor control.

Vocalizing itself is common in vertebrates, but far less common is vocal learning, the production of vocalizations which are not available to an animal in the absence of prior exposure. A common example of a non-learned (or innate) vocalization is the meow of the domestic cat (*Felis catus*), which is stereotyped and yet identifiable across individuals, and readily produced by juveniles, even if they have been removed from conspecifics from an early age (Moelk, 1944; Bradshaw, 2000). (Note that the domestic cat has been under artificial selection pressure for millennia since its domestication (Vigne et al, 2004; Driscoll et al, 2007; Ottoni et al, 2017), and the persistence of juvenile-like vocal behavior into adulthood may be one result of this pressure. Acclimation to human contact may be a factor as well – feral adult cats generally do not vocalize (Jensen, 2009).) Such vocalizations are referred to as "calls." Most vertebrate species produce a range of distinct call types, with the choice of call dependent on the animal's motivational state and its social context (Smith, 1977). Ethologists envisioned call production as driven by so-called sign stimuli that act as call-specific releasers (Tinbergen, 1951). The context of call production can be

developmentally labile, however: one example is vervet monkey alarm calling, which has been claimed to be referential (Seyfarth et al, 1980; Cheney and Seyfarth, 1988; Seyfarth and Cheney, 1990; Price et al, 2015). Learning the context of which call to produce and when to produce it is a plausible intermediate step to full vocal learning behavior (Takahashi et al, 2015; Margoliash and Tchernichovski, 2015).

There is robust evidence for vocal learning in numerous animal taxa. In birds, oscine passerines (songbirds), psittacines (parrots) (Bradbury, 2013; Walløe et al, 2015), and trochilids (hummingbirds) (Snow, 1968; Wiley, 1971; Nottebohm, 1972) all display vocal learning. These three groups do not form an evolutionarily distinct clade themselves, suggesting the parallel evolution of vocal learning in these groups, but passerines and psittacines do form a clade, opening the possibility that they are examples of the evolution of vocal learning from a common ancestor, with subsequent loss of the behavior in their non-vocal-learning sister clades. The existence of mixed patterns of vocal learning in different species of suboscine species arguably supports this hypothesis (Liu et al, 2013; Kroodsma et al, 2013; Amador et al, 2008; Kroodsma and Konishi, 1991; ten Cate, 2021), but some authors have suggested instead that vocal learning arose independently in a subset of suboscines (Kroodsma et al, 2013; ten Cate, 2021).

Vocal learning is clearly a feature of human behavior, and there is also strong evidence for vocal learning in other mammals, including multiple species of bats (Knörnschild, 2014), cetaceans (both whales (Noad et al, 2000; Janik, 2014) and dolphins (Richards et al, 1984; Fripp et al, 2005; King et al, 2013; Janik, 2013; consider particularly the Fripp and King studies, which describe vocal learning in ethologically-relevant contexts and between dolphin conspecifics, rather than experimenter-instigated trials), and seals (Reichmuth and Casey, 2014; Stansbury and Janik, 2021; Sanvito et al, 2007). The patchwork pattern of vocal learning and its apparent absence across mammalian species makes a case for parallel evolution of vocal learning in mammals, as does the presence of vocal

learning in the taxonomically distant birds and mammals.

Weaker evidence has been reported for vocal learning in many other mammals, including elephants (Poole et al, 2005; two case studies, neither of which occurs between conspecifics or in a naturalistic context, but one must bear in mind the likely extreme difficulty in obtaining elephant data), some nonhuman primates (Takahashi et al, 2015; Pika et al, 2003; Russell et al, 2013), mice (Arriaga and Jarvis, 2013), and goats (Briefer and McElligott, 2012). Jarvis has proposed (Arriaga and Jarvis, 2013; Jarvis, 2019) that a continuum of vocal behavior exists, with different species and clades displaying varying degrees of ability to modulate vocalizations based on experience. While this suggestion is not implausible, the evidence in support of the arguments put forward to date is not compelling. The rodent evidence, on which Jarvis leans heavily, is rather weak. Indeed, Janik and Knörnschild (2021) propose several alternative explanations for the mouse results and summarize evidence arguing against vocal learning as commonly understood occurring in mice. Like many discussions which hinge on distinguishing a continuum from a sharp delineation, it is largely philosophical – currently unresolvable, and likely to remain so, particularly given the difficulty and expense of a systematic effort to survey vocal learning across taxa.

Choice of animal model

Practical considerations rule out many unambiguous vocal learners as useful experimental subjects for the study of vocal behavior – humans, seals, and cetaceans have long generation times and are rather large. They raise important ethical concerns as well. If one intends to probe the neurophysiological basis of vocal behavior and learning, these groups are thus likely to provide primarily either case-study evidence or non- or lightly-interventional evidence.

Vocal learning birds are, then, a likely target for study. Choosing a species on which to focus

efforts is a challenge, given the diversity of the avian vocal learning clades (particularly songbirds), and has largely occurred by a combination of chance, availability, and convenience. Parrots are skilled imitators, but temperamentally somewhat poorly suited to captivity, anxiety-like disorders being common even in pet parrots (Luescher, 2006). Hummingbirds are highly challenging to maintain in captivity, and so studies have often been performed in large aviaries, which has constrained the types of experiments which can be performed, as well as placing limits on the feasibility of setting up a facility (Meadows et al, 2012; Jarvis et al, 2000). Many songbird species are also difficult to maintain in captivity. Another consideration is the ability to induce successful breeding in captivity, which has both practical and scientific benefits – ready access to juveniles allows an entire panoply of developmental studies involving experimental manipulation that can provide unique insight that studies restricted to adult animals do not. This further limits the passerine species available for study.

One songbird species which overcomes these constraints has come to dominate the birdsong literature (but see below): the zebra finch (*Taeniopygia castanotis*), an estrildid finch native to Australia. ("Finch," here, is used colloquially – estrildids are not true finches (Beecher, 1953; Ericson et al, 2000).) Zebra finches are widely available from commercial breeders (and pet stores) throughout the world, tolerate a range of temperatures and feeding regimes, and will breed readily in captivity, with little to no seasonality if day-night cycles are held constant to match that of their breeding season in the wild (Zann, 1996).

Zebra finches are also attractive experimental subjects for conceptual reasons. Their songs are spectrally rich and contain subunits (syllables) which fall into recognizable categories (e.g., stack, sweep, high note, etc.). They will readily sing even when acoustically isolated, producing hundreds to thousands of songs per day. Adult male zebra finches are determined singers and continue singing even under experimental conditions that other species do not adapt to as well. Adult

females do not sing at all. This strong sexual dimorphism in singing behavior and in the neural control of song in this species may be related to unusual features of feedback regulation in males (see data chapters). Perhaps their most significant behavioral feature from the point of view of the experimenter, however, is that once a male reaches adulthood, he sings the same sequence of the same syllables throughout his life (Zann, 1996). This naturally divides what might otherwise be a fluid and difficult-to-classify behavior into a "trial" structure similar to that imposed on rodents and monkeys in behavioral experiments, and allows the experimenter to more easily perform averaging over song renditions to increase the signal-to-noise ratio of electrophysiological recordings. This convenient behavioral feature has enabled many powerful experiments, but has likely also limited the possible outcomes, and therefore the hypotheses that arise, as I will consider in the Discussion.

Other songbird species

The prominence of the zebra finch model notwithstanding, the study of birdsong learning is enormously enriched by results obtained in numerous additional species. A large body exists of studies of song learning behavior and its constraints, especially developmental constraints (see, for example, Kroodsma et al, 1982). Many species have also been investigated from a neuroethological perspective (see, for example, Zeigler and Marler, 2004). Beyond zebra finches, the most prominent model species is the so-called Bengalese finch (which is neither Bengalese, nor a finch, nor even, strictly speaking, a species), *Lonchura striata domestica*. Another pet store staple (known in North America as the "society finch"), it was domesticated in Japan (where it is not native – it seems that it was imported from China and then bred extensively in captivity in Japan and later Europe), likely from another estrildid finch species, the white-rumped munia (*Lonchura striata*) (Svanberg, 2008; I note in passing that this paper includes the German common name for the species – *Japanische*

Mövchen, a name which at least reflects part of its actual provenance).

Bengalese finches share many of the convenient traits of zebra finches, including easy captive breeding and readiness to sing. Their behavioral distinction most significant for neuroethology is that, rather than singing a single syllable sequence in adulthood like the zebra finch, Bengalese finches sing what might be termed a branching sequence. They maintain a repertoire of syllables that does not appear to change after maturation, but the path a given song takes through this syllable repertoire may vary from instance to instance – on one rendition a bird might sing "ACDF", and then follow that by "ABCD" (as common in the field, each syllable is here identified by a letter, which is assigned in order of appearance, and has no relation to syllable phonology). The structure of Bengalese finch song is reasonably well characterized by a first-order Markov chain (Honda and Okanoya, 1999). This trial-to-trial variability has allowed some novel experiments to be performed in Bengalese finches (and has been used to make claims about "syntax" in birdsong which may be premature), making them a useful adjunct to zebra finches. A long period of selective breeding (likely at least two centuries, and possibly longer (Svanberg, 2008; O'Rourke et al, 2021)), however, has made comparative arguments using Bengalese finch data questionable.

A number of other species contribute to the songbird literature, albeit more sparsely than do the zebra finch or Bengalese finch. Focusing just on electrophysiological studies, a partial list includes swamp sparrows (*Melospiza georgiana*) (Prather et al, 2008), canaries (*Serinus canaria*) (Boari et al, 2022), mockingbirds (*Mimus polyglottos*, among others) (Zollinger et al, 2008), starlings (*Sturnus vulgaris*) (Meliza et al, 2010), plain-tailed wrens (*Pheugopedius euophrys*) (Fortune et al, 2011), white-crowned sparrows (*Zonotrichia leucophrys*) (Margoliash, 1983), and many others. Relatively seldom, however, has the same question been asked of more than one or two species at a time, which has, I believe, been much to the field's detriment. The atypical aspects of zebra finch brain and behavior (partially in relation to the accentuated sexual dimorphism in this

species) further constrain theories of birdsong learning and production imposed by these limitations in the diversity of experimental species.

The work presented here was performed entirely on zebra finches, but I will try to emphasize throughout where I think the peculiarities of the species are relevant to the results observed, and to mention related work in other species where it exists.

Anatomy of song control structures

A number of structures in the songbird brain have been identified as contributing to song production and vocal learning. These structures are collectively termed the "song system," though the term is not always precisely defined. They include HVC, in the forebrain (the avian forebrain is not laminated, as the mammalian cortex is, but organized into spatially coherent clusters termed "nuclei" which are arranged without immediately obvious structure within the forebrain).

HVC is critical for both song learning and song production – when it is lesioned, birds do not sing (McCasland and Konishi, 1981). It projects primarily to three targets: the robust nucleus of the arcopallium (nucleus RA), the avian basal ganglia analogue area X, and a small nucleus in the auditory forebrain called Avalanche (Av) (Dutar et al, 1998; Akutagawa and Konishi, 2010). These projections are all ipsilateral – the avian brain lacks a corpus callosum, and midline-crossing projections do not appear in the telencephalon. HVC receives projections from nucleus interfacialis (NIf), an auditory nucleus that receives projections from the auditory forebrain region CLM (Vates et al, 1996), nucleus uvaeformis (Uva), a thalamic nucleus that receives complex and likely multisensory input (Wild, 1994), from Av (i.e., HVC and Av are reciprocally connected) (Akutagawa

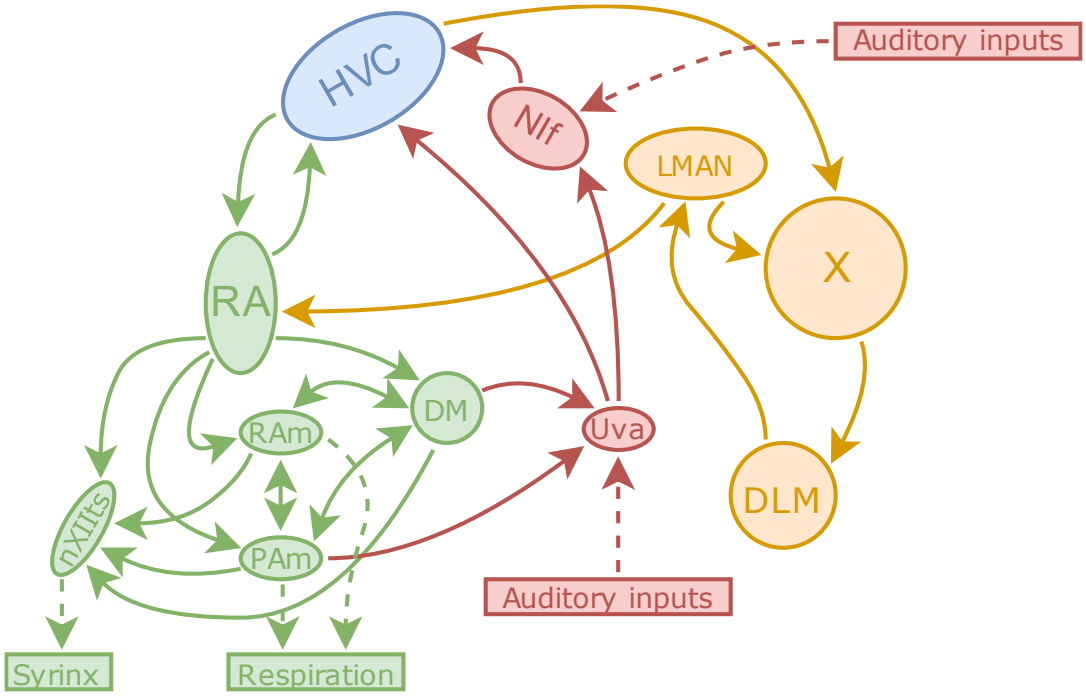


Figure 1.1. Diagram of the song system nuclei. Sensory and feedback pathways are drawn in red, the anterior forebrain pathway (AFP) in gold, and the motor pathway in green. None of the nuclei are drawn to scale.

and Konishi, 2010), from MMAN (Vates et al, 1997), and possibly from the "shelf" region directly ventral to HVC (but see Wang et al (2001), which suggests that the projection is unidirectional from HVC to the HVC shelf).

RA is also critical for song production (Nottebohm et al, 1976). It has been shown in comparative studies to be analogous to, and share gene expression patterns with, layer 4 pyramidal neurons in mammalian motor cortex (Dugas-Ford et al, 2012; Briscoe et al, 2018). RA projects to several structures, including strong projections to the midbrain dorsomedial nucleus of the intercollicular complex (DM) (Vicario, 1991), which is involved in producing calls (Simpson and Vicario, 1990), two brainstem respiratory regions nucleus retroambigualis (RAm; largely controlling expiration (Wild, 1993b)) (Vicario, 1991) and nucleus paraambigualis (PAm; largely controlling inspiration (Reinke and Wild, 1998)) (Vicario, 1991) which act on spinal motor neurons, the

tracheosyringeal part of the hypoglossal nucleus (nXIIts), which controls the spinal motor neurons that innervate the muscles in and around the syrinx, the bird's vocal organ (analogous to the mammalian larynx, but differing somewhat in structure), and weaker but distinct projections to two other brainstem respiratory control regions and to DLM of the thalamus (Vicario, 1991; Wild, 1993a; Foster et al, 1997). All of these projections from RA are ipsilateral (Vicario, 1991). There is also a sparse direct feedback projection from RA to HVC; the significance of this projection has not been determined (Roberts et al, 2008), but it has been reported elsewhere (Basista et al, 2014).

The brainstem targets of RA, as stated above, drive activity in spinal motor neurons involved in respiration and syringeal movements; it is these concerted movements which work together to produce vocalization, including song (see below for more detail). There are also interconnections between some of these nuclei – between RAm and PAm certainly (mirroring reciprocal connectivity between inspiratory and expiratory control nuclei in the mammalian brainstem (Smith et al, 2013)) (Reinke and Wild, 1998). DM also projects to nXIIts and to all the brainstem respiratory nuclei, but the connection is unidirectional, whereas its connections to RAm and PAm are bidirectional (Wild et al, 1998). Each of the respiratory nuclei has a unidirectional projection to nXIIts, while nXIIts only projects to the syringeal musculature. Thus, respiratory nuclei are in a position to control the final common output to the vocal musculature, whereas (in the brainstem) the reverse control condition cannot be achieved.

PAm (the brainstem inspiratory nucleus) also projects ipsilaterally to Uva, the thalamic nucleus which sends projections to HVC (Reinke and Wild, 1998). There is also evidence for a bilateral projection from DM to Uva (Striedter and Vu, 1998). (But note that the existence and nature of this projection is under some dispute, arising from the depth of these targets and their small sizes, making tracer studies particularly challenging (M. Wild, personal communication).)

Uva, as mentioned, projects to HVC, but it also projects to NIf and, sparsely, to its contralateral

correlate (i.e., Uva-Uva) (Wild, 1994).

Thus, a feedback loop exists involving HVC, RA, the brainstem, and Uva, back to HVC. I will discuss the implications of this feedback loop later.

Returning to HVC's projection targets, area X is the avian basal ganglia analogue. It occupies a large portion of the forebrain, and receives input from many areas (Bottjer et al, 1989). The projections that have been investigated most thoroughly with respect to song learning and production are those to ventral pallidum neurons that in turn project to dopaminergic regions in the midbrain (Person et al, 2008; Gale and Perkel, 2010; Goldberg and Fee, 2010) and those to a thalamic nucleus DLM (the medial portion of the dorsolateral nucleus of the anterior thalamus) (Bottjer et al, 1989; Luo and Perkel, 1999), which itself projects to a forebrain nucleus LMAN (the lateral magnocellular nucleus of the anterior neostriatum), which then projects to RA, as well as projecting back to area X (Bottjer et al, 1989). This X-DLM-LMAN loop, together with its HVC input and its RA output, is known as the anterior forebrain pathway (AFP), and has received a great deal of attention due to the role it plays in song learning during development.

Little is known about the function of Av, HVC's third projection target. The neurons which project to Av are large but sparsely distributed within HVC, and Av itself is quite small, and reportedly difficult to find within the auditory forebrain (Roberts et al, 2017). The sparseness of HVC_{Av} neurons and Av's small size together make thorough investigation of the purpose of their interconnection difficult. HVC's location integrating auditory information into a motor pathway makes a reciprocal connection with an auditory area intriguing, however. Indeed, targeted genetic ablation of HVC_{Av} neurons during development disrupted the juvenile bird's specific ability to acquire the tutor song (Roberts et al, 2017). Changes in both HVC and Av could contribute to this phenomenon.

Architecture and cellular features of HVC

Neurons in HVC can be broken into two groups: interneurons, which produce only local projections, and projection neurons, which may project to one or more of HVC's downstream target nuclei – Av, Area X, or RA (styled here as HVC_{Av}, HVC_{X}, and HVC_{RA}). There are three types of HVC interneurons as defined by the combinations of calcium binding proteins parvalbumin, calbindin, and calretinin that each type expresses (Wild et al, 2005). HVC interneurons appear to be exclusively inhibitory and GABAergic (more specifically, $GABA_{A}$) (Mooney and Prather, 2005). HVC projection neurons appear to be exclusively glutamatergic. Most HVC projection neurons have a number of intrinsic projections but project to a single extrinsic target. However, a recent study has provided evidence for a small population (2 neurons out of a dataset of 40; 5%) of projection neurons which also project to Area X (Benezra et al, 2018; this paper notes a prior report of a dual-projecting cell in a doctoral dissertation (Kittelberger, 2002), which has not been published). The importance or function of these dual-projecting neurons is unknown. Further subdivision of the HVC_{RA} into two distinct classes is well established morphologically (DeVoogd and Nottebohm, 1981; Fortune and Margoliash, 1995) with at least two electrophysiological classes reported as well (Shea et al, 2010; Daou and Margoliash, in prep). Additional subdivisions of these neuron types (and of course wholly new neuron types) are possible, and perhaps even likely. Interestingly, however, in a study of HVC_{X} recorded in a whole-cell patch slice preparation, only a single class of neurons was described (Daou and Margoliash, 2020).

Regardless of external target, HVC projection neurons also produce local collaterals within the nucleus, producing glutamatergic synapses onto interneurons and other projection neurons. One study of *in vitro* paired whole-cell patch recordings reported modest numbers of excitatory monosynaptic connections from HVC_{X} to HVC_{RA}, with smaller numbers of inhibitory monosynaptic

connections from interneurons to HVC_{RA}, HVC_X, and other interneurons, excitatory monosynaptic connections from HVC_{RA} and HVC_X to interneurons, and excitatory monosynaptic connections from HVC_{RA} to HVC_X (the study included no HVC_{Av} cells) (Mooney and Prather, 2005). The same study also reported a high prevalence of disynaptic inhibition (with an interneuron in the middle) between HVC_{RA} and HVC_X neurons, proposing that such triplets form "a dominant pattern of synaptic connectivity" among HVC projection neurons (Mooney and Prather, 2005). One cannot lean too heavily on the relative proportions of the synaptic connections in this paper, however, as whole-cell patch clamp recordings are easiest to obtain from the relatively large HVC_X neurons (a sampling bias which the authors take care to point out). An additional and significant source of bias in paired cell recordings is that randomly chosen cells in close proximity are more likely to be connected than are randomly chosen cells that are distant from each other, and known long-distance connections within HVC suggest that this source of sampling bias may have a substantial impact on HVC paired recordings (though see below). Another paired intracellular recording dataset contained comparable numbers of monosynaptic connections, focusing solely on HVC_{RA} neurons: very few (1/140) HVC_{RA} to HVC_{RA} synapses, and larger numbers of HVC_{RA} to interneuron (34/43) and interneuron to HVC_{RA} (28/42) synapses (Kosche et al, 2015). This study, too, suffered from the same intracellular recording sampling biases. An electron microscopy study of HVC local connectivity found a higher prevalence of monosynaptic HVC_{RA} to HVC_{RA} connections than these slice electrophysiology studies, however (Kornfeld et al, 2017). In that paper arguments were presented for inflating the number of actually observed synapses; these seem at least somewhat contrived. It is noteworthy, however, that even the most conservative estimates of HVC_{RA} to HVC_{RA} connections in Kornfeld et al (2017) were far higher than reported in the intracellular studies: at least 1% of all synapses onto any given HVC_{RA} neuron were from another HVC_{RA} neuron (note, though, that direct comparison of synaptic connectivity is somewhat difficult, as the two studies

report somewhat different measures). These monotypic connections preferentially occurred at distal axonal segments within HVC (>200 μm from the soma), which may explain why the earlier intracellular studies missed them (Kornfeld et al, 2017). This study also reported that proximally, HVC_{RA} neurons mostly synapsed onto interneurons, and conversely, HVC_{RA} dendrites received the preponderance of their inputs from interneurons. Direct connectivity between HVC_{RA} neurons and the absence of modulating activity from other inputs (other than proximal interneurons) figures critically in one theory of HVC functional activity, as will be described later.

Earlier tracer injection studies of HVC reported that major projections to HVC from NIf, Uva, and MMAN lacked any apparent topographic organization (Fortune and Margoliash, 1995; Bottjer et al, 1989; Foster and Bottjer, 1998), with the implication that HVC's functional organization is similarly not topographic (Margoliash et al, 1994). Further support is provided by the fact that both HVC_{RA} and HVC_{X} are found throughout the cytoarchitectonic subdivisions of HVC (Fortune and Margoliash, 1995). However, more recent work has cast some doubt upon these conclusions. A partial lesion study targeting different subregions of HVC found different effects on song production depending on lesion location (Basista et al, 2014). Lesions of medial HVC were accompanied by an increase in atypical song syllable sequences (involving omissions, premature truncation, and repetition of syllables, sometimes within the same bout of song), while lesions of lateral HVC were accompanied by early truncations of the song (omission of syllables in the middle of song was also observed, but the effect was not statistically significant). This suggests the possibility of a functional organization within HVC along a mediolateral axis, a suggestion supported by a report of a lack of dramatic song degradation after mechanical transection of HVC across the mediolateral axis (Poole et al, 2012), and an apparent rostrocaudal bias for intra-HVC connectivity (Stauffer et al, 2012). (While this study did not investigate the question, it is possible that the regions in question correspond to those identified in Fortune and Margoliash, 1995.) This

study also double-labeled HVC in the medial and lateral regions, and found no evidence for projections to both parts of HVC from either RA, NIf, Uva, or MMAN – i.e., the authors report no double-labeled cells in these nuclei afferent to HVC, though there was also no topographic organization of single-labeled cells in these afferent nuclei, which could explain earlier reports of a lack of topography in HVC. (However, they do report widespread double-labeling across the whole of RA, indicating that inputs from the two regions of HVC converge in this nucleus.) The authors conclude that inputs to HVC are broadly organized into medial and lateral "streams," but it is still unknown whether or how responsibilities for song production and properties are distributed between these and other regions of HVC.

Intrinsic properties and singing

As discussed above, within the broad categories of HVC neurons – HVC_{RA}, HVC_X, and interneurons (the newly discovered HVC_{Av} neurons have yet to be characterized electrophysiologically) – there are several subclasses, based on functional (Daliparthi et al, 2019; Kozhevnikov and Fee, 2007; Rajan, 2018) and morphological (Fortune and Margoliash, 1995; Nixdorf et al, 1989) properties, and cellular and electrophysiological properties have been shown to define several subclasses of HVC_X and interneurons (Wild et al, 2005; Daou et al, 2013).

Intrinsic electrophysiological properties of HVC_X neurons were found to be related to song learning and maintenance (Daou and Margoliash, 2020). The firing properties (including firing frequency, spike timing within a burst, and spike morphology) of HVC_X neurons, when recorded in *in vitro* slice preparations, were reflected in firing patterns in response to current injection that appeared similar within an individual animal, but varied more widely across animals. These intrinsic properties can be recapitulated using Hodgkin-Huxley models of a number of membrane

ion channels (Daou et al, 2013), and the ionic conductances predicted by these models can then be examined and related to behavior. Zebra finches, even those reared with the same tutor, naturally sing different songs, and the variation in the song motif can be quantified to produce a measure of distance. This distance measure was positively correlated with variation in the modeled conductances of HVC_X neurons. Birds which sang similar songs occupied regions within a state space defined by the modeled ionic conductances that were closer together than those of birds which sang songs which were less similar. Moreover, a given bird's neurons tended to occupy a small region of this overall multidimensional space, and one bird's neurons tended to be closer to one another within this space than to another bird's neurons. This result ties the outcome of a complex behavior – song learning – to cellular properties of neurons. The relationship between song similarity and intrinsic properties is not fixed after the song is learned, however. When birds were exposed to delayed auditory feedback (DAF), a perturbation of the normal auditory feedback birds naturally receive when they hear themselves sing, while their songs did not appreciably change (at least in a manner captured by the similarity metric used; changes to song structure above the level of the motif were almost certainly occurring – see Chapter 2), the modeled ionic conductances within a given bird's HVC_X neurons began to lose their tight clustering. This effect occurred with as little as four hours of DAF exposure, and the effect became more pronounced with longer periods of DAF exposure. Thus, not only are intrinsic properties of HVC_X neurons related to the song a bird learns, they exist under active regulation, modulated by auditory feedback, even after the bird has successfully learned the song he will sing in adulthood. Plasticity in these neurons as a result of artificial aberrant feedback aligns with their known connectivity to the anterior forebrain pathway, and suggests their involvement in producing or calculating error signals, rather than simply passing information about feedback or ongoing motor activity to their efferent targets.

There is evidence that intrinsic properties of both HVC_{RA} and HVC_X change over the course of

development, gaining the features which characterize their adult behavior, while interneurons maintain more consistent intrinsic properties across development (Ross et al, 2017). This phenomenon is presumably related to the process of song learning. That interneurons do not show the same type of changes in intrinsic properties is interesting, and suggest a role in shaping the developmental trajectory of the projection neurons as vocal learning progresses, dovetailing with reports of interneuron-mediated inhibition "protecting" stable portions of song once they are successfully learned (Vallentin et al, 2016). This idea may even align with a report of HVC projection neurons lacking day-to-day stability in their activity during singing, instead disappearing and reappearing within a framework of relatively fixed interneuron firing patterns during singing (Liberti et al, 2016). These suppositions are also consistent with observations that interneuron firing rate minima correlate with projection neuron burst times (Amador et al, 2013).

Temporal dynamics of HVC activity

HVC projection neurons are challenging to record in singing birds, particularly given how sparsely they fire (and for HVC_{RA}, their small size); this necessitates very high impedance electrodes (10 MΩ) and greatly limiting the duration of the recordings (typically 15 minutes or less). The first recordings of HVC_{RA} and HVC_X neurons positively-identified (via antidromic stimulation) in singing birds demonstrated a striking feature: HVC_{RA} neurons fired a burst at a single time during the motif (the canonical sequence of syllables that the bird sings), while HVC_X neurons produced bursts at one or more (up to 4 or 5) times during the motif (Hahnloser et al, 2002). For those neurons which are active during singing, the burst times were remarkably precise across motif renditions (with 0.66-0.73 ms (HVC_{RA}) and 1.0 ms (HVC_X) of jitter on average; these values were derived by piecewise linear time warping to account for small variations in timing between motifs)

(Hahnloser et al, 2002; Kozhevnikov and Fee, 2007).

There has been a tendency to ignore the role of HVC activity prior to singing. There are also HVC_{RA} and HVC_X neurons that do not fire during song. At least some of these are active during calling (McCasland and Konishi, 1981) or during the preparatory period preceding song (Daliparthi et al, 2019). These topics are discussed further below.

The HVC_{RA} burst times in Hahnloser et al (2002) exhibited "no obvious timing relation to the onset or offset of song syllables" – they appeared to be distributed randomly across the duration of the motif. (Of course, with 16 neurons across 3 birds, a relationship with any aspect of song timing would have to be extremely obvious for chance to be discarded as an explanation. A lack of such evidence is not a compelling argument for there being no possible relation.) The extreme sparsity of HVC_{RA} bursts, together with the lack of obvious patterning within the motif, was immediately argued to support the idea that each HVC_{RA} burst represents a "transient event or 'state'" within the motor sequence producing the motif, and that the entire set of HVC_{RA} firing times constitutes a "canonical sequence in HVC" which can be viewed as a "time code" that serves as a metronome to the "muscle map" in RA (Hahnloser et al, 2002). This hypothesis had, on its face, theoretical simplicity to recommend it. It also fit neatly into an emerging literature in theoretical neuroscience on the abilities of so-called "synfire chains" (sequentially active neurons, each of which drives the next neuron in the chain via an excitatory synapse) to reliably produce precise sequential activity (e.g., Herrmann et al, 1995).

Thus was born the "clock model:" HVC projection neuron (HVC_X were incorporated into the synfire chain theory along the way – see below) burst times represent time within the motif, and nothing else. As an immediate consequence of this time-locking, they can have no relation to muscle contractions. Muscle activation is severely underdetermined with respect to song – the same syringeal and air sac muscles are activated during very different types of song elements. It is only

through the concerted patterning of muscle activations across time that diverse acoustic features can be produced. Thus, if one HVC_{RA} neuron represents only one point in time within the motif, the muscles active during its burst (or soon thereafter, to account for a premotor delay) will also be active at other moments when the neuron is silent. Variations have been proposed, including that multiple synfire chains may be active at once, competing in a winner-take-all manner to be active next (Jin, 2009; primarily in the context of more variable Bengalese finch song), or that each syllable has a separate synfire chain or chains associated with it, which are joined loosely by local HVC synapses or possibly through recurrent loops through RA, the brainstem, and Uva (Danish et al, 2017). Despite these various proposals, the core premise has remained largely the same: HVC represents time. Clock model proponents have also seemed to rather freely shift between "hard" and "soft" versions of the model as circumstances demand, proposing syllable-specific clocks which ignore gaps one moment and returning to the strict uniformity of projection neuron burst times the next. (Synfire chains seem to have fallen somewhat out of favor since the early 2000s, perhaps because nobody has ever actually seen one *ex silico*.)

The clock model has seen its share of support, but it has been largely circumstantial. The EM study described above (Kornfeld et al, 2017) provided evidence that the physical substrate required for a synfire chain of HVC_{RA} neurons did exist in HVC. It does not, however, itself constitute proof that the HVC_{RA} neurons that are connected in fact form a synfire chain – structural connectivity cannot reliably predict functional connectivity, particularly in light of the immense inhibitory input that same study found HVC_{RA} neurons to receive. Another set of studies used a clever technique of local cooling in HVC (which will slow molecular processes within and between cells, including synaptic transmission and, to a lesser extent, axonal propagation) and reported linear stretching of song duration as temperature varies, which is consistent with the idea that it is solely local connectivity within HVC that determines the tempo of song (Long and Fee, 2008; Andalman et al,

2011). However, even were HVC indeed the sole source of timing for song, this would provide no support whatsoever for a synfire "clock" hypothesis as the mechanism generating that timing. Furthermore, later studies using the same cooling technique demonstrated that cooling zebra finch HVC slowed local synaptic connectivity much more than it slowed the song, and additionally that cooling HVC had temperature-dependent effects elsewhere in the brain due to heat flow (Hamaguchi et al, 2016). A study of cooling HVC in canaries showed a linear regime but also non-linear breaking at syllable boundaries (Goldin et al, 2013). These results sharply undercut the conclusions drawn from the earlier cooling studies. Another prominent study (Lynch et al, 2016) supporting the clock model will be discussed at length in Chapter 3. There has also been extensive work in birdsong in the last two decades purporting to be aligned with the clock model, but whose primary experimental results are far more interesting and useful in their own right than any (sometimes tenuous) connection to the clock model.

Unfortunately, there is also a vast literature of experimental results which are difficult to align with the clock model, or which contradict its premises or predictions entirely. While reviewing this is beyond the scope of this document, here I focus on a handful of such issues.

Syllables as fundamental units

It appears that syllables represent a fundamental unit of singing behavior for songbirds. An external startling stimulus (a flash of light) will disrupt ongoing singing in zebra finches, but predominantly during gaps between syllables – i.e., a flash of light in the middle of a syllable is more likely to lead to song truncation at the next gap, rather than immediately (Cynx 1990; Franz and Goller, 2002). A similar phenomenon is observed in Bengalese finches, with the additional interesting feature that sequences of song which are generally static across renditions are not truncated by a light flash until their completion, whereas sequences of song which are more variable

(this increased sequence variability being, as stated above, a principal difference between zebra finch and Bengalese finch song) are usually cut short immediately, implying that not only do periods of relatively "ballistic" vocalization vary across species and individuals within a species (the Bengalese finches in this study did not all sing the same song, and thus did not have the same static and variable portions of song), but they must also be defined in a more complex manner than simply "syllable vs gap" (Seki et al, 2008). Consistent with these results are studies of song learning, which reliably report that juvenile songbirds learn song in roughly syllable-sized chunks (Tchernichovski et al, 2001; Lipkind et al, 2013; Okubo et al, 2015; Vallentin et al, 2016; Lipkind et al, 2017). (Let me not overstate this point – there are without doubt other features of song which may be learned, possibly depending on species: overall tempo, bout structure, and sequence statistics, as well as non-vocal song-associated movements like posture, wing flaps, hops, etc. (Maguire et al, 2013; Perkes et al, 2019).) Additionally, respiratory requirements work to impose a natural intermittence onto vocalization – the bird cannot breathe out continuously without breathing in, for reasons of both gas pressure and oxygen needs. In the gaps between syllables songbirds take inspiratory "mini-breaths" (Hartley and Suthers, 1989), which are only rarely accompanied by phonation (Goller, 2016). Respiratory dynamics also appear to be learned (i.e., a juvenile's respiratory dynamics will closely resemble the tutor's, more so when his song is a close copy), and are similar across similar syllables (Franz and Goller, 2002). More recent evidence demonstrates the natural resonances at which birds breathe and how this interacts with the dynamics of singing (Fainstein et al, 2021). Thus, evidence for a fundamental role for syllabic structure is both extensive and varied.

Do syllables exist as natural categories? While there are specific examples in specific individual songs that present some challenge, in spectrographic analysis of zebra finch song (and in numerous other species) a great many of the syllables clearly stand out as distinct categories, and are characterized as such by independent observers. When a zebra finch (as in other species) fails to

complete his song, he terminates at syllable boundaries, rather than mid-syllable. Additionally, in challenging juvenile zebra finches to learn sequence and phonology changes simultaneously, Lipkind et al (2017) found that most followed a learning trajectory in which individual syllables were morphed into the target they already best matched, at the cost of sequence errors, which were only corrected later. The evidence from the periphery also supports the idea that syllables are meaningful independent of human assumptions – mini-breaths in the gaps are a fundamentally different behavioral output from a vocalization, and are accompanied by qualitatively different syringeal muscle activation (Goller and Suthers, 1996).

A fundamental claim of the clock model is that moments within the motif are not privileged in HVC – each moment is represented by some number of HVC projection neuron burst times, even if more recent versions have relaxed the insistence upon strict uniformity to allow some modulation of burst density. Reconciling this idea with the evidence above that syllables are in fact of behavioral importance to the bird, particularly during song learning, is challenging. (This is also partly what makes the lack of obvious correspondence between HVC activity and song structure so surprising.) It is not impossible – Okubo et al (2015) argue that the more uniform distribution of HVC projection neuron burst times across the song in adult zebra finches is the result of a loss of heterogeneity as a consequence of development and crystallization of song. This is a point I will return to in the Discussion.

Interhemispheric coordination

Another fundamental problem in the study of birdsong is that of interhemispheric coordination. As described above, much of the connectivity within the song system is ipsilateral – the only midline-crossing projections are in the brainstem and diencephalon. And yet many species

can control their two syringes as independent sound sources (Suthers, 1998). In zebra finches, activity in HVC in each hemisphere remains remarkably synchronized during song, particularly at the beginning of syllables (Schmidt, 2003). Unilateral electrical stimulation in HVC during song will disrupt song at short latency (without the syllable-based inertia discussed above), and will additionally rapidly disrupt HVC activity in both hemispheres (Vu et al, 1994). But then it cannot be the case that a synfire chain of HVC_{RA} neurons (one per hemisphere) operates in a ballistic manner oblivious to feedback – there must be connectivity, robust enough to disrupt HVC activity when engaged, that implicates HVC circuitry in a more dynamic, responsive regime than that seemingly required by the clock model. After all, this feedback circuit must carry motor-related information – if it is not auditory (thus reflecting recent motor activity), it is almost certainly derived from Uva, which receives robust motor-related projections from the brainstem.

This interhemispheric coordination problem is further complicated by a study using electrical stimulation during song which reported not only tight coordination of HVC activity across hemispheres, but a remarkable rapid alternation of which hemisphere appears to dominate motor production (Wang et al, 2008). When one hemisphere's HVC is dominant or leading, electrical stimulation there will disrupt song, while stimulation in HVC in the other hemisphere will have little or no effect on song. After a short period (mean interval between swaps 35 ms, with a range of 4 ms (!) to 150 ms), the hemispheres switch, with the other hemisphere becoming sensitive to stimulation and the originally sensitive one becoming insensitive. They found no obvious relationship between switching times or hemispheric dominance and song features. The mechanism by which this phenomenon occurs remains unknown, but it seems certain to involve recurrent loops through subcortical structures, again including Uva. Subsequent birdsong studies have largely declined to grapple with the implications of this interhemispheric switching result, but it is far easier to imagine incorporating it into a model of song production that already

acknowledges the importance of feedback and recurrent connections within the song system.

Hemispheric dominance and interhemispheric switching must also interact with lateralized control of the peripheral musculature – while the syrinx is controlled unilaterally, with birds able to produce different sounds with each side of the syrinx, respiration is controlled bilaterally, with patterns of activation of respiratory muscles similar across the midline (Goller and Suthers, 1999).

Syllable combination selectivity

Combination-selective HVC neurons were first identified in white-crowned sparrows. Some HVC neurons with phasic firing profiles (suggesting that they were projection neurons; the cell types within HVC had not been elucidated at the time of that study) were highly selective only for relatively long sequences within the bird's own song (Margoliash, 1983). Similar neurons have also been identified in zebra finches. Margoliash and Fortune (1992) reported HVC neurons (type unidentified, but based on firing rates some appeared to be interneurons and some projection neurons) that responded with high selectivity to specific combinations of song syllables. In some cases, the neurons responded robustly to portions of the (or the entire) syllable that evoked a response from them during whole-song playback. In others, neurons required two or more syllables to be played back to evoke a response similar to that during whole-song playback; in some cases the required portion of song spanned hundreds of milliseconds. The authors termed these neurons "temporal combination selective" (TCS) cells. These response types complicate the simple concept of BOS playback selectivity – it is not the case that the entire song is necessary for HVC to respond. It is also not required for the beginning of the motif to be played back for HVC to respond robustly, which means that if HVC is a clock, it is a curious one, with the ability to shortcut past large sections of the song to start in the middle. Other studies have reported similar long integration times for HVC neurons which respond to song playback (Lewicki and Arthur, 1996; Lewicki, 1996).

One study in Bengalese finches reported that many HVC interneurons (putatively identified based on firing rate) were broadly responsive to pairwise combinations of syllables from the bird's own song, including combinations which did not appear within the song. The authors took this as evidence for a population encoding of song structure in these HVC interneurons, and suggested that the TCS neurons described above may be similar (I question their reasoning in this respect – they cited high tonic firing rates, but many of the TCS neurons in Margoliash and Fortune (1992) had relatively low firing rates outside of song playback) (Nishikawa et al, 2008). Such a broad selectivity, and a shared encoding, within the interneurons strongly implies similar (or at least related) features within HVC projection neurons, especially given the robust interconnectivity between interneurons and HVC_{RA} and HVC_{X} neurons. (It is also possible, however, that they found a primarily species-related phenomenon, resulting from the statistical properties of Bengalese finch song.)

Finally, it must be acknowledged that robust HVC responses to BOS playback (which might be seen as merely a very long selectivity for syllable combinations) are a troublesome phenomenon to explain. If one is unwilling to stipulate that sensory feedback may shape HVC activity during singing, one is left no choice but to wave the phenomenon away as a curious, experimentally convenient, but ultimately irrelevant coincidence or relict. Needless to say, this explanation hardly satisfies. TCS neurons are highly selective for playback of the individual's own song, which incorporates learned acoustic features (Margoliash and Fortune, 1992).

Direct correspondence between HVC and syringeal dynamics

Many of these more complicated features of HVC dynamics are what might be considered higher-order effects – they strongly suggest motor-related activity in HVC, but do not constitute

direct evidence themselves. And the lack of a clear correspondence between HVC dynamics and any spectral or temporal feature of song examined so far has been taken as support for the idea that HVC is entirely divorced from motor actions (i.e., except as a time-keeping mechanism). This stance is only possible, however, by way of two critical misconceptions – that the proper representation of motor behavior is the spectrogram of the song, and that a motor role for HVC must take the form of a simple representation of behavioral output. Neither of these assumptions holds much water. The spectrogram – i.e., the spectrotemporal features of song – is the result of a complicated interlocking set of time-varying muscle programs, involving tight synchrony between respiration, the syringeal muscles controlling phonation, and additional (less well studied but still critical for proper sound production (Amador et al, 2013)) muscles controlling parts of the upper vocal tract. The spectrogram remains the default output used as a measure of motor actions in birdsong largely for convenience, requiring only consumer-grade microphones and amplifiers, acoustic analysis that is well-understood and relatively simple, and sound recordings that can be gathered and stored without perturbing the bird. Measuring biophysical parameters on the bird itself is more difficult (but certainly not impossible), and examining muscle activation directly via implanted EMG more challenging still (but again, doable) – the problem of implant weight is inescapable, although great strides have been made on this front in the past decade. And the question of how neuronal activity in motor regions produces and relates to motor behavior is far from clear – simple representations of kinetics, kinematics, or individual muscle activation have been found to a limited degree in motor cortical regions but have so far failed to yield much clarity in studies in mammals, and they leave important theoretical questions unanswered (Evarts, 1968; Georgopoulos et al, 1982). The reason for this may lie in the fundamental problem with looking for representations in motor networks: representation of reality is not the purpose (if one will pardon the teleological word choice) of a motor system – producing a motor action is (Scott, 2008; I must thank Kyler Brown for

pointing me towards this simple expression of this distinction). More complicated conceptions of motor behavior as being produced by dynamical systems formed from motor networks have shown promise, but their complexity makes application to experimental results, and generation of hypotheses, somewhat more difficult.

One step in bringing birdsong research into line with the mammalian motor literature would be demonstrating some correspondence between neuronal activity in HVC and a more appropriate or direct measure of motor output than the spectrogram. A prior student in this laboratory, Kyler Brown, did just that. He simultaneously recorded neuronal activity in HVC and EMG from syringeal muscles (primarily vS) in head-fixed sleeping birds during spontaneous replay events (Brown, 2017). He found correlations between syringeal EMG and HVC activity (measured as multiunit activity and single units) in both preparations. In the sleeping bird, the strength of the correlation depended somewhat upon the inferred behavioral state – in three of Brown's birds he was able to elicit only weak selectivity to presentation of BOS playback, which is most likely a consequence of poor-quality sleep in these animals. Regardless, in one of these three birds the strong correlation between HVC activity and syringeal muscle activation persisted, suggesting that the dependence of this motor synchronization on sleep is not all-or-nothing, but may arise in intermediate behavioral states. In several cases in the birds which did show good BOS selectivity in HVC, he reported a large spike-triggered averages (STA) in the EMG signal dependent on spikes produced by single units in HVC, a clear kinematic representation. (Many HVC neurons he recorded, however, did not have appreciable EMG STAs, which may either simply indicate that they were involved in activation of other muscles, or that they play a different role in generation of a motor sequence.) This was the first evidence of such a motor representation in HVC.

Brown went on to record simultaneous syringeal EMG and HVC neuronal activity in awake, singing birds, and reported a similar relationship between the two signals, as measured by STA and

by fitting a generalized linear model. (That this correlation was present in both paradigms further confirms the usefulness of the head-fixed sleeping preparation, though it must be noted that he was not able to directly compare the correlations in the two paradigms within individual birds.) In both the sleeping and singing birds, the delay between HVC activity and syringeal activation was in the range of 7-25 ms (mean 15-16 ms), which is consistent with a premotor delay. Brown's recordings primarily arose from interneurons and multiunits. Multiunits in HVC are themselves usually dominated by interneurons, given their high firing rates. (While many of Brown's multiunits clearly contained bursts from HVC projection neurons, he was not able to examine these events on their own.) Given the wealth of evidence discussed above for robust interconnectivity between interneurons and projection neurons in HVC, it is difficult to imagine that a motor representation in the interneurons is not also reflected in the activity of projection neurons. In any case, no such hypothesis based on a plausible cellular circuit has been proposed to date.

Goals of this thesis

HVC neuronal dynamics are poorly understood, and are clearly critically important for song production (among other roles related to vocal behavior). Two approaches seemed promising: first, the appearance of intrinsic property changes in HVC_X neurons after short periods of DAF exposure raised the possibility of behavioral changes on timescales of a similar duration. Such a conjunction between a behavioral effect and a neuronal change could provide clarity into the functional role and evolving activity of HVC. And second, directly examining the concerted behavior of large numbers of HVC projection neurons. This latter approach has been attempted before, but prior attempts have suffered from small datasets or an unwillingness to question prior assumptions.

I explored the first question using established techniques to expose birds to delayed auditory

feedback, which produces changes in intrinsic properties in HVC projection neurons after short periods of exposure. I sought evidence of behavioral changes by examining song structure at the motif and higher levels. My results demonstrate clear behavioral changes as a result of DAF, which are among the shortest latency behavioral changes reported in the literature. Further, they argue for a role for auditory feedback in song regulation in the zebra finch over short timescales, which has largely not been reported previously.

I explored the second question by reanalyzing an existing rich dataset of HVC projection neuron bursting activity during song, which was provided by the authors of a prominent paper advocating for the clock model of HVC. I found clear evidence for statistical regularities in burst times, specifically for neurons which fire multiple bursts per song (which are nearly exclusively HVC_X), which had been overlooked in the paper originally reporting on this dataset.

Finally, I attempted to address both questions by examining HVC neuronal activity during presentation of auditory stimuli that mimic DAF exposure. I observed striking changes in HVC neuron responses to these stimuli compared with their responses to the bird's own song, providing the beginnings of a potential mechanism for behavioral change as a result of auditory feedback perturbation.

CHAPTER 2

MOTOR INITIATION DEFICITS RESULTING FROM DELAYED AUDITORY FEEDBACK

Introduction

Motor preparatory activity has been studied extensively in nonhuman primates performing reaching tasks, and is believed to function to bring the premotor and motor cortices into a suitable state in a neural activity state space to carry out a successful movement (Tanji and Evarts, 1976; Riehle and Requin, 1993; Churchland et al, 2006; Afshar et al, 2011; Churchland et al, 2012). Interfering with this preparatory activity in premotor cortex via electrical stimulation lengthened reaction times, suggesting that this activity is important for motor initiation (Day et al, 1989; Churchland and Shenoy, 2007).

There is evidence that introductory notes in birdsong are in fact preparatory in function. Their acoustic structure becomes more stereotyped as song initiation approaches, and more introductory notes are produced if the first in a sequence is acoustically less similar to this stereotyped end state. Neuronal activity in song-related neurons in a premotor region mirrors this progression to stereotypy as the introductory note sequence progresses (Rajan and Doupe, 2013). These observations fit nicely into a model of the song motor system acting in concert with feedback from brainstem motor nuclei as a dynamical system responsible for song production (Amador et al, 2013; Alonso et al, 2015).

Prior work on neuronal motor preparatory activity, primarily in primates, has focused on behavioral paradigms that preclude or overlook the possibility of overt movement performed to prepare for an upcoming complex movement, and thus the treatment of introductory notes as

preparatory movement here and elsewhere is somewhat distinct from the mammalian preparatory motor literature. Refer to the discussion for a consideration of the implications of this distinction.

Work in our lab has previously demonstrated that several days of continuous DAF (cDAF) treatment induces gross abnormalities in song sequencing in zebra finches, including perseveration on some song syllables, a phenomenon resembling stuttering in humans (Fukushima and Margoliash, 2015; Fukushima, 2008; Manderscheid, 2014). This behavioral change did not occur in all birds exposed to DAF, and in birds that did begin to produce this stuttering-like behavior, normal song sequences were also produced. Its appearance is nonetheless dramatic. Behavioral changes as a result of auditory feedback perturbation have been reported using a handful of paradigms in zebra finches, including the playback of disruptive white noise when the bird attempts to sing (Leonardo and Konishi, 1999; this perturbation regime is also termed "DAF" by the authors of this report). This disruptive feedback took weeks to induce sequent abnormalities in zebra finches, whereas delayed auditory feedback can produce sequence abnormalities as quickly as the second day (Fukushima and Margoliash, 2015), and, as I will show, changes in higher levels of song organization within hours or less.

When applied to non-stuttering humans, DAF immediately induces a variety of speech dysfluencies, including something resembling stuttering, a curious phenomenon given that DAF is also an effective treatment (though only transiently so) for adults who stutter (Lee, 1950; Yates, 1963; Stuart et al, 2002; Stuart et al, 2004).

Stuttering occurs in humans most often during speech initiation (Quarrington, 1965; Taylor, 1966; Saltuklaroglu et al, 2009). Interestingly, Saltuklaroglu et al (2009) found that altered auditory feedback, including DAF, was more effective in reducing stuttering in the middle of speech than during speech initiation; the authors suggest that this may result from DAF providing an approximation to choral speech (i.e., speaking in unison with others), a condition which is

stunningly effective in eliminating stuttering *in situ*, though obviously impractical as a long-term intervention (Kalinowski and Saltuklaroglu, 2003).

The initial work on DAF effects on zebra finch singing focused on longer-term effects, especially on the subset of birds that exhibited syllable repetition akin to stuttering. More recent work has produced evidence for dramatic short-latency changes in intrinsic neuronal properties in HVC_X neurons, and so far only preliminary behavioral correlates related to such feedback manipulation (Daou and Margoliash 2020). In what follows, I characterize the initial, rapid effects of DAF on singing behavior. I hypothesize that these changes are driven by feedback processing via HVC_X activity, an idea which will be further developed in subsequent chapters.

Methods

All animal care and procedures were approved by the University of Chicago Institutional Animal Care and Use Committee.

Housing

Eight adult male zebra finches, age at least 120 days post-hatch, were taken from our aviary and housed individually in sound attenuation chambers (Industrial Acoustics Corp.). Birds had unrestricted access to food and water and were kept on a 12h:12h light:dark cycle. Acoustic recordings were made using a C-2 studio condenser microphone (Behringer) and U-PHORIA UMC1820 amplifier (Behringer). I used custom software (https://github.com/margoliashlab/jill) based on the Linux JACK audio toolkit to digitize the signals (24-bit at 22050 Hz through an M-Audio Delta 1010LT sound card)), confirming a bird sang robustly after a period of acclimation (up to one week). Two birds did not sing robustly post-surgery, and were removed from the experiment

before DAF administration.

Accelerometer preparation

The lightweight accelerometers (BU-7135, Knowles Acoustics) come with solder pads pre-wetted (Figure 2.1), making connector attachment easy. At first, I used the leftovers of a discontinued, obsolete connector type with three pins at roughly the correct pitch for the accelerometer solder pads (as these had been the standard part used by previous lab members). I found these unsatisfactory, and switched to using Omnetics PZN-12-DD connectors instead. The 12 straight tails of these connectors are easily bent with forceps so that their tips converge into three groups of four, with a pitch again close to that of the accelerometer's solder pads. (This also allows the use of off-the-shelf Intan lightweight tethers rather than tethers built by hand.) After soldering the connector to the accelerometer, I reinforced and insulated the connection with quick-setting epoxy. This construction was easily recovered from animals after euthanasia, and withstood reuse well.

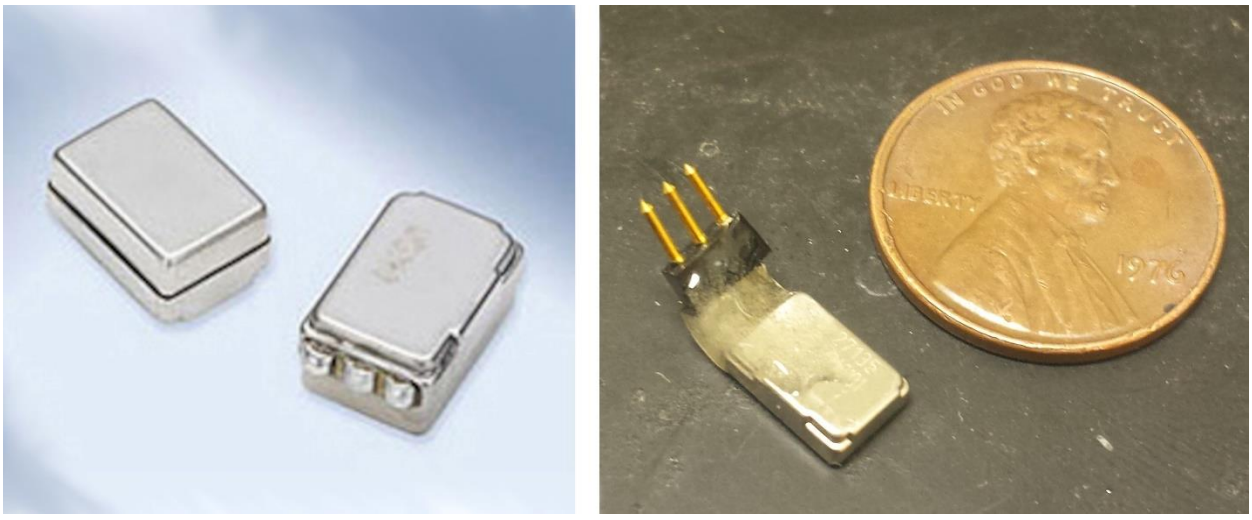


Figure 2.1 Lightweight accelerometers. Left: Knowles accelerometer before preparation. (Image from Knowles promotional materials.) Right: accelerometer after preparation, with the older style of connector attached, and penny for scale.

Implantation surgery

After 7 days of acclimation to isolation housing, birds were food-deprived for one hour, then anesthetized using isoflurane. Most of the procedure is standard, and very similar to that for implanting a pin for head fixation (Chapter 4), including pharmaceutical dosages (analgesic and antibiotic). The scalp was plucked, sterilized, cut, and retracted. I initially followed prior procedures used in the lab and shaved the top layer of skull in an ellipse of major axis roughly 9 mm centered on the midline and extending just caudal of the bifurcation of the midsagittal sinus (λ). The accelerometer then sat in the middle of this circle and was fixed in place using dental acrylic (a heavy coating beneath the device fixing it to the bottom layer of skull, and fingers of acrylic extending over the top of the device for extra stability), with curing accelerated with cyanoacrylate glue. I found, however, that implants of this style, at least in my hands, had a shorter lifetime than I would have liked – it was not uncommon for the accelerometer signal to degrade within a week or two of surgery, and detach from the skull within a month. After experimenting with several different approaches (including fixation on the side of the skull, following Fukushima and Margoliash (2015), which I never successfully accomplished), I realized that regrowth of skull and vascular tissue underneath the accelerometer was responsible for the loss of signal – once the mechanical coupling between the bottom layer of skull and the accelerometer was sufficiently cushioned by softer tissue pushing up and around it, no high frequency vibrations could be transmitted to the device for recording. Removing some of the top layer of skull is necessary for stable attachment of any head-mounted devices, though. I settled on leaving the top layer of skull intact beneath the accelerometer but cutting it away in small sections around the device's footprint (Figure 2.2). This arrangement allows both excellent and long-lasting mechanical coupling for signal fidelity as well as a robust mounting for the device to prevent detachment under the strain of tethering. With this technique I had no trouble maintaining implants on birds for many weeks,

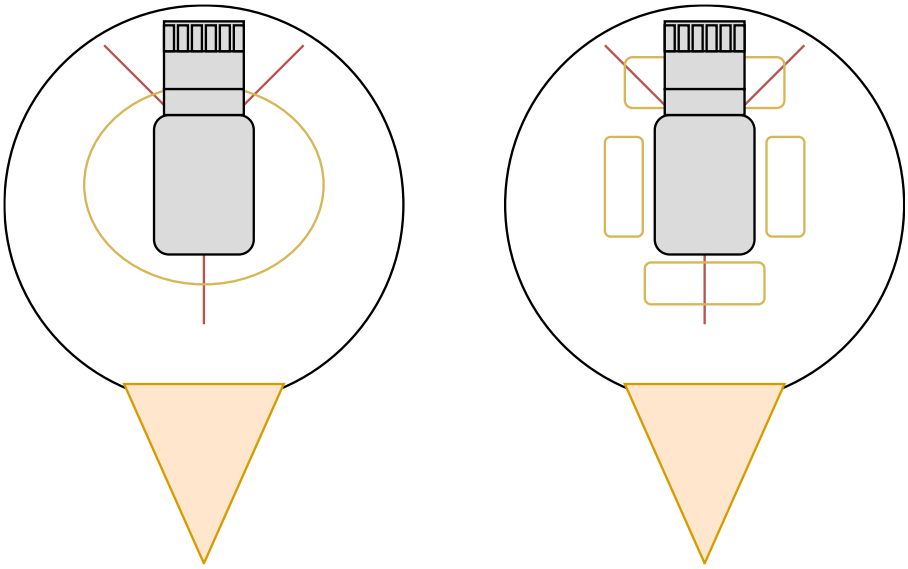


Figure 2.2. Skull incision strategies for accelerometer attachment. Left: original strategy of shaving the top layer of skull in a single contiguous shape and affixing the accelerometer to the bottom layer of skull with dental acrylic in the center of this window. Right: later strategy of affixing the accelerometer to the top layer of skull and shaving the top layer of skull in small patches around its footprint for a strong acrylic bond. This arrangement extends the lifetime of the implant considerably. In both diagrams, the view is that of the surgeon positioned above the bird, with the beak at bottom. The accelerometer is the gray assembly in the center. The midsagittal sinus and transverse sinuses are partially depicted by red lines. The retracted scalp is not depicted. Gold ellipse and rectangles depict the regions where the top layer of skull is shaved.

and in two cases implants still produced excellent transmission of vocalizations 6 months after implantation.

After surgery, birds were monitored for one hour and then returned to their home cage, where they were provided unrestricted access to food, water, grit, and cuttlebone. To avoid tangling tethers, perches were moved to floor-height.

Delayed auditory feedback and recordings

After singing had returned to pre-surgery levels (which occurred after between 2 and 6 days), the implanted accelerometer was attached to a custom-built low-torque passive commutator (https://github.com/margoliashlab/passive_commutator; for a discussion of the design of this device, see Brown, 2017) with a tether I built (later, purchased from Intan Technologies; given the

heavy wear tethers receive (leading to frequent failures) and the tedium and difficulty of building lightweight tethers by hand, I would strongly advise using off-the-shelf tethers whenever possible). A final acclimation period (2-3 days) ensured stable song production while a bird was continuously tethered.

The accelerometer records high-fidelity signals of the bird's singing, as well as movement sounds and artifacts induced by movement. I set up a continuous DAF environment by relaying the accelerometer signal through the amplifier used for the microphone and then broadcasting the signal back to the bird via a speaker in the sound chamber, after imposing a delay. Throughout these experiments, I used a delay of 100 ms, as this value has been shown to be most effective at eliciting changes in song in prior work (Manderscheid, 2014; Cynx and von Rad, 2001). Playback volume was chosen to be 5-10 dB louder than the bird's own song at the microphone in the sound isolation box.

I initially recorded a bird's singing for several days after it acclimated to isolation using the microphone. Once a bird recovered from the implantation surgery, I recorded both microphone and accelerometer signals on the days prior to and throughout the period of DAF exposure. Birds experienced continuous DAF beginning on the morning on the first day and continuing for seven full days (or until the bird became untethered). Both the accelerometer and microphone signals were recorded using JACK/jill.

A sample accelerometer trace containing singing may be seen in Figure 2.3. An implanted accelerometer will also pick up bone-conducted noise resulting from a bird's movement – mastication (particularly prominent as the birds are fed seeds), eye blinks, head bobs, hops (together with the attendant cage noise), and wing flaps. (While themselves silent or nearly so, head bobs and hops are usually present in the form of artifacts induced by rapid acceleration.) These noises were also played back to the birds over the speaker during DAF. Because of this, all

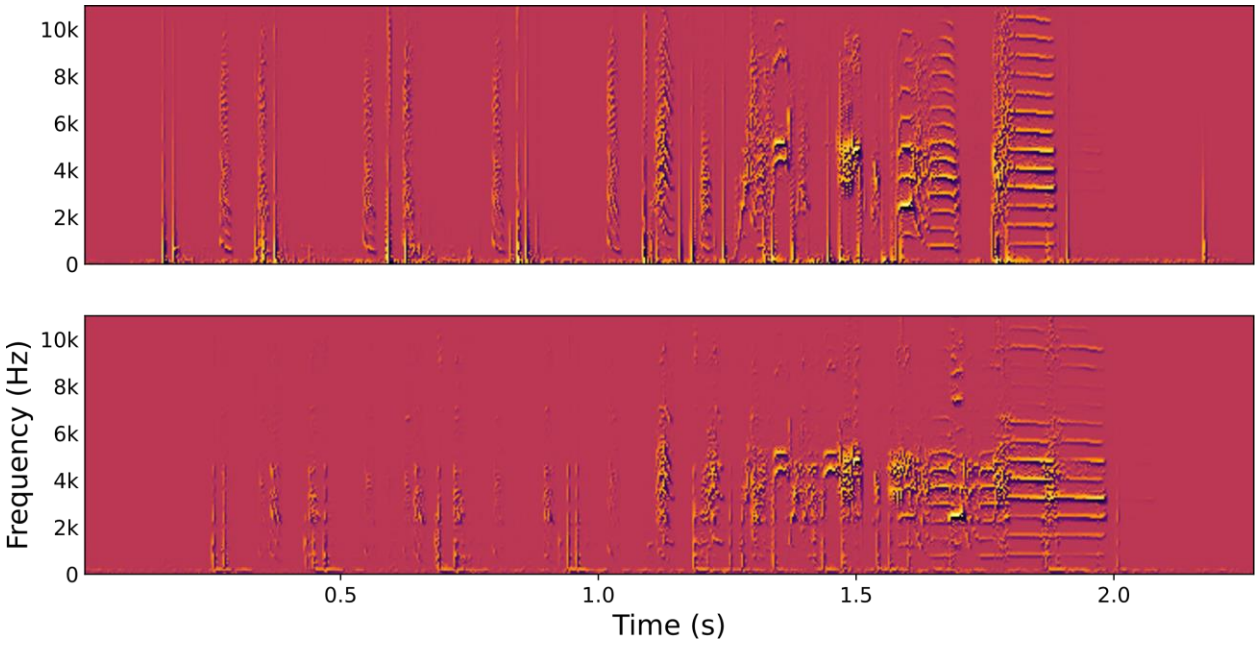


Figure 2.3. Example DAF spectrograms. Top: accelerometer signal. Bottom: microphone signal for the same time interval. Note the doubled elements in the microphone signal – the echo results from playing the accelerometer signal back over a speaker at 100 ms delay. The almost-punctate events in the accelerometer trace result from head movements, which are also picked up by the device.

song analysis was performed on the accelerometer recordings rather than the microphone recordings. Fortunately, birds remained relatively still while singing, and so these artifactual noises rarely overlapped with song. While singing, birds received two versions of auditory feedback – the natural feedback, both bone- and air-conducted, and the speaker-produced feedback 100 ms later.

Song analysis

All analyses were performed using custom software written in Python.

Accelerometer traces were first processed with a three-step segmentation routine (implemented in the lab's custom bark data framework): (1) segments were initially identified by amplitude threshold crossings; (2) segments separated by no more than 20 ms were joined; and (3) segments shorter than 30 ms were discarded. (This procedure follows one described in Koumura

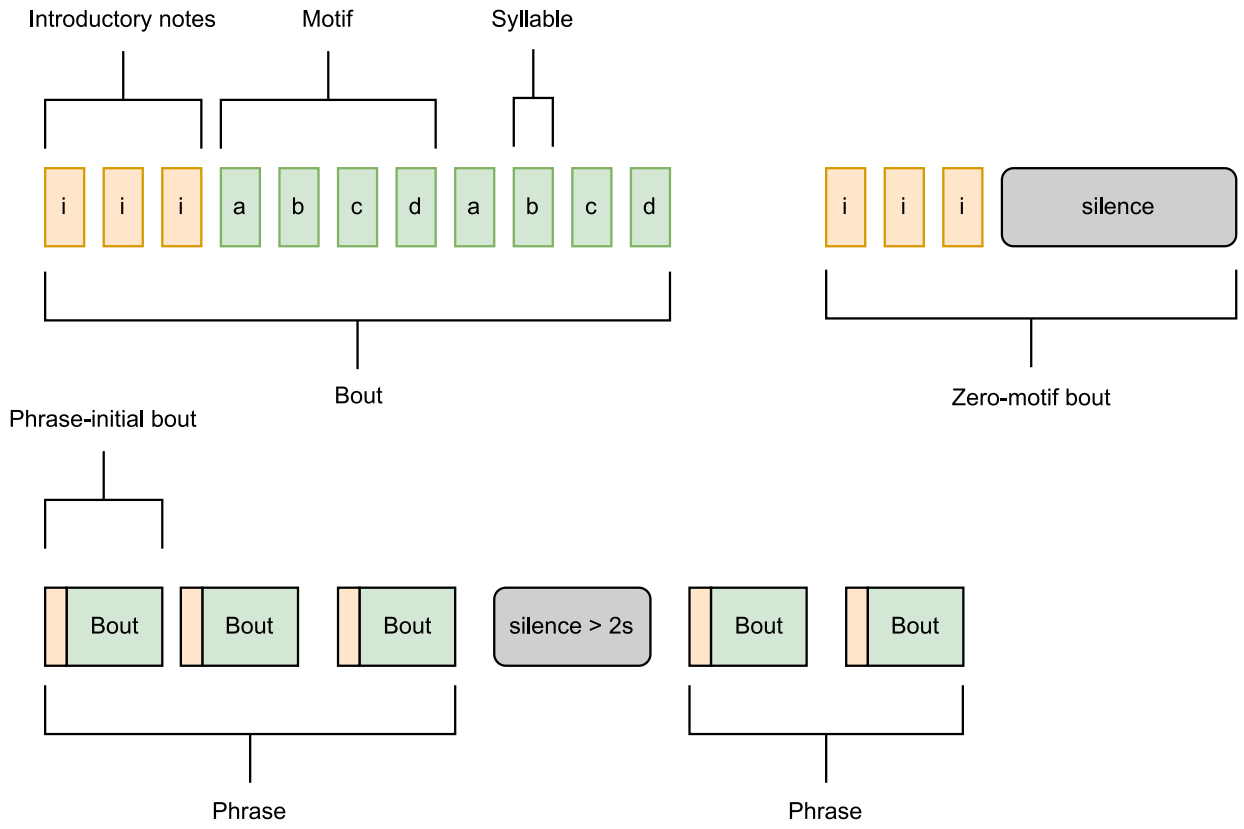


Figure 2.4. Diagram of song structural element definitions.

and Okanoya, 2016.) These segments were manually labeled by examining oscillograms and spectrograms; in the process, vocalization onsets and offsets were corrected, and non-vocalization segments (e.g., movement or cage noise) were discarded. Given the primacy of behavioral analysis for this study, I (reluctantly) determined that the error rate of a deep learning syllable labeling algorithm used previously in our lab (see Brown, 2017) would have required such intensive human correction that labeling by hand from scratch would represent little additional work.

Song structure above the syllable level was reconstructed using a simple state machine. Runs of introductory notes and syllables were identified independently, and combined with adjacent runs of the same type if separated by no more than 500 ms. For each bird, a canonical complete "motif" (a sequence of syllables produced in a stereotyped sequence; the most basic higher-order unit of

zebra finch song) was identified manually and served as the template for the automated breakdown of syllable runs into motifs. Within a run of syllables, motif start times were identified by the appearance of the first syllable label or a call, with end times determined by the end of the last syllable in the motif; manual checks confirmed that none of the birds sang out-of-order syllable sequences that would complicate this approach. Syllable and introductory note runs were split if they contained silent periods of 800 ms or longer. A motif was joined to a preceding introductory note run if separated from it in time by 500 ms or less. I identified as a "bout" any sequence of motifs, preceded or not by a run of introductory notes, and separated in time from other motifs by at least 800 ms. Runs of introductory notes without a following sequence of motifs I termed "zero-motif" bouts. Finally, bouts separated in time by no more than 2000 ms I grouped into "phrases." For a graphical explanation of these terms, see Figure 2.4

Using this strict hierarchical definition of zebra finch song, I was able to characterize essentially all vocalizations produced by these birds.

Statistical analysis

I believe my analyses of the song data fall into two families: those which confirm previous results reported in birds subjected to DAF over longer periods of time, and those which are new to this study, and which were devised and applied *a posteriori*.

These post hoc analyses of behavioral changes under DAF grew out of exploratory analysis of the labeled singing data. As will become apparent below, measures of these behavioral changes do not lend themselves well to summary statistics – none of the distributions is normally distributed. This complicated attempts to select appropriate statistical tests. At the same time, it is particularly important with post hoc analysis to bound the family-wise error rate to guard against spurious conclusions due to multiple comparisons. To address these points, I defined a measure of effect size

per behavioral change and per bird, creating an aggregated family of 18 tests, and applied the Bonferroni-Holm sequential correction to this pseudo-family (Aickin and Gensler, 1996). In discussing the results, I will follow Holm's approach of comparing the uncorrected p-values to progressively more conservative thresholds (for 18 tests, these thresholds scale from 0.05 down to 0.00278, using a family-wise target α of 0.05), and I will also report corrected p-values obtained by multiplying the uncorrected p-values by the appropriate integer, which can be directly compared to the family-wise target α of 0.05, in the hope that this method of presentation is clearer to the reader.

I used two tests for these comparisons: the (nonparametric) two-sample Kolmogorov-Smirnov (K-S) test to compare two distributions directly, and a bootstrapped estimate of the mean of the relevant distribution on the pre-DAF day to compare with that observed under DAF. The K-S test can produce unintuitive results for small sample sizes (it is very sensitive to tail values), and so I did not want to rely on it exclusively. For the bootstrap, for each bird the mean of the true distribution from the first DAF morning was compared with a distribution of estimates of the mean of the relevant pre-DAF distributions calculated for each of 10,000 resamples. The p-values obtained from these comparisons were compared for significance (again using Bonferroni-Holm correction for an 18-member family) against a target α of 0.025, as the tests are one-sided.

Further analytical investigation of the behavioral changes underlying these coarse-grained outcomes were not part of the second family proper, and so were only corrected for multiple comparisons across six birds.

Code and software accessibility

`bark` is a data storage and analysis framework developed in our lab. The bark analysis framework (version 0.2) may be found here: https://github.com/margoliashlab/bark.

The song structure analysis library (`zfparse`, version 1.1) may be found here: https://github.com/gfetterman/zfparse.

Results

I recorded singing from six birds before and during continuous DAF. My analysis focuses on the first four hours of DAF administration (the birds' subjective morning), and all comparisons are to the corresponding period of the day before the beginning of the DAF protocol. Birds sang alone – I did not introduce female cagemates, to eliminate the potential confounding factor of the female's own response to DAF.

Every bird showed changes in singing behavior in response to DAF. Some changes I observed in all six birds, some in most, and a few appeared to be unique to one bird or another. I will first describe these responses qualitatively, then proceed to statistical measures of the changes.

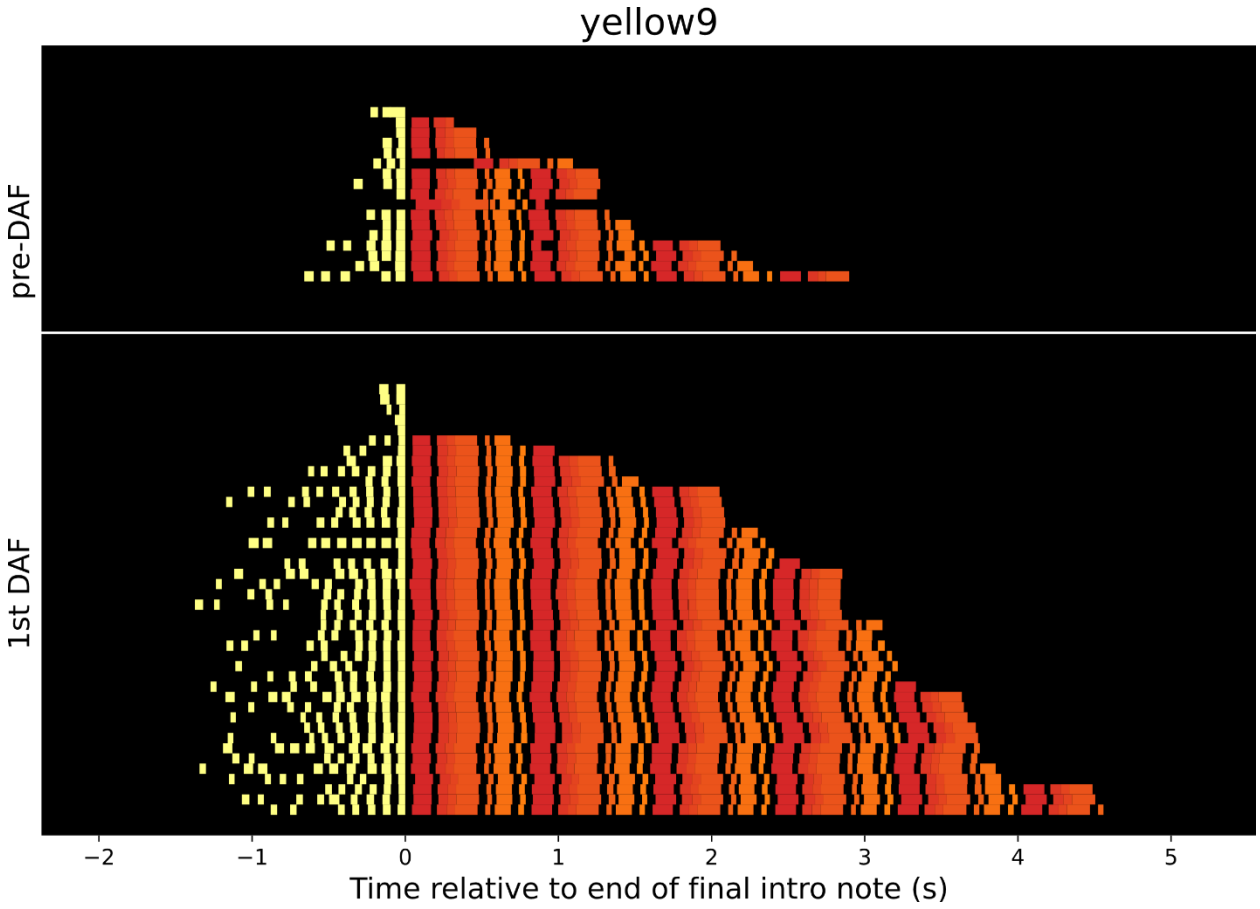


Figure 2.5. Cakeplot for yellow9. Raster plot of all bouts sung. Top panel: songs from the pre-DAF morning. Bottom panel: songs from the first DAF morning. The duration of the bouts in seconds is labeled along the horizontal axis. Bouts are arranged within each day by the number of motifs in the bout. Bouts are aligned by the end of the last introductory note, unless no introductory notes were sung preceding the bout – in this case, the beginning of the first syllable of the first motif in the bout is aligned to the mean beginning of the first syllable in the first motif of all the bouts from that bird. Introductory notes are colored light yellow (nearly white in grayscale), and successive syllables in a motif are colored in shades of orange (darker grays in grayscale). This layout scheme applies to Figures 2.6 through 2.10 below as well.

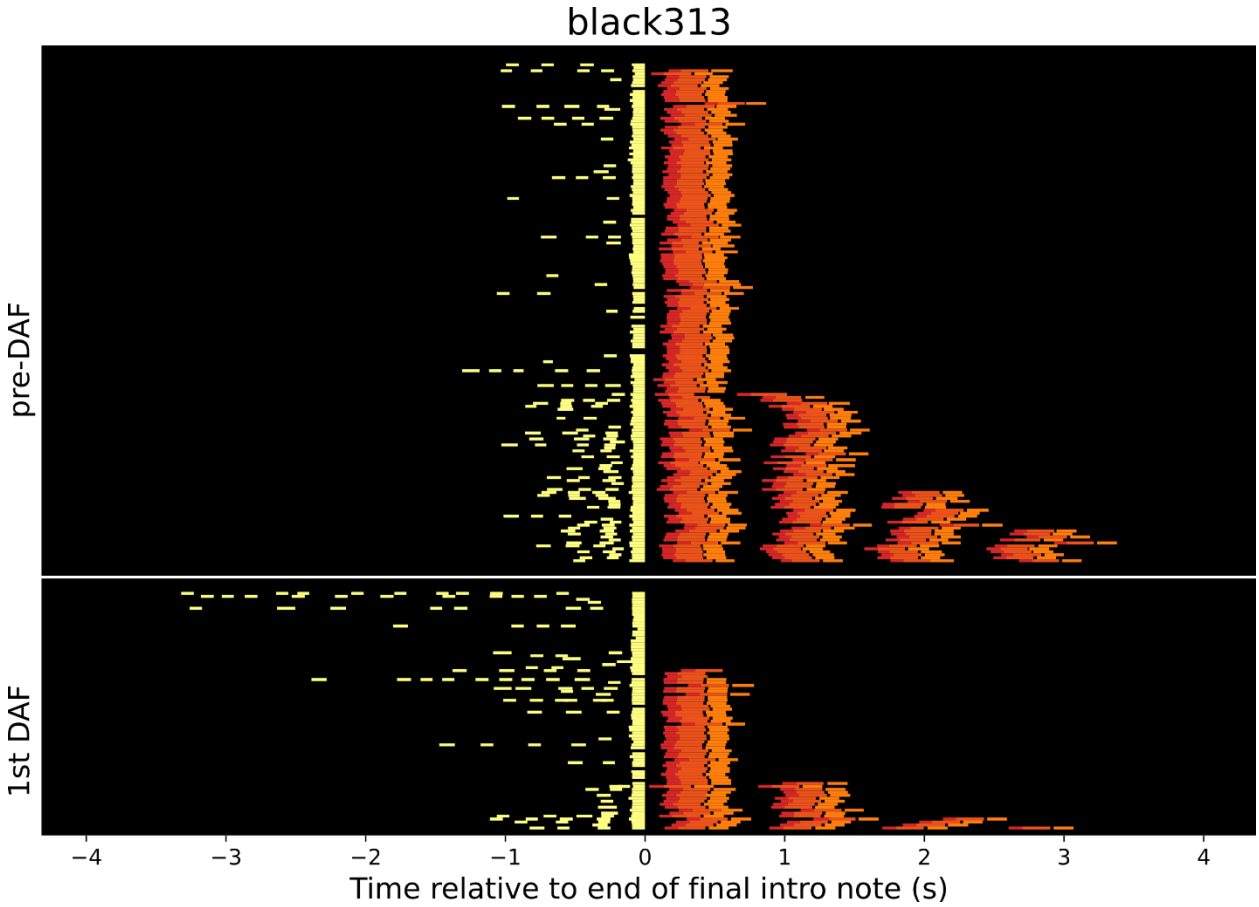


Figure 2.6. Cakeplot for black313. Refer to caption for Figure 2.5 for description of layout.

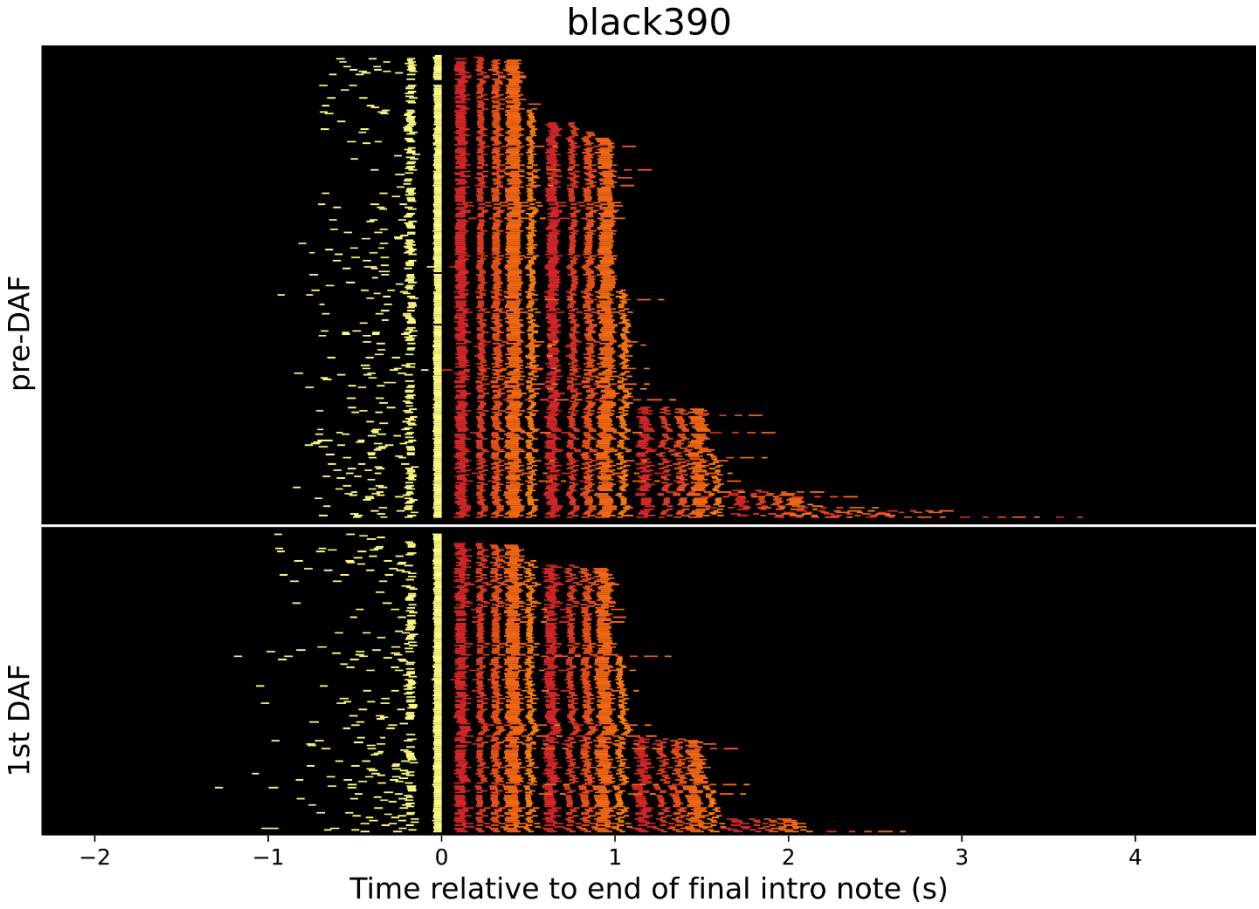


Figure 2.7. Cakeplot for black390. Refer to caption for Figure 2.5 for description of layout.

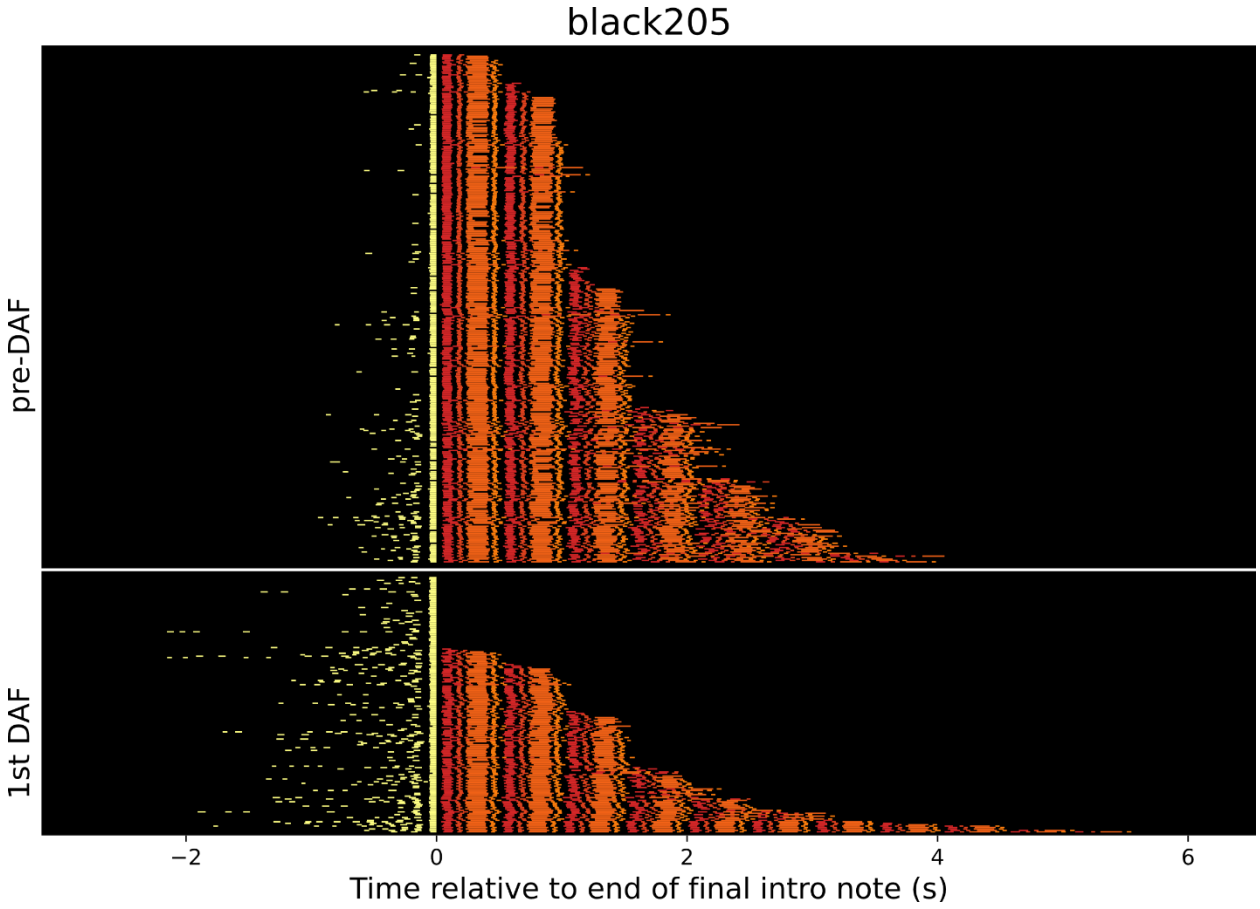


Figure 2.8. Cakeplot for black205. Refer to caption for Figure 2.5 for description of layout.

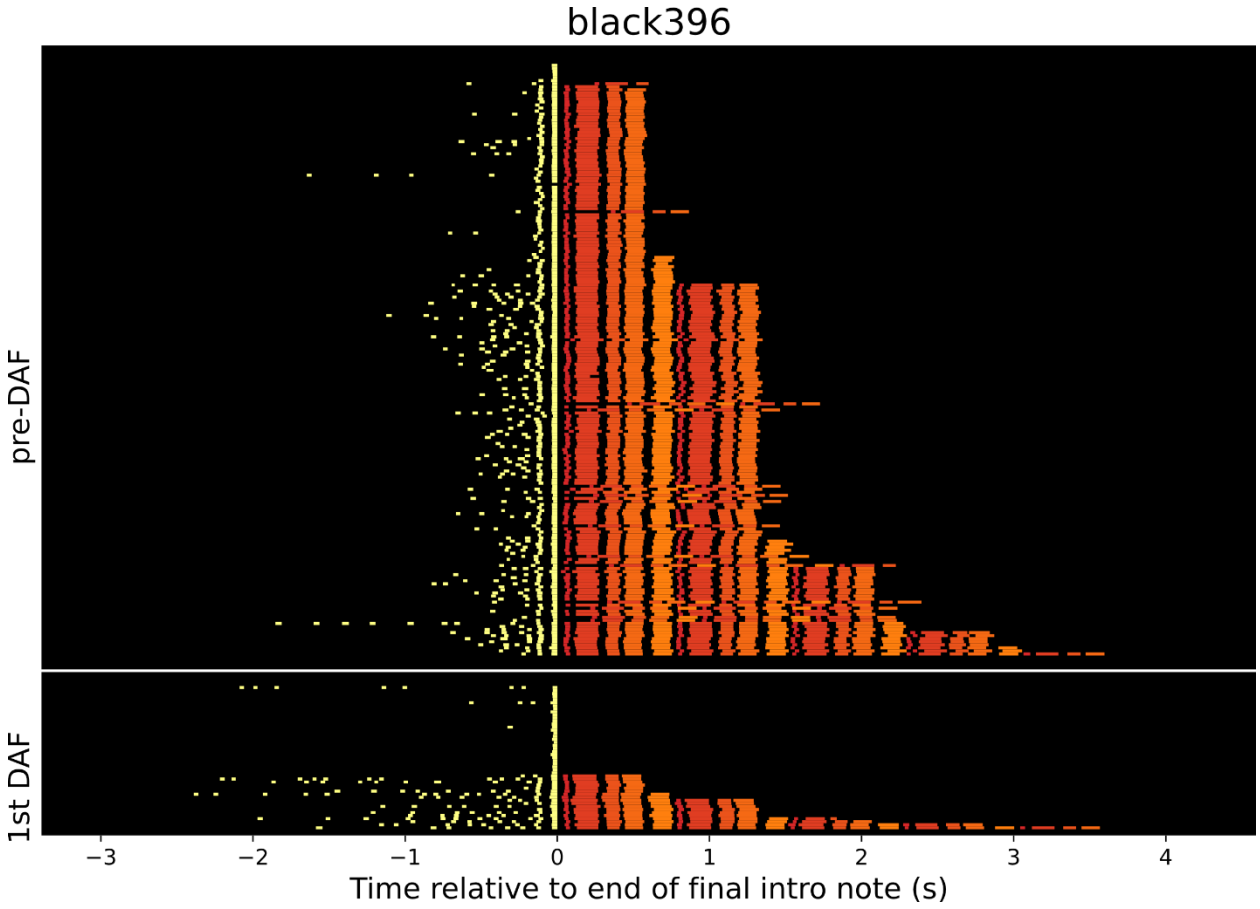


Figure 2.9. Cakeplot for black396. Refer to caption for Figure 2.5 for description of layout.

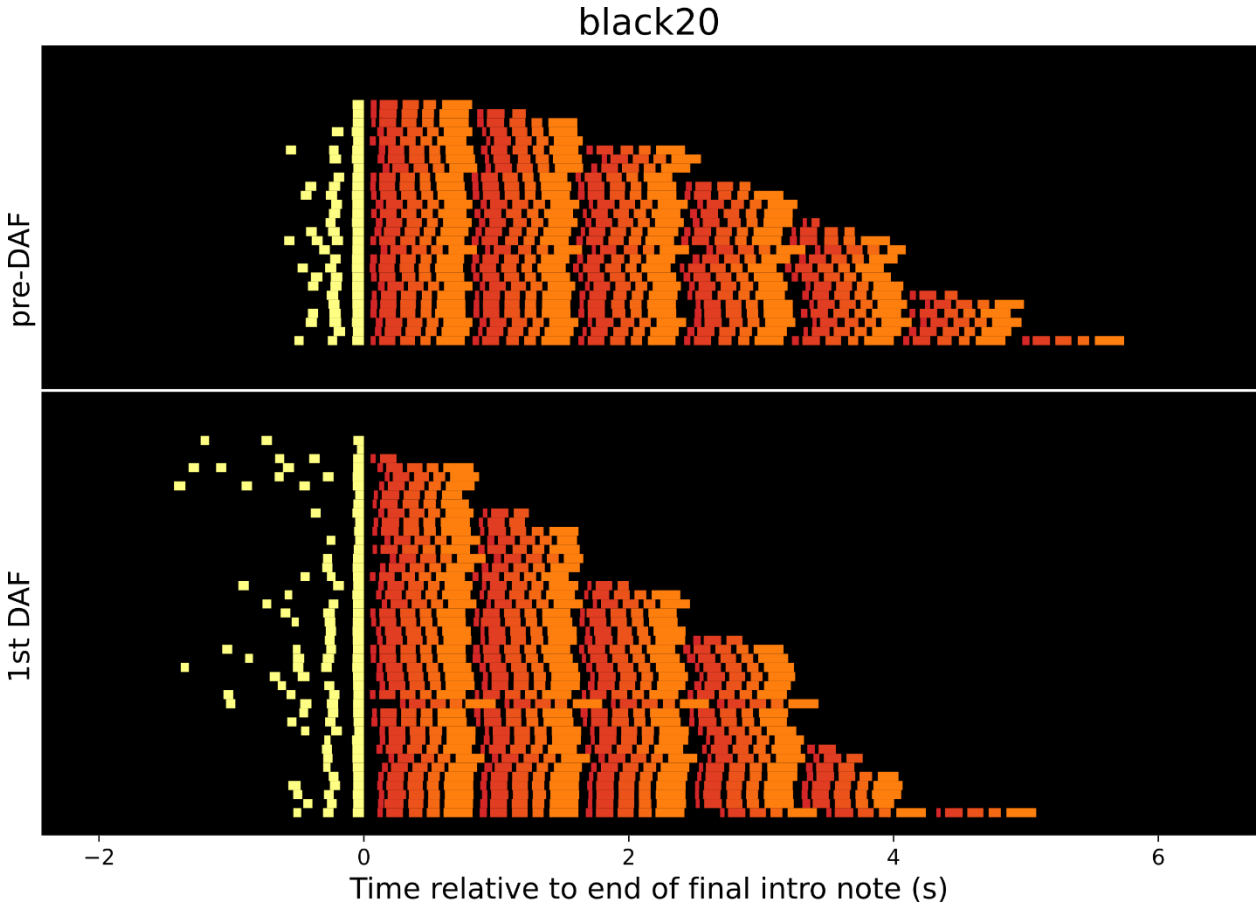


Figure 2.10. Cakeplot for black20. Refer to caption for Figure 2.5 for description of layout.

Qualitative observations

Most birds sang robustly (Figures 2.5 – 2.10) both before (top panels) and during (bottom panels) DAF. (I modeled this graphical depiction of total song activity, which we have taken to calling "cakeplots," after similar figures in Bruno and Tchernichovski, 2017.) Each row of colored bars represents one bout of song, arranged top-to-bottom within each condition (before or during DAF) by the number of motifs in the bout. Introductory notes are shown in yellow (near-white in grayscale) and song syllables are shown in shades of orange and red (light to dark grays in grayscale). I chose the colors manually, rather than tying them to the average frequency of the syllable in question (as was done in Bruno and Tchernichovski, 2017; while such a labeling scheme

Bird	Pre-DAF morning		First DAF morning	
	Motifs	Bouts	Motifs	Bouts
yellow9	32	17	146	42
black313	258	168	75	80
black390	644	296	427	191
black205	1104	356	442	179
black396	349	194	35	17
black20	114	27	129	42

Table 2.1. Song produced per bird, pre-DAF vs DAF.

makes sense for their purposes, it adds nothing here). Each bout of song is aligned to the end of the final introductory note prior to the motif beginning. The abscissa of the figures indicates time relative to this alignment. Those few bouts without introductory notes are aligned so that the beginning of their first motifs are situated at the mean value of first motif onset for the bouts with introductory notes; this was done purely for display purposes, and had no effect on numerical analyses or statistical outcomes below.

Prior to the onset of DAF, all but one bird sang between 100 and 1100 motifs per morning (for most of them, this number had been higher before the accelerometer implantation surgery). The exception was yellow9, which sang only 32 motifs on the pre-DAF morning. This bird sang more in the afternoon of the pre-DAF day (110 motifs within 4 hours), and so I did not exclude him from my analyses. Another bird, black20, sang 114 motifs on the pre-DAF morning, but the rest of the birds sang between 259 and 1104 motifs during this period (Table 2.1).

As expected for zebra finches, songs were highly stereotyped. Transitions outside of the canonical motif order were rare (Figure 2.11). The birds sang their motifs to completion at rates between about 50% (yellow9 and black396, at 47% and 52%, respectively) and 90% or more (black205 (89%), black20 (96%), and black313 (98%)) (Figure 2.12). There appeared to be a correspondence between song complexity and prevalence of complete motifs, with yellow9's song

being particularly long and complex and black313's song short and simple, but the relationship was not especially reliable (as an example, black396's song is similar in complexity and length to black205 and black390). Motifs not sung to completion were predominantly likely to occur at the end of a bout (Figure 2.12).

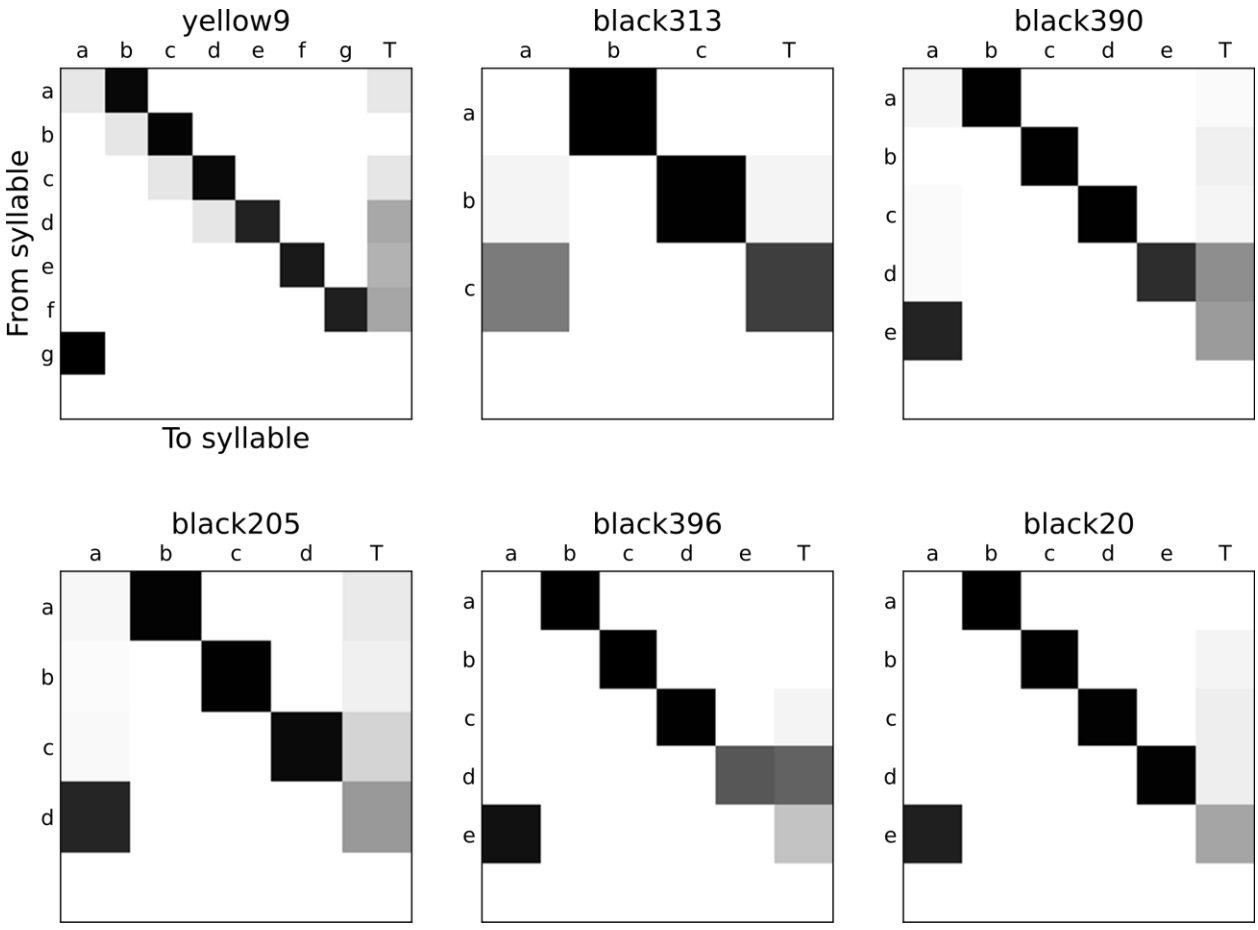


Figure 2.11. Transition probabilities in unperturbed song. Darker colors indicate higher probability of a transition. The symbol 'T' denotes bout termination. Probabilities in a row sum to unity.

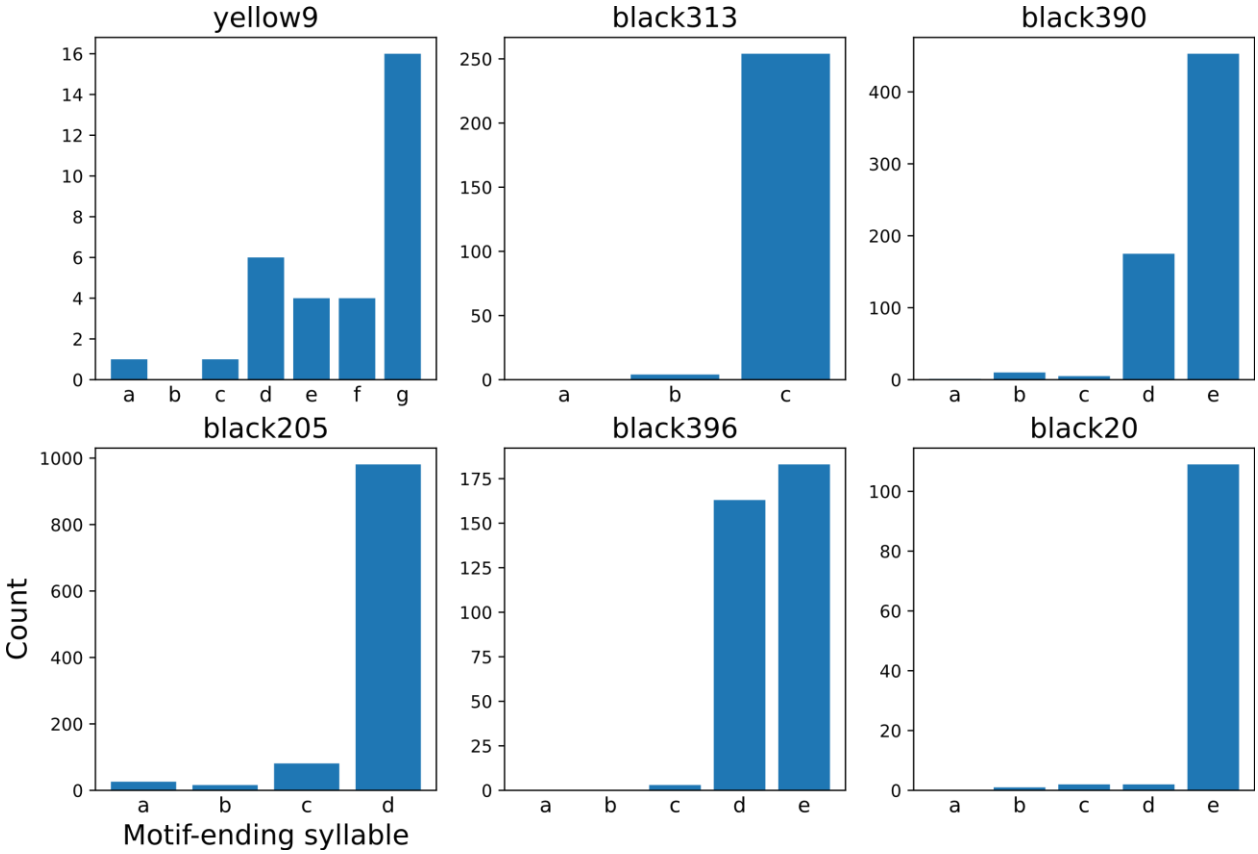


Figure 2.12. Motif-ending syllables in unperturbed song. These histograms represent the final column (labeled 'T') in Figure 2.11 (but in some cases do not match the color map for Figure 2.11, as its probabilities are normalized across a row, not along a column).

The average number of introductory notes preceding a bout of song varied among birds between 1.4 ± 1.0 and 3.0 ± 1.4 (mean ± standard deviation, but note that the distribution for a given bird is usually non-normal under visual inspection – see Figure 2.13), though the maximum number of introductory notes was more variable (4 per bout for black20, and 10 per bout for black396). My suggestion of the phrase as a meaningful unit of behavior was initially based on the histogram of inter-bout intervals across birds (with all exhibiting a strong peak under 2 s) during normal song (Figure 2.14), but I also observed dramatically different introductory note sequence lengths between phrase-initial bouts and later bouts in a phrase (Figures 2.13, 2.14). Phrase-initial bouts were preceded by 1.56 more introductory notes on average than are later bouts. This difference

between phrase-initial and later bouts was statistically significant for every bird (Kolmogorov-Smirnov test: yellow9 p=7.4e-4; black313 p=2.22e-16; black390 p=1.77e-71; black205 p=7.77e-16; black396 p=1.43e-30; black20 p=5.70e-4) and for the birds in aggregate (p=0.0312, Wilcoxon signed-rank test on mean intro notes per bout type). This effect also appeared in the maximum-length introductory note sequences, which increased by 3.67 introductory notes on average. These results reinforce prior reports of a relationship between inter-bout interval and the number of introductory notes before the next bout (Rajan and Doupe, 2013; Rajan 2018); I believe that introducing a phrase structure provides a conceptual framework for this phenomenon.

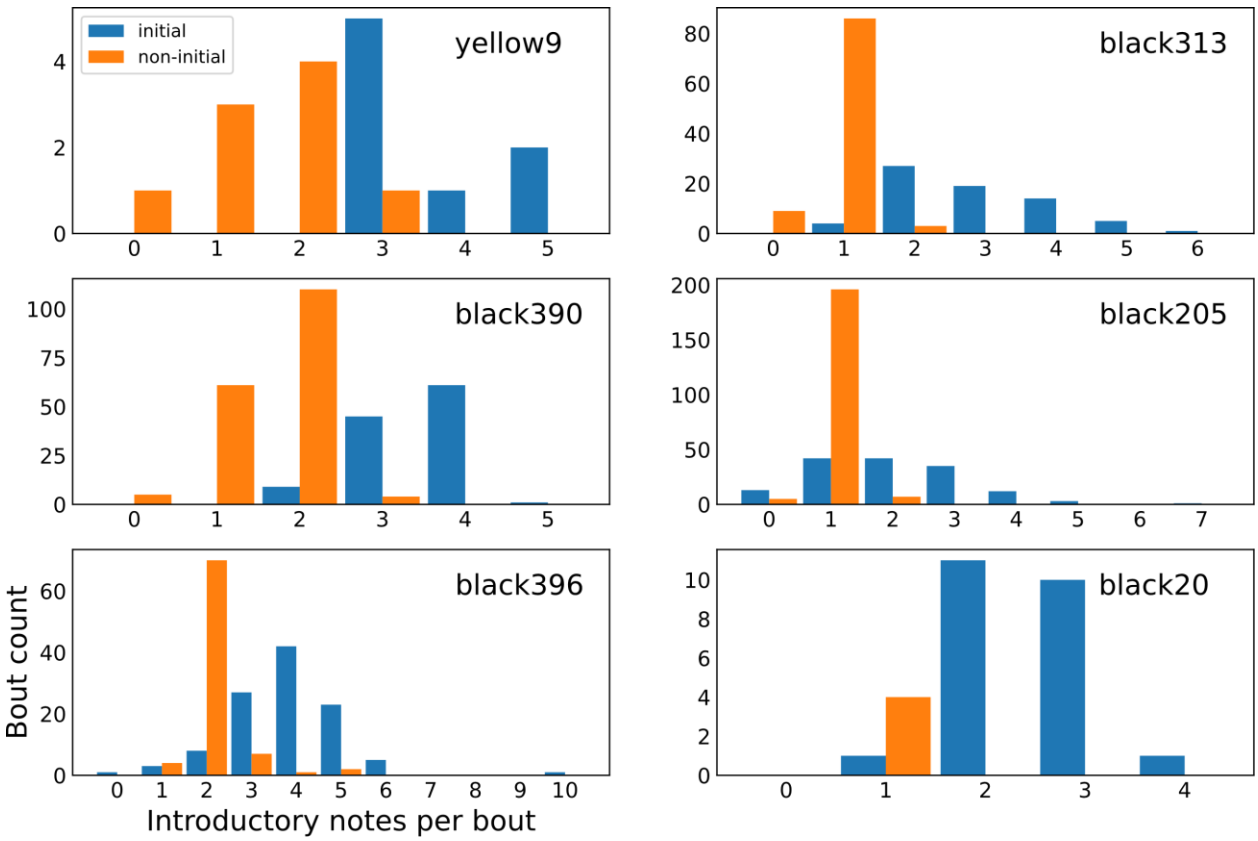


Figure 2.13. Introductory notes per bout by type, in normal song. For every bird, phrase-initial bouts are preceded by more introductory notes, on average, than non-phrase-initial bouts. Note also that the aggregated histograms would, for most birds, appear non-normal.

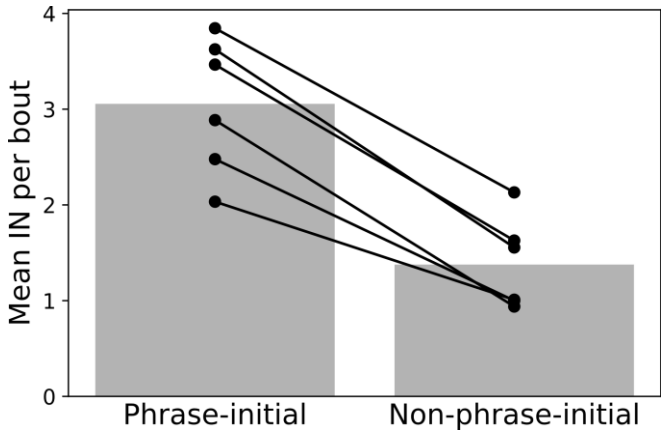


Figure 2.14. Mean introductory notes per bout type in normal song. The mean across birds is shown in gray bars, with values for individual birds in black.

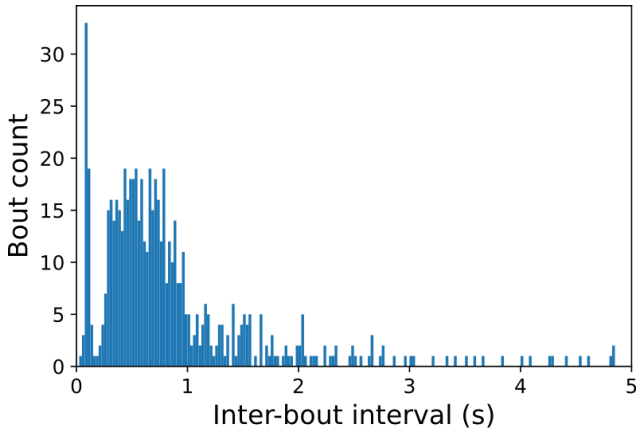


Figure 2.15. Inter-bout intervals, aggregated across all birds, in normal song.

(I feel compelled to point out that I posited the existence of a structure of zebra finch song above the level of the bout, and settled on both the name "phrase" (chosen more or less arbitrarily from musical terminology) and the definition (separation between bouts by 2 seconds or less), around 2016, and presented these results in a poster at the Society for Neuroscience conference that year. Another group of researchers have since used the same name and structure and nearly the same definition in a recent paper (Daliparthi et al, 2019), in what is certain to be pure

coincidence – there are, after all, only so many terms in music, and the peak in the inter-bout interval distribution has been well-established since at least 2013 (seen in, for example, Rajan and Doupe, 2013). I make this point only to suggest that this structure has been independently remarked upon by other researchers, who defined it similarly to my choices here. (There is a small difference between the definitions: mine requires bouts in a phrase to be separated by no more than 2 seconds, measured between the end of the final motif of one bout and the beginning of the introductory note sequence of the next, whereas theirs requires the separation interval to be measured between the end of one bout's last motif and the beginning of the next bout's first motif. My definition more closely follows the measurements reported in Rajan and Doupe (2013), and so I think is a better choice, but the difference is small, and not worth perseverating upon.) Thus, the phrase structure seems likely to be a useful tool in the future, so long as we are ready to acknowledge evidence that may undermine it.)

The birds also exhibited variation between individuals in the temporal arrangement of song. Birds produced bouts of a characteristic length, ranging from a single motif per bout (black313) to three (black205) to five motifs per bout (black20). Introductory note run structure also showed variation across individuals. Some birds virtually always preceded non-phrase-initial bouts with a single introductory note (black20, black205), while black396 more characteristically sang two introductory notes before bouts later in a phrase (Figure 2.14). (This result echoes similar observations in Bruno and Tchernichovski (2017) regarding what they term "connectors.") Inter-intro-note intervals may also be characteristic of individuals, ranging from an average of 100 ms (black396) to 150 ms (black390).

During the first morning of DAF treatment, 4 out of the 6 birds sang less (measured by numbers of both bouts and total motifs) than they had the morning before. The 2 birds who did not sing less (yellow9 and black20) were the two who sang the least on the pre-DAF morning (32 motifs in the

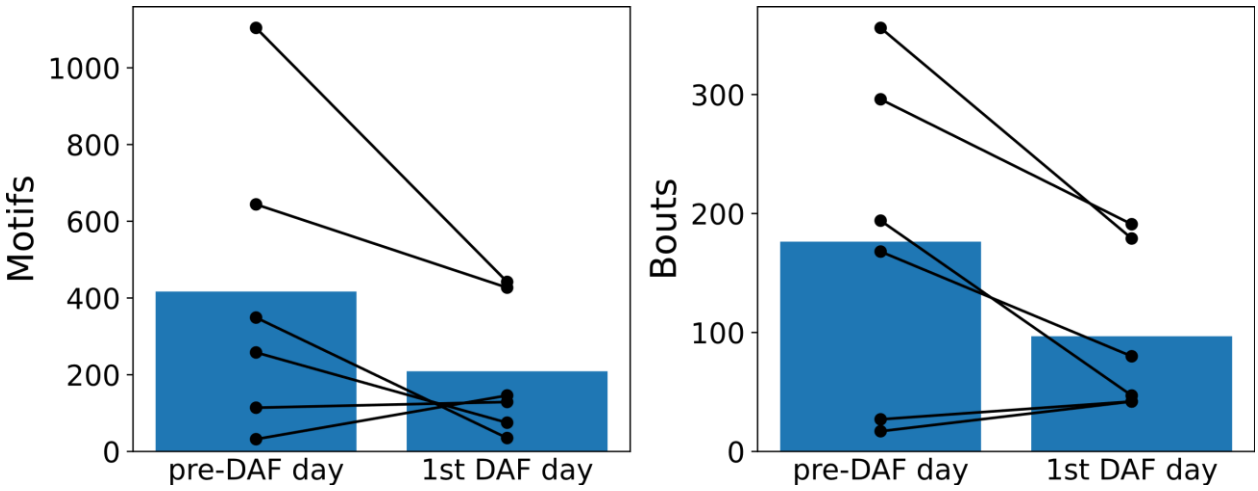


Figure 2.16. Song production before and during DAF. Mean values for all birds are shown as blue bars, with individual birds' values in black.

morning and 114 motifs in the morning, respectively). It is thus possible that, had I waited to begin DAF treatment until they sang at the same level as the other 4 birds, they too would have sung less under DAF (Figure 2.16).

The most striking change observable in these data, to my eye, was the change in the distribution of motifs per bout. The first DAF morning saw the appearance of introductory notes, both individually and in runs, without accompanying motifs. These "zero-motif" bouts were rare in unperturbed song (never occurring for black205, and black20, and comprising 1% or less of all bouts for black313 and black390, 3% of all bouts for black 396, and 6% of all bouts for yellow9). For 5 out of 6 birds, the prevalence of zero-motif bouts increased dramatically (Figure 2.17, 2.18), reaching over 60% of all bouts for black396 (Figure 2.9); for the remaining bird, black20, they also increased, but only to roughly 5% of all bouts (Figure 2.10). (At the risk of dissecting the frog: I chose the term "zero-motif bout" with deliberate irony – a bout is everywhere defined in the literature as a sequence of motifs, and a bout containing zero motifs is … not a bout. But something is happening here.)

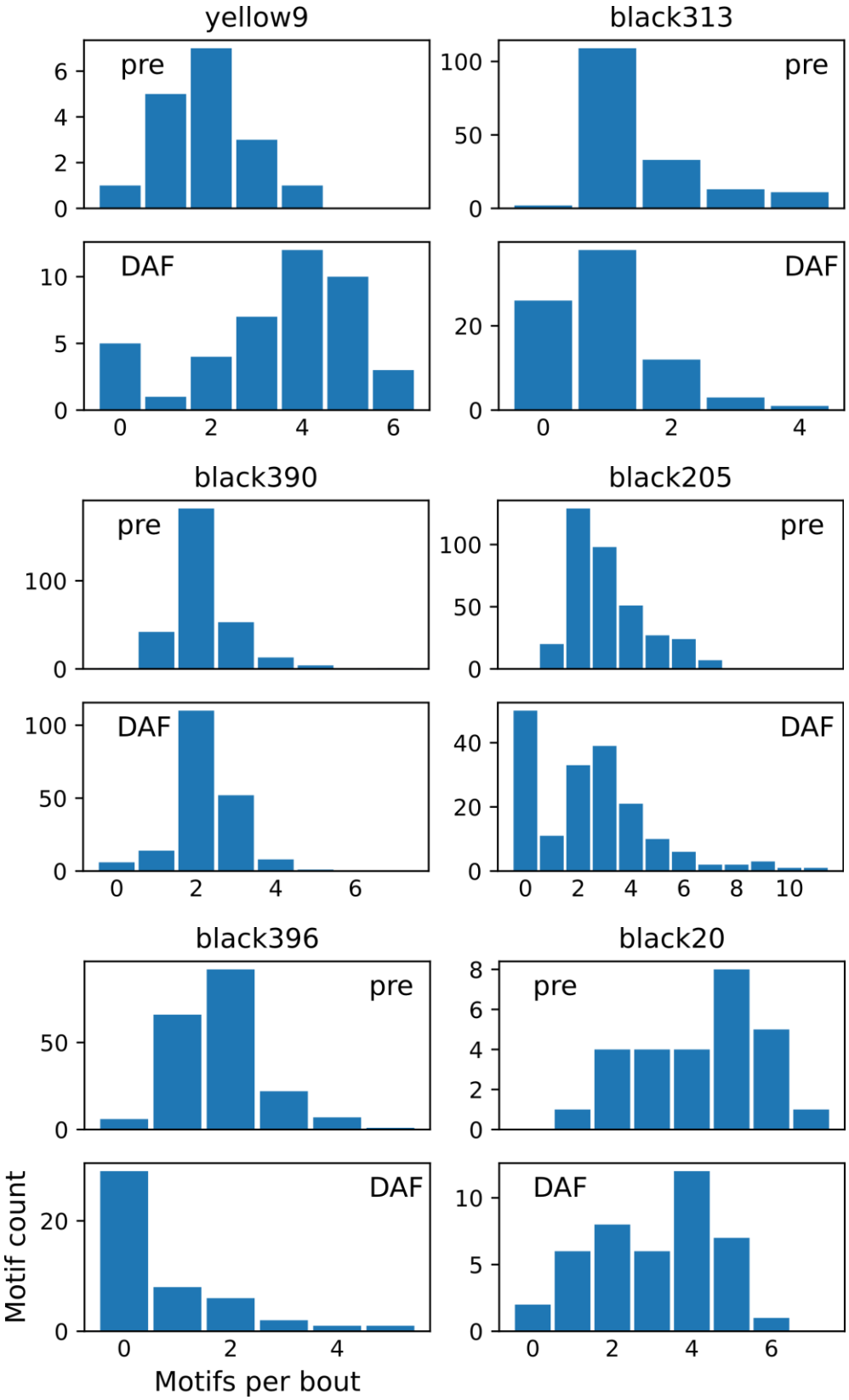

yellow9
black313
black390
black205
black396
black20
pre
DAF
Motif count
Motifs per bout

Figure 2.17. Motifs per bout, pre-DAF and first DAF day, all birds. (previous page)

I also particularly noted an apparent increase in the length of time separating introductory notes under DAF compared to normal song (Figure 2.19). In Figures 2.5 – 2.10, note the appearance of introductory notes in the bottom halves of the figures at greater delays prior to the onset of song than in the top halves. This appears to accompany an increase in the number of introductory notes preceding bouts (Figure 2.20). It should be noted that in the birds with more introductory notes under DAF, the longer introductory note runs do not appear to preferentially occur within zero-motif bouts.

I would also point out that, in Figures 2.5 – 2.10, the small number of introductory notes that appear to be shorter than the average are not an artifact of raster graphics or faulty labeling – these vocalizations were indeed shorter than their neighbors (sometimes by as much as 50% – an introductory note which is normally 50 ms in duration might appear in this manner to be only 25-28 ms in duration), as confirmed by eye from the audio oscillograms; I cannot currently explain this phenomenon. It is possible that these introductory notes were being subvocalized as another response of the bird to DAF, with the non-phonated portions not showing up on the spectrogram.

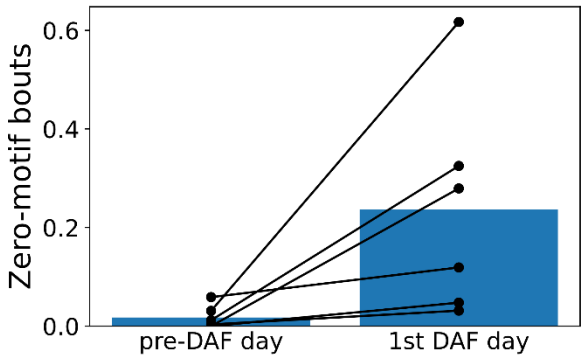


Figure 2.18. Proportion zero-motif bouts per morning. Mean values for all birds are shown as blue bars, with individual birds' values in black. Every bird produced more zero-motif bouts as a proportion of their bouts under DAF than under normal conditions.

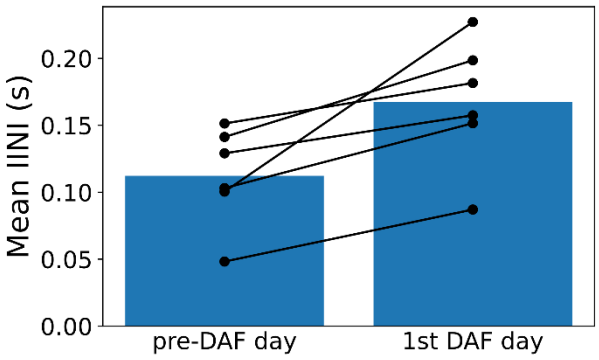


Figure 2.19. Inter-introductory-note intervals. Mean values for all birds are shown as blue bars, with individual birds' values in black.

I was unable to find evidence for this by manipulating the spectrogram color maps (i.e. to observe more of the signal near the noise floor). If it is the case, then the bird was truly subvocalizing, which could only be confirmed by recording syringeal EMG.

Statistical tests of behavioral changes

From these data, I initially identified three principal behavioral changes that occurred under DAF.

Bout duration changes

The distribution of the number of motifs in a bout changed under DAF. While this was prominently driven by a much higher prevalence of zero motif bouts in song produced under DAF, two birds (yellow9 and black390) tended to sing longer bouts, in addition to singing more zero motif bouts (Figure 2.17). Kolmogorov-Smirnov tests between the pre-DAF and first DAF morning motif-per-bout distributions indicated significant differences for 4 out of 6 birds (raw p-values: yellow9 p=0.00104; black313 p=3.29e-5; black205 p=5.34e-9; black396 p=1.05e-12; after Bonferroni-

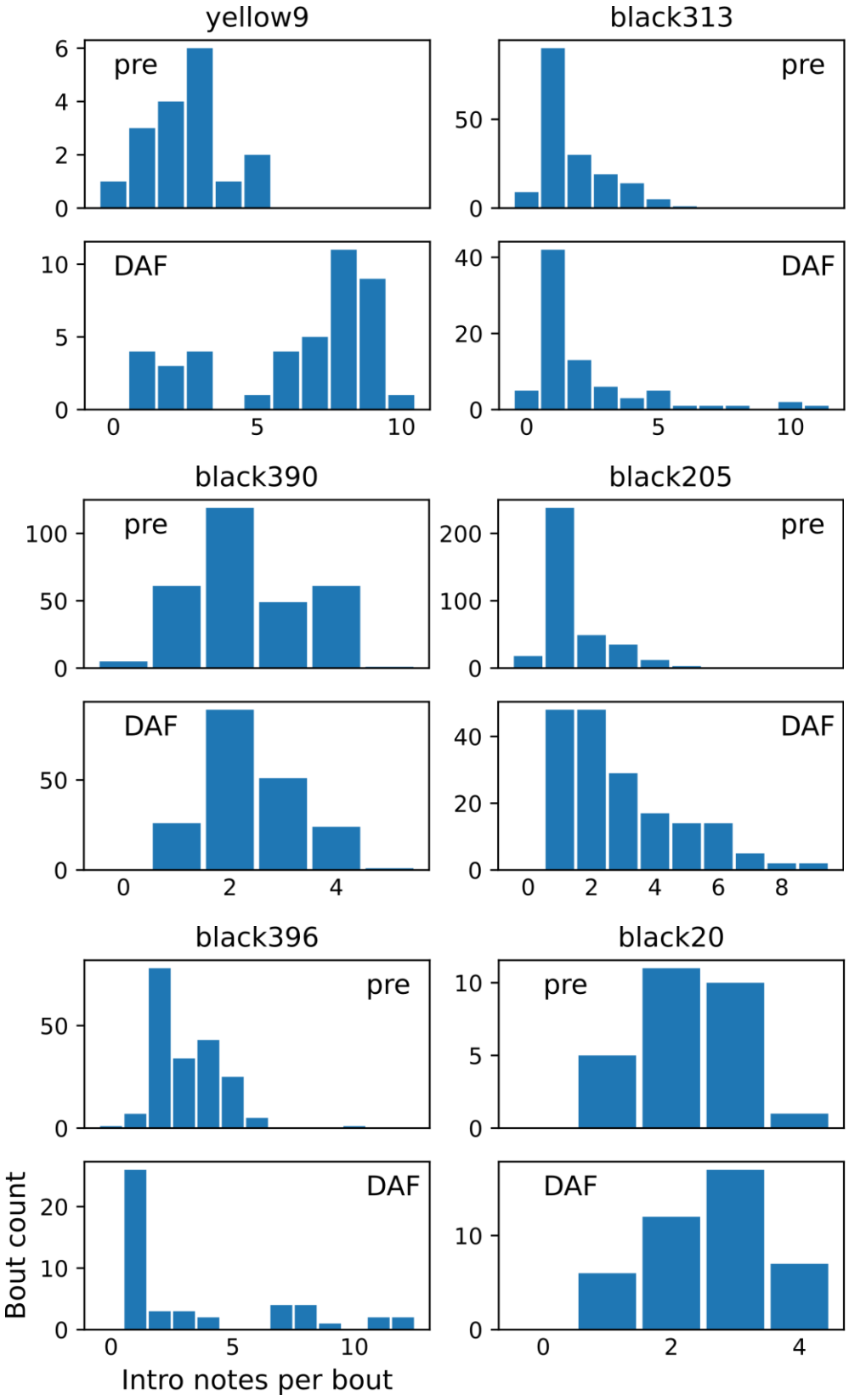

yellow9
black313
black390
black205
black396
black20
pre
DAF
Bout count
Intro notes per bout

Figure 2.20. Introductory notes per bout, pre-DAF and first DAF day, all birds. (previous page)

Holm correction (treating all 18 comparisons in these three sections as one family, as described in Methods), p=0.00728, 0.000329, 7.48e-8, and 1.785e-11, respectively), and a nonsignificant result for 2 birds (raw p-values: black390 p=0.427; black20 p=0.0445; after Bonferroni-Holm correction, p=1.28 and 0.22, respectively).

I also tested the significance of this effect under DAF via bootstrap resampling of the bird's pre-DAF distribution of motifs sung per bout, comparing the means of the motif-per-bout distribution on the first morning of DAF and the bootstrap resamples of the motif-per-bout distribution on the pre-DAF morning. 5 out of 6 birds showed significant differences in mean motifs per bout between the two days under this test (raw p-values: yellow9 p=0.0001; black313 p=0.0001; black205 p=0.0001; black396 p=0.0001; black20 p=0.0001; after Bonferroni-Holm correction, p=0.0018, p=0.0017, p=0.0016, p=0.0015, and p=0.0014, respectively), with a nonsignificant result for black390 (raw p=0.114, Bonferroni-Holm corrected p=0.228).

Bouts gain more introductory notes

The number of introductory notes preceding bouts increased under DAF (Figure 2.20). No dramatic explanation was obvious from the distributions – the effect appeared to be the result of simply shifting weight to the right in the distributions. The Kolmogorov-Smirnov test only indicated significant differences between the introductory note per bout distributions for 3 out of 6 birds (raw p-values: yellow9 p=1.74e-6; black205 p=8.88e-16; black396 p=1.22e-9; after Bonferroni-Holm correction, p=2.09e-5, 1.60e-14, and 1.83e-8, respectively), and could not distinguish the distributions for the other 3 birds (raw p-values: black313 p=0.586; black390 p=0.321; black20 p=0.700; after Bonferroni-Holm correction, p=1.17 and 0.7, respectively).

I also tested the significance of this effect under DAF via bootstrap resampling of the bird's pre-DAF distribution of intro notes produced per bout, comparing the means of the intro-note-per-bout distribution on the first morning of DAF and the bootstrap resamples of the intro-note-per-bout distribution on the pre-DAF morning. 4 out of 6 birds showed significant differences in mean motifs per bout between the two days under this test (raw p-values: yellow9 p=0.0001; black313 p=0.0001; black205 p=0.0001; black396 p=0.0001; black20 p=0.0001; after Bonferroni-Holm correction, p=0.0013, p=0.0012, p=0.0011, and p=0.001, respectively), with a nonsignificant result for 2 out of 6 birds (raw p-values: black390 p=0.217; black20 p=0.0109, Bonferroni-Holm corrected p=0.217 and p=0.0327, respectively; note that the target α for these bootstrap tests is 0.025, as the tests are two-sided).

Inter-introductory note intervals increase in duration

The silent periods between introductory notes in a series (i.e., two or more in a row) typically become shorter as motif initiation approaches, as has been previously reported (Rajan and Doupe, 2013; Rajan, 2018). This tendency persisted under DAF. However, the average duration of these silent periods between introductory notes increased significantly, from 0.115 s ± 0.025 s to 0.167 s ± 0.042 s under DAF (Figure 2.19). This effect was present in the distributions of inter-introductory note intervals for 5 out of 6 birds, as determined by the Kolmogorov-Smirnov test (raw p-values: yellow9 p=1.61e-8; black313 p=0.000319; black390 p=3.09e-6; black396 p=0.000196; black20 p=2.34e-11; after Bonferroni-Holm correction, p=2.09e-7, 0.00255, 3.40e-5, 0.00176, and 3.74e-10, respectively), with a nonsignificant result for the remaining bird (raw p-value: black205 p=0.0372; after Bonferroni-Holm correction, p=0.223).

I also tested the significance of this effect under DAF via bootstrap resampling of the bird's pre-DAF distribution of inter-introductory note interval durations, comparing the means of the inter-

introductory note duration distribution on the first morning of DAF and the bootstrap resamples of the inter-introductory note duration distribution on the pre-DAF morning. All 6 birds showed significant differences in mean motifs per bout between the two days under this test (raw p-values: yellow9 p=0.0001; black313 p=0.0001; black390 p=0.0001; black205 p=0.0001; black396 p=0.0001; black20 p=0.0001; after Bonferroni-Holm correction, p=0.0009, p=0.0008, p=0.0007, p=0.0006, p=0.0005, and p=0.0004, respectively).

Summary of post-hoc statistical tests

Of the 18 comparisons in the post-hoc family, using a Kolmogorov-Smirnov test 12 out of 18 are significant (less than 0.05) after correction for multiple comparisons, indicating that these effects are unlikely to result if the distributions of the relevant behavioral measures were unchanged under DAF. This conclusion also holds when considering a bootstrap resample of the means of the relevant distributions, with 15 out of 18 comparisons significant (less than 0.025) after correction.

I now turn to a more detailed examination of zero motif bouts.

Zero motif bouts

All 6 birds sang more zero motif bouts as a proportion of their overall bouts under DAF than before exposure (Figure 2.18). These increases in proportion were significant as a group (p=0.0312, Wilcoxon signed-rank test). This increase in zero motif bout production was the primary driver of the change in the motif-per-bout distributions. Comparing the distributions of motifs per bout excluding zero motif bouts using a K-S test produces a significant result for only 1 out of 6 birds (yellow9: raw p=0.00016, corrected p=0.00096, after Bonferroni-Holm correction for 6 tests, one per bird), with the rest nonsignificant (raw p-values: black313 p=0.976; black390 p=0.305; black205 p=0.588; black396 p=0.996; black20 p=0.0555).

If a zero motif bout appears within a phrase, it is likely to be in the first position, at a greater likelihood than a random bout would be ($p=0.0312$, Wilcoxon signed-rank test between proportion of zero motif bouts appearing at the beginning of a phrase and the proportion of all bouts that appear at the beginning of a phrase – i.e., the total number of phrases divided by the total number of bouts). Somewhat surprisingly, a zero motif bout is no more likely to terminate a phrase than a bout containing motifs ($p=0.844$, Wilcoxon signed-rank test between proportion of zero motif bouts appearing at the end of a phrase and the proportion of all bouts that appear at the end of a phrase). In fact, all 6 birds sang phrases under DAF containing a zero motif bout followed by a bout that included motifs, suggesting that the initiation difficulty indicated by a zero motif bout is not irrecoverable.

Zero motif bouts were more likely to occur earlier in the morning of the DAF exposure day than the average bout. The mean occurrence time for zero motif bouts was 85 minutes post lights-on, while mean occurrence time for all bouts was 103 minutes post lights-on. The earlier likelihood is more apparent when all bout occurrences are plotted (Figure 2.21). From this figure it is apparent that zero motif bouts comprised some of the first bouts sung by all 6 birds. Some birds (yellow9, black390, black205, black20) sang many fewer zero motif bouts later during the morning, suggesting that they may have been able to modify their singing behavior to escape whatever deficit in initiation ability had led them to produce zero motif bouts earlier in the morning. Black396, on the other hand, seems to have been unable to do the same, singing a large number of zero motif bouts with no successful motif initiations before ceasing singing entirely for the morning.

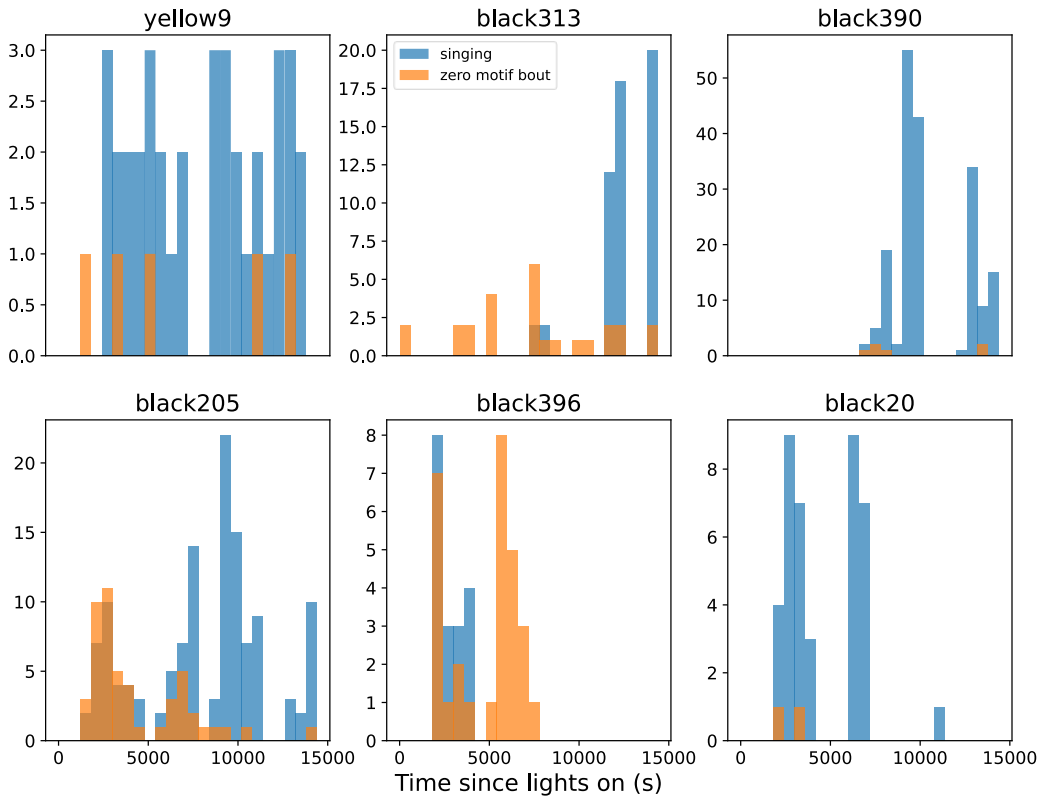


Figure 2.21. Latencies of bout occurrence relative to lights-on. Zero motif bouts are shown in orange, bouts containing motifs in blue.

Unchanged features of song

As has been previously reported (Fukushima and Margoliash, 2015), motifs sung under DAF remained highly stereotyped. The acoustic structure of syllables was not altered, as measured by a number of commonly-used metrics including spectral entropy variance, mean pitch, and pitch goodness. Syllables within a motif were neither omitted nor rearranged, and the occurrence of truncated motifs (common in normal song, but variable from bird to bird) did not change significantly. Neither overall motif duration, nor inter-syllable gap duration, nor syllable duration, changed significantly. While prior work has reported stuttering-like behavior appearing after

several days of continuous DAF (Fukushima and Margoliash, 2015; Manderscheid, 2014), I found no indication of this on the first morning, consistent with those reports.

Discussion

These results provide evidence for rapid changes in singing behavior in zebra finches under a continuous DAF paradigm. These changes occur within the first four hours of DAF exposure – much earlier than changes reported under other feedback perturbation protocols, such as deafening or white noise playback (Leonardo and Konishi, 1999; Zevin et al, 2004). The behavioral changes I observed occurred above the motif level of song organization, in the structure of phrases and bouts – I did not observe changes in motif sequence or phonology. This is to be expected, given the characteristic regularity of motif structure in adult zebra finches (Price, 1979; Sossinka and Böhner, 1980). It also aligns with prior work in our lab that found no changes in the motif itself until much later during DAF exposure (Fukushima and Margoliash, 2015). A recent study informed of my results made preliminary observations on the changes in the total number of introductory notes following DAF exposure in four birds (Daou and Margoliash, 2020), finding a reduction in introductory notes overall after durations of DAF exposure from 4 hours to 2 days. Interestingly, the increase in introductory notes did not appear in that study to be more dramatic after longer DAF exposure, suggesting that there may be an upper bound on the capacity of the system to compensate for aberrant feedback.

I interpret my results to indicate a difficulty in song initiation under DAF. Introductory notes preceding the beginning of a motif are believed to bring the song-control regions in the forebrain and the peripheral muscles controlling vocalization into register, allowing precise and accurate performance of the subsequent motif(s) (Rajan and Doupe, 2013; Rajan, 2018; Rao et al, 2019;

Daliparthi et al, 2019). I observed changes in introductory note production itself – longer inter-introductory-note intervals and more introductory notes produced on later bouts in a phrase – implying that DAF interferes with the generation of this preparatory behavior. Also telling was the dramatic increase in the number of introductory note sequences that failed to lead to song, or "zero-motif" bouts, which suggests that disruption of the preparatory activity in the song system is so severe as to prevent the intended behavior from being carried out at all.

There is one prior study which reports rapid changes in singing behavior under altered auditory feedback. Its authors report variability in syllable order, more introductory notes, premature song termination, skipped early syllables, and phonological changes immediately upon application of their DAF protocol (Cynx and von Rad, 2001). Prior work in our lab, however, has shown no immediate effects of DAF on syllable order, song termination, or skipped syllables (Fukushima and Margoliash, 2015; Fukushima, 2008; Manderscheid, 2014). I did not observe them in this study, either. A detailed comparison of our results with those reported by Cynx and von Rad is unfortunately difficult, as, with the exception of the increase in introductory notes, they aggregate the sequence changes they observe into a "% errors" metric. Thus, I cannot definitively account for these differences, though methodological variation seems the most likely culprit.

To the best of my understanding, Cynx and von Rad applied their DAF protocol only to individual singing events (bouts, in their terminology, but likely corresponding to phrases in mine, given the interior introductory notes present in their example spectrograms), and began DAF semi-randomly during a bout/phrase, occasionally allowing unperturbed singing. I applied DAF continuously, including outside of song, which may allow the birds to verify its persistence via calls. If this reduces the "surprise" of the effect beginning mid-song, it could account for Cynx and von Rad's reported sequencing abnormalities, particularly truncation of motifs, in light of evidence that brief but intense visual stimuli can startle songbirds into early song termination (Cynx, 1990; Franz

and Goller, 2002; Seki et al, 2008). The uncontrolled nature of DAF onset in Cynx and von Rad (2001) report suggests that trying to replicate the result may not be fruitful.

Regardless of the methodological differences, and the divergence in results, Cynx and von Rad do not focus on song initiation in their study. The behavioral changes that I observed argue strongly for interference with introductory notes and song initiation, and therefore that neuronal activity that prepares the song system and periphery to produce precise and accurate song is abnormal or impaired under DAF.

An obvious and unmistakable difference between our work and the bulk of the motor preparatory activity literature is the presence of an overt behavior during the motor preparatory period: introductory notes. Nonhuman primate studies focus nearly exclusively on pre-movement trial epochs. However, there is evidence that preparatory and movement-related motor activity can co-occur in a task requiring rapid movement correction, and that even in less complicated tasks movement-related activity begins to manifest before preparatory activity fully subsides (Ames et al, 2019; Lara et al, 2018). In addition, reliable preparatory muscle activation prior to an arm movement while seated or standing has been reported in humans, likely for postural maintenance (Bouisset and Zattara, 1981; Diener et al, 1989). It thus seems possible that preparatory muscle activation or movement, in addition to neuronal activity, is a feature common to many behaviors.

I believe that there is real value in a behavioral paradigm that includes preparatory activity. Given the challenges inherent in teaching an animal to carry out an artificial behavior in an uncomfortable setting, it is understandable that motor preparation studies focus on simple movements that are most amenable to operant training, often variants on reaches. These movements are additionally easier to track and quantify. As a result, however, the animals in these studies perform the movements with no difficulty – the major obstacle to training usually seems to lie in communicating what movement is appropriate in what situation. But if the movement is easy,

if it does not tax the animal's physical dexterity, coordination, or strength, it should hardly surprise us that, once the animal understands *when* to perform the motion, *doing so* is a matter of course, with little fanfare. These movements simply may not engage the full panoply of motor cortical activity, and thus may not require complex preparation. Birdsong enjoys several advantages as a model behavior to probe motor control: it can be extremely reliable trial-to-trial (depending on species), it is easy to record and quantify, and, crucially, it is a behavior performed at the very limit of the animal's physical capabilities, making it likely to engage motor control circuitry to its fullest extent (Podos, 1996; Gil and Gahr, 2002).

Beyond its experimental value as a motor behavior, birdsong is a rare, and certainly the most extensively studied, example of vocal learning in the animal kingdom. It may thus help shed light on the neural mechanisms of human speech production (Doupe and Kuhl, 1999).

Finally, while it must be acknowledged that the usefulness of DAF-induced stuttering in humans as a model for endogenous stuttering is a matter of debate in the field (Stuart et al, 2002), I find the correspondence between the heavy concentration of stuttering events at speech initiation and the song initiation difficulties I found to be induced by the DAF paradigm too suggestive to ignore. Recent work in our lab has produced evidence for changes in intrinsic properties in song premotor neurons under DAF *in vitro* (Daou and Margoliash, 2020). Further work elucidating the neuronal correlates of these DAF effects in behaving songbirds will, I hope, provide insight into both motor preparation as a whole and human stuttering in particular.

These results can also be interpreted in the context of the song control circuitry. Later chapters will provide a better perspective to develop such an argument, but briefly: DAF is known to induce changes in HVC_X intrinsic properties within 4 hours (Daou and Margoliash, 2020), and so the aberrant feedback may disrupt the expected input to these neurons, driving plasticity of their intrinsic properties. While ablation of HVC_X neurons does not appear to markedly impair song

production (Scharff et al, 2000), ablation or lesion studies can only indicate how a network behaves without some members, not rule out a function for those members in the intact network. It is entirely possible that simply removing HVC_X neurons has little effect on ongoing song production precisely because it is their abnormal activation which is significant. One consequence of unusual activity in HVC_X neurons may be an alteration of the song generating network activity so as to attempt to escape the disruption, driving changes to preparatory behavior; alternatively, altered feedback may drive a variety of changes in the song system, of which those in HVC_X are only a part. Further, it may be that the disruption of the preparatory behavior itself causes later sequencing deficits.

CHAPTER 3

CO-OCCURRING SEQUENCES IN HVC_X NEURONS

Introduction

Primary motor cortex in mammals has two principal outputs. Intratelencephalic (IT) pyramidal neurons in layers 2/3 and in layer 5 include neurons projecting to the basal ganglia whereas pyramidal tract (PT) neurons in layer 5 project to the spinal cord, brainstem, midbrain, and thalamus. Different aspects of motor control may arise from recruitment of the different pathways. Studies in rodents and primates suggest that IT and PT neurons have overlapping but distinct patterns of activity in regulating skilled movement control. For lack of a good animal model, how these pathways contribute to skilled vocal behavior is not well established in mammals.

The songbird model provides an excellent opportunity to address this question albeit with differences in some organizational details as compared with mammals. In the forebrain nucleus HVC (used as a proper name) the two most numerous classes of projection neurons (PNs) are anatomically distinct (DeVoogd and Nottebohm, 1981; Fortune and Margoliash, 1995)): most or all HVC_X neurons project to the basal ganglia (Area X) and most or all HVC_{RA} neurons project to the primary motor cortex analog nucleus robustus arcopallium (RA) (Nottebohm et al, 1982; Nottebohm et al, 1976). (In zebra finches a recent study reports a small population (~5%) of neurons that project to both Area X and RA (Benezra et al, 2018).) RA neurons have major projections to brainstem nuclei regulating song output, as well as related nuclei in the midbrain and thalamus, thus expressing a similar pattern of projections as seen in mammalian PT neurons.

HVC_{RA} and HVC_X also have distinct functional roles, in analogy to IT/PT distinctions in mammals. In adults, HVC_{RA} are necessary for singing whereas lesions of HVC_X have minimal long-

term disruption of singing (Sanchez-Valpuesta et al, 2019; Scharff et al, 2000). This can be viewed as an extreme form of functional segregation as suggested for IT and PT neurons. Nevertheless, it remains to be resolved if the two classes of HVC neurons express similar or different patterns of activity during singing. Recent studies sampling zebra finch HVC populations during singing with large numbers of sequential single neuron recordings (Lynch et al, 2016) or using optogenetic methods to simultaneously sample the populations (Picardo et al, 2016) have been interpreted to indicate that both HVC_{RA} and HVC_X are organized so that the precisely timed spike bursts they emit during singing cover the duration of the motif (canonical sequence of syllables) in a continuous representation and are largely (Lynch et al, 2016) or entirely (Picardo et al, 2016) independent of the vocal movements a bird makes during singing. The PNs are envisioned as contributing to a “clock” governing song production with the time of each burst a "tick" in that clock. That is, in the clock model HVC PN encode time itself, not vocal movements.

The posited uniformity of functional properties comparing HVC_{RA} and HVC_X and the posited lack of information about vocal movements contrasts with what is described for mammalian IT and PT neurons driving limb movements. Perhaps this arises from some unspecified constraints of learned vocal behavior (not the focus of the mammalian studies), but a wealth of data in songbirds is also difficult to reconcile with the clock model. In many species examined to date (e.g., white-crowned sparrow (Margoliash and Konishi, 1985), canary (Del Negro et al, 2005), Bengalese finch (Nishikawa et al, 2008), and zebra finches (Margoliash and Fortune, 1992)) song playback experiments demonstrate that HVC neurons are selective for each individual bird's song compared to conspecific songs or other stimuli. These responses in HVC and throughout the song system reflect sensitivity to learned features of song (Coleman et al, 2021; Doupe and Konishi, 1991; Margoliash, 1986; Nick and Konishi, 2005; Shank and Margoliash, 2009). They are thought to arise from sensorimotor interactions, for example selectivity for BOS increases comparing neurons of

nucleus interfacialis (that are presynaptic to HVC) with HVC neurons (Janata and Margoliash, 1999), and HVC auditory selectivity changes continuously during song development to remain tightly matched to the developing song (Nick and Konishi, 2005; Volman, 1993). A subset of HVC neurons require the sequence of notes in a syllable or the sequence of syllables in a motif to express excitation (Lewicki, 1996; Margoliash, 1983; Margoliash and Fortune, 1992). These observations are inconsistent with HVC neurons being uninformed of song features.

In adult zebra finches a behavioral state change (sleep) is required to observe robust auditory responses in the song system (Dave et al, 1998; Schmidt and Konishi, 1998). Perhaps this yields completely different HVC properties comparing the sleeping and singing bird. Yet single zebra finch RA neurons show spontaneous activity and response to song playback at night in a similar fashion to their activity during daytime singing (Dave and Margoliash, 2000), the RA activity is driven by HVC activity (Elmaleh et al, 2021; Fee et al, 2004), and studies promoting the clock hypothesis have relied in part on analysis of song system activity in sleeping zebra finches (Elmaleh et al, 2021; Hahnloser et al, 2002). Furthermore, in contrast to zebra finches most songbird species examined exhibit HVC neurons with auditory responses in awake animals (e.g., white-crowned sparrows (Margoliash, 1986), Bengalese finch (Sakata and Brainard, 2008), swamp sparrows (Prather et al, 2008), canaries (Boari et al, 2022). Although feed-forward activity driving singing and feedback activity in response to song playback should in principle be separated by at least 35–50 ms, in the subset of species tested HVC PNs exhibit near-zero time lag comparing timing of bursts emitted by the same neurons during song playback and during singing (zebra finches: Dave and Margoliash, 2000; swamp sparrows and Bengalese finches: Prather et al, 2008). This implies complex dynamical interactions between auditory and motor modalities expressed in HVC activity. Near-zero time lag was observed examining auditory responses and also singing activity of HVC PNs (albeit not differentiating between HVC_{RA} and HVC_{X}) or interneurons and a “gesture” model

of syringeal movements during singing (Amador et al, 2013). While recent papers supporting the clock model concept have vigorously responded to the gesture model (but focusing only on the activity during singing while ignoring the equivalent auditory responses (Lynch et al, 2016; Picardo et al, 2016)) those papers and the clock model do not address any of these other observations. A simple model uniting the actual observed responses in a very broad range of song system studies with the hypothesized clock model of HVC activity remains to be described.

Towards resolution of this dilemma, I have observed unrecognized structure in the data, and attendant statistical errors in the analyses, of Lynch et al (2016) (and by extension, Picardo et al (2016)). Instead, in this work I have identified novel song-related structured activity of the patterns of populations of HVC_X neurons recorded during singing. The patterns change from bird to bird, as would be expected in representations sensitive to learned features of song. The results suggest a solution to the zero-lag issue and provide for a new model of sensorimotor integration in vocal production.

Methods

Data collection

These data were collected by the authors of Lynch et al (2016) and provided by them at D. Margoliash's request. I restate the salient details here; for more detail on collection, refer to their methods. All HVC projection neuron data were collected using microdrive-mounted high-impedance single electrodes surgically implanted into HVC of male zebra finches. Electrophysiological recordings were made while the birds sang in the presence of a female zebra finch (directed song) and neurons were identified as interneurons or PNs according to firing properties. PNs were classified as HVC_{RA} or HVC_X based on antidromic stimulation using electrodes

implanted into nucleus RA or Area X, or were classified as “putative PNs” if they demonstrated song-related firing patterns characteristic of PNs but could not be definitively identified by antidromic stimulation.

All RA projection neuron data were collected similarly, with projection neurons identified by their firing properties. These data were previously reported in Leonardo and Fee (2005) and were provided to D. Margoliash by Anthony Leonardo. For more detail, refer to that paper's methods.

Data received

I received both spike times and inferred burst times from the Lynch et al (2016) authors. In both cases, the values provided were those obtained after having been warped to a canonical motif for each bird (piecewise linearly by syllable start and stop times, following Glaze and Troyer, 2006).

Lynch et al (2016) determined burst times by producing 1 ms binned spike rates pooled over all motif renditions, smoothing by convolution with a 9 ms square window, and identifying candidate burst onset and offset times using a 10 Hz threshold. Only candidate windows which contained spikes on greater than 50% of all song renditions were considered. Burst times were calculated from these windows by taking the mean spike time within the window.

In all analyses involving burst timing, I used the burst times provided, rather than recalculating them myself (beyond confirming that I was able to), to maintain continuity with the original data and analyses.

Burst timing plots

For all burst time plots (such as Figure 3.6), neurons were arranged on the ordinate by the order within the canonical motif of their first burst, paralleling how Lynch et al (2016) arranged bursts by their order within the motif in their Figure 1.

Population IBI

The population IBI distribution was constructed by arranging all burst times in order, regardless of neuron identity, and taking the difference between them.

Surrogate uniform distributions

Surrogate uniform distributions for comparison with the real data were generated by parametric bootstrap as follows: to generate a single surrogate dataset for one bird, new burst times were generated for each neuron from that bird, matching the number of bursts that neuron produced. These burst times were drawn uniformly from an interval with length equal to that of the bird's canonical motif, with the requirement that for a given neuron, no two bursts occur within 9 ms of one another (as the burst identification analysis described above precludes this possibility in the real data). Burst times were generated independently across neurons. Population IBIs were generated as for the real data. (Lynch et al (2016); G. Lynch, personal communication).

For each bird, 10,000 of these surrogate datasets were generated.

Population IBI CDFs and Kolmogorov-Smirnov tests

The primary statistical test performed on the population IBIs was a two-sample Kolmogorov-Smirnov (K-S) test between the real data and the median of the 10,000 surrogate uniform distributions for each bird. The K-S statistic generated was compared with the distribution of K-S statistics calculated between the median and each of the 10,000 surrogate distributions. The p-values quoted in the results express the proportion of this bootstrapped distribution which is greater than or equal to the K-S statistic obtained between the real data and the median of the surrogates. The K-S test was applied using the `ks_2samp` method from the Python `scipy.stats`

library.

Application of population IBI analysis to RA data

The RA data provided by Anthony Leonardo included spike rates in 0.5 ms bins, averaged over all motif renditions (after piecewise linear warping to a canonical motif). As the firing properties of RA projection neurons differ markedly from those of HVC PNs, I could not apply the burst identification method used by Lynch et al (2016). Following Leonardo & Fee (2005) I identified burst onsets and offsets with a 125 Hz threshold on this spike rate.

Surrogate random data were generated largely following the procedure described above, except that the prohibition on bursts from a single neuron closer together than 9ms was removed.

Otherwise, all analyses performed on these data were identical to those described above for the HVC PN data.

Multiburst neuron IBI analysis

The multiburst neuron IBI distributions were generated by first calculating the IBIs between the bursts for each neuron, and then pooled into a single distribution per bird. That is, if a bird had two multiburst neurons, each of which fired two bursts, then each neuron contributed one IBI to that bird's multiburst neuron IBI distribution. For histograms displaying these distributions, a bin size of 10 ms was used for all birds.

Multiburst neuron autocorrelation analysis

Autocorrelograms were created from the 1 ms binned spike rates for each neuron. They included all spiking activity (averaged across trials and linearly warped to a canonical motif) for a neuron, not just that confined to bursts.

Pooled autocorrelograms were generated by rescaling each per-neuron autocorrelogram to the unit interval and summing across all neurons for a given bird.

Projection of subsequent burst times onto a single axis

For a given bird, the curves approximated by the first burst times from all multiburst neurons appear to be separated into pieces, each of which is roughly linear. I identified the knot points separating these pieces by eye prior to any further analysis; subsequent analyses are not particularly sensitive to the precise location of these knot points.

To allow simpler analysis, it is desirable to reduce the subsequent burst times to a single dimension. I did this by defining the axis normal to the piecewise first-burst fit line (determined by linear regression on the first bursts in question) and projecting the later burst times onto this axis.

Clustering along this normal axis was evaluated using Ripley's L metric (Ripley, 1976).

Periodic gratings

Periodic gratings along the normal axis described above were produced with spacings between one quarter of the minimum IBI for a bird and twenty times the minimum IBI for a bird, with a step size picked to be small in comparison. (The spacing values for Bird 1 are representative: from 0.005 to 0.4, in steps of 0.005. Note that these values are reported in arbitrary units along this normal axis, but they are close to the values that would obtain if the axis were parallel to the abscissa.)

A burst is "captured" by a grating if its location on the normal axis falls within a fixed window of 0.01 of one tine of the grating (clipped appropriately if the grating is spaced more densely than 0.02; i.e., a grating spaced more densely than 0.02 will capture all later bursts).

Set cover analysis

The set cover solutions described above are optimal with respect to the total number of offsets required to capture all of a bird's later bursts, but from a biological perspective this seems too stringent a requirement – projection neurons in HVC may be doing several things during song, and so there is no reason to expect that every single later burst will be involved in a sequence mirroring that of the first bursts. Some consideration will also suggest, as I later observed, that the solutions generated by the greedy set cover algorithm do not maximize the area under the capture curves discussed above – the greedy solution will ignore larger potential capture sets if they leave more bursts to be picked up singly and thus require more offsets to capture all later bursts.

Results

Initially, D. Margoliash identified apparent inconsistencies in the data for Birds 1-5 presented in Lynch et al (2016) (described below). This motivated preliminary reanalysis of the veridical burst time data from Bird 1 ("Bird 2" of Kozhevnikov and Fee (2007), available from data previously provided to D. Margoliash by A. Kozhevnikov) and approximations of the data from Birds 2-5 I extracted from the Lynch et al (2016) pdf, that tended to confirm the apparent inconsistencies. I then re-analyzed the data as originally reported in Lynch et al (2016), provided by the authors; what follows are new findings from these data. For a brief summary of the relevant data collection procedures, see Methods; for more detail refer to Lynch et al (2016).

Reevaluating analyses for uniform distributions

Reproduction of HVC population IBI results

Following the procedure of Lynch et al (2016), bursts from all projection neurons (PNs)

recorded from a bird during singing were pooled and reduced to a burst time within a canonical song motif. The pooling procedure aggregates, for each bird, the single bursts from all identified HVC_{RA} with all of the one or more bursts from each of all identified HVC_X, and all of the one or more bursts from each of all putative PNs, either HVC_{RA} or HVC_X (see Lynch et al, 2016). The resulting vector of burst times is processed irrespective of the PN type, the ordinal burst position (for "multiburst" neurons – those neurons emitting two or more bursts per motif), the penetration (hence location in HVC), or any other constraint – issues I will return to later. Lynch et al (2016) graphically represented such data with the time of each burst represented on the abscissa and the ordinal position of each burst within the canonical motif represented on the ordinate (Figures 3.1A–3.1E, insets). This "burst cascade" representation defines a space that is necessarily not truly two-dimensional (i.e., each row contains a single burst, and a column contains multiple data points only if there are two or more bursts with identical times). This depiction yields a strong tendency for typical experimental data to flow from top left to bottom right, giving the visual impression of an approximately straight line. The graphical decision to represent the data as spike rates tends to smear the time base of each burst causing overlap between bursts, thus reinforcing visual continuity. The graphic also excludes spikes that did not qualify as bursts, which eliminates any background noise. The degree to which burst times represented as a point process in this space are described by a continuous and uniform distribution, the so-called "clock" model, is a central question explored in Lynch et al (2016).

Following Lynch et al (2016), for each bird I calculated inter-burst intervals (IBIs) from consecutive pairs of bursts to produce what Lynch et al (2016) termed a "population IBI" and compared this to 10,000 population IBIs produced from a synthetic uniform distribution over the song duration via parametric bootstrap (see Methods). These results are visualized as cumulative density functions (CDFs) of the real data, one per bird (black lines, Figures 3.2A–3.2E), the mean of

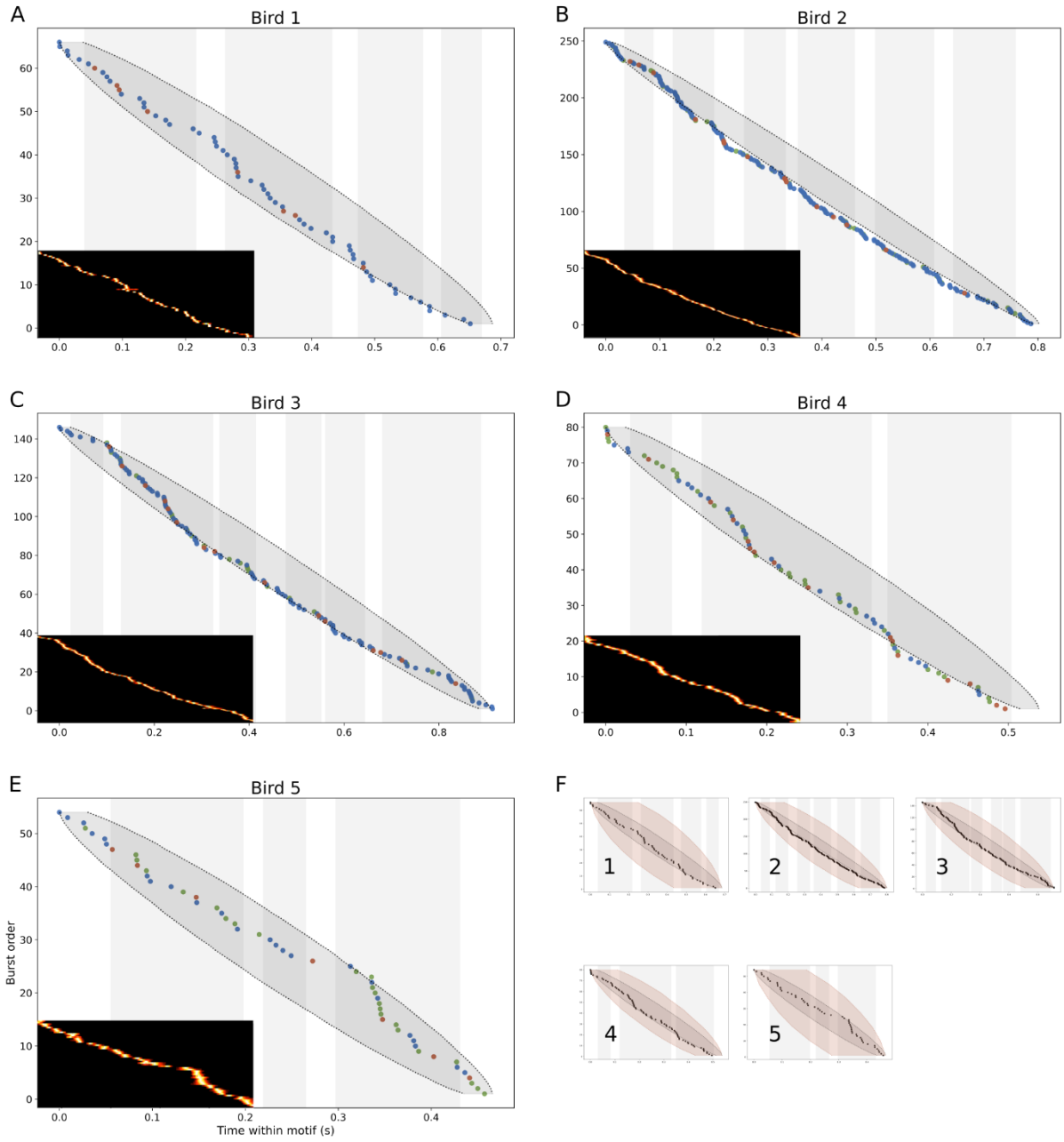


Figure 3.1. Burst cascades for HVC projection neurons. A-E: Burst times for each bird, with bursts pooled across neurons and ordered vertically by time within the song. Burst times are denoted by circles, with HVC_{RA} in red, HVC_X in blue, and putative PNs in green. Syllables are marked by shaded gray bars. The dark gray shaded region denotes where 95% of random uniformly distributed surrogate datasets fall. F: The same burst times for each bird as in panels A-E, with all bursts in black. The red shaded region denotes the 95% confidence band for random uniformly distributed surrogate datasets of the same size as the positively-identified HVC_{RA} dataset for each bird.

the bootstrapped data (dark blue lines) and the 95% confidence intervals of the bootstrapped data (blue shading delimited by light blue lines). my plots reasonably accurately reproduce the equivalent plots reported in Lynch et al (2016) (see insets, Figures 3.2A–3.2E) while showing additional detail since I did not apply 1 ms binning to the data as did Lynch et al (2016). Note that the real data stay confined within the confidence intervals at virtually all points for all birds. For the IBIs I calculated I then compared the real data CDF distribution with the surrogate population CDF distributions using a Kolmogorov-Smirnov test, obtaining Bird 1 p=0.640, Bird 2 p=0.015, Bird 3 p=0.421, Bird 4 p=0.714, Bird 5 p=0.811. Lynch et al (2016) reported p=0.014 for Bird 2 and p>0.28 for the other four birds. These values closely align with those reported by Lynch et al (2016). Moreover, the remaining variation (for Bird 2) lies within the distribution of p-values obtained by running the overall parametric bootstrap procedure many times. Lynch et al (2016) observed that after Bonferroni correction for five birds the K-S test is not significant for all birds, as did I. The similarity in the two sets of results gives confidence that I have reproduced the Lynch et al (2016) method accurately.

I also plotted these results for all five birds similarly to the organization of Figures 3.1A–3.1E insets but with each burst time represented by a point (the weighted average of all spike times within the burst), with HVC_{RA}, HVC_{X}, and putative PNs distinguished by color (red, blue, and green, respectively, and with 95% confidence intervals calculated for the population of PNs irrespective of identity marked by shading around the path from first to last burst (Figures 3.1A-3.1E). Considering all bursts irrespective of their PN identity, note that there are regions where the path of the sequential bursts exceeds the limits of the confidence intervals for each bird, in some cases substantially in magnitude (Bird 2; Figure 3.1B) or over an extended interval of the path (Birds 2, 4; Figures 3.1B, 3.1D). This contrasts with the behavior of the same data when viewed as a CDF (Figure 3.2), where the real data rarely or never violates the confidence intervals. This calls into question

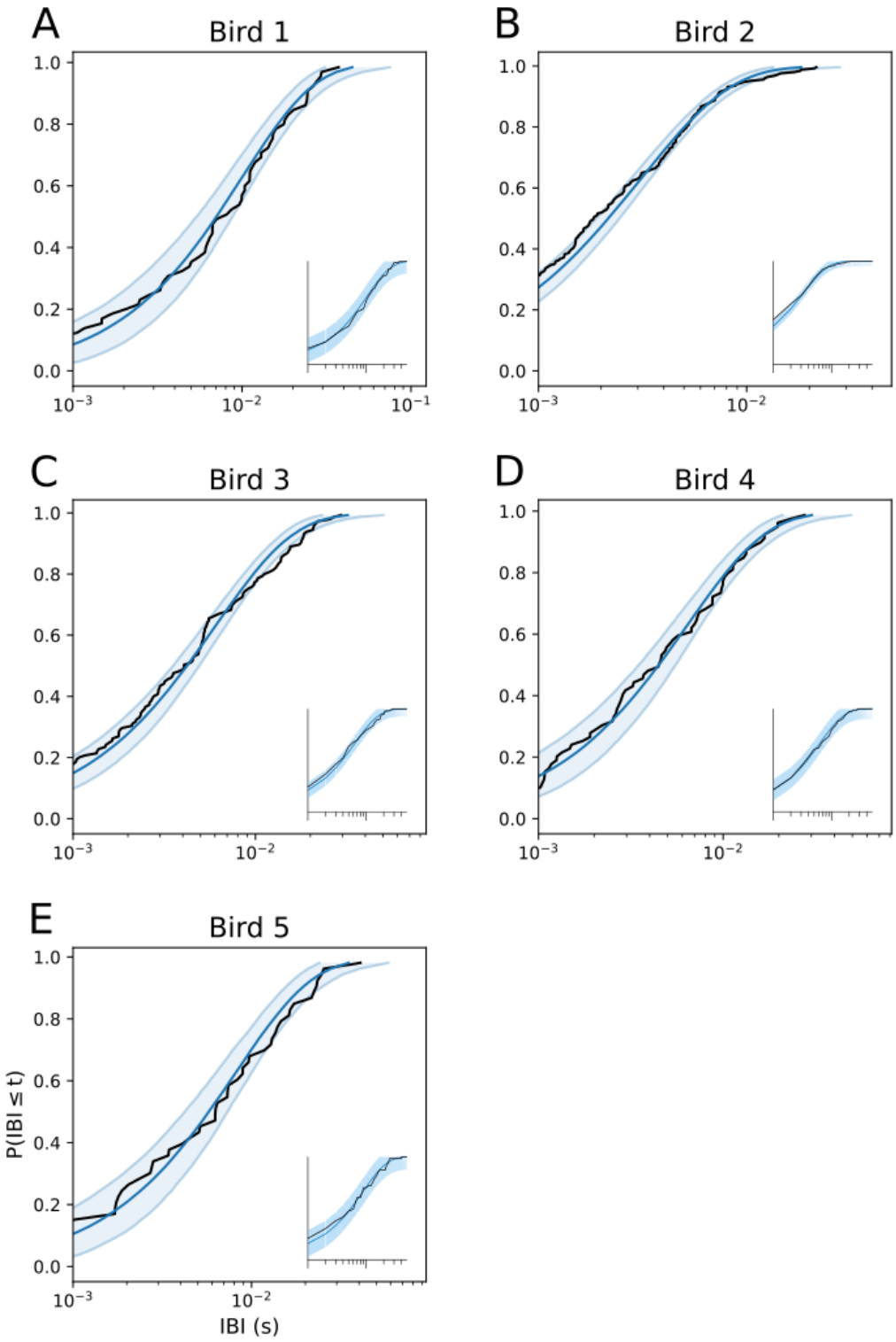

A
Bird 1
B
Bird 2
C
Bird 3
D
Bird 4
E
Bird 5
P(IBI ≤ t)
IBI (s)

Figure 3.2. Population IBI distributions for HVC projection neurons. (previous page) Cumulative density functions for the HVC projection neuron IBI distribution for each bird. Real data denoted by black lines, surrogate data by dark blue line (median for surrogates) and light blue shaded region (95% confidence interval). The original CDFs from Lynch et al (2016) for each bird are displayed as insets.

the sensitivity of the evaluation of the population IBI distribution to violations of uniformity as assessed by the K-S statistic, which relies on the CDFs.

Population IBI analysis yields unsound conclusions

To address this question, I devised a biologically meaningful test to evaluate the utility of the population IBI approach, asking what the result of such analysis would be when applied to other neurons. To this end, I examined the activity of RA projection neurons, which like HVC PN also emit very high frequency bursts that are precisely aligned to specific times in the motif during singing (albeit with many more bursts per motif per neuron), using data from Leonardo and Fee (2005) (kindly provided by A. Leonardo). Restricting myself to the three Leonardo birds with the largest data sets, I conducted population IBI analysis on Leonardo Bird i9 (34 neurons, 267 bursts), Bird i10 (10 neurons, 134 bursts), and Bird i12 (26 neurons, 329 bursts) – a data set with larger numbers of bursts than those for Birds 1, 4, and 5 and similar numbers of bursts as for Birds 2 and 3 of Lynch et al (2016). Plotting these data in the burst cascade approach (Figures 3.3A, 3.3C, 3.3E) evinced clear deviations of the burst time paths from the confidence intervals, with prominent deviations for Birds i10 and i12 (Figures 3.3C, 3.3E). Yet, as I observed for the HVC data, even with such violations of the confidence intervals the CDFs constructed from the population IBIs for all three birds hewed closely to the CDF of the uniform distribution for each of the birds (Figures 3.3B, 3.3D, 3.3F). The K-S test verified this (Bird i9, p=0.129, Bird i10, p= 0.271, Bird i12, p=0.038); with Bonferroni correction for multiple comparisons all p-values are above 0.05. By the logic of Lynch et al (2016) I have demonstrated that RA PNs are also organized in a continuous uniform

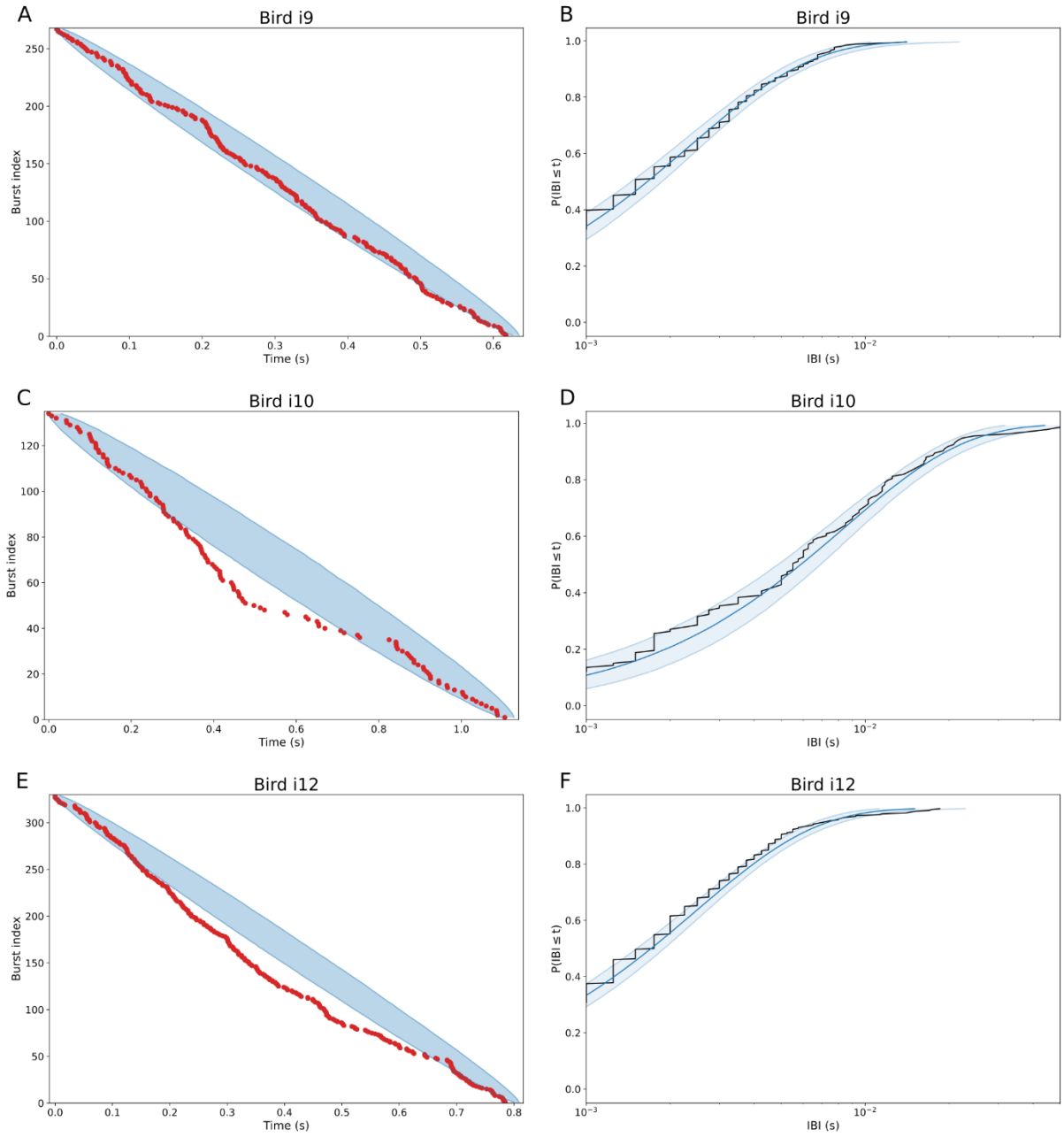


Figure 3.3. Population IBI analysis for RA data. A, C, E: Burst times for each bird in the RA dataset with large numbers of neurons, with bursts pooled across neurons and ordered vertically by time within the song. Burst times are denote by red circles. Black dotted line runs from the first neuron and beginning of song to the last neuron and the end of song. Blue shaded region denotes where 95% of random uniformly distributed surrogate datasets fall. B, D, F: Cumulative density functions for the RA neuron IBI distribution for each bird in the RA dataset. Real data denoted by black lines, surrogate data by dark blue line (median for surrogates) and light blue shaded region (95% confidence interval).

distribution as in a "clock", a conclusion at odds with extensive literature including discussions of the clock model that envision RA as encoding movements in muscle-related coordinates. One alternative explanation for the RA results is that the statistical approach chosen by Lynch et al (2016) lacks sufficient power to address the question of uniformity given the sample sizes of bursts available for the RA recordings, and hence for the sample sizes of bursts achieved from the heroic HVC recordings of Lynch et al (2016). Another non-mutually exclusive alternative explanation is that for purposes of statistical analysis each burst of a multiburst neuron cannot be treated as an independent event.

Fundamentally important for evaluating the conclusions of Lynch et al (2016), the failure to distinguish the real HVC data from the null hypothesis (uniform distribution) gives no insight into the validity of the null. Reliance on failure to reject a null hypothesis does not preclude whatsoever the existence of overlooked non-uniform structure in the data. Such structure may motivate different statistical assumptions and permit the normal, more powerful form of statistical argumentation which relies on gaining insight from falsifying, not failing to falsify, the null. For example, in a population of N rhythmic oscillators that all share the identical inter-burst interval the only additional information in the population is the phase of each oscillator (i.e. the relative timings of the initial spike bursts). The timing of the other bursts carries no additional information. Applying the Lynch et al (2016) statistical approach to such a population of bursts would conflate the N-1 oscillator phase relationships with the much larger sample of bursts. In the next section, I explore such issues, focusing on the HVC_X data.

Structured timing of HVC_X bursting

While the RA data differed from the HVC data in that on average there were fewer RA neurons each with more bursts, the HVC data sets also included many multiburst neurons. Analyzing the

timing of each HVC PN (or RA PN) as independent of the others represents a statistical assumption but also precludes searching for structure in the burst timing of multiburst neurons. I investigated this by examining the timing of bursts emitted by multiburst neurons. In HVC the vast majority of these neurons are thought to be HVC_X as identified using antidromic stimulation techniques (Kozhevnikov and Fee, 2007; Lynch et al, 2016). In a population of 103 identified HVC_X recorded from 7 adult birds, 18 emitted zero bursts per motif, 45 emitted one burst per motif, and 42 emitted between 2–4 bursts per motif (Kozhevnikov and Fee, 2007). In contrast, to date no multiburst HVC_{RA} have been described.

The Lynch data set was strongly biased towards HVC_X (positively identified HVC_X / total positively identified PNs (i.e., not including putative PNs) = 80%, 87%, 82%, 62%, 80%, Birds 1-5, respectively; see Table 3.1). (Any bias towards one or the other class of identified PN in the putative PNs is not known.) Furthermore, the later (i.e., non-initial) bursts from multiburst neurons contribute substantially to the Lynch et al (2016) data (later bursts/total bursts = 39%, 45%, 33%, 20%, 20%, Birds 1-5, respectively; see Table 3.1), and considering just HVC_X (i.e. eliminating identified HVC_{RA} from the analysis) increases these percentages (later bursts/total bursts = 45%, 48%, 37%, 24%, 22%, Birds 1-5, respectively; see Table 3.1). In what follows I initially focus solely on multiburst neurons and investigate the structure in the timing of the multiple bursts. I include the three multiburst HVC_{RA} from the Lynch et al (2016) data (one neuron in each of Birds 2, 4, and 5) given that such HVC PNs are not described elsewhere and that I cannot exclude the possibility that like HVC_X these neurons also project to Area X (the basal ganglia) (see Benezra et al, 2018). This sample of the data also includes multiburst putative PNs, whose projection target was not confirmed. Thus, the sample is highly enriched in HVC_X but may include very small numbers of other as yet poorly defined cell classes. In the rest of the Results for ease of exposition I refer to these data as HVC_X.

	Bird 1	Bird 2	Bird 3	Bird 4	Bird 5
HVC_X	32	93	69	25	20
HVC_{RA}	8	14	15	15	5
putative PN	0	29	16	24	19
Total PN	40	136	100	64	44
Later bursts	26	113	48	17	11
Total bursts	66	249	146	85	54

Table 3.1. Projection neuron types in the Lynch dataset. HVC_X and HVC_{RA} were identified by antidromic stimulation. Putative PNs had characteristic features of HVC PNs but were not identified by antidromic stimulation (see Lynch et al, 2016).

The timing of HVC_X bursts shows bird-specific structure absent from the population IBIs

IBI histograms. HVC_X with multiple bursts evinced simple low-order structure in the timing of the bursts. I first constructed histograms from the pooled inter-burst intervals from each multiburst neuron within a bird. These histograms were not appreciably affected by bin size or boundary location choice (see Methods). Bird 1 (26 IBIs from 18 neurons) had an almost uniform distribution of IBIs (Figure 3.4A), as did Bird 5 (11 IBIs from 10 multiburst neurons) (Figure 3.4I). In contrast Bird 2 had a massive concentration of IBIs between 80 – 170 ms rising asymmetrically to a peak at 150 ms that accounted for 62% (70/113) of the IBIs (from 68 multiburst neurons) (Figure 3.4C). Bird 3 evinced perhaps three separate concentrations of IBIs, with peaks at 110 and 220 ms, and a smaller peak at 470 ms, and scattered IBIs throughout (48 IBIs, 44 multiburst neurons) (Figure 3.4E). Finally, 71% (12/17) of Bird 4's IBIs were concentrated in a narrow interval (40 ms) centered around a single tall peak at 170 ms (14 multiburst neurons) (Figure 3.4G). Thus, there was evident first-order structure in the timing of HVC_X multiple bursts and variation from bird to bird. This was a strong effect and was apparent even in the birds with the smaller sample sizes. Simple scaling or truncation of the time axes (to account for different motif durations) could not account for these differences.

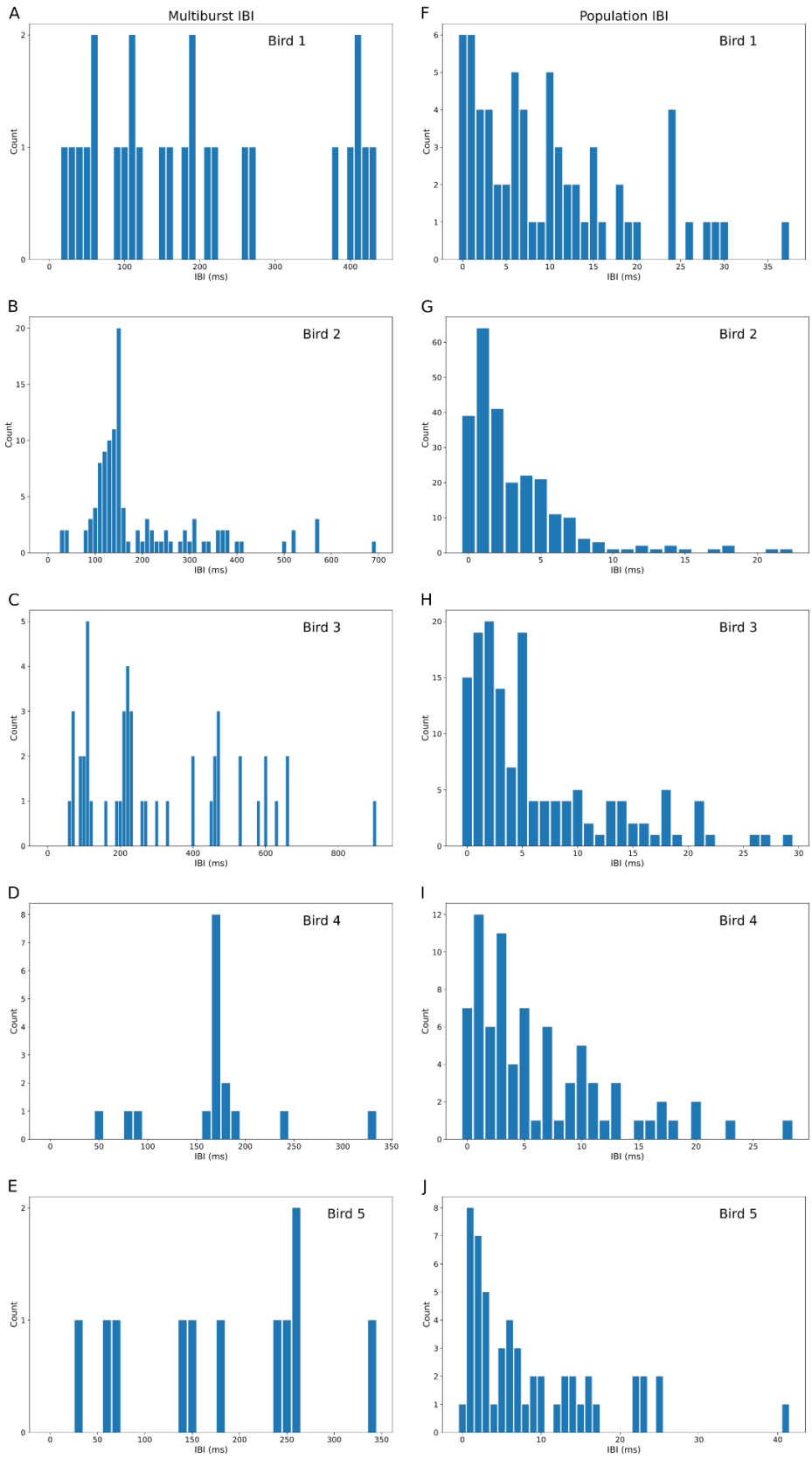

A
Multiburst IBI
Bird 1
Count
IBI (ms)
F
Population IBI
Bird 1
Count
IBI (ms)
B
Bird 2
Count
IBI (ms)
G
Bird 2
Count
IBI (ms)
C
Bird 3
Count
IBI (ms)
H
Bird 3
Count
IBI (ms)
D
Bird 4
Count
IBI (ms)
I
Bird 4
Count
IBI (ms)
E
Bird 5
Count
IBI (ms)
J
Bird 5
Count
IBI (ms)

Figure 3.4. Multiburst neuron IBIs compared with population IBIs. (previous page) A, C, E, G, I: Histograms of the HVC multiburst projection neuron IBIs for each bird, with a bin size of 10ms. IBIs for all bursts produced by a single neuron were calculated, and these numbers pooled to produce the distribution for a given bird. These plots only consider neurons which produced more than one burst, and include HVC_X neurons and putative PNs, as well as very small numbers of HVC_{RA} neurons (1 each for Birds 2, 4, and 5). B, D, F, H, J: Histograms of the HVC projection neuron population/global IBI for each bird. All bursts from all projection neurons (single-burst and multiburst, and regardless of subtype) were pooled, sorted by time within the song, and an IBI distribution produced from this sequence directly, following Lynch et al (2016). A bin size of 1ms was used, to better capture the predominance of IBIs under 10ms.

It is useful to compare the structure of these standard IBI histograms from multiburst HVC_X with histograms constructed from the population IBIs. The population IBI histograms are of course all concentrated at very short time intervals, with peaks near 0 ms, but also with similar exponential-like decrease within 10-20 ms (Figures 3.4B, 3.4D, 3.4F, 3.4H, 3.4J). To quantify this difference, for each bird I constructed 10-bin histograms with the ordinate normalized by the peak height of each histogram, both for standard IBI and population IBI representations. Comparing pairs of histograms by computing an RMS value summed over the difference between bin 1 pairs, bin 2 pairs, etc. yielded smaller differences (i.e. greater similarity) in the within-population IBI comparisons than the within-standard IBI comparisons, with the two distributions completely non-

Multi-burst neurons

Bird	1	2	3	4	5
1		0.11	0.1	0.16	0.12
2			0.1	0.12	0.13
3				0.12	0.1
4					0.14
5					

All neurons

Bird	1	2	3	4	5
1		0.089	0.049	0.067	0.076
2			0.048	0.046	0.047
3				0.04	0.046
4					0.032
5					

Table 3.2. Distances between multiburst neuron IBI distributions. Root mean square of bin differences from multiburst neuron IBI histograms (top) and population IBI histograms (bottom).

overlapping (Table 3.2). There was no statistically significant trend between the two distributions of values. This confirms that the distributions of population IBIs are more similar to each other than are distributions of the multiburst neuron IBIs, and that the population IBI concept suppresses bird-specific structure that exists in the actual recorded activity of the individual neurons.

The existence of preferred intervals in the IBIs in multiburst neurons of Birds 2, 3, and 4 indicates that for these birds the timing of bursts in one neuron is not independent of the timing of bursts in other neurons. Extensive violation of sample independence and its interaction with sample size renders the results of the (Lynch et al, 2016) statistical analysis potentially uninterpretable. The failure of the simple IBI analysis to show dependence for Birds 1 and 5 (the latter having the smallest sample of later bursts (11) of all the birds) provides no relief for the Lynch et al (2016) approach, as shown below.

Firing rate autocorrelograms. Peaks in the multiburst neuron IBIs result from common intervals across neurons and do not explore the structure in the timing of multiple bursts within individual HVC_X PNs. To explore this, I calculated the autocorrelograms of the firing rates of each multiburst PN. An example is shown in Figure 3.5A. The mean spike rate of a neuron (Figure 3.5A, top panel) is autocorrelated yielding a principal peak at zero lag and burst-interval-related peaks at positive and negative lags (Figure 3.5A, middle panel). Focusing just on the positive lags I defined the numerically largest values as the peaks (Figure 3.5A, bottom panel).

To search for common second order patterns in the neurons of each bird, I then scaled the autocorrelograms to have heights in the unit interval and summed them, resulting in a single plot per bird (Figures 3.5B–3.5F). These summed autocorrelograms show distinct peaks within a given bird, indicating that the intervals between periods of high firing rate were similar across neurons within a bird. For each autocorrelogram the two tallest peaks are labeled with the number of neurons whose spike rates contribute to the peaks. The summed autocorrelograms of Bird 2 and

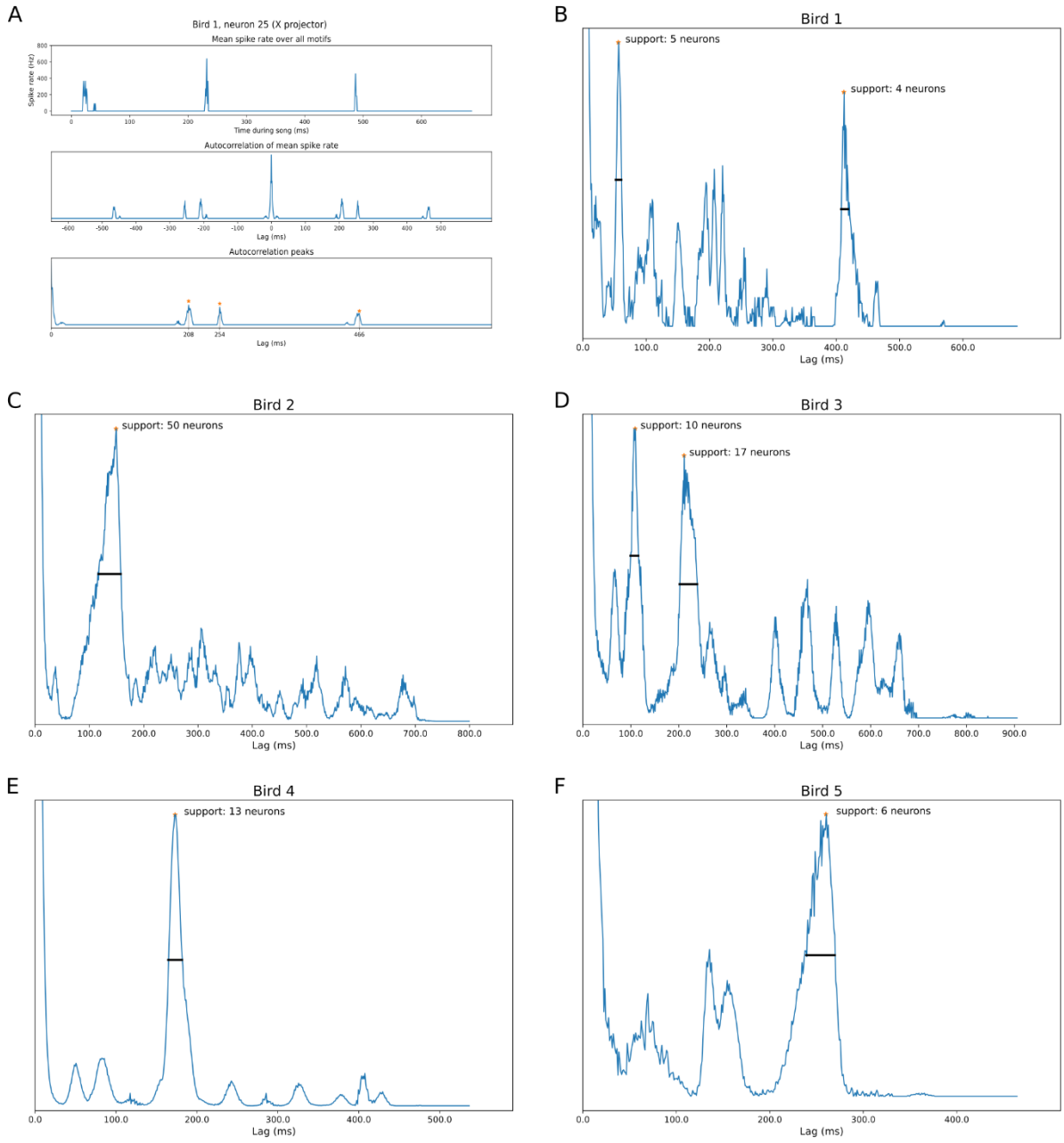


Figure 3.5. Firing rate autocorrelations for multiburst neurons. A. Example autocorrelation for one HVCX neuron from Bird 1. Top panel: 1ms-binned spike rate averaged over all motifs. Middle panel: autocorrelation of averaged spike rate for lags up to the motif duration. Bottom panel: positive domain of autocorrelation, with peaks labeled indicated by orange stars. B-F. Summed autocorrelograms for each bird. Per-neuron autocorrelograms were normalized to the unit interval and summed to produce these plots. In each plot, the largest peaks are annotated with the 50% height band and the number of neurons whose autocorrelograms contributed to the peak within this region.

Bird 4 are dominated by large single peaks, while those of Birds 1 and 5 are characterized by several clear peaks. Bird 3's summed autocorrelogram has a larger number of smaller peaks demonstrating many preferred intervals of burst timing as compared to the other birds. As would be expected, only a subset of these peaks (second order measure) appear in the multiburst ISI histograms (Figure 3.4) (first order measure). For example, both Birds 1 and 5 have relatively uniform multiburst ISI histograms but Bird 1's summed autocorrelograms show multiple distinct and narrow peaks whereas Bird 5 summed autocorrelograms show a few broad peaks.

The summed autocorrelograms also show clear differences between birds in the peak locations, the numbers of peaks, and the peak width and structure. Thus, both first and second order patterning of HVC_X bursts vary with individual birds. This motivated my exploration of the organization of the multiple bursts during singing.

Multiple sequences of HVC_X bursts during singing

Subsequent bursts define additional, bird-specific sequences parallel to the first burst sequence

Lynch et al (2016) (their Figure S1) provided a useful representation for examining the structure of the population of HVC activity in relation to the song motif. Indeed, it was observing structure in the patterns of bursts in that figure which first raised the possibility of an alternative explanation of their data that motivated the present analysis. In that "neuron cascade" representation (a modification of their Figure 1), bursts from the same neuron are plotted on the same row, and neurons are sorted similarly to their Figure 1 from top to bottom but here by the time of their first burst relative to the canonical song motif.

I calculated these representations for each of the five birds (Figures 3.6A–3.6E; first bursts in blue, later bursts in orange). Preserving the ordinal relationship of the timing of the first bursts while retaining the neuron identity of the subsequent bursts aligns all bursts in relation to when

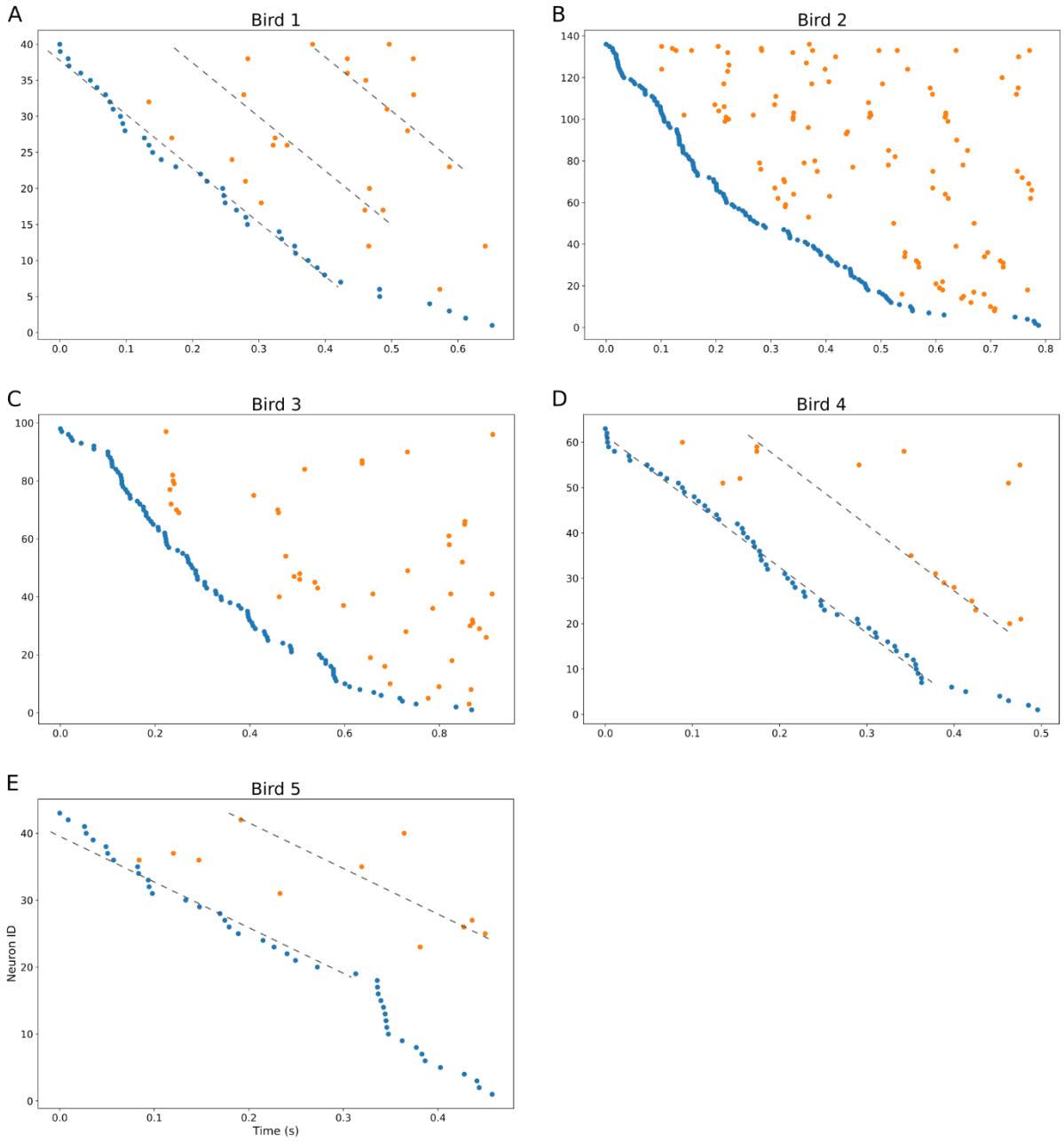


Figure 3.6. Neuron cascade plots for all HVC projection neurons. Burst times for each bird, with bursts for a single neuron kept on the same elevation, and neurons ordered vertically by the time of their first bursts within the song. First burst times are in blue, later burst times in orange. Dotted gray lines describe an approximate linear fit to the first bursts for Birds 1, 4, and 5 (leftmost lines), along with visually-apparent later burst sequences. These lines were placed manually. Within a single bird, all lines are parallel.

they occurred during the song motif while representing the flow of neurons as they were initially recruited during the motif. Inspecting these plots, one observes a clearly defined slope of initial bursts (or piecewise fits where appropriate; see below), and linear segments of later bursts parallel to the slope of the initial burst. A striking example is found in Bird 4, with over 50% (10/17) of later bursts arranged around a single line (Figure 3.6D, dotted line). Other clear example are the latter bursts of Birds 1 and 5 (Figures 3.6A, 3.6E, dotted lines).

To quantitatively analyze for significance of clustering of subsequent bursts along lines parallel to initial burst line segments, I first projected the later burst times onto an axis normal to the piecewise first burst fit line (Figure 3.7A). A single line was defined as the best fit to all the initial bursts, or piecewise fits of initial bursts were defined where there were places of clear slope discontinuity in sequences of initial bursts (see Methods). (The only piecewise fit associated with significant numbers of subsequent bursts was in Bird 2.) Parallel lines were then drawn through each of the subsequent bursts. An axis normal to the lines was then defined. (While the normal axis in Figure 3.7A appears graphically to be far from normal to the first-burst fit line, this is due to the different axis scales (roughly 100 to 1).) The intersection of later burst lines with the normal axis yields a one-dimensional representation of later bursts while doubling as a measure of how well the burst times line up together. Bursts that lie close to the same line (parallel to the line fitted to the first bursts of these neurons) will lie close together on this axis (Figure 3.7B).

I verified that the bursts do in fact cluster on this axis, using Ripley's L statistic (Figures 3.7C–3.7H; Bird 2 has two first-burst piecewise fits with later bursts, and so contributes two plots). I compared the clustering along this axis with clustering observed in synthetic data produced so as to maintain the locations of each neuron's first bursts, but to uniformly randomly place its subsequent bursts between its first burst and the end of the canonical song motif (maintaining a 9 ms interval between any two bursts, per Lynch et al (2016).

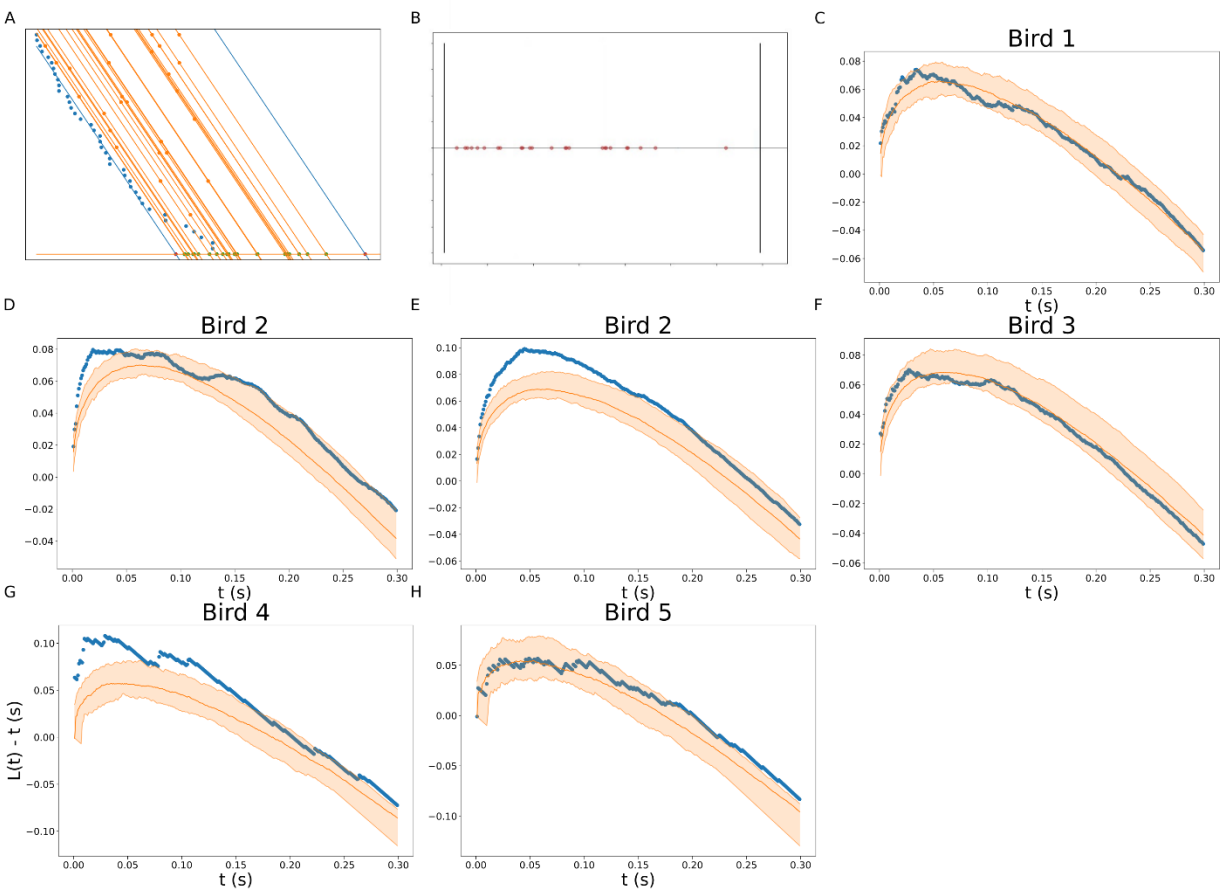


Figure 3.7. Normal axis intersections for later projection neuron bursts. A. Burst times for Bird 1, plotted as in Figure 3.6A, with first bursts in blue and later bursts in orange. The piecewise linear regression fit to the first bursts is shown as a blue line, and repeated to indicate the line of the same slope passing through the last moment of song (i.e., the top-right point of Figure 3.6A. The line normal to this linear fit is the nearly-horizontal orange line at bottom. (It appears to be anything but normal to the blue lines, but this is a visual artifact of the differing axis scales.) Lines projecting each later burst into the normal line are shown in orange, and their intersection points with the normal are shown as green circles. The intersection points of the blue lines with the normal line are depicted as red circles, and define the bounds of possible projections of later bursts onto the normal line. B. The normal line from panel A, with the lower and upper bounds marked by black vertical lines, and projected later burst times shown as red circles. C-H. Plots of Ripley's L spatial clustering metric for each bird. Bird 2 contributes 2 plots because there were two regions of its first burst line with different slopes and whose neurons produced large numbers of later bursts. Real data is in blue, with the median surrogate values denoted by the dark orange line and the 95% confidence interval for the surrogate values denoted by the light orange shaded region. Plots are restricted to timescales between 0 and 300ms as it is unclear how to interpret clustering above this length scale given the short durations of the birds' songs.

For Birds 1–4 the normal-axis intersection clustering exceeds the 95% confidence interval calculated via this bootstrap procedure at short cluster distances. This was limited to brief intervals for Birds 1 and 3 (Figures 3.7C, 3.7F, respectively), while there were dramatic departures from the 95% confidence intervals for Bird 2 (both fits) (Figures 3.7D, 3.7E) and Bird 4 (Figure 3.7G). Bursts from Bird 5 (Figure 3.7H) exceed this confidence interval only for clustering at long cluster distances; however, given that Bird 5 had very few later bursts (n = 11 later bursts; Table 3.1) it exhibited wide confidence intervals at short cluster distances. Bird 4, with the next-fewest later bursts (n = 17; Table 4.1), was the only other bird to share these features of the shape of the confidence intervals: wide intervals at short cluster distance and exceeding the confidence interval at long cluster distance.

Short-range clustering on this axis indicates that the later bursts of multiburst neurons tend to describe sequences during singing that proceed at the same rate as the sequence described by the first bursts of these neurons. For Bird 2 this clustering appears for each of two regions of first burst fits with different slopes. This could arise from more than one "rate" at which burst sequences evolve during singing throughout HVC or from sampling biases when recording neurons in different regions of HVC that are expressed at different rates throughout the motif, each suggesting a different mechanism for the result.

These results refute a central implication or tenet of the clock model, that there is a single sequence of bursts that defines how activity unfolds over the time course of a song motif. I now turn to better defining the structure and number of observed sequences.

Later sequences mirror the temporal evolution of the first burst sequence

To analyze sequences of later bursts, I first developed an approach using periodic grating of lines which demonstrated that at some grating intervals (that varied from bird to bird) the real data

lay outside of the 95% confidence interval for bootstrapped data. There is no *a priori* reason, however, to believe that a burst sequence should be well-characterized by a linear fit. Thus, I developed an analysis to examine the arrangement of subsequent bursts without the parameterization imposed by the linear regression and normal-axis projection method.

Starting with a neuron cascade representation for a given bird, e.g., Bird 1, (Figure 3.8A) I shifted each neuron's burst times so that its first burst occurred at zero, while preserving all within-neuron IBIs (Figure 3.8B, top panel). All later bursts could then be projected onto the time axis (Figure 3.8B, bottom panel), and subsequent analysis was restricted to this one-dimensional representation. Bursts close together on this axis will be arranged in the original burst time plots in such a way as to follow a similar progression in time as the first bursts.

Using Ripley's L metric, I probed the distribution of later bursts on this axis and found that, as for the normal-axis projection described above, Birds 1-4 displayed clustering at short timescales that lay outside the 95% confidence band for the surrogate data. Bird 5 did not exhibit clustering greater than that displayed by its surrogate data, but this is likely explained by its small set of later bursts (n = 11) (Figure 3.7).

Having demonstrated the existence of clustering of later bursts I then determined how many clusters existed and the timing of those clusters for each bird. The question can be seen to be equivalent to the set cover problem in combinatorics, with a set defined by an interval on the time axis containing one or more burst times. While this problem is NP-hard in the general case, the present one-dimensional geometric version admits of a simple greedy polynomial-time solution (see Methods). I applied this algorithm to the burst locations on the time axis to generate a collection of offsets, together with the later bursts that they capture. A given offset captured all later bursts that fell within a fixed window of the offset. When mapped back onto the original burst time plots, this set of bursts can be seen to fall along a portion of an offset first burst line (Figure

3.8A). The examples shown in Figure 3.8 all used a half-window of 15 ms around each offset. I confirmed that they are representative of the effect using a range of half-window sizes from 10 ms to 30 ms. Each bird's later bursts were covered by a collection of offsets smaller than the number of bursts, often dramatically so (Bird 1: 9 offsets captured all 26 later bursts; Bird 2: 20 offsets captured all 113 later bursts; Bird 3: 16 offsets captured all 48 later bursts; Bird 4: 7 offsets captured all 17 later bursts; Bird 5: 6 offsets captured all 11 bursts).

To evaluate the statistical significance of these results I compared the performance of the set cover algorithm on real data against its performance on the surrogate uniformly-placed data. I constructed a plot for a set cover solution, similar to a scree plot (but inverted around the unit slope line), of the proportion of the later bursts captured versus the number of offsets, once they had been sorted in descending order by proportion of later bursts captured. A given solution will describe a monotonically increasing curve, whose slope is everywhere nonincreasing (Figure 3.8C). To produce a single value for comparison between the real and surrogate data, I took the area under the curve (AUC). For Birds 1–4, the AUC for the real data was significantly greater than that for the surrogate data (Bird 1, $p=0.042$; Bird 2, $p=0.0002$; Bird 3, $p=0.0023$; Bird 4, $p=0.0001$; p-values derived via bootstrap). The set cover solution for Bird 5, which had the fewest later bursts, did not rise to significance ($p=0.11$). With Bonferroni correction for five birds, Bird 1 becomes non-significant. These values are largely insensitive to the choice of window size between 1 ms and 50 ms.

The set cover solutions described above capture all of a bird's later bursts using the fewest offsets but might overlook biologically significant solutions in which some offsets capture more bursts. I therefore extended the analysis above using a modified algorithm that preferentially selects large capture sets, rather than minimizing the total number of sets (see Methods). The same pattern of values obtained, with Birds 1-4 having significantly greater AUC than their surrogates, and

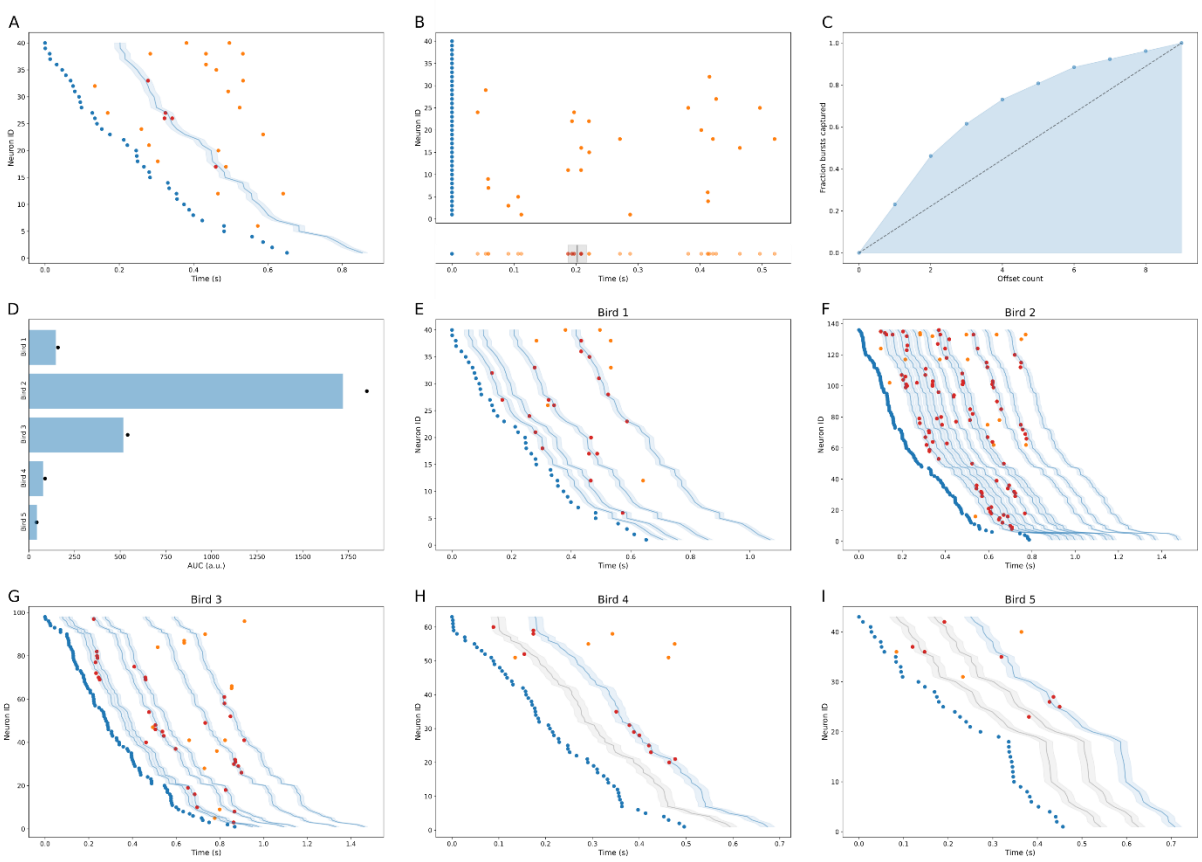


Figure 3.8. Set-cover clustering analysis for later projection neuron bursts. A. Burst times for Bird 1, plotted as in Figure 3.6A, with first bursts in blue and later bursts in orange. An example offset line at 202 ms is shown in darker blue, with a 15 ms capture window on either side of it shaded B. Top panel: burst times for Bird 1, adjusted per-neuron so that each neuron's first burst occurs at 0, with within-neuron IBIs preserved. First bursts are again in blue and later bursts in orange. Bottom panel: projection of all bursts from the top panel onto the time axis. An example offset line is placed at 202 ms (black vertical line), with its capture region (15 ms half-window) shaded in gray. Bursts captured by this offset are shown in red, with uncaptured later bursts still in orange and first burst in blue. C. The cumulative proportion of later bursts captured by increasing numbers of offset regions for Bird 1, using the algorithm prioritizing large capture sets (see Methods). Proportion of later bursts captured is depicted by blue points, with the shaded blue region the area under this curve (AUC). The diagonal black line shows the capture result if every offset had only captured a single burst. D. Summary of AUC resulting from the large-capture-set coverage algorithm for each bird. The blue bars denote the 95% confidence interval for the surrogate random uniformly-distributed data. Black points represent the AUC for the real data. E-I. As in panel A, the first (blue) and later (orange and red) bursts for each bird during song. The offset regions depicted (middle dark blue line, capture region light blue shading) are those produced by the large-capture-set coverage algorithm which capture more than two later bursts. Their captured later bursts are in red, and uncaptured later bursts are in orange. For Birds 4 and 5, the additional offsets which captured two bursts are plotted in gray, with their captured bursts also in red.

Bird 5 not being significantly different from its surrogates (Bird 1, p=0.0032; Bird 2, p=0.0001; Bird 3, p=0.0016; Bird 4, p=0.0001; Bird 5, p=0.14; p-values derived via bootstrap) (Figure 3.8D). Birds 1–4 retain significance following Bonferroni correction for five birds. These values, too, were insensitive to the choice of half-window size.

The offsets capturing three or more bursts for all birds are plotted in Figures 4.8E–4.8I, providing visual confirmation of the fits described above. It is immediately apparent that different offsets are emphasized in different birds. At any given moment in song there can be multiple sequences engaged simultaneously – bursts occurring later in one sequence occurring at the same time interval as bursts occurring earlier in another sequence. For example, for Bird 3 different bursts occurring at roughly 250 ms contribute to three distinct sequences. A similar effect for Bird 3 is also observed at 550 ms and 925 ms (Figure 3.8G). This pattern of activity is distinct from the single sequence reproduced in all birds as envisioned in the clock model. Across the birds the number of sequences scaled with the sample size (number of later bursts), thus from this data set there was no evident upper bound to the number of sequences. A remarkable degree of structure was observed for Bird 4, with a single 30 ms offset window accounting for 59% (10/17) of later bursts. Average numbers of bursts/offset were highest in Birds 2 and 4 (9.3 and 10 bursts/cluster, respectively), with intermediate values for Birds 1 (5.5) and 3 (5.7), and Bird 5 having the lowest numbers of bursts/offset (4). Thus, sample size was not directly related to the success of the clustering procedure. It was observed, however, that the two birds with small sample sizes had the greatest proportion of solitary bursts (on the transformed axis) (Bird 4: 0.41 (7/17), Bird 5 0.55 (7/11)). I conservatively defined sequences as having at least three bursts (see Methods), but this may overly penalize the birds with the smallest sample sizes. Lowering the criterion to two bursts per sequence decreases the proportion of solitary bursts (Bird 4: 5/17; Bird 5: 3/11), bringing them more in line with the other three birds.

Note that the timing and numbers of neurons per burst described here is related to the structure seen in the autocorrelogram peaks (Figure 3.5), but here it is conveyed as one or more correlated sequences per bird evolving over the duration of the motif. In each case, the later sequence structure is defined by the structure of the primary (first burst) sequence.

Higher-order structure of HVC_X sequences and song

Additional sequence offsets show no obvious relation to song intervals

The above results demonstrate a set of time intervals recovered from HVC_X later bursts. I wondered whether there existed a relationship between these neuronal sequence intervals and temporal features of song. Here I used the set cover intervals, one set per bird. I defined a second set of intervals from each bird's canonical song motif by calculating the durations between onset and offset for each syllable in the song, as well as for each pair of syllables. To measure the correspondence between these two sets of interval durations, I defined a distance between them by the mean of the minimum distances from each neuronally-derived interval duration to the set of song-derived interval durations.

A parametric bootstrap on a uniform distribution of neuronal burst times produces mean-minimum-distances not significantly different from those calculated from the real neuronal data (Bird 1, p=0.08; Bird 2, p=0.21; Bird 3, p=0.58; Bird 4, p=0.42; Bird 5, p=0.13). I next compared the mean-minimum distances between neuronal interval duration and song interval duration within-bird against those across-birds. Only for one bird (Bird 2) were the within-bird mean-minimum distances smaller than all of those produced by the cross-comparisons.

Thus, while neuronal intervals are consistent within a bird but different across birds, as is each bird's learned song, I find no clear relationship between neuronal intervals and a simple characteristic of song timing. This should not be taken as evidence of the absence of a

representation of song structure within HVC, but if such a representation exists it is more complicated than my simple characterization.

The timing of HVC bursts shows consistent structure across birds relative to song

While I were not able to relate burst sequence timing to timing of syllables in each bird's song, I did find relations of HVC PN to overall timing features of song. Examining all HVC PN I observed non-uniform distributions of first bursts over the duration of motifs, so that first bursts (blue dots in Figure 3.6) tended to be expressed at lower density towards the end of song. Splitting the motif durations in half showed that first bursts fell more heavily in the first half of song for four out of the five birds (70%, 75%, 76%, 67%, 49%, Birds 1–5, respectively), see Table 3.3 for details. To assess if this was due to chance alone, I used the surrogate random data described above. On average, the surrogate random data demonstrated similar proportions of first bursts occurring in the first half of song as the real data (median 65%, 66%, 62%, 57%, 58%, Birds 1–5, respectively), with 95% confidence intervals containing the values for the real data for Birds 1, 4, and 5. (Bird 2's real data has first bursts more concentrated in the first half of song than 98.55% of the surrogates, and Bird 3's real data has first bursts more concentrated in the first half of song than 99.86% of surrogates. If I take these values as pseudo-p-values, only Bird 3 retains significance above a Bonferroni-corrected alpha of 0.05.) The tendency in Birds 1–4 for the real data to have more first bursts in the first half of song than for the surrogate data median is probably related to additional constraints on the real data – the preferred long IBIs (Figure 3.4) further bias first bursts of multiburst neurons to occur earlier in song. Correspondingly, the later bursts (non-first bursts) of multiburst neurons tended to occur later within song (62%, 60%, 81%, 71%, 55% in the second half of song, Birds 1–5, respectively), with confidence intervals for the surrogate data containing the values obtained for the real data.

	Bird 1	Bird 2	Bird 3	Bird 4	Bird 5
First bursts in first half of song	28	102	74	42	21
Total first bursts	40	136	98	63	43
Later bursts in second half of song	16	68	39	12	6
Total later bursts	26	113	48	17	11
Single bursts in first half of song	12	50	38	30	11
Total single bursts	22	68	54	49	33

Table 3.3. Dispersal of multiburst neuron bursts across song. Distribution of first and latter bursts from multiburst neurons and bursts from single burst neurons.

The combined effects are quite noticeable. Note that first bursts of multiburst HVC_X are at reduced density and later bursts are at an increased density towards the end of motifs for Birds 1-4 (Figure 3.6). If different physiological mechanisms give rise to first and later bursts, then this too would be inconsistent with the clock hypothesis.

The above constraints on the distribution of burst timing in multiburst HVC_X do not apply to single burst neurons, however – neither HVC_X nor HVC_{RA}. If there are no other constraints on burst times, then the bursts of single-burst neurons should distribute themselves uniformly over the duration of the motif. Nonetheless, I observed a bias for single bursts to appear in the first half of song in some birds (tested against binomial distribution, p = 0.42, 0.000065, 0.0019, 0.076, 0.98, Birds 1–5, respectively; see Table 3.3); the sample of neurons evaluated include putative PN that only emitted a single burst per motif. Further analysis showed that, considering only identified HVC_X, a significant result remained for Bird 2 (p = 0.00024, 30/38 single bursts in the first half of song) and Bird 3 (p=0.021, 21/30 single bursts in the first half of song) but the trend for Bird 4 was eliminated (p=0.13, 13/20 single bursts in the first half of song). The existence of these distribution biases in the birds with the largest samples hint at possible additional mechanisms related to network stabilization (see Discussion).

Discussion

Little evidence to date in support of the clock model

Our results demonstrate fundamental weaknesses in the analysis of Lynch et al (2016), including the assumptions that neuronal identity of bursts is irrelevant and that HVC_X burst times are independent even though many of these neurons are known to burst more than once per song motif. This assumes rather than tests a basic premise of the clock hypothesis. A similar error was made in another paper (Picardo et al, 2016) that concluded "HVC projection neurons are divorced from the kinematics of ongoing movements." In that study optogenetic techniques were used to image large numbers of HVC PNs neurons during singing. This included recording from multiburst neurons but in that paper too the timing of each burst was treated as an independent event. As with the analyses of Lynch et al (2016) this statistical error in the analyses performed by Picardo et al (2016) makes uninterpretable the results of their analyses of the optical recordings. Both Lynch et al (2016) and Picardo et al (2016) argued that their data indicates that HVC PN are organized in a continuous representation across the song motif. my results refute this for HVC_X PN.

The clock model concept was initially described as characterizing the activity of small samples of HVC_{RA} recorded during singing, correctly noting a lack of obvious relation of HVC_{RA} burst timing with syllable onsets and offsets (Hahnloser et al, 2002). Given that HVC_{RA} are part of the descending motor pathway for song production, this motivated the original conception of the clock model as describing properties of the motor pathway for song production. Later the activity of HVC_X during singing was first reported (Kozhevnikov and Fee, 2007), also interpreted as a clock (note that Bird 1 from their Fig 1A derives originally from Kozhevnikov and Fee (2007)). (Song structure remains relatively intact following long-term recovery from lesions directed exclusively to HVC_X (Sanchez-Valpuesta et al, 2019; Scharff et al, 2000), and the contribution of HVC_X to

moment-to-moment control of song remains unresolved.) In contrast, more recent clock papers (Lynch et al, 2016; Okubo et al, 2015) have "hitched" very small samples of recordings from HVC_{RA} onto much larger samples of HVC_X when analyzing the temporal structure of bursts from HVC PNs. This presumes there are no functional differences between the different PN classes, which is inconsistent with the lesions results, and is an untested assumption that my results for HVC_X further refute. Evaluating the small HVC_{RA} samples together with the roughly 5-10x larger samples of HVC_X (yielding even larger divergences in the numbers of bursts from the two classes of neurons) can be highly misleading. For example, the 95% confidence intervals for the identified HVC_{RA} in Birds 1–5 (Figure 3.1F) are wide because the sample sizes are small (n = 4–15 HVC_{RA}). Such small samples of HVC_{RA} characterize all birds reported in almost any study to date (Hahnloser et al, 2002; Kozhevnikov and Fee, 2007; Lynch et al, 2016; Okubo et al, 2015; while Daliparthi et al (2019) reported on a larger dataset of 223 positively-identified HVC_{RA} neurons in 6 birds, this impressive result was obtained using genetically-encoded calcium indicators, and so its temporal resolution is necessarily lower – and the authors were interested in different questions than those here). While these sample sizes are remarkable and laudatory given how difficult it is to record HVC_{RA} with single electrode techniques (which provide the desired high temporal resolution), still these recordings can offer little information about the sequential organization of populations of HVC_{RA} in relation to singing. In contrast, the optical recordings of Picardo et al (2016) included presumptive large samples of both HVC_{RA} and HVC_X, but for reasons of technical limitations they could not distinguish between the two classes of PN, and therefore those recordings provide no resolution to this problem. Finally, remarkable intracellular recordings of identified HVC_{RA} have been reported (Long et al, 2010; Picardo et al, 2016). Not surprisingly, in each case there are small numbers of neurons reported per bird. These sample sizes are also insufficient to provide statistically compelling evidence in support of a continuous representation throughout the motif

predicted by the clock model. Given these difficult sampling size problems, no paper to date has provided any compelling evidence that HVC_{RA}, the proper object of the clock model, are organized in the continuous representation envisioned in the clock model hypothesis. Thus, the temporal organization of HVC_X bursts are inconsistent with the clock model, and it remains entirely in the realm of speculation whether HVC_{RA} are.

Both Lynch et al (2016) and Picardo et al (2016) were principally focused on forcefully refuting an alternative "gesture" hypothesis (Amador et al, 2013) by conducting analyses purported to demonstrate that their recordings were more consistent with the clock model than their interpretation of the gesture model. (There are significant issues with how Lynch et al (2016) and Picardo et al (2016) evaluated the gesture model, but I will not discuss them in this document.) As per above, I suggest that this focus was not to their advantage, diverting attention from basic tests of the assumptions of their own model, and diverting attention from searching for possible structure inconsistent with their model. Here, I have intentionally avoided engaging in this form of argumentation – i.e., of which model is more consistent with the data. Even were the analyses in Lynch et al (2016) able to demonstrate that their data was more consistent with one model than the other, this would be at best very weakly informative as to the actual validity of either the clock model or the gesture model. There is a universe of potential models of HVC population activity, including the new one reported here, and both Lynch et al (2016) and Picardo et al (2016) compared just two. They need to provide direct positive support in the form of falsifiable hypotheses for the clock model, not relative support for one model over another. If it is not possible to do so, then it is reasonable to ask if the clock model is amenable to scientific verification at all.

To rescue the clock model as applied to HVC_X it would need to be reformulated to incorporate all the bird-specific, multisequence, and burst-position-within-motif results I have reported. How network activity (see below) can give rise to these results but be blind to song related features of

activity in the distribution of bursts within individual sequences would remain to be addressed.

Furthermore, the standard approach is to use rejection of a model's null as providing insight into the model hypothesis under investigation. As presently formulated the clock model requires inverting this standard logic of statistical testing, attempting to gain insight from lack of rejection of the null hypothesis. This is perilous. In addition, the null model for these tests is a uniform distribution, which shares a lack of structure or regularity with common sources of noise. Noise introduced at any level, from attributes of individual data points, experimental measurements, or analyses, will tend to erode structure, and would then be confused as supporting the model hypothesis. Until it is reformulated so as to provide explicit hypotheses that are falsifiable, statistical testing of the clock model will likely remain unreliable. Furthermore, as I have shown, testing for a uniform distribution requires large samples of neurophysiological recordings (with high temporal resolution), beyond any reported to date. Perhaps optical recordings can ultimately provide such data, but the optical signals will need to have comparable temporal resolution to the fine variation in singing structure on a trial-by-trial basis.

Alternatively, examining other species with more favorable song structure may alleviate some of these technical challenges. Canaries, for example, have songs that vary on a longer timescale (i.e., on the level of what is called a "phrase" in the canary literature – note that this name refers to a different organization than does my usage with respect to zebra finches) than do those of zebra finches, with different phrase types well modeled by different dynamical regimes (Alonso et al, 2010; Trevisan et al, 2006). Canary song also has more heterogeneous distributions of vocal gesture timing than zebra finch song (Lassa Ortiz et al, 2019). Results of HVC recordings in canaries provide independent support for the gesture model (Boari et al, 2022; Herbert et al, 2020), consistent with the zebra finch results (Amador et al, 2013).

Finally, I note that if HVC PN exhibit variation in the rate at which burst timing evolves based

on topography, regional specification, or non-spatially uniform connectional specification within HVC, then aligning bursts or first bursts (i.e., neurons) based on their timing alone could mask this variation. Alternatively, could local spatial uniformity in the sampled neurons explain results such as the remarkable proportion of secondary bursts aligned to a single narrow time window for Bird 4? The entire enterprise of aligning bursts based on timing while ignoring recording location (and numbers of cells recorded) assumes spatial uniformity within HVC. There is some evidence in zebra finches in support of this assumption: response to song playback has shown uniform selectivity for the bird's own song throughout HVC (Margoliash, 1986; Sutter and Margoliash, 1994). Whether this covers all subdivisions of HVC was not assessed, nor does it speak directly to the issue here, though. HVC has three cytoarchitectonically distinct regions (Fortune and Margoliash, 1995) with the large central region showing local rostro-caudal connectivity (Nottebohm et al, 1982; Stauffer et al, 2012). In zebra finches, medial, central, and lateral regions of HVC have complex relationship to song structure and sequence which likely arises in part from complex patterns of regionally localized parallel inputs to HVC (Basista et al, 2014). Given this anatomy, there remains considerable room for functional specialization within and between HVC regions.

Multiple sequences of HVC_X: a new sensorimotor integration model

Exploring the organization of HVC_X bursting has uncovered a wealth of novel structure. Some of these observations, such as bird-specific patterns of activity, align with prior observations (see Introduction) and help to resolve the apparent discrepancy between motor and auditory activity implied by the clock hypothesis, while rejecting the implicit assumption that patterns of activity in zebra finch HVC are different from those reported in a broad range of species.

Other observations such as the multiple sequences are entirely new. One question is: how do the multiple sequences arise? In slice preparations HVC_X show rebound excitation based on the

interactions of hyperpolarization-activated cyclic nucleotide-gated channels (I_h) and T-type calcium channels ($I_{Ca\text{-}T}$) (Daou et al, 2013). Recent results in adult birds demonstrate that the magnitudes of these and the other principal ion currents driving HVC_X spiking are uniform in the HVC_X recorded from a given bird but that these vary from bird to bird. The variation in these HVC_X intrinsic properties (IPs) is related to the similarity of the songs of the birds and is maintained by plastic mechanisms in adult birds (Daou and Margoliash, 2020). Rebound excitation is one of several mechanisms that could give rise to HVC neurons that in song playback experiments have been shown to require sequences of syllables for excitation (Lewicki, 1996; Margoliash, 1983; Margoliash and Fortune, 1992). Here, I propose that a similar mechanism is active when a bird is singing – that the common IBIs observed here in individual birds (Figures 4A, 4C, 4E, 4G, 4I) arise in part from their sharing common properties of rebound excitation. In this model both network properties and IPs are similar within birds and vary between birds. Juvenile zebra finches in the plastic stage of song development show differences in their IPs (Ross et al, 2017) and far less uniformity in IPs (Daou and Margoliash, 2020) as compared with adults, suggesting that network and IPs develop conjointly towards a common solution. Daou and Margoliash (2020) analyzed pairwise comparisons from 18 birds and found a relationship between HVC_X IPs and the acoustic features of birds' songs. my failure here to relate the network properties of the five Lynch et al (2016) birds (bird-specific number and timing of multiple sequences) to their songs may simply be a sample size issue. Alternatively, relating HVC_X IPs to the timing of HVC_X sequences may provide insight.

What might the multiple sequences encode? Sequences of HVC_X later bursts reflect the timing of sequences of HVC_X prior bursts. This could be a potent organization to enhance detection of small variations in HVC activity during singing. Birds presented with continuous delayed auditory feedback (cDAF), showed increases in the number of introductory notes preceding song in as little

as 4 hours (Daou and Margoliash, 2020) hence effects on the preparatory period prior to song, and changes in HVC_X IPs. (A subset of HVC_X are active during the preparatory period (Woolley et al, 2014), as are HVC_{RA} (Daliparthi et al, 2019), and, during the introductory note sequence, HVC interneurons (Rajan, 2018).) Thus, changes in singing behavior can modify HVC_X IPs, either directly (via local projections of movement–related HVC_{RA}) and/or through feedback (extrinsic projections to HVC). Modification of HVC_X IPs were detected in slices prepared as little as four hours after the onset of exposure to cDAF, and a fit to the data across birds with different durations of cDAF exposure suggested that IPs start to change within seconds (Daou and Margoliash, 2020). Thus, HVC_X are well positioned for rapidly responding to fluctuations in singing behavior. Zebra finches typically sing songs in bouts of motifs. Thus, for a given HVC_X, changes at one point in song in one motif might be reflected in modification of HVC_X activity at a later point (a following sequence) in the same motif or the same point in song in a following motif.

The robust representation of the sequential timing of HVC_X initial bursts in sequences of subsequent bursts can serve to provide multiple copies of HVC_X song representation throughout the motif. This suggests a neural implementation of a solution to the well-known problem of feedback delay in the context of internal models of motor control (Wolpert et al, 1995). Instead of posing the problem as to how feedback that is delayed relative to the motor act can "find" and alter a neuron whose motor-related activity earlier in time gave rise to that feedback I envision the problem as mapping a sequence of feedback events onto a sequence of delayed copies of motor-related events. In HVC, this preserves the temporal precision and sparseness of individual HVC_X bursts while organizing them in continuous (dense) representations. I propose that feedback signals from respiratory, auditory, and somatosensory activity representing information about singing behavior are transmitted to HVC (Akutagawa and Konishi, 2010; Ashmore et al, 2005; Suthers et al, 2002; Wild, 1994). A given feedback signal would synapse into a local circuit

(including HVC_X and HVC interneurons) with the HVC_X emitting a non-primary burst at the time that compensates for the feedback delay. Perhaps depending on the strength, structure, and timing of the feedback spikes this would tend to increase, decrease, or leave unaltered the strength, structure, or timing of the HVC_X burst, depending on excitatory/inhibitory local interactions. For a given feedback modality, such an organization would distribute the feedback signal across a sequence of HVC_X spike bursts representing the song while limiting feedback error-driven changes in HVC_X spike bursts to those driven by the appropriately timed previous motor-related activity.

For different modalities, feedback signals would tend to arrive asynchronously in HVC, for example respiratory signals from brainstem centers may feed back onto HVC more quickly than would auditory signals (that involve transduction and many more synapses back to HVC). These differences should be relatively small, however. In anesthetized white-crowned sparrows the shortest sensory-neural delay for auditory noise or tone bursts was 18 ms (Margoliash, 1983). One resolution to this problem would require that each modality access the HVC_X sequence with the modality-appropriate temporal offset. This would have a given HVC_X representing feedback signals of different modalities with different time delays. However, if local circuits receive segregated feedback signals, they can compensate for these small differences in modality-specific delays so that a given HVC_X represents a unified, distinct moment in time. Alternatively, perhaps there are distinct HVC_X sequences for different feedback modalities. Note that any mechanism for distinct processing of different feedback modalities could also address the structural credit assignment problem – associating different processing with different forms (modalities) of feedback. In this model of HVC_X activity each multiburst neuron would integrate forward in time over each interburst interval and perhaps differentiate laterally over representations of feedback. Such hypothesized holistic representations of song and feedback at the level of single cells are consistent with the observation that somatic ion currents of HVC_X neurons are related to the individual's song

and responsive to changes in feedback (Daou and Margoliash, 2020).

Such models can help to explain how auditory and motor representations of song are almost precisely temporally aligned in zebra finches (Amador et al, 2013; Dave and Margoliash, 2000) as well as in other species (Prather et al, 2008). I envision HVC_X activity as an error signal first calculated in HVC that is then transmitted to the basal ganglia (Daou and Margoliash, 2020). If so, this would require reformulation of models that hypothesize that error signals are first computed in the basal ganglia circuit (e.g., Fee and Goldberg (2011)).

Distributions of burst activity: network models and network constraints

If HVC_X have "lateral" interactions with each other, this could increase the acuity of spike burst timing for the initial spike bursts, hence improving the acuity of the song motor representation. For HVC_X with multiple bursts such lateral interactions will also improve the acuity for detecting feedback errors through differential comparisons of different parts of a sequence that were receiving different feedback values. This would arise dynamically, limiting feedback interactions to later bursts of HVC_X with other near-synchronously bursting HVC_X.

The data suggest that a given sequence of HVC_X later bursts can include neurons throughout the temporal extent of the motif (Figure 3.8). This opens the possibility that activity from the earlier parts of the sequence carries information about the intended target of the motor behavior. This envisions HVC_X organized as a "forward" model (Amador et al, 2013). Feedback is also evaluated throughout the sequence, giving rise to information about the consequences of the motor behavior. Both features of motor control can be accommodated.

What would give rise to a distribution of HVC_X having one or more bursts? A reverberant network will tend to accumulate local activity (hence feedback). During development, those HVC_X that happen to receive increased levels of local feedback over the duration of a motif may tend to

be more likely to emit multiple bursts. I also observed a trend for single bursting HVC_X to fire earlier in song. This can balance the necessity for the later bursts of multiburst neurons to be emitted later in song. This could arise if there is a network-level constraint to balance excitatory activity (in the HVC_X) across the song motif. In evaluating developmental scenarios it is noteworthy that early in development HVC_X and the basal ganglia (but not HVC_{RA}) drive song (Aronov et al, 2008).

One obvious limitation in these models is the failure to incorporate HVC_{RA}, for which there is insufficient information. If there are significant lateral interactions from HVC_X to HVC_{RA} or features of HVC_{RA} that mimic what I have reported for HVC_X this too would have profound effects on HVC models. In any case, rather than being unusually difficult to invalidate I believe the totality of these speculative suggestions are primed for easy invalidation, in part or whole, motivating experimental testing and new HVC model building. This would be a positive result.

CHAPTER 4

HVC NEURONAL RESPONSES TO PSEUDO-DAF PLAYBACK

Introduction

In many songbird species, it has been observed that acoustic stimuli can provoke neuronal responses throughout the song system, including in motor-related nuclei in the absence of motor activity. These responses have been reported in white-crowned sparrows (Margoliash, 1983; Margoliash and Konishi, 1985; Margoliash, 1986), zebra finches (Margoliash and Fortune, 1992; Dave and Margoliash, 2000; Rauske et al, 2003), Bengalese finches and swamp sparrows (Prather et al, 2008), and others. In Bengalese finches, swamp sparrows, and white-crowned sparrows, among others, these responses can be elicited by playback of auditory stimuli to awake animals, but in zebra finches these responses are suppressed during wakefulness, and can only be reliably elicited while the bird is asleep (Dave et al, 1998; Rauske et al, 2003).

Typically the most potent stimulus to provoke this activity in many species is the bird's own song (BOS), and this is true in zebra finches as well. In the motor nucleus RA, BOS playback elicits neuronal responses in sleeping zebra finches reliably and robustly, and the pattern of neuronal activity evoked closely resembles that observed in the nucleus during singing (Dave and Margoliash, 2000). This similarity between HVC activity evoked by playback (during wakefulness) of elements of a bird's own song and activity produced during singing that same song has also been observed in Bengalese finches and swamp sparrows (Prather et al, 2008). Such similarity has not been directly assessed in zebra finch HVC, but auditory responses in RA are likely to arise from HVC activity (Dave and Margoliash, 2000; Shea and Margoliash, 2003; Hahnloser et al, 2006).

Modifications of BOS produce playback responses that are attenuated compared to unmodified

BOS. In white-crowned sparrows and zebra finches, neurons that responded selectively to sequential elements of the bird's own song generally responded less robustly, if at all, to a wide variety of manipulations of their maximal stimulus, including sequential shuffling, changing inter-element intervals, modifying the frequency content of song elements, or constructing artificial approximations of these song elements (Margoliash, 1983; Margoliash and Fortune, 1992). Similarly, attempts to produce synthetic reconstructions of a zebra finch's song for BOS playback using a biomechanical dynamical systems model resulted in stimuli that could elicit responses from BOS-specific neurons in HVC, but they were significantly attenuated (attaining at most roughly 65% of the response to BOS) (Amador et al, 2013). To date, there have been no reports of stimuli which produce super-normal responses in HVC – that is, responses greater than those to BOS. Examining HVC responses in anesthetized zebra finches, in 328 pair-wise comparisons between responses to BOS and responses to conspecific song (CON), only approximately 20 (~6%) showed marginally stronger responses to CON. The peak of the distribution was at zero – i.e., no overall response to CON (Margoliash et al, 1994).

The behavioral changes I observed in response to short-term DAF in Chapter 2 indicate that DAF is inducing some change in song control circuits at short latency. A prior report of changes to HVC projection neuron intrinsic properties as a result of similarly brief DAF treatment points to one part of a mechanism for such a change (Daou and Margoliash, 2020), but how changes to auditory feedback drive changes in intrinsic properties in individual neurons is unknown. Playback of auditory stimuli to sleeping birds to elicit HVC responses was a convenient paradigm to approach this question, as it allows complete control over what a bird hears. Responses in HVC to auditory stimuli resembling what a bird hears during DAF treatment should thus be informative as to changes in neuron activity during aberrant auditory feedback.

Methods

Housing

A population of adult male zebra finches (at least 120 days post-hatch) were acclimated to a reversed 14:10 light:dark cycle in a large isolation box where they were housed communally, thus adjusting the bird's subjective night to start when the experiments began in the solar morning. From this population, ten birds were placed in sound isolation boxes (maintaining light-reversal) with unrestricted access to food, water, grit, and cuttlebone. I recorded and monitored their singing behavior until they sang robustly (2-5 days, typically), and identified recordings of bouts of song containing good renditions of their canonical motif after examining typically approximately 100 bouts of song. The recording hardware comprised a Behringer C-2 studio condenser microphone attached to either an M-Audio Delta 1010LT or Behringer U-PHORIA UMC1820 amplifier/sound card for amplification and digitization. The digitized signal was recorded using the Linux JACK audio framework and an in-house recording suite, jill, as in Chapter 2.

Surgical implantation

Once I had song recorded, I surgically implanted a pin for head-fixation under isoflurane anesthesia (2% delivered in compressed air, at 2 LPM; the concentration of anesthetic was reduced steadily while ensuring maintenance of anesthesia and steady breathing at a rate of 1-2 Hz, typically being stabilized around 1%; for more detailed discussion of anesthesia technique, see Canavan, 2019). This procedure has been described elsewhere (Margoliash and Fortune, 1992; Sutter and Margoliash, 1994; Janata and Margoliash, 1999), but briefly: the bird is given intramuscular injections (in the pectoral muscle) of 0.2mg/kg meloxicam (analgesic; for a 15g bird this works out to 30 µL of 0.1 mg/mL solution) and 5 mg/kg enrofloxacin (antibiotic, brand name Baytril; for a 15g

bird this works out to 3-4 μL of 22.7 mg/mL solution) before being placed in a stereotaxic apparatus. The scalp is sterilized, cut, and retracted, the top layer of skull is removed with a scalpel in a circle of roughly 9 mm diameter (to ensure that HVC in both hemispheres is available during an experiment), centered on the midline, with its back edge lying 3-4 mm caudal of the bifurcation of the midsagittal sinus (λ). The pin is placed far enough caudal of λ that access to HVC will be unimpeded later – this typically means that the rostral face of the pin is 1-2 mm caudal of λ. The pin is cemented in place with dental acrylic (with curing accelerated with cyanoacrylate glue), which is also used to create a "well" around the window in the top layer of skull for deposition of sterile saline during the experiment. The bird was allowed to wake up, monitored for 1 hour during recovery, and returned to his home cage thereafter. Birds typically resumed normal robust singing (a good proxy for overall health and comfort, along with daily visual inspection) within 2 days of surgery.

Stimulus preparation

While the bird was recovering, I prepared experimental stimuli from the recorded song. Similar to what has become standard practice (Janata and Margoliash, 1999; Brown, 2017), I selected a single bout of song containing 3-5 motifs sung to completion (the exact number chosen depended on the bird's behavioral peculiarities – for some birds, bouts longer than 3 motifs are rare), plus the bout's introductory notes. As stimulus responses in acute playback experiments often increase in strength and reliability as a bout progresses (i.e., the response to the second or third motif in a bout is typically both stronger and more consistent across trials than that to the first motif, though this is not always the case, and it is not uncommon when presenting longer bouts for the last motifs in a bout to produce weaker responses again), more than one motif is desirable. This bout constituted the BOS stimulus.

Once I had the BOS stimulus identified, I generated further stimuli by manipulating this recording using custom Python software. Every bird received BOS as well as a reversed BOS (the entire BOS stimulus played backwards in time), termed REV. Every bird also received a stimulus I termed pseudo-DAF, wherein the BOS signal is duplicated and overlaid on itself at 100 ms delay, producing a stimulus that resembled the auditory feedback experienced by a bird undergoing DAF treatment as in Chapter 2. In pseudo-DAF, the first copy of BOS represents the auditory feedback normally perceived by the bird when it sings, and the 100 ms echo represents the delayed copy, in this case both signals presented to the bird over the speaker. Throughout this chapter, I refer to stimuli of this sort as "DAF" for simplicity, but note that these birds have never been exposed to DAF during singing. The hypothesis under test is that these stimuli mimic the effect of DAF during singing in a sufficient manner (and under better controlled and more tractable conditions) so as to produce meaningful results. Work in bat echolocation where one stimulus mimics the emitted pulse and another the returning echo gives support for this approach (O'Neill and Suga, 1979; Suga and O'Neill, 1979; Suga et al, 1979; Macías et al, 2016). I also presented a variety of other stimuli, which will be described as they are presented in results. Other than BOS, REV, and DAF, not every stimulus was presented to every bird, as the limitations of the head-fixed sleep experiment and likelihood of holding a neuron for an extended period of time mean that restricting a full run of experimental stimuli to a little over 1 hour is advisable; this typically corresponds to 5-7 stimuli in total.

Multielectrode arrays

For all but one bird, I used silicon multielectrode arrays that D. Margoliash received from Tim Harris at the Janelia Research Campus. (Similar designs are available from a number of vendors, including Cambridge NeuroTech and Diagnostic BioChips.) Two designs were used: a single-shank

probe, and a two-shank probe (with an inter-shank pitch of 250 μm); both designs had a total of 64 recording sites. For the remaining bird, I used a Diagnostic BioChips P64-8 array; this device is very similar to the 2-shank array above, and also had a total of 64 recording sites. The primary difference between the Diagnostic BioChips array and the others is its inclusion of a 64-channel Intan amplifier headstage pre-attached via a flexible ribbon cable; otherwise, recording using this device was set up in the same manner as the others. The other arrays were connected to two Intan 32-channel amplifier headstages (model 2132). In all cases, the Intan chips were attached to an Intan USB evaluation board (Intan Technologies) for display and recording on a Linux desktop using the Intan GUI recording software. Data were recorded at 30 kHz (some birds at 20 kHz due to operator error; this had no bearing on results, and was accounted for throughout data analysis), with a bit depth of 16 bits per sample. Stimuli were played using JACK and jill, and copies of the stimuli being played along with a trigger pulse were sent to the Intan board and recorded along with the electrophysiological signals.

Experimental paradigm

Recordings commenced after several days following the surgery, once the bird had recovered based on its appearance and behavior (including rate of singing). Just before the subjective night, I removed a bird from his home cage and lightly anesthetized him with isoflurane (same parameters as described for surgery above) while injecting meloxicam and enrofloxacin (same dosage as for surgery above), as well as 1 mg/kg dexamethasone (corticosteroid; for a 15g bird this works out to 38 μL of 0.4 mg/mL solution in 70% EtOH), to reduce inflammation and the potential of brain swelling or herniation in response to electrode array insertion. The bird was placed in a stereotaxic apparatus (using a threaded hole in the head-fixation pin) in a darkened sound isolation chamber large enough to hold the bird, the experimental apparatus, and the experimenter (during the initial

surgical portion). I located HVC using stereotaxic coordinates (initial coordinates: 400 μm rostral of the back of λ and 2100 μm lateral of the midline), removed the bottom layer of skull in a small craniotomy with a scalpel, and cut and retracted as much as possible of the dura mater under the craniotomy using dura hooks.

I inserted a silver ground wire by hooking it under the lip of the skull window, and then inserted the electrode array at the desired coordinates deep enough that I was confident it would not "dimple" the brain upon further insertion, but shallow enough that it was still likely dorsal of HVC. I filled the acrylic well around the recording site with sterile saline to prevent drying of and damage to the tissue. I then applied 50 μL melatonin jelly (0.2 mg/mL in petrolatum) to the skin of the back of the neck to encourage sleep (S. Seltmann, personal communication; protocol later published in Seltmann et al, 2016). At least one previous study has reported using melatonin injections to promote sleep (Hahnloser et al, 2006), but topical application works well, and additionally can be reapplied without removing the bird from the head-fixation apparatus. In longer experiments, I reapplied melatonin at hour 5.

I then left the sound isolation booth and slowly (no faster than 50 μm per minute) lowered the array further into the brain while monitoring the electrophysiological signals from the array, periodically playing the BOS and REV stimuli to assess whether the neurons being recorded responded differentially. At all points, I monitored the bird's sleeping condition via webcam (with IR filter removed; the booth was illuminated with an IR light so as not to wake the bird), a microphone inside the booth which would pick up any sounds of movement, and the neuronal signals I was observing. If I did not see a neuronal response I believed to the HVC by the time I had lowered the array by 1000 μm, I retracted it out of the brain and tried a new set of coordinates, picking out a grid across the brain surface with a spacing of 200-300 μm until I either found HVC or assessed that the experiment for that day was unlikely to succeed, because of the advanced hour

or poor sleeping condition of the bird. During successful experiments, I typically required 1-3 penetrations to find a neuronal response to BOS (and lack of a response to REV) that I recognized from experience as being that of HVC.

Stimulus presentation

Once I was confident I had found HVC and had a good set of neuronal signals on the electrode array, I waited 20-30 minutes to allow the brain tissue time to relax from the pressure placed on it from advancing the electrode. I verified that this relaxation had not removed the array from HVC, and then played the set of prepared stimuli. Stimuli in a set were repeated 20 times each. Stimuli were played in pseudo-random order to guard against potential entrainment of auditory responses to an expected stimulus. The interval of silence between each stimulus varied between 20 and 30 seconds. Early in the experiment I relied on a complete shuffle of the stimuli, but found that this could interact poorly with the nature of a sleep experiment. It was normal for a bird to wake up briefly once or twice during a run of a stimulus set, either in response to an arbitrary stimulus or for no apparent cause, before returning to sleep. A completely random shuffle of stimuli could (and on one occasion, did) mean that a large number of presentations of one particular stimulus were played during a few-minute period when the bird had awoken, and thus produced no usable data. To avoid this in subsequent experiments, I changed the shuffle procedure to break the stimuli into sets or epochs (e.g., BOS-REV-DAF) and shuffle only the stimuli within an epoch (e.g., BOS-REV-DAF, followed by REV-BOS-DAF, etc), so that any given period of wakefulness would cover a roughly evenly distributed sample of the stimuli. Once a stimulus set was over, I would repeat this procedure, either with a new stimulus set or a new electrode location.

Experiment termination

I halted an experiment either when I had assessed that the quality of the bird's sleep had degraded, or when his subjective dawn was approaching. After retracting the array and sealing the craniotomy with silicone elastomer (Kwik-cast), I removed the bird from the stereotaxic apparatus and returned him to his home cage. I monitored the bird for an hour post-experiment, verifying that he was comfortable and was eating and drinking. Birds typically recovered from an experiment the next day, and would be usable for subsequent experiments thereafter. The shortest period between experiments for a given bird was 72 hours.

Data analysis

Post-experiment, data was converted into a lab-standard format ("bark") and separated into different pipelines: one for stimulus data, including trigger pulses and stimulus playback, and one for electrophysiological data. Stimulus data were used to extract trial windows within the experiment. Electrophysiological data were band-pass filtered between 300 Hz and 8 kHz (Butterworth filter, 2^{nd} order) and both examined as-is as well as spike-sorted. Spike-sorting was done using Kilosort 2 (https://github.com/MouseLand/Kilosort). I chose sorting parameters within a range suggested by the developers for silicon array recordings. Within that range, I used trial-and-error to find a best sort for HVC data. Typical sorting parameters were: high threshold 10-16, low threshold 6-12, minimum firing rate 0.01, lambda 10, AUC 0.9 (the higher-than-default thresholds appear to be critical to get good isolation of units out of multielectrode HVC recordings), and a final manual unit curation step using Phy (https://github.com/cortex-lab/phy). I then evaluated units by their response to BOS, their IBI distribution, their mean spike shape, and their background firing rate. A stimulus whose response properties are well-understood goes a long way towards removing some of the ambiguities and uncertainty in spike-sorting. I was able to easily

distinguish HVC interneurons from projection neurons by their firing rate (interneurons are tonically firing, projection neurons are phasically firing). Although I did not perform antidromic stimulation to definitively identify projection neuron targets, the phasic-response projection neurons which produce two or more bursts per motif are almost certainly HVC_X (Kozhevnikov and Fee, 2007).

Results

I recorded responses to playback of natural and manipulated song in an acute preparation in 10 birds; in 1 bird (lilac60) preliminary analysis of the data indicated that BOS selectivity was much lower than it had appeared during the experiment, indicating that either the electrode array was not in HVC, or that the quality of the bird's sleep was very low. I therefore excluded this bird from further analysis, leaving 9. Each bird received BOS, REV, and pseudo-DAF (see Methods), and received in addition between 2 and 6 other stimuli (Table 4.1), some of which will be described below.

Bird	Stimuli presented
lilac51	BOS, REV, DAF, intro note DAF, CON+BOS
lilac52	" , " , " , intro note DAF, CON+BOS
lilac59	" , " , " , intro note DAF, amplitude-matched BOS
lilac17	" , " , " , intro note DAF, primed DAF, scrambled DAF
lilac194	" , " , " , intro note DAF, primed DAF, scrambled DAF
lilac154	" , " , " , intro note DAF, partial DAF
lilac4	" , " , " , intro note DAF, scrambled DAF
lilac20	" , " , " , assorted delay DAF
lilac199	" , " , " , partial DAF, assorted delay DAF

Table 4.1. Summary of experimental subjects and stimuli presented.

Robust BOS responses in HVC in sleeping birds

All birds produced reliable neuronal responses to presentation of BOS during sleep. A small set of units (~8%) did not show specificity for BOS playback over playback of REV, and were not analyzed further. The majority of spike-sorted units showed marked responses to BOS presentation, and either no response to REV or attenuated responses compared to BOS. (The high proportion of units showing BOS selectivity over REV in this sleeping paradigm highlights its value – similar experiments in urethane-anesthetized animals report much higher proportions of nonselective responses (e.g., Margoliash, 1983, though note that this study examined another species, the white-crowned sparrow).) The datasets for all birds contained units clearly representing both putative interneurons (Figure 4.1) and putative projection neurons (Figure 4.2), identified as such by their firing properties, with interneurons having tonic activity, even outside of stimulus periods, and projection neurons having phasic responses to BOS and no or low (and inconsistent, not tonic) activity outside of stimulus periods, as is well established. I have termed these neurons "likely projection neurons" and "likely interneurons," respectively. Each bird's dataset also included a number of less obviously clear units – some have firing properties not clearly meeting the criteria for either interneuron or projection neuron, but have what appears to be good isolation (less than 1% of inter-spike intervals (ISI) under 1.5 ms) (Figure 4.3). Many of these units are probably interneurons, but many others, across multiple birds, show clear phasic firing properties, and so likely represent projection neurons contaminated by spikes from one or more interneurons. I have termed these "probable projection neurons," to distinguish them from both multiunits and from those units whose identity is clear and which are not visibly contaminated. Each bird's dataset also contained a large number of multiunits, which are likely dominated by interneuron activity (Figure 4.4). Units returned by Kilosort which were obviously noise (determined by spike waveform, ISI,

	likely PN			likely IN			probable PN			multiunits		
Bird	total	up	down	total	up	down	total	up	down	total	up	down
lilac51	4	2	1	7	1	5	20	3	12	32	1	31
lilac52	3	2	1	9	2	5	36	3	31	62	0	62
lilac59	1	1	0	6	1	5	14	5	9	23	0	23
lilac17	2	1	1	15	7	6	17	5	12	63	1	62
lilac194	3	1	2	10	3	7	15	3	12	41	0	41
lilac154	2	0	2	4	2	2	24	3	21	13	2	11
lilac4	1	0	1	2	1	1	8	3	2	54	15	39
lilac20	4	1	3	0	0	0	5	0	5	20	12	8
lilac199	0	0	0	0	0	0	10	1	9	90	13	77

Table 4.2. Summary of spike-sorting results for all birds. All identifications are putative. Likely PN and IN are clearly identifiable by firing properties. Probable PN show phasic BOS responses but also likely contamination by tonically-firing neurons. "Up" and "down" refer to whether the neuron's response to DAF involved potentiation or suppression, respectively, determined by mean firing rate per trial. A small number of neurons did not show a clear difference in response between BOS and DAF, and so for any given category in a bird, "up" and "down" may not sum to "total."

and "firing" properties) were discarded. A summary of the units recovered via spike-sorting from each bird can be found in Table 4.2.

Pseudo-DAF playback triggers differential responses

Playback of pseudo-DAF stimuli produces very different responses than does playback of BOS. The responses fall into three categories – one set of neurons shows a strong suppression of response when pseudo-DAF is played (example in Figure 4.5), and another set shows a strong increase in response compared to BOS (example in Figure 4.6). The third set, which for all birds contains a small fraction of the neurons, have unclear or not obviously potentiated or suppressed responses. (That is, nearly all neurons from all birds show clear potentiation or suppression of response.) The first set, suppressed neurons, would not be hugely surprising were it to appear in isolation – as discussed above, all manipulations of BOS assayed to date appear to produce weaker responses in HVC neurons than these neurons produce in response to unmodified BOS. A uniform suppression of response to DAF playback could also be readily explained by poor sleep, either in a single bird or

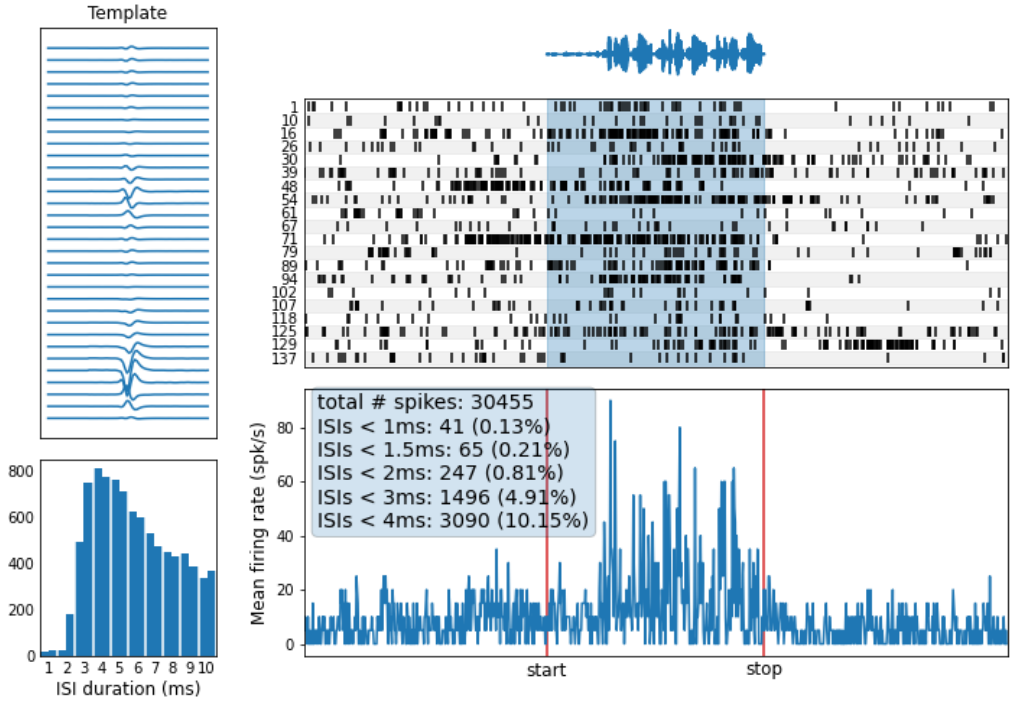


Figure 4.1. Example BOS-responsive likely HVC interneuron. This unit is from bird lilac52. Top-left: the average spike template produced by Kilosort. (Note that in reality the top and bottom halves of this panel are arranged side-by-side, as the array's shanks have two columns of pads next to each other.) Bottom-left: histogram of the ISI distribution over short intervals. Top-right: oscillogram of the stimulus played – here, the bird's own song. This is a three-motif song. Middle-right: per-trial spike raster. Shaded blue region is the stimulus period. The entire stimulus is roughly 4 seconds in duration. Bottom-right: peri-stimulus time histogram (PSTH) showing the firing rate for this unit averaged over all trials. Red vertical bars denote stimulus beginning and end. Inset is the refractory period violation data. The threshold for good isolation was chosen to be less than 1% of all ISIs below 1.5 ms. The high tonic firing rate outside of the stimulus period marks this unit as a likely interneuron.

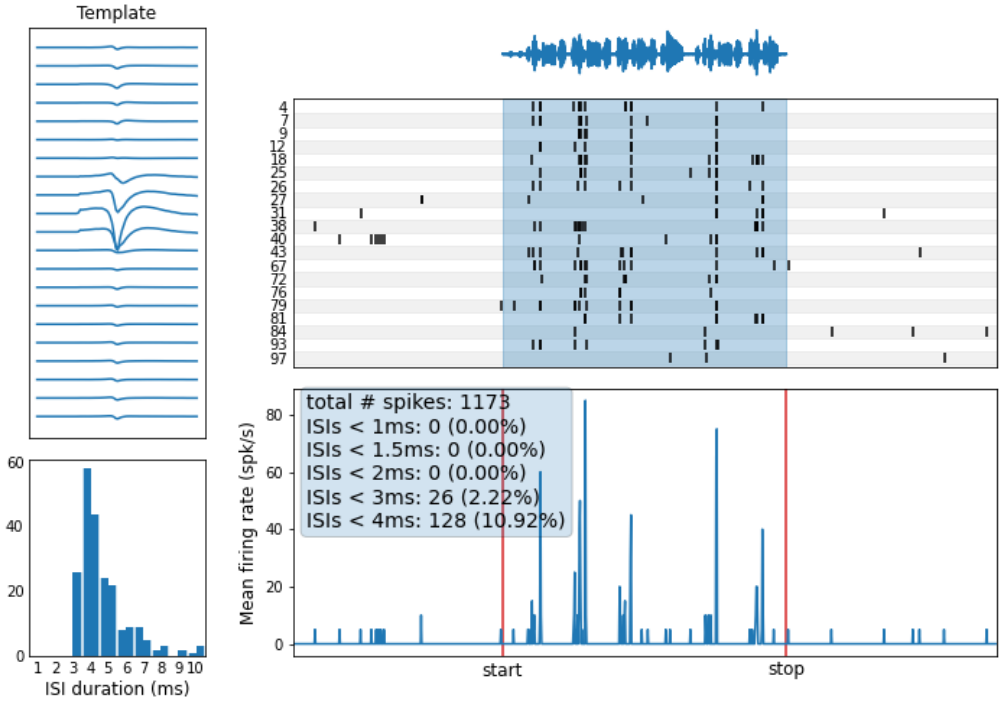


Figure 4.2. Example BOS-responsive likely HVC projection neuron. This unit is from bird lilac59. Top-left: the average spike template produced by Kilosort. (Note that unlike Figure 4.1, this template is veridical – the array used in this experiment had only a single column of recording sites.) Bottom-left: histogram of the ISI distribution over short intervals. Top-right: oscillogram of the stimulus played – here, the bird's own song. This is a five-motif song with a call note between the 3rd and 4th motif. Middle-right: per-trial spike raster. Shaded blue region is the stimulus period. The entire stimulus is roughly 5 seconds in duration. Bottom-right: peri-stimulus time histogram (PSTH) showing the firing rate for this unit averaged over all trials. Red vertical bars denote stimulus beginning and end. Inset is the refractory period violation data. The threshold for good isolation was chosen to be less than 1% of all ISIs below 1.5 ms. The near-zero firing rate outside of the stimulus period together with its phasic response to BOS marks this unit as a likely projection neuron.

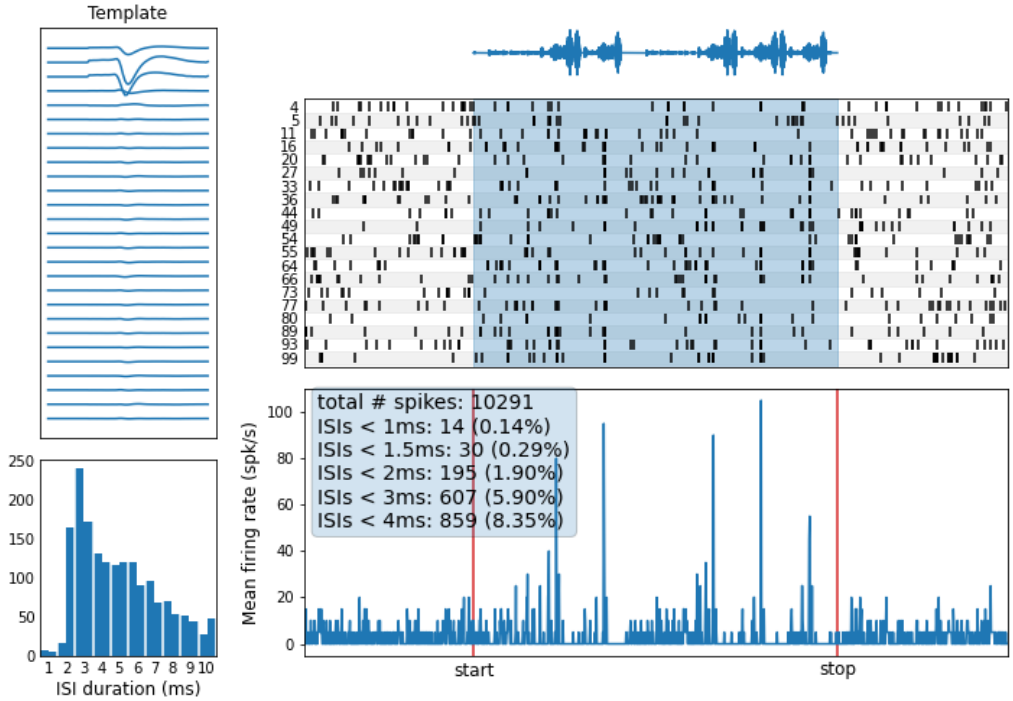


Figure 4.3. Example BOS-responsive probable HVC projection neuron. This unit is from bird lilac17. Top-left: the average spike template produced by Kilosort. Bottom-left: histogram of the ISI distribution over short intervals. Top-right: oscillogram of the stimulus played – here, the bird's own song. Middle-right: per-trial spike raster. Shaded blue region is the stimulus period. The entire stimulus is roughly 7 seconds in duration, as it includes two bouts. The first bout has three motifs and the second bout has two motifs. Bottom-right: peri-stimulus time histogram (PSTH) showing the firing rate for this unit averaged over all trials. Red vertical bars denote stimulus beginning and end. Inset is the refractory period violation data. The threshold for good isolation was chosen to be less than 1% of all ISIs below 1.5 ms. The highly phasic BOS response of this unit suggests it is a projection neuron, but the higher tonic firing rate, both during and outside of stimulation, indicates that it is likely contaminated by spikes from one or more interneurons, despite the favorable ISI distribution.

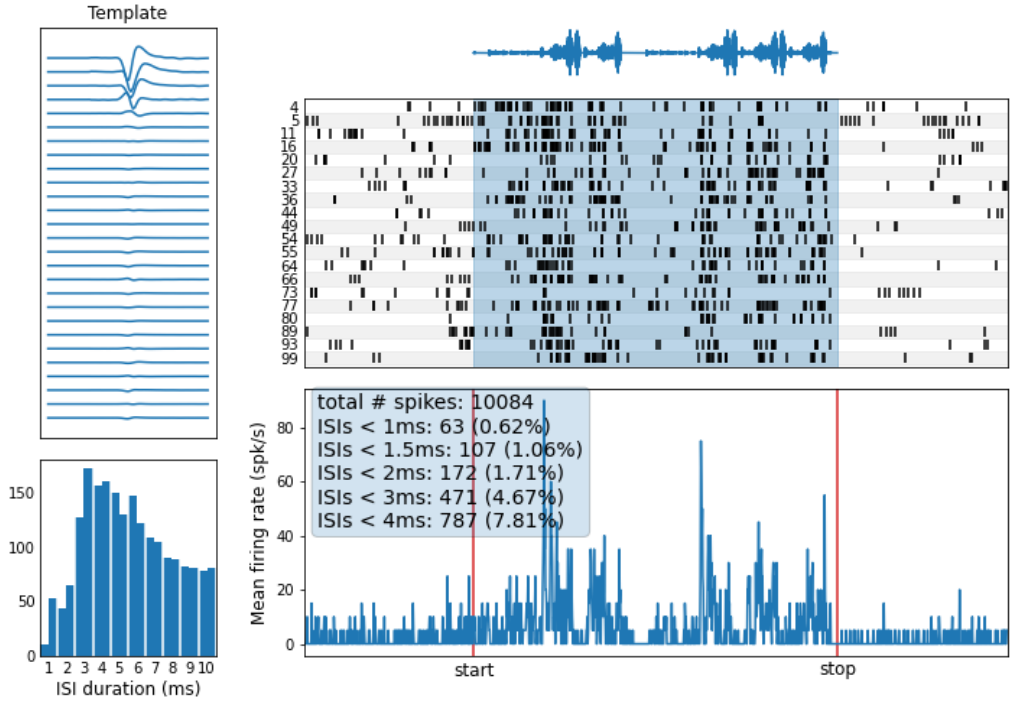
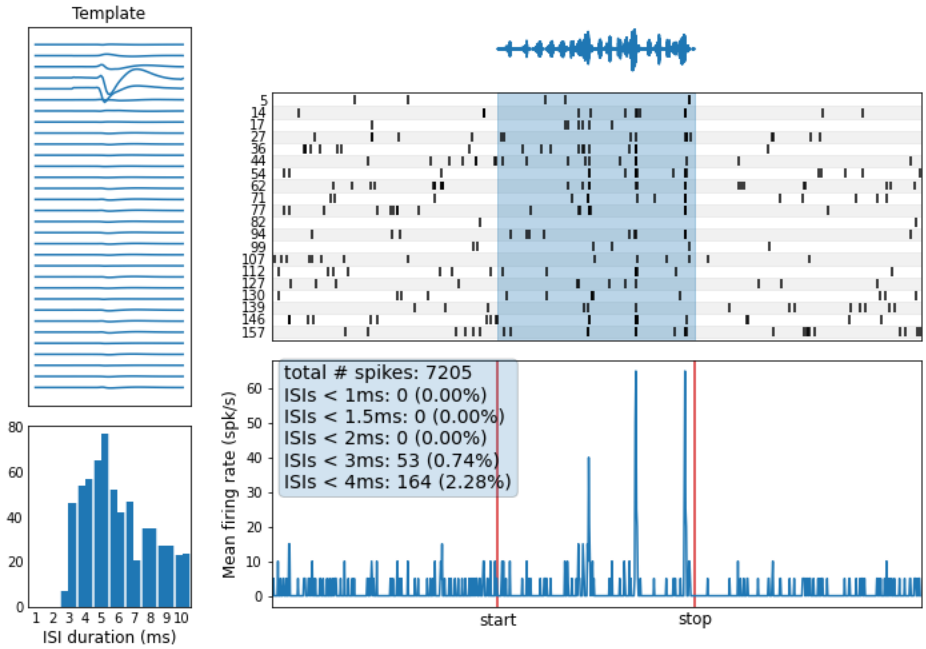


Figure 4.4. Example BOS-responsive HVC multiunit. This unit is from bird lilac17. Top-left: the average spike template produced by Kilosort. Bottom-left: histogram of the ISI distribution over short intervals. Top-right: oscillogram of the stimulus played – here, the bird's own song. Middle-right: per-trial spike raster. Shaded blue region is the stimulus period. The entire stimulus is roughly 7 seconds in duration, as it includes two bouts, the first with two motifs and the second with three motifs. Bottom-right: peri-stimulus time histogram (PSTH) showing the firing rate for this unit averaged over all trials. Red vertical bars denote stimulus beginning and end. Inset is the refractory period violation data. The higher number of refractory period violations (greater than 1% of ISIs below 1.5 ms) indicates that this unit should be classified as a multiunit. It is clearly BOS-responsive, however.

in response to the DAF stimulus in particular – it would be possible that some feature of the DAF stimulus consistently awakened the birds (but see below the discussion of presenting amplitude-matched BOS to a subset of birds). I monitored the birds throughout the experiments via video and microphone, however, and did not note a consistent effect of the DAF stimuli on any of the measures of sleep quality, whether background neuronal firing rate or the likelihood of a bird opening an eye or moving in the apparatus. But it is the appearance of a set of neurons which respond more strongly to DAF than to BOS, and moreover the appearance of neurons of each type in each bird in a single experiment (eliminating day-by-day variation as a potential cause), which argues even more firmly against such explanations (see Table 4.2).


Template
ISI duration (ms)
total # spikes: 7205
ISIs < 1ms: 0 (0.00%)
ISIs < 1.5ms: 0 (0.00%)
ISIs < 2ms: 0 (0.00%)
ISIs < 3ms: 53 (0.74%)
ISIs < 4ms: 164 (2.28%)
Mean firing rate (spk/s)
start
stop

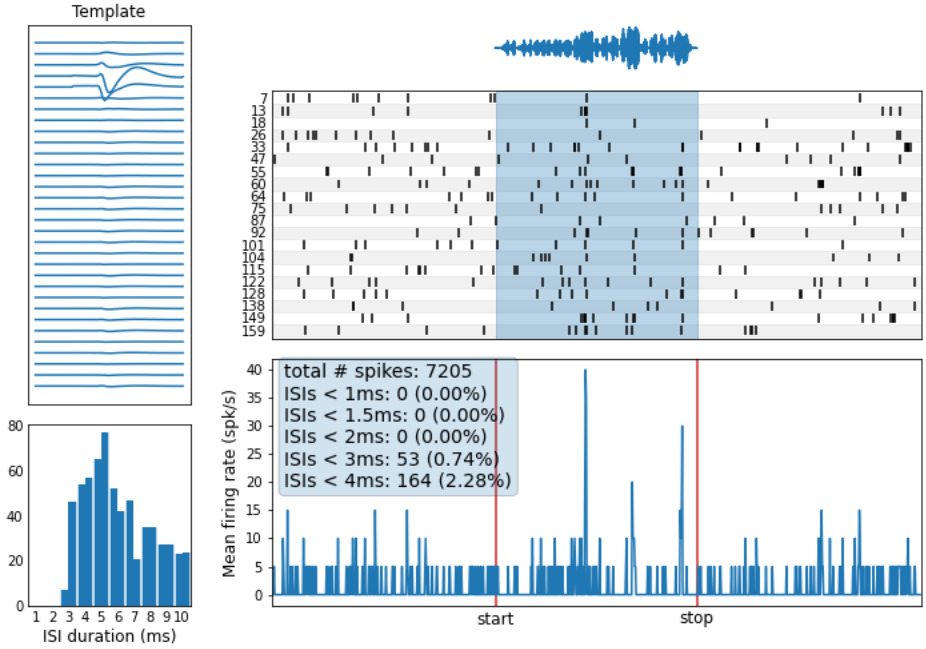
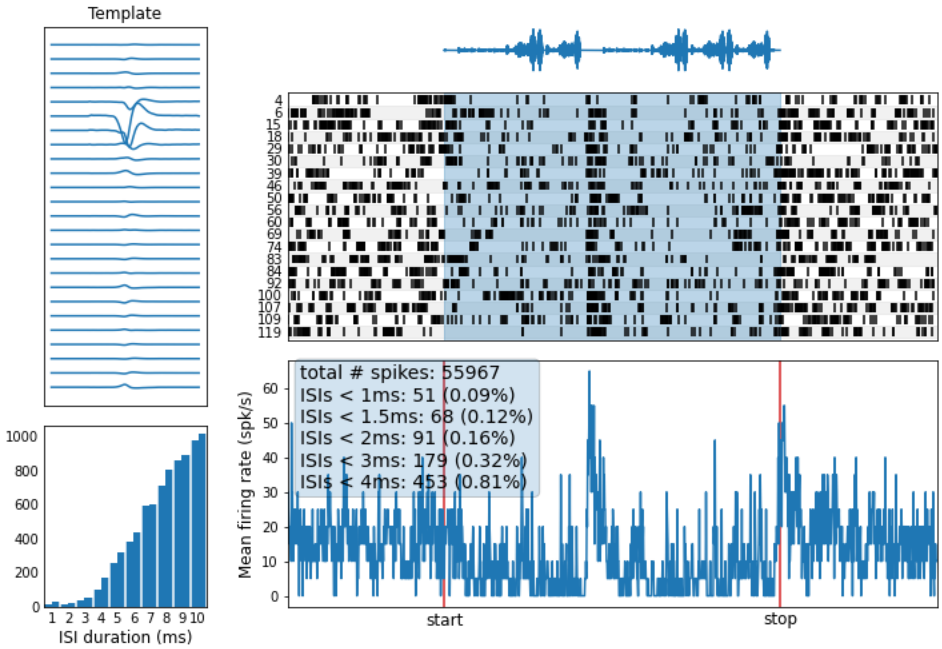

Template
ISI duration (ms)
total # spikes: 7205
ISIs < 1ms: 0 (0.00%)
ISIs < 1.5ms: 0 (0.00%)
ISIs < 2ms: 0 (0.00%)
ISIs < 3ms: 53 (0.74%)
ISIs < 4ms: 164 (2.28%)
Mean firing rate (spk/s)
start
stop

Figure 4.5. Probable projection neuron suppression in response to pseudo-DAF. (previous page) This unit was recorded from lilac154. Top half: unit summary as in Figure 4.1-4.4 for the response to BOS playback. Bottom half: unit summary for the response to pseudo-DAF playback. Note that the firing rate spikes in response to pseudo-DAF occur at the same locations as they do for BOS playback.


Template
ISI duration (ms)
total # spikes: 55967
ISIs < 1ms: 51 (0.09%)
ISIs < 1.5ms: 68 (0.12%)
ISIs < 2ms: 91 (0.16%)
ISIs < 3ms: 179 (0.32%)
ISIs < 4ms: 453 (0.81%)
Mean firing rate (spk/s)
start
stop

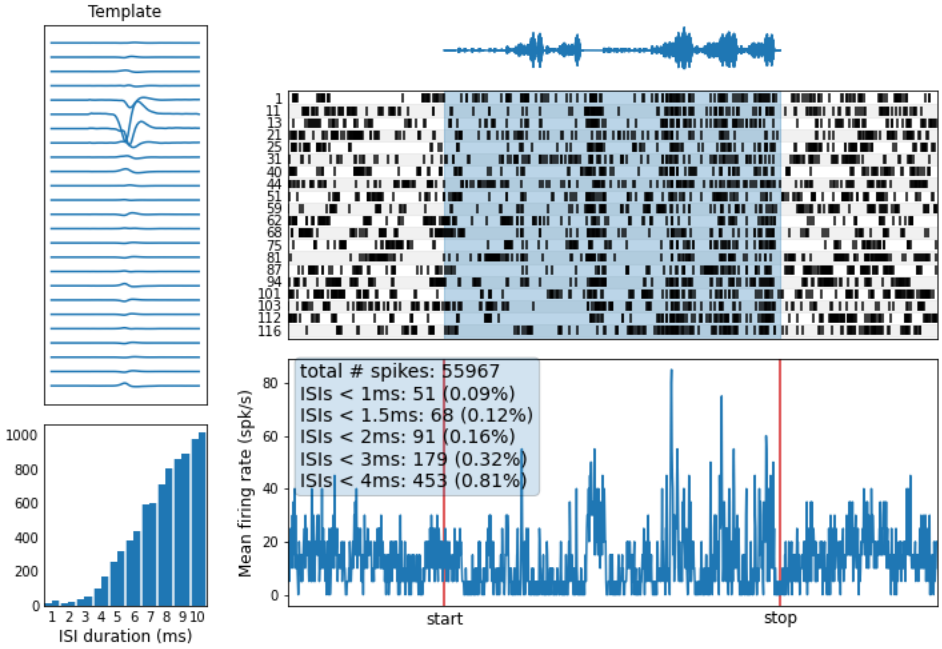

Template
ISI duration (ms)
total # spikes: 55967
ISIs < 1ms: 51 (0.09%)
ISIs < 1.5ms: 68 (0.12%)
ISIs < 2ms: 91 (0.16%)
ISIs < 3ms: 179 (0.32%)
ISIs < 4ms: 453 (0.81%)
Mean firing rate (spk/s)
start
stop

Figure 4.6. Likely interneuron potentiation in response to pseudo-DAF. (previous page) This unit was recorded from lilac17. Top half: unit summary for the response to BOS playback. Bottom half: unit summary for the response to pseudo-DAF playback. Note that the pseudo-DAF stimulus at bottom only contains DAF overlaid on top of BOS for the second half of the stimulus. This allows additional comparison of BOS and DAF responses within a single stimulus.

Likely projection neurons were identified in 8 out of 9 birds. For the remaining bird, lilac199, while it was not obvious during the experiment, it is likely that a grounding issue led to lower recording quality, which presented a larger challenge to the spike-sorting algorithm. Probable projection neurons were identified in this bird, as well as a large number of multiunits. In the birds with likely projection neurons, 6 out of the 8 had 1 or 2 projection neurons whose mean firing rate over a trial increased under DAF compared to BOS ("up" neurons in Table 4.2). Spike-sorting identified probable projection neurons in all birds , and in all birds but lilac20, these probable projection neurons included at least 1 neuron whose firing rate increased in response to DAF playback. 7 out of 9 birds had likely interneurons among their spike-sorted units, and of these, all had at least 1 which increased its firing rate in response to DAF playback. All birds had large numbers of multiunits, and in most, the preponderance of multiunits showed response suppression during DAF playback.

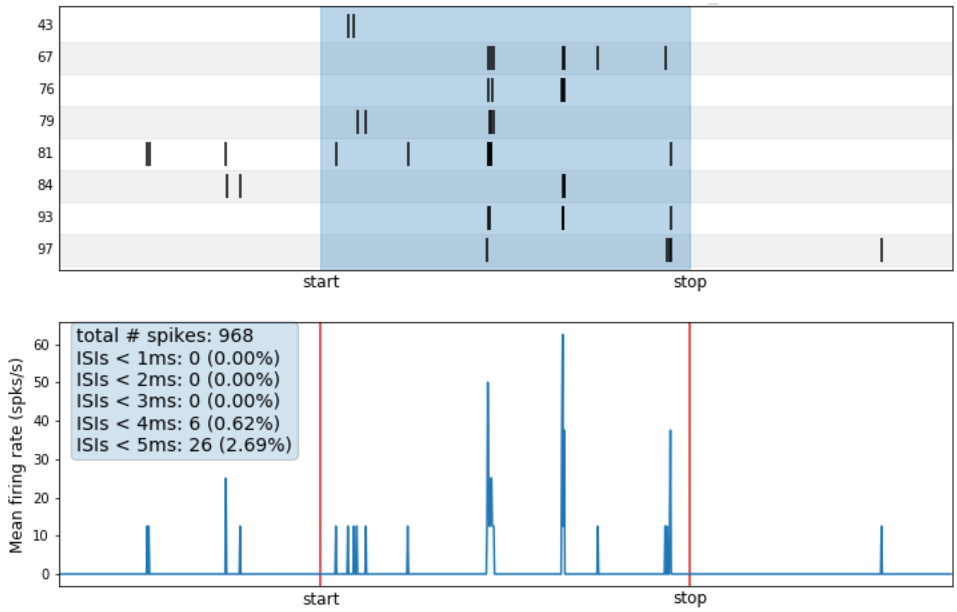
Pooling neuron types across birds, 8 out of 20 (40%) likely projection neurons responded more strongly to DAF playback than to BOS. 17 out of 53 (32%) total likely interneurons also showed response potentiation to DAF playback. Probable projection neurons were less likely to show response potentiation – 26 out of 149 (17%) of all units. Multiunits were the most likely overall to show response suppression, with 354 out of 398 (89%) multiunits having lower firing rates in response to DAF playback. It is possible that these differences reflect different populations of neurons in HVC. A more parsimonious explanation, however, is that, given that more than half of well-isolated units were suppressed by DAF playback, any unit which contained spikes from more

than one neuron was likely to encompass at least one neuron which showed response suppression, even if another neuron contributing spikes to the unit showed response potentiation. It seems reasonable to take the response breakdown of likely projection neurons and interneurons an estimate of the response profile of the population of HVC neurons at large.

The changes in playback response between BOS and DAF were assessed statistically. I used variance in firing rate during playback response to compare response specificities in HVC (Mooney, 2000). For the 8 likely projection neurons across all birds that showed response potentiation under DAF playback, the variance in firing rate during the stimulus increased dramatically, from 5.82 s^{-2} to 9.7 s^{-2} on average, and this effect was statistically significant ($p=0.0039$, Wilcoxon signed-rank test). For the 11 likely projection neurons that showed response suppression under DAF, firing rate variance dropped from 9.23 s^{-2} to 1.99 s^{-2} on average. This effect, too, was statistically significant ($p=0.00049$, Wilcoxon signed-rank test). The response variances for likely interneurons that showed response potentiation (17 units) increased significantly from 45.88 s^{-2} to 74.64 s^{-2} ($p=0.0011$, Wilcoxon signed-rank test), and for likely interneurons that showed response attenuation (31 units) decreased significantly from 50.13 s^{-2} to 20.8 s^{-2} ($p=0.00034$, Wilcoxon signed-rank test).

The structure and arrangement of response peaks for likely projection neurons that showed response potentiation to DAF playback was interesting, but varied across animals. One unit in lilac51 (Figure 4.7) fired one burst per motif in the stimulus, and the average location of these bursts within the stimulus playback period was virtually unchanged under DAF playback (mean time between corresponding peaks in response to BOS and DAF was 0.94 ms, comparable in magnitude to the mean jitter in burst times from trial to trial). Another unit, from lilac52 (Figure 4.8), maintained in response to DAF the single burst per motif it produced in response to BOS, and added a precisely-timed burst to its response, at an offset of 104 ms from the first. A third unit, from lilac17 (Figure 4.9), maintained the BOS response peak locations when responding to DAF, but

displayed increased firing around the peaks, in effect "smearing" them. The half-height width of the peaks produced in response to BOS was 5 ms on average, but the peaks produced in response to DAF had half-height widths of 170 ms on average. Curiously, the peak locations remained consistent across the two stimuli – i.e., the smearing effect involved an increased response of this neuron to DAF earlier in the song than in response to BOS.


43
67
76
79
81
84
93
97
start
stop
total # spikes: 968
ISIs < 1ms: 0 (0.00%)
ISIs < 2ms: 0 (0.00%)
ISIs < 3ms: 0 (0.00%)
ISIs < 4ms: 6 (0.62%)
ISIs < 5ms: 26 (2.69%)
Mean firing rate (spks/s)
0
10
20
30
40
50
60
start
stop

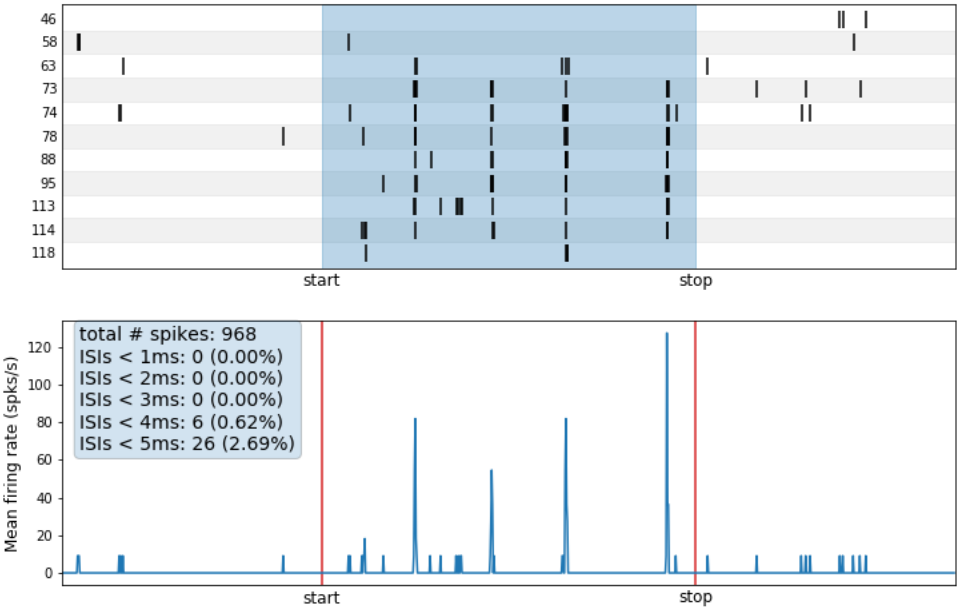
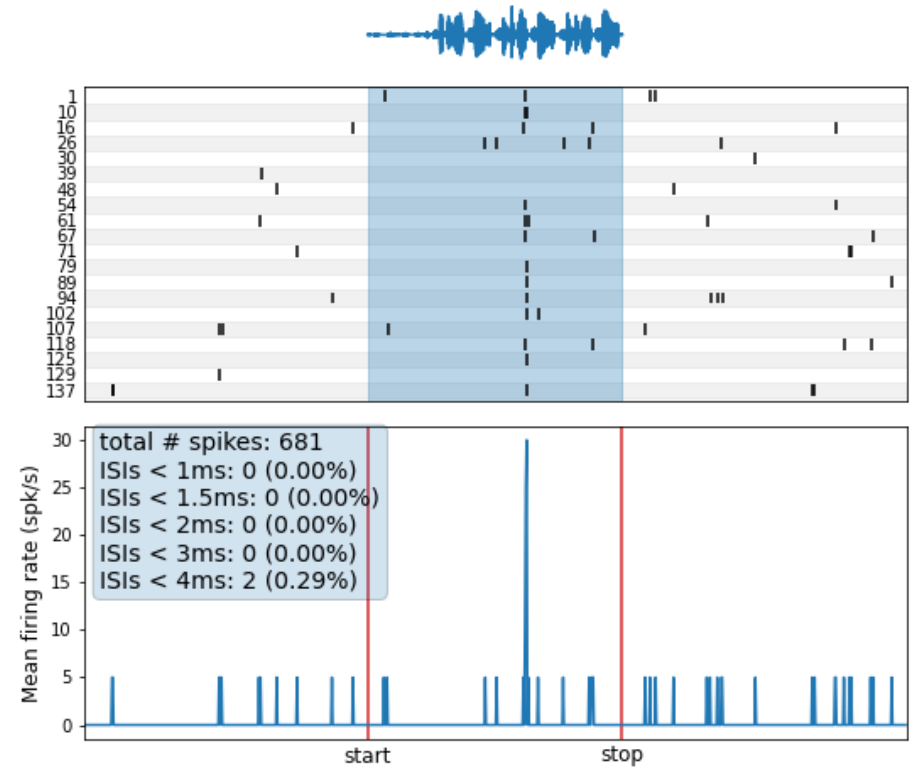

46
58
63
73
74
78
88
95
113
114
118
start
stop
total # spikes: 968
ISIs < 1ms: 0 (0.00%)
ISIs < 2ms: 0 (0.00%)
ISIs < 3ms: 0 (0.00%)
ISIs < 4ms: 6 (0.62%)
ISIs < 5ms: 26 (2.69%)
Mean firing rate (spks/s)
0
20
40
60
80
100
120
start
stop

Figure 4.7. Likely projection neuron potentiation in response to pseudo-DAF. (previous page) This neuron was recorded from lilac51. Top: response to BOS playback. Bottom: response to pseudo-DAF playback. The locations of the peaks in the two PSTHs are nearly identical.


1
10
16
26
30
39
48
54
61
67
71
79
89
94
102
107
118
125
129
137
total # spikes: 681
ISIs < 1ms: 0 (0.00%)
ISIs < 1.5ms: 0 (0.00%)
ISIs < 2ms: 0 (0.00%)
ISIs < 3ms: 0 (0.00%)
ISIs < 4ms: 2 (0.29%)
Mean firing rate (spk/s)
0
5
10
15
20
25
30
start
stop

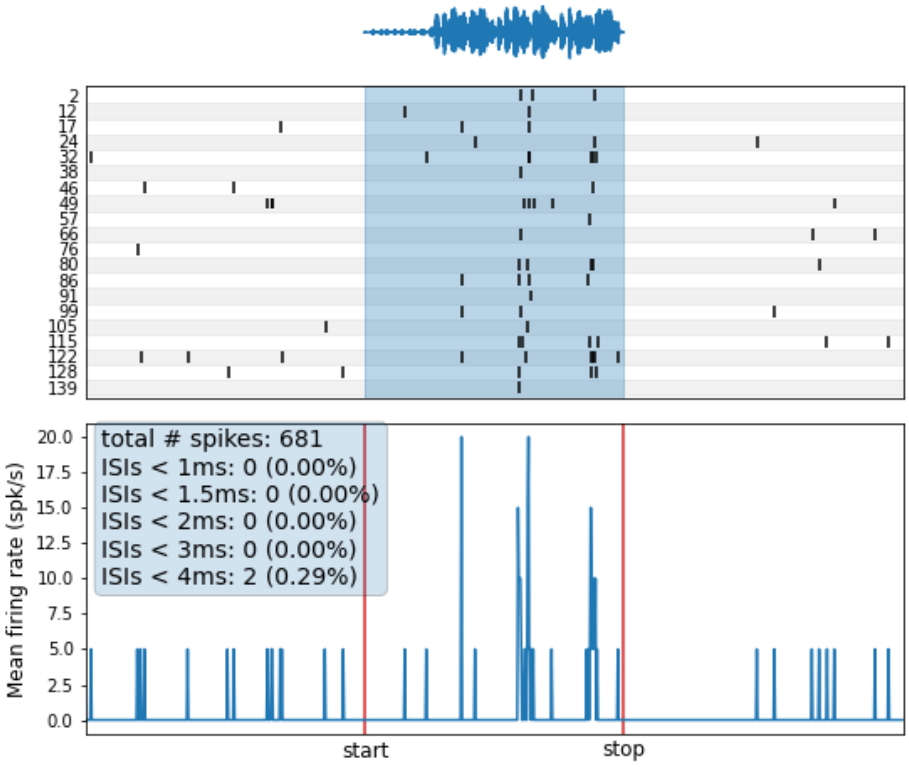
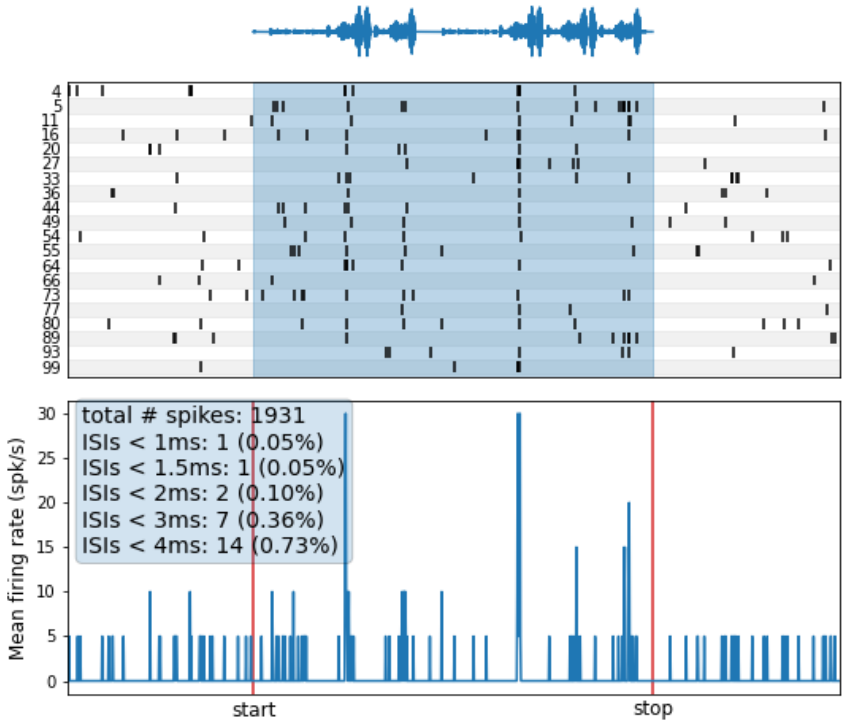

2
12
17
24
32
38
46
49
57
66
76
80
86
91
99
105
115
122
128
139
total # spikes: 681
ISIs < 1ms: 0 (0.00%)
ISIs < 1.5ms: 0 (0.00%)
ISIs < 2ms: 0 (0.00%)
ISIs < 3ms: 0 (0.00%)
ISIs < 4ms: 2 (0.29%)
Mean firing rate (spk/s)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
start
stop

Figure 4.8. Likely projection neuron potentiation in response to pseudo-DAF. (previous page) This neuron was recorded from lilac52. Top: response to BOS playback. Bottom: response to pseudo-DAF playback. The peak in the BOS PSTH appears in the DAF PSTH as well, as the first of a pair of peaks separated by ~100 ms.


4
5
11
16
20
27
33
36
44
49
54
55
64
66
73
77
80
89
93
99
total # spikes: 1931
ISIs < 1ms: 1 (0.05%)
ISIs < 1.5ms: 1 (0.05%)
ISIs < 2ms: 2 (0.10%)
ISIs < 3ms: 7 (0.36%)
ISIs < 4ms: 14 (0.73%)
Mean firing rate (spk/s)
0
5
10
15
20
25
30
start
stop

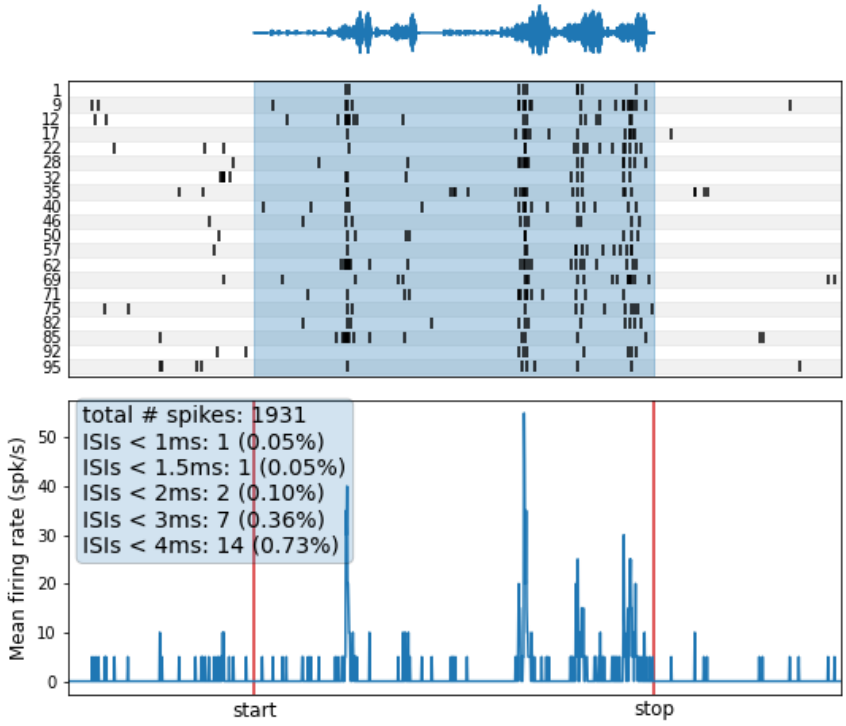

1
9
12
17
22
28
32
35
40
46
50
57
62
69
71
75
82
85
92
95
total # spikes: 1931
ISIs < 1ms: 1 (0.05%)
ISIs < 1.5ms: 1 (0.05%)
ISIs < 2ms: 2 (0.10%)
ISIs < 3ms: 7 (0.36%)
ISIs < 4ms: 14 (0.73%)
Mean firing rate (spk/s)
0
10
20
30
40
50
start
stop

Figure 4.9. Likely projection neuron potentiation in response to pseudo-DAF. (previous page) This neuron was recorded from lilac17. Top: response to BOS playback. Bottom: response to pseudo-DAF playback. The peaks in the BOS PSTH appear in the DAF PSTH as well, but they also gain increased response before and after the peak.

Pseudo-DAF responses cannot be explained by stimulus amplitude

One possible explanation for a neuron to respond more strongly to DAF playback is the higher amplitude of the stimulus – it is louder than the equivalent BOS, typically by 2-3 dB (the precise value will depend on average amplitudes of the stimulus and its time-shifted echo. (I deliberately did not attenuate the DAF stimulus to match that of the BOS stimulus.)

To test this possibility, for one bird (lilac59) I produced a version of BOS with maximum amplitude scaled to that of the DAF stimulus for the bird and included it in the stimulus set. As expected, the increase in stimulus amplitude of this "DAF-matched BOS" produced a small increase in response over BOS (Margoliash, 1986) (firing rate variance during playback increased for most likely projection neurons, likely interneurons, and probable projection neurons, from 20.32 s^{-2} to 23.7 s^{-2}, and this result was statistically significant ($p=0.0097$, Wilcoxon signed-rank test)), but this was lower than the response to DAF itself in neurons which responded more strongly to DAF than to BOS, and was not accompanied by appreciable suppression of response for any neurons (Figures 4.10-4.12; this unit's responses are representative of those from lilac59 which showed potentiation of response under pseudo-DAF). It is therefore unlikely that the effect observed for DAF playback can be attributed to stimulus amplitude.

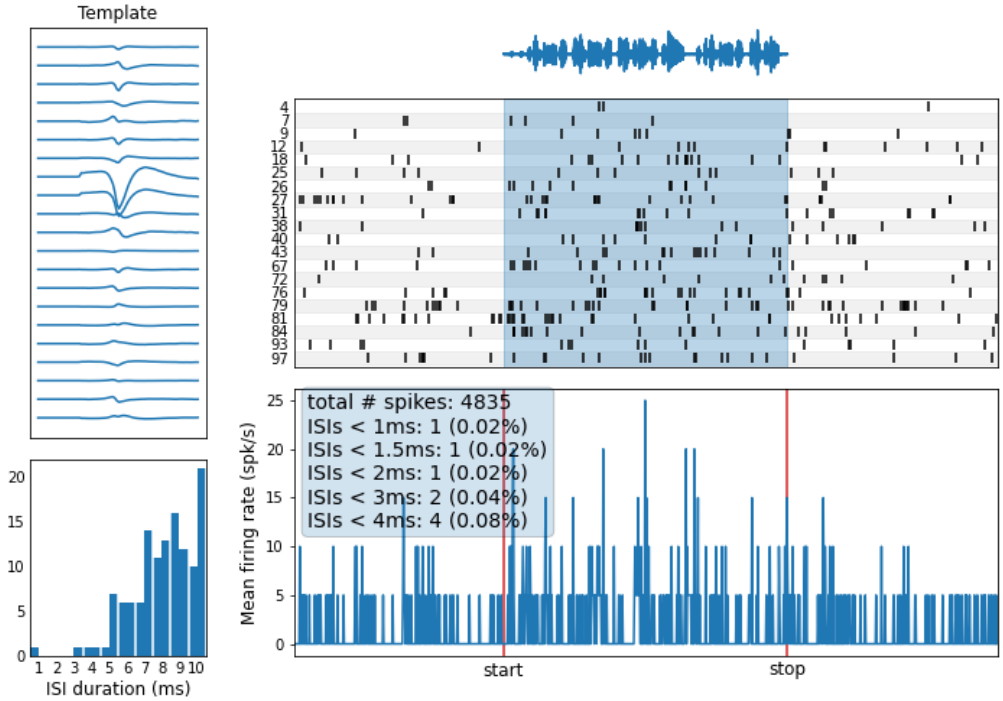


Figure 4.10. Likely interneuron response to BOS. Compare the somewhat weaker response of this unit in this figure to its response to BOS amplitude-matched to DAF in Figure 4.11, and then to pseudo-DAF in Figure 4.12.

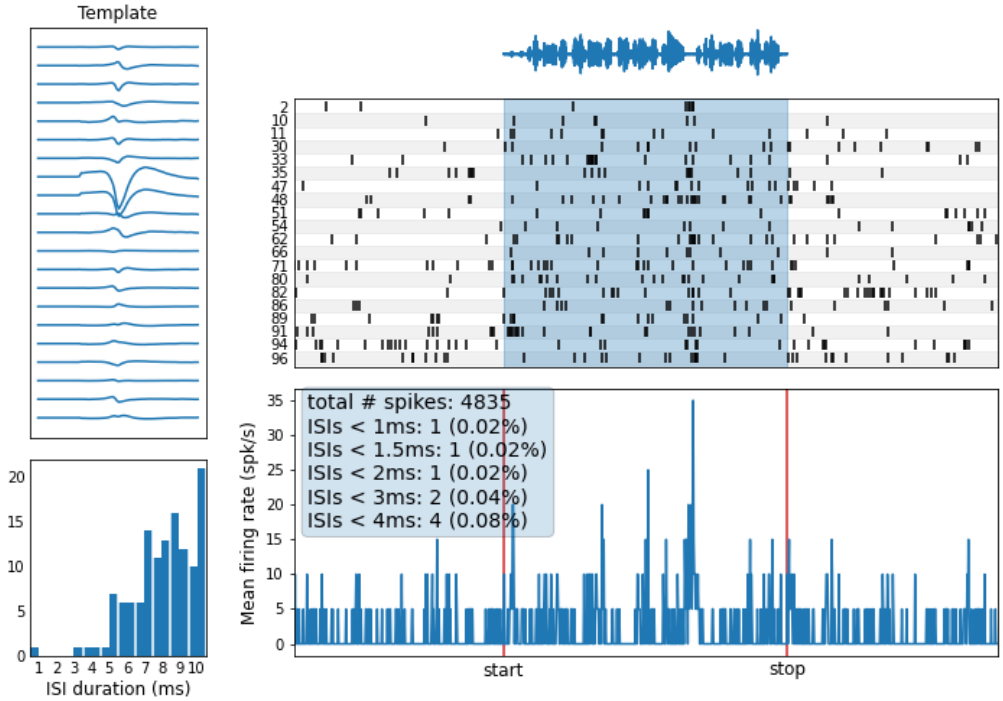


Figure 4.11. Likely interneuron response to BOS amplitude-matched to DAF. The response of this unit to amplitude-matched BOS is stronger than that to unmodified BOS (Figure 4.10), but lower than that to pseudo-DAF (Figure 4.12).

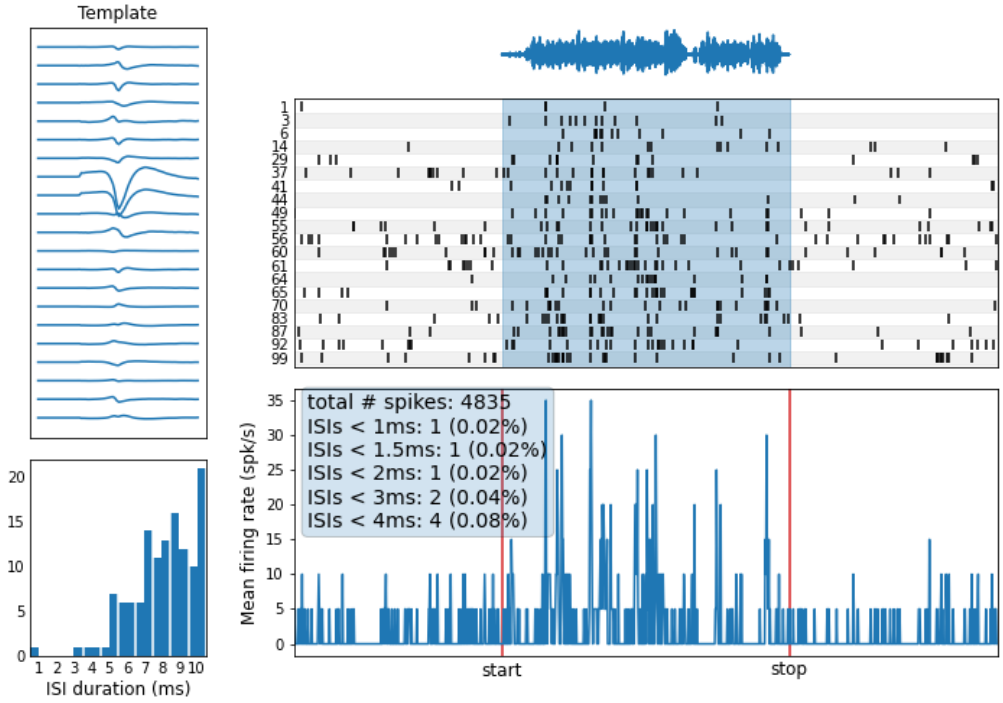


Figure 4.12. Likely interneuron response to pseudo-DAF. Compare the much stronger response of this unit to pseudo-DAF with the weaker response to BOS amplitude-matched to DAF in Figure 4.11, and its still weaker response to unmodified BOS in Figure 4.10.

Pseudo-DAF responses are not due to arbitrary sound overlay

Another possible explanation for the effect I observed is that simply contaminating BOS playback with any other signal will produce abnormal responses in HVC. This is a more remote possibility, for two reasons. First as discussed above, prior playback response work has included manipulations of BOS, and the effect has consistently been a suppressed neuronal response – many neurons in HVC appear to be extremely specifically tuned to a bird's own song (which is not to say they may not also be activated by other stimuli, including a given bird's tutor's song during development (Nick and Konishi, 2004) and in adulthood (Prather et al, 2010)). Second, zebra finches naturally live gregariously, with many bonded pairs and their offspring in close proximity,

thus song overlap in the cacophony of the colony is unavoidable (Zann, 1996). While males will, when observed in pairs, attempt to avoid singing over one another, and both males and females will modulate their call timing relative to another bird's vocalization (Benichov et al, 2016; Benichov and Vallentin, 2020), in the wild strict turn-taking is not possible. It is thus unlikely that HVC activity in zebra finches is extensively disrupted by the presence of any arbitrary external noise while they sing.

Nevertheless, I proceeded to test this possibility by another playback manipulation. For two birds, I constructed stimuli consisting of each bird's BOS, and layered over it, at a 100 ms delay, the song of a conspecific bird (not closely related to the bird in question, and with a song not obviously similar). I observed a general suppression of response to this stimulus, as expected (a representative unit can be seen in Figure 4.13), sometimes accompanied by onset responses when the first overlaid motif begins (these onset responses rapidly extinguish as the stimulus progresses). I quantified the suppression of response of these units again (n=25 pooled likely projection neurons, likely interneurons, and probable projection neurons) using the response variance, which decreased significantly from a mean of 37.05 s^{-2} to a mean of 22.68 s^{-2} ($p=0.00761$, Wilcoxon signed-rank test). It therefore seems unlikely that the choice of sound overlaid onto BOS is unimportant to produce responses like those to DAF playback.

Note that this stimulus type exists in a large parameter space, however, and can be as varied as any zebra finch song, as the conspecific song layered over BOS could be chosen arbitrarily, with any arrangement of temporal pattern and any set of syllables. I did not attempt to explore this large parameter space, as an exhaustive search would have required infeasible resources. It is entirely possible that for a given BOS a conspecific song could be found that produces responses in HVC that mirror or rival those for DAF. The existence of such a song would be intriguing, and a comparison of its features with the BOS and pseudo-DAF in question would likely be fruitful.

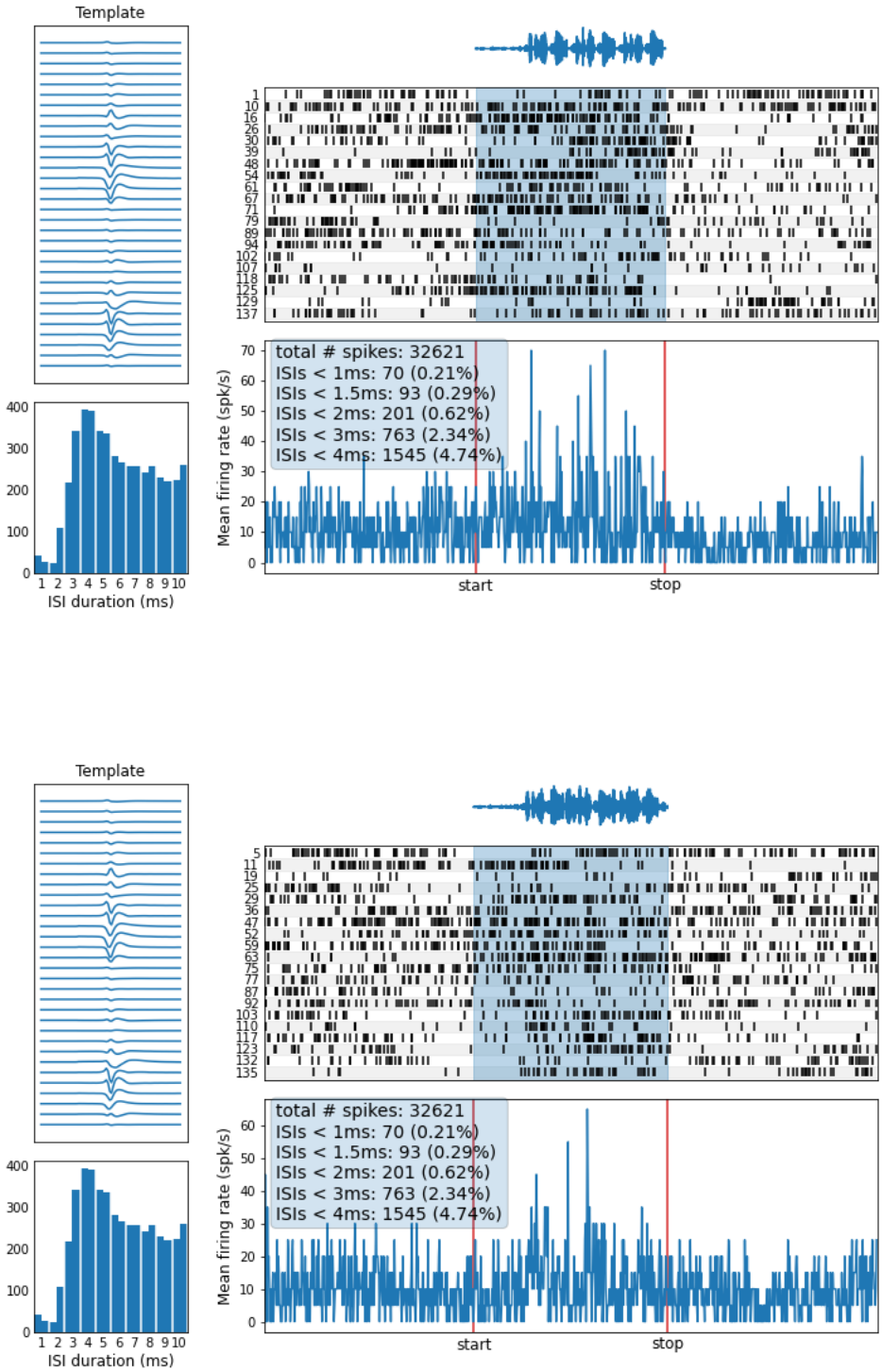


Figure 4.13. Suppression of BOS response by overlaying a conspecific song. Unit taken from lilac59.

Discussion

As mentioned above, auditory stimuli which induce responses in HVC equivalent to or greater than those induced by BOS playback are very rare. Extensive modifications of BOS have been presented to birds, and the consistent pattern of outcomes has been attenuated responses. That simply echoing BOS onto itself at a short delay would consistently produce significantly stronger responses in even a subset of HVC neurons is surprising.

It is consistent, however, with one primary outcome of the work in Chapter 2. Most attempts at behavioral manipulation in zebra finches, such as deafening or white noise playback during singing, have been reported to take at the very least days, and often weeks or more, to produce observable changes in singing. DAF, however, can provoke behavioral changes within no more than hours, and by some measures nearly immediately.

A simple unidirectional response to pseudo-DAF playback would have been more equivocal. A general suppression of response would have been entirely in line with prior work on BOS playback – DAF would simply be another less effective stimulus. Such an outcome might have suggested that HVC is not primarily engaged in responding to artificially aberrant auditory feedback, or that a playback experiment is ill-suited to investigate what role it might have. A general increase in response to DAF might have been relatively easily explained as a stimulus amplitude effect, again somewhat inconclusive as to whether HVC activity is important to process unexpected auditory feedback. This could have been evaluated with a more extensive analysis of the neuronal sensitivity to changes in BOS amplitude, which could be a useful future experiment. A network being activated as normal but at a higher intensity could have little effect on its outputs. But a response such as that reported here, in which different parts of the HVC network – indeed, different portions of both local inhibitory and local and projecting excitatory populations – are differentially affected by

pseudo-DAF playback suggests that this stimulus is interfering with normal activity within HVC. Further investigation of this effect is necessary.

Analysis of the behavior of firing rate peaks (corresponding to reliable burst time locations) of likely projection neurons revealed a variety of responses, including a simple increase in burst reliability (producing a higher mean firing rate across trials), the addition of a subsequent peak (at a delay very close to that used in producing the pseudo-DAF stimulus – that this could be a coincidental outcome would appear unlikely), and the addition of spikes in response to pseudo-DAF both before and after the neuron's precise response peak. These latter two responses, in particular, tie plausibly into a model of HVC activity encompassing comparisons between expected and actual sensory feedback, such as that proposed in prior work in our lab (e.g., Daou and Margoliash, 2020) and in Chapter 3. This will be expanded upon in Chapter 5.

CHAPTER 5

GENERAL DISCUSSION

Summary of results and implications

I have shown, first, that manipulation of acoustic feedback produces behavioral changes in song in the zebra finch, potentially nearly immediately. These changes are related to motor preparatory activity – the introductory notes preceding the beginning of song – and appear to reflect the disruption of the evolution of the neuronal system involved in singing towards a state in which singing can begin. The appearance of zero motif bouts suggests that when this shift in neuronal network state is disrupted, singing cannot begin. The increased number of introductory notes is consistent with the idea that introductory notes are preparatory in function, as more are needed to overcome the challenge that DAF poses to the song system network. And the lengthening of the periods of silence between introductory notes suggests that a specific interval between them is key to their preparatory function – and that the interval chosen here for DAF (100 ms) may determine the degree to which DAF interferes with this activity. Ultimately, as well, these results argue (as have others in the past) unambiguously for a role for auditory feedback in the production of song in the zebra finch, despite prior theories otherwise. They also argue, though perhaps less strongly, that auditory feedback during the introductory note sequence is actively used in modulating the progression of the sequence. A role for feedback during introductory note production runs contrary to one theory of ballistic production of introductory notes; this contradiction is likely due to the different types of feedback probed in this study and that one (Rao et al, 2019).

I then demonstrated that basal ganglia-projecting neurons in HVC produce bursts which comprise sequences during singing that appear to act as copies of one another, allowing

information about neuronal activity in the past to be projected forward into the future. Over the course of a single song, multiple sequence copies are produced at a range of offsets from the first burst sequence, even for very short songs, implying that the same information is used multiple times over the course of the song. This would provide an elegant mechanism to explain the long integration times reported for some song-specific neurons in HVC (Margoliash and Fortune, 1992). Another possible function for this phenomenon is to allow comparison between a prior motor command (i.e., an efference copy) and the sensory feedback arising from that motor command. This would situate the recognition and calculation of performance errors, a function critical for the learning and maintenance of complex behavior, at least partially in HVC, rather than entirely downstream of it within the anterior forebrain pathway (AFP). These results also indicate that these sequences of recurring activity may be specific to a given bird. Given that the song a specific bird sings constitutes a major behavioral feature differentiating him from other birds, these recurring sequences may reflect both constraints on network structure as a consequence of the song learned as well as the learning trajectory taken during development.

Finally, I demonstrated that neurons in HVC, both interneurons and projection neurons, respond differentially to stimuli resembling altered auditory feedback (pseudo-DAF). These altered responses involve both suppressed firing as well as potentiated activity. This may be the first report of auditory stimuli that produce greater responses in HVC than a bird's own song. The varied nature of the responses to pseudo-DAF, rather than simple suppression, suggests that these stimuli are perturbing a complex network including HVC rather than simply representing a lack of recognition of an unusual stimulus. Such a dramatic response may help to explain how DAF exposure during singing can produce immediate effects, when other auditory feedback perturbations have not been shown to produce effects until much more time has passed. It is possible that the changes in introductory note sequence production reported in Chapter 2 are related to these unusual

responses of HVC neurons to pseudo-DAF: alterations to the introductory note sequence, whether by adding more introductory notes or spacing them farther apart, could represent an attempt to bring the song system into a state in which auditory feedback, in this case DAF, gives stable (and therefore predictable) responses that will not trigger unusual HVC activity. That is, the introductory note sequence serves as a probe and corrective for the state of the song system and motor periphery at song onset, and it is the continued failed attempts to correct for DAF exposure (which cannot be corrected or "escaped" by behavioral modification) that produce the changes in introductory note production that I reported in Chapter 2. The recurring sequences reported in Chapter 3 provide a potential mechanism by which perceived errors in behavioral output (such as DAF) are recognized in HVC and propagated onwards through the rest of the song system. A suppressed response to auditory feedback will leave gaps in one or more of these sequences, allowing precise targeting of correction. (This hypothesis is agnostic to how the neurons comprising such sequences are connected, functionally or synaptically.) A potentiated response to auditory feedback will also produce extraneous activity in one or more of these sequences, and possibly induce entirely abnormal sequences to occur.

Taken together, the results discussed here make a case for the importance of auditory feedback to ongoing song production in the zebra finch, demonstrate likely features of neuronal responses to vocal production errors, and suggest the beginning of a network-level mechanism (informed by cellular circuitry) by which errors (possibly extending beyond auditory feedback) are recognized and propagated within HVC.

Future directions

Use of other species

It seems to be customary for a birdsong thesis discussion to reiterate the importance of doing more work in species other than the zebra finch, and to lament the extent to which electrophysiological studies have focused on this species which has been found time and again to be rather unusual among its fellows. I cannot help but join in this tradition. Zebra finches are without doubt immensely useful for the study of birdsong – the regularity of their motif alone is a point of such convenience, allowing as it does the segmentation of a potentially immensely variable behavior into a well-defined trial-like structure, thus rendering it amenable to traditional electrophysiological signal averaging (not to mention semi-automated behavioral labeling), that it should hardly be surprising to find that so much of what is known about the learning and production of birdsong arises from this single species.

Yet zebra finches represent a particular adaptation, so represent a limited range of diversity within the songbird clade, and an extreme adaptive form in their high degree of sexual dimorphism of singing behavior and brain anatomy as compared to most other songbird species. This specialization is expressed in features of zebra finch singing, which necessarily limit the type of experiment that can be performed, and therefore the universe of answers that are possible to obtain. To take as a single example, the idea that each projection neuron burst within HVC represents a moment in time within the motif (the fundamental basis of the "clock" model) would not be an unreasonable supposition to make given sufficient data after noting the remarkable sparsity of projection neuron bursting during song. But the zebra finch's stereotyped song prohibits falsifying some of the obvious hypotheses that might grow out of this supposition – e.g., (1) in a song with repeated elements, the same projection neuron (let it be an HVC_{RA} neuron) does not burst more

than once, or (2) in a song with a stochastic structure, a projection neuron will burst at the same time within song regardless of which branch of the variable sequence is being followed. A dynamical system which follows a non-looping trajectory is necessarily difficult to distinguish from a ballistic controller. These questions could be readily addressed in any number of songbird species – including in the Bengalese finch, which produces song with both repeated elements and nondeterministic syllable transitions (Honda and Okanoya, 1999).

(There are, as I have discussed, many other reasons to doubt this model's validity, but even were none of those other objections valid, this singular focus on the zebra finch would still cast doubt on the usefulness of the clock model. If the zebra finch is unusual within the songbird clade, it would be possible that the clock model only held in the zebra finch – and, despite the charm that the species possesses, there is little of broad interest to be found in producing an exquisitely refined model of song production in a single species if the model does not extend even so far as other songbirds, much less across the phylogenetic gulf separating the zebra finch and the human.)

Additionally, the spectral complexity of zebra finch song presents challenges to the experimenter. One such challenge is in identifying correspondences between acoustic output and electrophysiological measurements, both in muscle and brain. Another more pragmatic concern is the difficulty in producing automated labeling algorithms for vocalizations with complex spectral structure. Closely related to this point is the additional difficulty of labeling vocalizations in real time, an absolutely critical requirement for the more complex perturbation schemes that are called for to properly study sensory feedback responses in the brain. A species which sings a spectrally simpler song would address all of these issues. In contrast, in zebra finches almost all experiments that attempt real time analysis of song focus on the spectrally simple harmonic stacks. One fervently hopes – but it is just hope – that the field has not unwittingly restricted itself by exploring features of song production limited to harmonic stacks or the immediately following syllable.

Kyler Brown (Brown, 2017) proposed two candidate species, the canary (*Serinus canaria*), and the brown thrasher (*Toxostoma rufum*), and his arguments are persuasive. Simply repeating some of my experiments in one or the other of these species would provide insight. In the brown thrasher, which commonly sings two copies of a syllable in a row, continuous DAF should be particularly illuminating. Modulating the feedback delay (there is, of course, no reason to assume that the 100 ms delay found to be effective in the zebra finch will generalize to other songbird species) to present the echo of the first syllable copy during production of the second copy, while recording HVC activity (this could also be done with pseudo-DAF playback) would allow testing of hypotheses regarding reactivation of multiburst neurons. Such work would additionally represent the first recordings of HVC neurons in the brown thrasher, and an increase in the number of species examined in the context of HVC activity can only be to the good. There have been recordings made in HVC of canaries, some of which have provided evidence for encoding of the history (seconds into the past) of a branching syllable sequence in HVC activity (Cohen et al, 2020), which extends the results on temporal combination selective (TCS) cells in zebra finch HVC, which may have integration times of several seconds (Margoliash and Fortune, 1992), into the canary. Other work in canaries has reported reflection of vocal rhythms in aggregate HVC activity (Boari et al, 2022). These results are intriguing, and represent only the beginning of what might be discovered by applying the same focus to another species as has been granted the zebra finch up to this point.

The future of the clock model

In demonstrating regularity within HVC_X burst timing data, these results constitute yet another challenge to a view of HVC activity during singing that is restricted to timekeeping. From a strictly analytical point of view, lack of independence of the HVC_X bursts in a given bird, together with logical errors in prominent papers advocating for the clock model, suggest that further attempts to

lump HVC_{RA} and HVC_X neurons together for the purposes of population analysis would be unwise at best.

None of these results preclude the possibility that HVC_{RA} neurons keep time in some way, though many other results, as discussed previously, suggest that even if they do, every single moment within the motif is not represented equivalently. The methodological difficulties of recording from large numbers of HVC_{RA} neurons at once do mean, however, that it will likely be challenging to prove that this is the case. Certainly an approach that must be abandoned is that of arguing for uniformity or continuity of burst timing distributions simply because real data cannot be distinguished from a uniform distribution. This line of argument is entirely unsound, as it relies on a misapplication of the logic of frequentist statistical testing to argue for the truth of the null hypothesis from the failure to reject it. The deficiencies of such analyses will only become more glaring if they are confined to smaller numbers of burst times (i.e., once HVC_X are excluded from consideration, as they must now be).

What is needed is a strong hypothesis arising from the clock model – and none have so far been offered. Such a hypothesis should be restricted to HVC_{RA} neurons, and will need to make a clear positive claim about their behavior during some behavioral state, or under a perturbation regime (such as DAF, or conditioned auditory feedback, or some other such manipulation). Some possible hypotheses might include the following: the appearance of additional HVC projection neuron bursts as a result of DAF exposure will produce the same combination of bursts in RA recipient neurons in response to the new burst as to that neuron's normal burst; or, playback of pseudo-DAF with the echo restricted to one syllable will abolish not only projection neuron bursts within that syllable but also projection neuron bursts after that syllable. These are only examples, but they are moving in the right direction. This effort may not prove fruitful, however. Recent results that at least one of the two major classes of zebra finch HVC_{RA} have intrinsic properties that vary among

birds but are similar within a bird (A. Daou, unpublished results) suggest similar aspects of functional organization for HVC_{RA} as has been demonstrated for HVC_X. This possibility may doom the representation of time within HVC_{RA} neurons just as it is unlikely within HVC_X neurons.

Further DAF experiments

The immediate and obvious extension of the work reported here is the recording of HVC neuronal signals from birds as they are singing and exposed to DAF. Responses to auditory playback during sleep throughout the song system are similar in structure but generally weaker than those produced during singing itself. Such an experiment would allow examining real time changes in any class of HVC neurons, but changes in the HVC_X are a likely first target. Beyond simple demonstration of the fact, it may provide compelling insight constraining how HVC networks rapidly adapt to abnormal feedback. One possibility is that HVC auditory responses are gated "on" by exposure to abnormal feedback. The release of auditory activity in the song system at the transition from wake to sleep has been described (Dave et al, 1998; Schmidt and Konishi, 1998). This would also reconcile results here and elsewhere on activity in HVC_X neurons with the observation that their ablation in adult birds produces few appreciable changes in singing behavior (Scharff et al, 2000). If it is only their abnormal activity during singing that produces noticeable changes in singing, then this result is entirely consistent with the lesion data – a lesioned neuron will never have the opportunity to produce abnormal activity to drive singing changes.

Another set of experiments involves modifying the pseudo-DAF stimulus. Of highest interest is the outcome of presenting to the bird recordings of BOS with only some portions echoed. These stimuli could be produced quickly while neuronal responses to unmodified BOS are available, allowing the tailoring of the pseudo-DAF stimulus to maximally affect a given neuron or set of neurons being recorded, probing the correspondence between neuron and behavior directly.

REFERENCES

Afshar A, Santhanam G, Yu BM, Ryu SI, Sahani M, Shenoy KV (2011) Single-trial neural correlates of arm movement preparation. Neuron 71:555–564.

Aickin M, Gensler H (1996) Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. Am J Public Health 86:726–728.

Akutagawa E, Konishi M (2010) New brain pathways found in the vocal control system of a songbird. J Comp Neurol 518:3086-3100.

Alonso LM, Alliende JA, Mindlin GB (2010) Dynamical origin of complex motor patterns. Nonlinear Dyn 70:943-960.

Alonso RG, Trevisan MA, Amador A, Goller F, Mindlin GB (2015) A circular model for song motor control in *Serinus canaria*. Front Comput Neurosci 9:41.

Amador A, Goller F, Mindlin GB (2008) Frequency modulation during song in a suboscine does not require vocal muscles. J Neurophysiol 99:2383-2389.

Amador A, Perl YS, Mindlin GB, Margoliash D (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. Nature 495:59–64.

Ames KC, Ryu SI, Shenoy KV (2019) Simultaneous motor preparation and execution in a last-moment reach correction task. Nat Commun 10:1–13.

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.

Aronov D, Andalman AS, Fee MS (2008) A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science 320:630-634.

Arriaga G, Jarvis ED (2013) Mouse vocal communication system: are ultrasounds learned or innate? Brain Lang 124:96-116.

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.

Basista MJ, Elliott KC, Wu W, Hyson RL, Bertram R, Johnson F (2014) Independent premotor encoding of the sequence and structure of birdsong in avian cortex. J Neurosci 34:16821-16834.

Beecher, WJ (1953) A phylogeny of the oscines. The Auk 70:270-333.

Benezra SE, Narayanan RT, Egger R, Oberlaender M, Long MA (2018) Morphological characterization of HVC projection neurons in the zebra finch (*Taeniopygia guttata*). J Comp

Neurol 526:1673-1689.

Benichov JI, Benezra SE, Vallentin D, Globerson E, Long MA, Tchernichovski O (2016) The forebrain song system mediates predictive call timing in female and male zebra finches. Curr Biol 26:309-318.

Benichov JI, Vallentin D (2020) Inhibition within a premotor circuit controls the timing of vocal turn-taking in zebra finches. Nat Commun 11:221.

Boari S, Mindlin GB, Amador A (2022) Neural oscillations are locked to birdsong rhythms in canaries. Eur J Neurosci 55:549-565.

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.

Bouisset S, Zattara M (1981) A sequence of postural movements precedes voluntary movement. Neurosci Lett 22:263–270.

Bradbury J. Vocal Communication in Wild Parrots. In: de Waal F, Tyack P (ed.) Animal Social Complexity: Intelligence, Culture, and Individualized Societies. 2013:293-316.

Bradshaw J, Cameron-Beaumont C. The signalling repertoire of the domestic cat and its undomesticated relatives. In: The domestic cat: The biology of its behaviour. 2000:67-93.

Briefer EF, McElligott AG (2012) Social effects on vocal ontogeny in an ungulate, the goat, *Capra hircus*. Anim Behav 83:991-1000.

Briscoe SD, Albertin CB, Rowell JJ, Ragsdale CW (2018) Neocortical association cell types in the forebrain of birds and alligators. Curr Biol 28:686-696.

Brocklehurst RJ, Schachner ER, Codd JR, Sellers WI (2020) Respiratory evolution in archosaurs. Phil Trans R Soc B 375:20190140.

Brown KJ (2017) "Vocal motor coding in a songbird premotor nucleus." PhD diss. University of Chicago.

Bruno JH, Tchernichovski O (2017) Regularities in zebra finch song beyond the repeated motif. Behav Processes 163:53-59.

Cheney DL, Seyfarth RM (1988) Assessment of meaning and the detection of unreliable signals by vervet monkeys. Anim Behav 36:477-486.

Churchland MM, Yu BM, Ryu SI, Santhanam G, Shenoy KV (2006) Neural variability in premotor cortex provides a signature of motor preparation. J Neurosci 26:3697–3712.

Churchland MM, Shenoy KV (2007) Delay of movement caused by disruption of cortical preparatory activity. J Neurophysiol 97:348–359.

Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, et al (2012) Neural population dynamics during reaching. Nature 487:51–56.

Cohen Y, Shen J, Semu D, Leman DP, Liberti WA, Perkins LN, et al (2020) Hidden neural states underlie canary song syntax. Nature 582:539-544.

Coleman MJ, Day NF, Rivera-Parra P, Fortune ES (2021) Neurophysiological coordination of duet singing. Proc Natl Acad Sci U S A 118:e2018188118.

Cynx, J. (1990). Experimental determination of a unit of song production in the zebra finch (Taeniopygia guttata). J Comp Psychol 104:3–10.

Cynx J, von Rad U (2001) Immediate and transitory effects of delayed auditory feedback on bird song production. Anim Behav 62:305–312.

Daliparthi VK, Tachibana RO, Cooper BG, Hahnloser RHR, Kojima S, Sober SJ, et al (2019) Transitioning between preparatory and precisely sequenced neuronal activity in production of a skilled behavior. eLife 8:e43732.

Danish HH, Aronov D, Fee MS (2017) Rhythmic syllable-related activity in a songbird motor thalamic nucleus necessary for learned vocalizations. PLoS ONE 12:e0169568.

Daou A, Margoliash D (2020) Intrinsic neuronal properties represent song and error in zebra finch vocal learning. Nat Commun 11:952.

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.

Dave AS, Margoliash D (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. Science 290:812-816.

Dave AS, Yu AC, Margoliash D (1998) Behavioral state modulation of auditory activity in a vocal motor system. Science 282:2250-2254.

Day BL, Rothwell JC, Thompson PD, Maertens de Noordhout A, Nakashima K, Shannon K, et al (1989) Delay in the execution of voluntary movement by electrical or magnetic brain stimulation in intact man: evidence for the storage of motor programs in the brain. Brain 112:649-663.

Del Negro C, Lehongre K, Edeline JM (2005) Selectivity of canary HVC neurons for the bird's own song: modulation by photoperiodic conditions. J Neurosci 25:4952-4963.

DeVoogd TJ, Nottebohm F (1981) Sex differences in dendritic morphology of a song control nucleus in the canary: a quantitative Golgi study. J Comp Neurol 196:309-316.

Diener HC, Dichgans J, Guschlbauer B, Bacher M, Langenbach P (1989) Disturbances of motor

preparation in basal ganglia and cerebellar disorders. Prog Brain Res 80:481-488.

Doupe AJ, Konishi M (1991) Song-selective auditory circuits in the vocal control system of the zebra finch. Proc Natl Acad Sci U S A 88:11339-11343.

Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. Ann Rev Neurosci 22:567–631.

Dugas-Ford J, Rowell JJ, Ragsdale CW (2012) Cell-type homologies and the origins of the neocortex. Proc Natl Acad Sci U S A 109:16974-16979.

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.

Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, et al (2007) The near eastern origin of cat domestication. Science 317:519-523.

Elmaleh M, Kranz D, Asensio AC, Moll FW, Long MA (2021) Sleep replay reveals premotor circuit structure for a skilled behavior. Neuron 109:3851-3861.

Ericson PGP, Johansson US, Parsons TJ (2000) Major divisions in oscines revealed by insertions in the nuclear gene c-*myc*: a novel gene in avian phylogenetics. The Auk 117:1069-1078.

Evarts EV (1968) Relation of pyramidal tract activity to force exerted during voluntary movement. J Neurophysiol 31:14-27.

Fainstein F, Geli SM, Amador A, Goller F, Mindlin GB (2021) Birds breathe at an aerodynamic resonance. Chaos 31:123132.

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.

Fortune ES, Margoliash D (1995) Parallel pathways and convergence onto HVc and adjacent neostriatum of adult zebra finches (*Taeniopygia guttata*). J Comp Neurol 360:413-441.

Fortune ES, Rodríguez C, Li D, Ball GF, Coleman MJ (2011) Neural mechanisms for the coordination of duet singing in wrens. Science 334:666-670.

Foster EF, Bottjer SW (1998) Axonal connections of the high vocal center and surrounding cortical regions in juvenile and adult zebra finches. J Comp Neurol 397:118-138.

Foster EF, Mehta RP, Bottjer SW (1997) Axonal connections of the medial magnocellular nucleus of the anterior neostriatum in zebra finches. J Comp Neurol 382:364-381.

Franz M, Goller F (2002) Respiratory units of motor production and song imitation in the zebra finch. J Neurobiol 51:129–141.

Fripp D, Owen C, Quinata-Rizzo E, Shapiro A, Buckstaff K, Jankowski K, et al (2005) Bottlenose dolphin (*Tursiops truncatus*) calves appear to model their signature whistles on the signature whistles of community members. Anim Cogn 8:17-26.

Fukushima M (2008) "Feedback learning in a sensorimotor system." PhD diss. University of Chicago.

Fukushima M, Margoliash D (2015) The effects of delayed auditory feedback revealed by bone conduction microphone in adult zebra finches. Sci Rep 5:8800.

Gale SD, Perkel DJ (2010) A basal ganglia pathway drives selective auditory responses in songbird dopaminergic neurons via disinhibition. J Neurosci 20:1027-1037.

Georgopoulos AP, Kalaska JF, Caminitii R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J Neurosci 2:1527-1537.

Gil D, Gahr M (2002) The honesty of bird song: multiple constraints for multiple traits. Trends Ecol Evol 17:133–141.

Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: implications for motor coding. J Neurosci 26:991-1005.

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.

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 (2016) Sound production and modification in birds – mechanisms, methodology and open questions. Comparative Bioacoustics: An Overview (C. Brown and T. Riede, Editors). Bentham Science Publishers, Sharjah, United Arab Emirates, 165-230.

Goller F, Suthers RA (1996) Role of syringeal muscles in gating airflow and sound production in singing brown thrashers. J Neurophysiol 75:867-876.

Goller F, Suthers RA (1999) Bilaterally symmetrical respiratory activity during lateralized birdsong. J Neurobiol 41:513-523.

Hahnloser RHR, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature 419:65-70.

Hahnloser RHR, Kozhevnikov AA, Fee MS (2007) Sleep-related neural activity in a premotor and a basal-ganglia pathway of the songbird. J Neurophysiol 96:794-812.

Hamaguchi K, Tanaka M, Mooney R (2016) A distributed recurrent network contributes to temporally precise vocalizations. Neuron 91:680-693.

Hartley RS, Suthers RA (1989) Airflow and pressure during canary song: direct evidence for mini-breaths. J Comp Physiol A 165:15-26.

Herbert CT, Boari S, Mindlin GB, Amador A (2020) Dynamical model for the neural activity of singing *Serinus canaria*. Chaos 30:053134.

Herrmann M, Hertz JA, Prügel-Bennett A (1995) Analysis of synfire chains. Network: Computation in Neural Systems 6:403-414.

Honda E, Okanoya K (1999) Acoustical and syntactical comparisons between songs of the white-backed munia (*Lonchura striata*) and its domesticated strain, the Bengalese finch (*Lonchura striata* var. *domestica*). Zoolog Sci 16:319-326.

Janata P, Margoliash D (1999) Gradual emergence of song selectivity in sensorimotor structures of the male zebra finch song system. J Neurosci 19:5108-5118.

Janik VM (2013) Cognitive skills in bottlenose dolphin communication. Trends Cogn Sci 17:157-159.

Janik VM (2014) Cetacean vocal learning and communication. Curr Opin Neurobiol 28:60-65.

Janik VM, Knörnschild M (2021) Vocal production learning in mammals revisited. Phil Trans R Soc B 376:20200244.

Jarvis ED, Ribeiro S, da Silva ML, Ventura D, Vielliard J, Mello CV (2000) Behaviourally driven gene expression reveals song nuclei in hummingbird brain. Nature 406:628-632.

Jarvis ED (2019) Evolution of vocal learning and spoken language. Science 366:50-54.

Jensen, P. The Ethology of Domestic Animals. Wallingford, England: Centre for Agriculture and Bioscience International; 2009.

Jin DZ (2009) Generation variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. Phys Rev E 80:051902.

Kalinowski J, Saltuklaroglu T (2003) Choral speech: the amelioration of stuttering via imitation and the mirror neuronal system. Neurosci Biobehav Rev 27:339–347.

King SL, Sayigh LS, Wells RS, Fellner W, Janik VM (2013) Vocal copying of individually distinctive signature whistles in bottlenose dolphins. Proc R Soc B 280:20130053.

Kittelberger JM (2002) “Neurotrophins, synaptic connectivity and the regulation of song plasticity

in the zebra finch." PhD diss. Duke University.

Knörnschild M (2014) Vocal production learning in bats. Curr Opin Neurobiol 28:80-85.

Koumura T, Okanoya K (2016) Automatic recognition of element classes and boundaries in the birdsong with variable sequences. PLoS ONE 11:e0159188.

Kornfeld J, Benezra SE, Narayanan RT, Svara F, Egger R, Oberlaender M, et al (2017) EM connectomics reveals axonal target variation in a sequence-generating network. eLife 6:e24364.

Kosche G, Vallentin D, Long MA (2015) Interplay of inhibition and excitation shapes a premotor neural sequence. J Neurosci 35:1217-1227.

Kozhevnikov AA, Fee MS (2007) Singing-related activity of identified HVC neurons in the zebra finch. J Neurophysiol 97:4271-4283.

Kroodsma D, Hamilton D, Sánchez JE, Byers BE, Fandiño-Mariño H, Stemple DW, et al (2013) Behavioral evidence for song learning in the suboscine bellbirds (*Procnias* spp.; cottingidae). Wilson J Ornithol 125:1-14.

Kroodsma DE, Konishi M (1991) A suboscine bird (eastern phoebe, *Sayornis phoebe*) develops normal song with auditory feedback. Anim Behav 42:477-487.

Kroodsma DE, Miller EH, Ouellet H, editors. Acoustic Communication in Birds: Song learning and its consequences. Academic press; 1982.

Lara AH, Elsayed GF, Zimnik AJ, Cunningham JP, Churchland MM (2018) Conservation of preparatory neural events in monkey motor cortex regardless of how movement is initiated. eLife 7:e31826.

Lassa Ortiz JN, Herbert CT, Mindlin GB, Amador A (2019) Significant instances in motor gestures of different songbird species. Front Phys 7:142.

Leonardo A, Fee MS (2005) Ensemble coding of vocal control in birdsong. J Neurosci 25:652-661.

Lee BS (1950) Effects of delayed speech feedback. J Acoust Soc Am 22:824–826.

Leonardo A, Konishi M (1999) Decrystallization of adult birdsong by perturbation of auditory feedback. Nature 399:466–470.

Lewicki MS (1996) Intracellular characterization of song-specific neurons in the zebra finch auditory forebrain. J Neurosci 16:5854-5863.

Lewicki MS, Arthur BJ (1996) Hierarchical organization of auditory temporal context sensitivity. J Neurosci 16:6987-6998.

Liberti WA, Markowitz JE, Perkins LN, Liberti DC, Leman DP, Guitchounts G, et al (2016) Unstable

neurons underlie a stable learned behavior. Nat Neurosci 19:1665-1671.

Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahasi M, et al (2013) Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. Nature 498:104-108.

Lipkind D, Zai AT, Hanuschkin A, Marcus GF, Tchernichovski O, Hahnloser RHR (2017) Songbirds work around computational complexity by learning song vocabulary independently of sequence. Nat Commun 8:1247.

Liu WC, Wada K, Jarvis ED, Nottebohm F (2013) Rudimentary substrates for vocal learning in a suboscine. Nat Commun 4:2082.

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.

Luescher AU (ed). Manual of Parrot Behavior. Blackwell Publishing, 2006.

Luo MM, Perkel DJ (1999) A GABAergic, strongly inhibitory projection to a thalamic nucleus in the zebra finch song system. J Neurosci 19:6700-6711.

Lynch GF, Okubo TS, Hanuschkin A, Hahnloser RHR, Fee MS (2016) Rhythmic continuous-time coding in the songbird analog of vocal motor cortex. Neuron 90:877-892.

Macías S, Mora EC, Hechavarría JC, Kössl M (2016) Echo-level compensation and delay tuning in the auditory cortex of the mustached bat. Eur J Neurosci 43:1647-1660.

Maguire SE, Schmidt MF, White DJ (2013) Social brains in context: lesions targeted to the song control system in female cowbirds affect their social network. PLoS ONE 8:e63239.

Manderscheid E (2014) "The influence of auditory feedback on vocal sequence production." PhD diss. University of Chicago.

Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. J Neurosci 3:1039-1057.

Margoliash D (1986) Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. J Neurosci 6:1643-1661.

Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. J Neurosci 12:4309-4326.

Margoliash D, Fortune ES, Sutter ML, Yu AC, Wren-Hardin BD, Dave A (1994) Distributed representation in the song system of oscines: evolutionary implications and functional

consequences. Brain Behav Evol 44:247-264.

Margoliash D, Konishi M (1985) Auditory representation of autogenous song in the song-system of white-crowned sparrows. Proc Natl Acad Sci U S A 82:5997-6000.

Margoliash D, Tchnernichovski O (2015) Marmoset kids actually listen. Science 349:688-689.

McCasland JS, Konishi M (1981) Interaction between auditory and motor activities in an avian song control nucleus. Proc Natl Acad Sci U S A 78:7815-7819.

Meadows MG, Roudybush TE, McGraw KJ (2012) Dietary protein level affects iridescent coloration in Anna's hummingbirds, *Calypte anna*. J Exp Biol 215:2742-2750.

Meliza CD, Chi ZY, Margoliash D (2010) Representations of conspecific song by starling secondary forebrain auditory neurons: toward a hierarchical framework. J Neurophysiol 103:1195-1208.

Moelk M (1944) Vocalizing in the house-cat: a phonetic and functional study. Am J Psychol 57:184-205.

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.

Nick TA, Konishi M (2004) Neural song preference during vocal learning in the zebra finch depends on age and state. J Neurobiol 62:231-242.

Nick TA, Konishi M (2005) Neural auditory selectivity develops in parallel with song. J Neurobiol 62:469-481.

Nishikawa J, Okada M, Okanoya K (2008) Population coding of song element sequence in the Bengalese finch HVC. Eur J Neurosci 27:3273-3283.

Nixdorf BE, Davis SS, DeVoogd TJ (1989) Morphology of Golgi-impregnated neurons in hyperstriatum ventralis, pars caudalis in adult male and female canaries. J Comp Neurol 284:337-349.

Noad MJ, Cato DH, Bryden MM, Jenner MN, Jenner KCS (2000) Cultural revolution in whale songs. Nature 408:537.

Nottebohm F (1972) The origins of vocal learning. Am Nat 106:116-140.

Nottebohm F, Kelley DB, Paton JA (1982) Connections of vocal control nuclei in the canary telencephalon. J Comp Neurol 207:344-357.

Nottebohm F, Stokes TM, Leonard CM (1976) Central control of song in the canary, *Serinus canarius*. J Comp Neurol 165:457-486.

Okubo TS, Mackevicius EL, Payne HL, Lynch GF, Fee MS (2015) Growth and splitting of neural

sequences in songbird vocal development. Nature 528:352-357.

O'Neill WE, Suga N (1979) Target range-sensitive neurons in the auditory cortex of the mustache bat. Science 203:69-73.

O'Rourke T, Martins PT, Asano R, Tachibana RO, Okanoya K, Boeckx C (2021) Capturing the effects of domestication on vocal learning complexity. Trends Cog Sci 25:462-747.

Ottoni C, Van Neer W, De Cupere B, Daligault J, Guimaraes S, Peters J, et al (2017) The palaeogenetics of cat dispersal in the ancient world. Nat Ecol Evol 1:0139.

Perkes A, White D, Wild JM, Schmidt M (2019) Female songbirds: the unsung drivers of courtship behavior and its neural substrates. Behav Processes 163:60-70.

Person AL, Gale SD, Farries MA, Perkel DJ (2008) Organization of the songbird basal ganglia, including area X. J Comp Neurol 508:840-866.

Picardo MA, Merel J, Katlowitz KA, Vallentin D, Okobi DE, Benezra SE, et al (2016) Population-level representation of a temporal sequence underlying song production in the zebra finch. Neuron 90:866-876.

Pika S, Liebal K, Tomasello M (2003) Gestural communication in young gorillas (*Gorilla gorilla*): gestural repertoire, learning, and use. Am J Primatol 60:95-111.

Podos J (1996) Motor constraints on vocal development in a songbird. Anim Behav 51:1061–1070.

Poole B, Markowitz JE, Gardner TJ (2012) The song must go on: resilience of the songbird vocal motor pathway. PLoS ONE 7:e38173.

Poole JH, Tyack PL, Stoeger-Horwath AS, Watwood S (2005) Elephants are capable of vocal learning. Nature 434:455-456.

Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature 451:305-310.

Prather JF, Peters S, Nowicki S, Mooney R (2010) Persistent representation of juvenile experience in the adult songbird brain. J Neurosci 30:10586-10598.

Price PH (1979) Developmental determinants of structure in zebra finch song. J Comp Physiol Psychol 93:260–277.

Price T, Wadewitz P, Cheney D, Seyfarth R, Hammerschmidt K, Fischer J (2015) Vervets revisited: a quantitative analysis of alarm call structure and context specificity. Sci Rep 5:13220.

Quarrington B (1965) Stuttering as a function of the information value and sentence position of words. J Abnorm Psychol 70:221–224.

Rajan R (2018) Pre-bout neural activity changes in premotor nucleus HVC correlate with successful initiation of learned song sequence. J Neurosci 38:5925–5938.

Rajan R, Doupe AJ (2013) Behavioral and neural signatures of readiness to initiate a learned motor sequence. Curr Biol 23:87–93.

Rao D, Kojima S, Rajan R (2019) Sensory feedback independent pre-song vocalizations correlate with time to song initiation. J Exp Biol 222:jeb199042.

Rauske PL, Shea SD, Margoliash D (2003) State and neuronal class-dependent reconfiguration in the avian song system. J Neurophysiol 89:1688-1701.

Reichmuth C, Casey C (2014) Vocal learning in seals, sea lions, and walruses. Curr Opin Neurobiol 28:66-71.

Reinke H, Wild JM (1998) Identification and connections of inspiratory premotor neurons in songbirds and budgerigar. J Comp Neurol 391:147-163.

Richards DG, Wolz JP, Herman LM (1984) Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. J Comp Psychol 98:10–28.

Riehle A, Requin J (1993) The predictive value for performance speed of preparatory changes in neuronal activity of the monkey motor and premotor cortex. Behav Brain Res 53:35–49.

Ripley BD (1976) The second-order analysis of stationary point processes. J Appl Probab 13:255-266.

Roberts TF, Hisey E, Tanaka M, Kearney MG, Chattree G, Yang CF, et al (2017) Identification of a motor-to-auditory pathway important for vocal learning. Nat Neurosci 20:978-986.

Roberts TF, Klein ME, Kubke MF, Wild JM, Mooney R (2008) Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. J Neurosci 28:3479-3489.

Ross MT, Flores D, Bertram R, Johnson F, Hyson RL (2017) Neuronal intrinsic physiology changes during development of a learned behavior. eNeuro 4:ENEURO.0297-17.2017.

Russell JL, McIntyre JM, Hopkins WD, Taglialatela JP (2013) Vocal learning of a communicative signal in captive chimpanzees, *Pan troglodytes*. Brain Lang 127:520-525.

Sakata JT, Brainard MS (2008) Online contributions of auditory feedback to neural activity in avian song control circuitry. J Neurosci 28:11378-11390.

Saltuklaroglu T, Kalinowski J, Robbins M, Crawcour S, Bowers A (2009) Comparisons of stuttering frequency during and after speech initiation in unaltered feedback, altered auditory feedback and choral speech conditions. Int J Lang Commun Disord 44:1000-1017.

Sanchez-Valpuesta M, Suzuki Y, Shibata Y, Toji N, Ji Y, Afrin N, et al (2019) Corticobasal ganglia projecting neurons are required for juvenile vocal learning but not for adult vocal plasticity in songbirds. Proc Natl Acad Sci U S A 116:22833-22843.

Sanvito S, Galimberti F, Miller EH (2007) Observational evidences of vocal learning in southern elephant seals: a longitudinal study. Ethology 113:137-146.

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.

Schmidt MF (2003) Pattern of interhemispheric synchronization in HVc during singing correlates with key transitions in the song pattern. J Neurophysiol 90:3931-3949.

Schmidt MF, Konishi M (1998) Gating of auditory responses in the vocal control system of awake songbirds. Nat Neurosci 1:513-518.

Scott SH (2008) Inconvenient truths about neural processing in primary motor cortex. J Physiol 586:1217-1224.

Seki Y, Suzuki K, Takahasi M, Okanoya K (2008) Song motor control organizes acoustic patterns on two levels in Bengalese finches (*Lonchura striata* var. *domestica*). J Comp Physiol A 194:533–543.

Seltmann S, Trost L, Ter Maat A, Gahr M (2016) Natural melatonin fluctuation and its minimally invasive simulation in the zebra finch. PeerJ 4:e1939.

Seyfarth Rm, Cheney DL (1990) The assessment by vervet monkeys of their own and another species' alarm calls. Anim Behav 40:754-764.

Seyfarth RM, Cheney DL, Marler P (1980) Vervet monkey alarm calls: semantic communication in a free-ranging primate. Anim Behav 28:1070-1094.

Shank SS, Margoliash D (2009) Sleep and sensorimotor integration during early vocal learning in a songbird. Nature 458:73-77.

Shea SD, Koch H, Baleckaitis D, Ramirez J, Margoliash D (2010) Neuron-specific cholinergic modulation of a forebrain song control nucleus. J Neurophysiol 103:733-745.

Shea SD, Margoliash D (2003) Basal forebrain cholinergic modulation of auditory activity in the zebra finch song system. Neuron 40:1213-1226.

Simpson HB, Vicario DS (1990) Brain pathways for learned and unlearned vocalizations differ in zebra finches. J Neurosci 10:1541-1556.

Smith JC, Abdala APL, Borgmann A, Rybak IA, Paton JFR (2013) Brainstem respiratory networks: building blocks and microcircuits. Trends Neurosci 36:152-162.

Smith WJ. The behavior of communicating: an ethological approach. Cambridge: Harvard University Press; 1977.

Snow DW (1968) The singing assemblies of Little Hermits. Living Bird 3:47-55.

Sossinka R, Böhner J (1980) Song types in the zebra finch *Poephila guttata castanotis*. Z Tierpsychol 53:123–132.

Stansbury AL, Janik VM (2021) The role of vocal learning in call acquisition of wild grey seal pups. Phil Trans R Soc B 376:20200251.

Stauffer TR, Elliott KC, Ross MT, Basista MJ, Hyson RL, Johnson F (2012) Axial organization of a brain region that sequences a learned pattern of behavior. J Neurosci 32:9312-9322.

Stuart A, Kalinowski J, Rastatter MP, Lynch K (2002) Effect of delayed auditory feedback on normal speakers at two speech rates. J Acoust Soc Am 111:2237–2241.

Stuart A, Kalinowski J, Rastatter MP, Saltuklaroglu T, Dayalu V (2004) Investigations of the impact of altered auditory feedback in-the-ear devices on the speech of people who stutter: initial fitting and 4-month follow-up. Int J Lang Commun Disord 39:93-113.

Striedter GF, Vu ET (1998) Bilateral feedback projections to the forebrain in the premotor network for singing in zebra finches. J Neurobiol 34:27-40.

Suga N, O'Neill WE (1979) Neural axis representing target range in the auditory cortex of the mustache bat. Science 206:351-353.

Suga N, O'Neill WE, Manabe T (1979) Harmonic-sensitive neurons in the auditory cortex of the mustache bat. Science 203:270-274.

Suthers RA (1997) Peripheral control and lateralization of birdsong. J Neurobiol 33:632-652.

Suthers RA, Goller F, Wild JM (2002) Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. Proc Natl Acad Sci U S A 99:5680-5685.

Sutter ML, Margoliash D (1994) Global synchronous response to autogenous song in zebra finch HVc. J Neurophysiol 72:2105-2123.

Svanberg I (2008) Towards a cultural history of the Bengalese finch (*Lonchura domestica*). Zool Gart 77:334-344.

Takahashi DY, Fenley AR, Teramoto Y, Narayanan DZ, Borjon JI, Holmes P, et al (2015) The developmental dynamics of marmoset monkey vocal production. Science 349:734-738.

Tanji J, Evarts EV (1976) Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. J Neurophysiol 39:1062–1068.

Taylor IK (1966) What words are stuttered? Psychol Bull 65:233–242.

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.

ten Cate C (2021) Re-evaluating vocal production learning in non-oscine birds. Phil Trans R Soc B 376:20200249.

Tinbergen, Nikolaas. The Study of Instinct. Cambridge: Oxford University Press; 1951.

Trevisan MA, Mindlin GB, Goller F (2006) Nonlinear model predicts diverse respiratory patterns of birdsong. Phys Rev Lett 96:058103.

Vallentin D, Kosche G, Lipkind D, Long MA (2016) Inhibition protects acquired song segments during vocal learning in zebra finches. Science 351:267-271.

Vates GE, Broome BM, Mello CV, Nottebohm F (1996) Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (*Taeniopygia guttata*). J Comp Neurol 366:613-642.

Vates GE, Vicario DS, Nottebohm F (1997) Reafferent thalamo-"cortical" loops in the song system of oscine songbirds. J Comp Neurol 380:275-290.

Vicario DS (1991) Organization of the zebra finch song control system: functional organization of outputs from nucleus *robustus archistriatalis*. J Comp Neurol 309:486-494.

Vigne JD, Guilaine J, Debue K, Haye L, Gérard P (2004) Early taming of the cat in Cyprus. Science 304:259.

Volman SF (1993) Development of neural selectivity for birdsong during vocal learning. J Neurosci 13:4737-4747.

Vu ET, Mazurek ME, Kuo YC (1994) Identification of a forebrain motor programming network for the learned songs of zebra finches. J Neurosci 14:6924-6934.

Walløe S, Thomsen H, Balsby TJ, Dabelsteen T (2015) Differences in short-term vocal learning in parrots, a comparative study. Behaviour 152:1433-1461.

Wang CZH, Herbst JA, Keller GB, Hahnloser RHR (2008) Rapid interhemispheric switching during vocal production in a songbird. PLoS Biol 6:e250.

Wang J, Sokabe M, Sakaguchi H (2001) Functional connections between the HVC and the shelf of the zebra finch revealed by real-time optical imaging technique. Neuroreport 12:215-221.

Wild JM (1993a) Descending projections of the songbird nucleus robustus archistriatalis. J Comp Neurol 338:225-241.

Wild JM (1993b) The avian nucleus retroambigualis: a nucleus for breathing, singing and calling. Brain Res 606:319-324.

Wild JM (1994) Visual and somatosensory inputs to the avian song system via nucleus uvaeformis (Uva) and a comparison with the projections of a similar thalamic nucleus in a nonsongbird, *Columba livia*. J Comp Neurol 349:512-535.

Wild JM, Li DF, Eagleton C (1998) Projections of the dorsomedial nucleus of the intercollicular complex (DM) in relation to respiratory-vocal nuclei in the brainstem of pigeon (*Columba livia*) and zebra finch (*Taeniopygia guttata*). J Comp Neurol 377:392-413.

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.

Wiley RH (1971) Song groups in a singing assembly of Little Hermits. The Condor *73*:28-35.

Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. Science 269:1880-1882.

Woolley SC, Rajan R, Joshua M, Doupe AJ (2014) Emergence of context-dependent variability across a basal ganglia network. Neuron 82:208-223.

Yates AJ (1963) Delayed auditory feedback. Psychol Bull 60:213–232.

Zann RA. The zebra finch: a synthesis of field and laboratory studies. Cambridge: Oxford University Press; 1996.

Zeigler H, Marler PE. Behavioral neurobiology of birdsong. In: Behavioral Neurobiology of Birdsong, 2002, Hunter College, City University of New York.

Zevin JD, Seidenberg MS, Bottjer SW (2004) Limits on reacquisition of song in adult zebra finches exposed to white noise. J Neurosci 24:5849–5862.

Zollinger SA, Riede T, Suthers RA (2008) Two-voice complexity from a single side of the syrinx in northern mockingbird *Mimus polyglottos* vocalizations. J Exp Biol 211:1978-1991.