THE UNIVERSITY OF CHICAGO

VOCAL MOTOR CODING IN A SONGBIRD PREMOTOR NUCLEUS

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 COMPUTATIONAL NEUROSCIENCE

BY
KYLER JOSHUA BROWN

CHICAGO, ILLINOIS
DECEMBER 2017

Only a Sith Lord deals in absolutes.

— Obi-Wan Kenobi

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My advisor Dan Margoliash has been a constant positive force throughout my training. Without his unwavering faith and long-term vision, I would have abandoned my experiments after the first week, month, or year of failure. Dan took a subtle approach to guiding my research, granting me the freedom to find my own path and build my confidence as an independent scientist. In any future mentorship role I assume, I'll do my best to channel my inner "Dan".

Franz Goller and Gabo Mindlin generously took me into their own labs, giving me the perspective and tools required to investigate the mechanics of song. Correspondence with Hamish Mehaffey and Derek Zaraza was essential for the experiments detailed in Chapter 4. I an indebted to them for their advice and instruction.

I have had the great fortune of working with many excellent scientists in the Margoliash lab. Ana Amador guided me through my first electrophysiology experiments and me gave my first experience listening to the satisfying "crackle" of a living neuron through the audio monitor. Ana also taught me, by example, the dedication and sacrifice required to conduct successful experiments. Dan Meliza gave me my first exposure to high quality software engineering, and a template for successfully balancing data analysis with data collection. Arij Daou was not only a great scientist, but a good friend. I miss our lunches together. Peter Malonis and Graham Fetterman were both fantastic help sharing the burden of software development in the lab. Sofija Canavan helped develop surgical techniques that improved my experimental outcomes. Mike Lusignan, Etienne Manderscheid, and Tim Brawn gave advice on bringing my thesis work to completion. Nelson Medina provided feedback on interpreting the results of my research. I am deeply grateful to Daniel Baleckaitis, who provided essential support building experimental equipment and a thousand other ways over the course of my time in the Margoliash lab.

ABSTRACT

Birdsong has become a widely studied model for vocal learning and motor behavior. It is well understood that sensorimotor area HVC (proper name) is essential for song learning and production, but our understanding of the role of HVC neurons remains incomplete. In one view, HVC serves only as a clock – a specialized tissue designed to propagate timing information to other areas of the brain. From this perspective, HVC is agnostic about the motor periphery, providing other brain areas with a timing signal from which to construct the carefully sequenced gestures that constitute birdsong. Another hypothesis is that the time-varying activity in HVC contains information about the state of the motor periphery. The gesture trajectory extrema (GTE) model is one of many models that take this second view. Neurons in HVC that match the GTE model were first discovered by using a biomechanical model of the periphery to estimate the motor state during song.

In this thesis, I took three approaches to investigate HVC's role in song. First, I used large, dense micro-electrode arrays to sample from large populations of HVC neurons. I recorded from HVC while the birds were sleeping, and used auditory playback of the bird's own song to entrain the motor system. I found HVC activity to be significantly modulated by syllable onsets and vary depending on the syllable type. These observations are consistent with HVC encoding the gestures of the motor periphery. The time varying population rate appeared consistent with the GTE model, but as yet this has not been validated statistically.

Second, I used similar electrode arrays in HVC, but now combined with simultaneous recordings in the muscles of the syrinx (N=3). By recording from the muscles of the syrinx, I was able to directly examine the motor output of the song system without using the biomechanical model. Perhaps due technical limitations, I was unable to find the expected reliable auditory response to song playback during sleep, yet during sleep the motor system engaged in short spontaneous replay events. Analyzing these replay events revealed a tight coupling between activity in HVC and activation of syrinx muscles. These data suggest that

HVC may directly encode features of the motor periphery, a prediction of the GTE model. However it was unclear whether this coupling also existed in the awake state, because the data were collected during sleep.

Finally, in a third set of experiments, I develop a system to record from the syrinx muscles and neurons in HVC simultaneously in a singing bird. I found that the tight coupling between the muscles of the syrinx and neurons in HVC observed during sleep, was also maintained during singing (N=7).

Taken together, the data collected in this thesis suggest that the songbird motor periphery is represented in HVC during song. While these results cannot rule out the existence of a "clock" subpopulation of HVC neurons, HVC may have a premotor role in singing, and is tightly locked to the activity of syringeal muscles. These results are consistent with some of the predictions of the GTE model, albeit the zero-lag prediction of the GTE model remains to be confirmed. The data provide compelling positive results refuting a central prediction of the "clock" model – that HVC activity is divorced from muscle dynamics.

CHAPTER 1

INTRODUCTION

1.1 The zebra finch as model of complex motor sequences

Motor control was the first identified function of mammalian cortex, from studies of traumatic injury and later from stimulation experiments (Taylor and Gross, 2003). Yet much less is known about the role of microcircuits in motor cortex than microcircuits in sensory cortex. Unlike seminal work in primary sensory areas (Hubel and Wiesel, 1959), single electrode experiments have not uncovered unifying principles of motor cortex organization. Perhaps this inequality is due to the fundamentally different requirements of the motor system – while the role of the sensory system is to *encode* sensory information, motor systems *generate* it, and the statistical inference used for sensory coding may not be applicable to motor patterns (Churchland et al., 2012). Additionally, many complex motor tasks do not require cortex (Sherrington, 1910) while other tasks require cortex for learning but not execution (Kawai et al., 2015). While human limb movements like reaching and grasping have a clear homology to reaching and grasping in animal models, spoken language, the most distinctive and complex learned motor behavior of our species, is more challenging to study. Learned vocal motor behavior in mammals is not as rare as is commonly thought, being well established for cetaceans and several species of michrochiropteran bats, and with anecdotal evidence in a number of other mammalian species (e.g. elephants). Yet within primates there is at best anecdotal and equivocal evidence for vocal learning in chimpanzees (Watson et al., 2015) and for no other primate species. Socially-contingent vocal development has been observed in mother-infant interactions in marmosets, which may be a precursor to vocal learning (Takahashi et al., 2015). In contrast, vocal learning in birds is common throughout the songbirds (order Oscine) and parrots (order Psittaciformes).

The Australian zebra finch (*Taeniopygia guttata castanotis*) is a perching songbird (Passeri

suborder of oscines) well suited to laboratory study. They live in large flocks, are highly social, and sing readily in the presence of humans and in captivity (Morris, 1954). They are opportunistic breeders, and can breed year round in laboratory conditions. Zebra finches have strong sexual dimorphism, with adult males exhibiting orange cheek patches, a black and white striped throat and chest, and a brilliant white abdomen. Adult females are gray with black and white cheek stripes. Both sexes communicate with calls, which have about a dozen variations with distinct contexts and functions (Zann, 1996). The timing and inter-bird synchronization of calls develops through learning (Benichov et al., 2016). Male zebra finches learn to sing, and zebra finch song is a widely studied model of learned vocal behavior. The presence of a female will strongly bias adult males to sing. Regulating male access to females is easily controlled in a laboratory setting, further enhancing the value of the model. Female zebra finches learn to regulate the timing of their calls during male singing but themselves do not learn to sing. Female song learning is known in a subset of songbird species whereas male song learning is thought to be ubiquitous.

Female mate choice and male-male competition drive sexual selection of song quality in adults, and is important for the evolutionary fitness of the animal (Catchpole, 1987; Searcy and Andersson, 1986). From other songbird species, we know females select males with larger song repertoires (Catchpole, 1987) and this preference is sustained even in species, such as grackles, with single song types, suggesting females of single-song species have preferences that evolved before the single song of the species (Searcy, 1992). Song complexity is also predictive of fitness across species (Read and Weary, 1992). For zebra finches, peak frequency and frequency range is correlated with brood size, while the number of syllables and phrase length of song is correlated with offspring survival (Woodgate et al., 2012). Theses studies suggest that zebra finch song is under strong evolutionary pressure.

Song, like language for humans, is central to the life of a zebra finch, and they are particularly well adapted for singing. Every component of the descending motor pathway is

optimized for song, from specialized nuclei in the brain, to specially optimized vocal muscle fibers (Elemans et al., 2008). Using a "champion" species – a species with exceptional natural talents – has proven fruitful in other domains: rodents and spatial memory (Grieves and Jeffery, 2017); owls and sound source localization (Knudsen, 1987); cats and vision (Hubel and Wiesel, 1959). The zebra finch and its song provide a powerful paradigm for the study of complex motor sequences. It is both an enormous technical advantage and a conceptual simplification, and a technical and conceptual limitation, that typically a given zebra finch sings only a single song type, with great precision. The study of other songbird species representing an almost endless catalog of vocal behavior can compellingly address such limitations, albeit with loss of some of the technical advantages that have promoted the zebra finch model system.

1.2 The development of song

Auditory input is essential for song learning (Konishi, 1965b). In a natural setting, zebra finches learn to sing by imitating their father (Bolhuis et al., 2013), and though the juvenile song will eventually reach a close approximation of the father's song, each zebra finch tends to have a unique song. Between populations of separated zebra finches, there are statistically distinct syllable repertoires (Lachlan et al., 2016), likely due to the local nature of song learning, which parallels dialects in human speech. Dialects are well known in other songbird species (Marler and Tamura, 1962; Payne et al., 1981).

If raised by foster parents, a juvenile zebra finch will acquire the foster song, not that of the biological father, which constitutes early evidence that song is learned rather than acquired through genetic inheritance alone (Marler et al., 1973). Song can be acquired from alternate or artificial tutors (Ravbar et al., 2012). If exposed to other auditory stimuli, some of these sounds are incorporated into song, such as female calls if a male is raised only by a female (Price, 1979). Zebra finches raised by Bengalese finches developed songs with

Bengalese finch syllables (Immelmann, 1969). If raised in isolation, zebra finches develop abnormal song, lacking the stereotypy and rythmicity of normal song (Price, 1979). The descendants of isolated zebra finches slowly return to typical wild-type song after four generations of father-son tutoring (Feher et al., 2009) indicating that song structure has both environmental and genetic influences.

Auditory feedback is also essential for song learning, and deafened birds not only fail to acquire normal song (Konishi, 1965a), but produce adult song that is even more dissimilar to normal song than birds raised in isolation (Price, 1979). The requirement for auditory feedback was once thought to close at the end of song learning, hence the old term "closed-end learner" for species such as zebra finches that do not acquire new songs as adults. However, it is now well established for zebra finches that maintaining normal adult song also requires intact hearing (Nordeen and Nordeen, 1992).

Zebra finches, like other songbirds, have a critical period for song learning during development (Marler, 1970). After this critical period zebra finches will fail to produce normal adult song, even after exposure to other zebra finch (conspecific) song (Price, 1979; Thorpe, 1958). The sensory stage of learning begins perhaps as early as 20 days after hatching (Adret et al., 2012) and ends at about 65 days, though only the first 10 days are required for accurate learning of the tutor song (Roper and Zann, 2006). The latter boundary for the sensory period can be adjusted by partial deprivation from tutor song (Eales, 1987). The sensorimotor stage, in which the zebra finch begins to vocalize, starts at 35 days post hatch and ends at day 90, when the zebra finch has entered adulthood. The first song-like vocalizations of juveniles were termed "sub-song" by Immelmann (Immelmann, 1969), characterized by long unstructured syllables, a high degree of variability, and an amplitude much lower than adult song (Roper and Zann, 2006). Sub-song transitions into plastic song about 50 days post hatch (Immelmann, 1969). Plastic song is characterized by repeated syllables of roughly 100 ms duration (Immelmann, 1969; Tchernichovski et al., 2001). These repeated syllables

slowly evolve until the completion of song learning at day 90 (Tchernichovski and Mitra, 2002). Adult zebra finches retain this crystallized song for the remainder of their lives.

Adult zebra finch song consists of a rapid sequence of syllables, which themselves may be subdivided into notes. Some simple syllables may have spectral similarities to the call repertoire, especially the male "long" or contact call, which is learned. Most syllables are unique to the song. Syllables occur in a standard, stereotyped sequence, called the motif. Zebra finch motifs are typically composed of 3-8 syllables. A motif is often preceded by one or more simple introductory notes, which may be similar to one of the bird's calls, usually a down-sweep. Introductory notes tend to be spectrally similar to each other, but the precision of production tends to increase as the onset of the motif approaches. Notes between motifs are sometimes called connecting notes, and may or may not be the same as the introductory notes. A cluster of motifs sung in succession is called a bout of song (Zann, 1996).

1.3 The biomechanics of song production

The vocal organ of songbirds is the syrinx. In most species, including zebra finches, the syrinx is positioned below the trachea at the joining of the primary bronchi within the thorax (Goller and Larsen, 1997; Suthers and Zollinger, 2004). The syrinx has two separate sound sources. Each sound source has two pads of tissue, called the lateral and medial labia (LL and ML), which are adducted rostrally into the bronchial lumen to produce vocalizations. In zebra finches the labia consist of a fairly uniform mass of collagen and elastin, giving the tissue its viscoelastic properties (Riede and Goller, 2010a).

Biophysical models of the syrinx have greatly informed our understanding of how birds sing. The syrinx is modeled as two oscillating masses interacting with each other and with the force from pressurized air (Perl et al., 2011; Amador and Margoliash, 2013), a model adapted and refined from early work modeling the human larynx (Titze, 1988). The discovery that similar principles underlie both human and bird vocalizations was surprising —

human speech has high harmonic content while many bird songs are pure tones. Alternately, some theories of birdsong production described an aerodynamic sound source, as in human whistling (Ballintijn and Cate, 1998).

Definitive evidence for both a mechanical sound source and extensive vocal filtering in song birds resulted from light-gas experiments, where birds sang in helium-oxygen gas mixtures (Nowicki, 1987). In this environment, an aerodynamic source for birdsong should create a higher fundamental frequency (pitch), while a mechanical sound source should be largely unaffected by the decreased gas density and the pitch should remain the same. It was found that the fundamental frequency did not change, but there was an increase in harmonic power. This finding suggests that acoustic filters in the birdsong vocal system act to remove harmonics. With decreased air density, the speed of sound is changed, and the filters are no longer tuned to the fundamental frequency, which underscores the importance of vocal tract filtering in birdsong. Thus, upper vocal tract filtering is important for birdsong production, much as it is important for human speech production. Vocal tract filtering in zebra finches has at least two components: the trachea and the oropharyngeal-esophageal cavity (OEC) (Perl et al., 2011; Fletcher et al., 2006). The trachea has a fixed length of about 35 mm. Modeling the trachea as a tube closed at one end (the syrinx) and open at the other (the OEC), creates a time-delay filter which emphasizes frequencies around 2.5 kHz and 7.5 kHz, the band of power most commonly seen in zebra finch song. The OEC is a distensible oral cavity connecting the trachea at one end and the beak at the other. By expanding the OEC volume and changing the area of the beak gape, the filtering properties of OEC can vary as a function of time. The OEC can be modeled as a Helmholtz resonator, a type of band-pass filter around a central frequency. Cardinals use the variable size of the OEC to emphasize frequencies ranging from 1 kHz to 9 kHz within a single syllable (Fletcher et al., 2006). In zebra finches, high-speed X-ray cinematography has shown OEC volume and beak gape are adjusted for each syllable, with increased OEC volume pushing the energy

to lower frequencies, while increased beak gape results in increased higher frequency content (Ohms et al., 2010). Some changes in the acoustic structure of syllables are associated with simultaneous changes in beak gape (Goller and Cooper, 2004), which follows results in other species (Westneat et al., 1993).

The mechanics of sound production in the syrinx have been extensively studied by Mindlin and colleagues (Perl et al., 2011; Gardner et al., 2001; Sitt et al., 2008). The air pressure created by the syrinx is a function of the medial position of the labium, which in turn is governed by two forcing parameters of the dynamical system: the pressure exerted by the air sac system to create subsyringeal pressure and the restitution force (tension) of the labia, likely controlled by the intrinsic and extrinsic muscles of the syrinx. These two parameters, pressure and tension, can be used to model much of the acoustic diversity of zebra finch song and can be used to generate realistic zebra finch song when combined with models of the trachea and OEC (Arneodo et al., 2012).

The normal form of the sound source equations are

$$\begin{aligned}\frac{dx}{dy} &= y \\ \frac{dy}{dt} &= -\alpha(t)\gamma^2 - \beta(t)\gamma^2 x - \gamma^2 x^3 - \gamma x^2 y + \gamma^2 x^2 - \gamma x y,\end{aligned}$$

where x is the medial position of the labium and γ is a time constant. This system of equations has two bifurcation lines, a saddle node in a limit cycle (SNILC) and a Hopf bifurcation, see Fig. 1.2. The time varying parameters $\alpha(t)$ and $\beta(t)$ represent pressure and tension respectively, and their trajectories over time are called gestures. Using this model, and the model of vocal tract filtering, it is possible to create synthetic zebra finch sounds that drive replay activity in the motor system (Amador et al., 2013). The task of recovering the gestures from a recorded song is a difficult processes, requiring some method of inverting

the Mindlin model. The current state-of-the-art is an iterative guess-and-check method that is time consuming but can create convincing synthetic songs (convincing to both the human investigator and the zebra finch motor system), see Fig. 1.1.

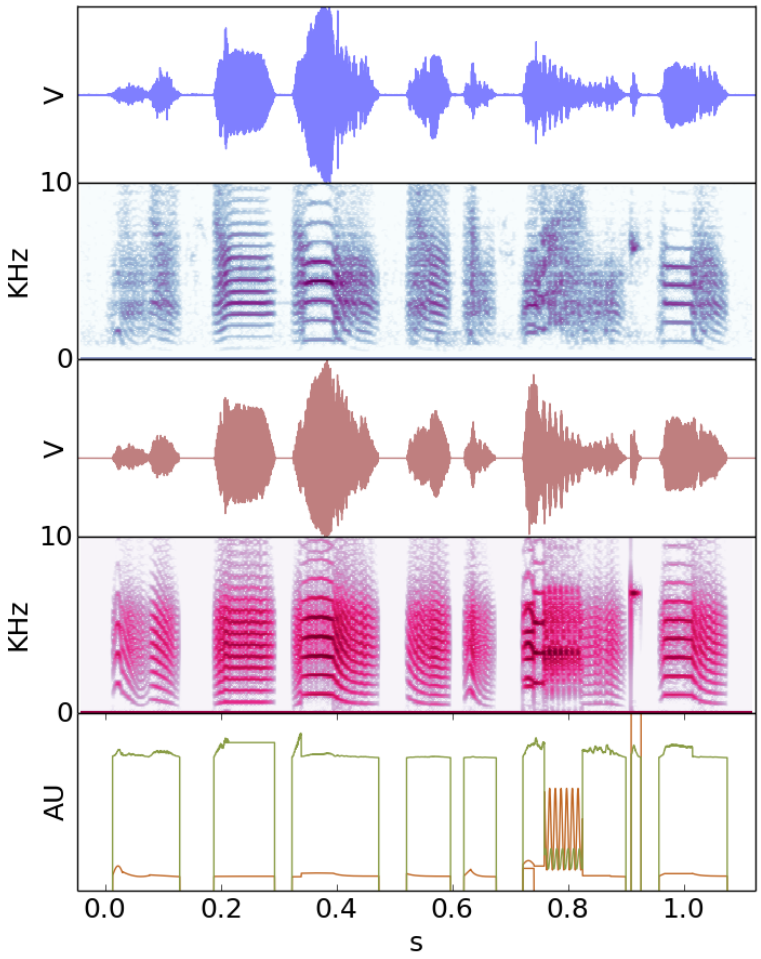

Figure 1.1. Example inversion of the Mindlin model Top: the oscillogram and spectrogram of a recorded song. Middle: the oscillogram and spectrogram of the synthetic song. Bottom: Gestures recovered by the iterative, manual processes of matching the synthetic song to the real song. Pressure (α) is in green, tension (β) is in yellow. The excellent oscillogram match is the result of multiplying the sound source signal by the true amplitude as part of the trachea model, not a result of the pressure and tension values.

Zebra finches can produce between 5 and 14 different syllable types (Sturdy et al., 1999; Zann, 1993), all of which can be modeled by variations of pressure and tension. Further validation of the Mindlin model arises from experiments transecting the tracheosyringeal

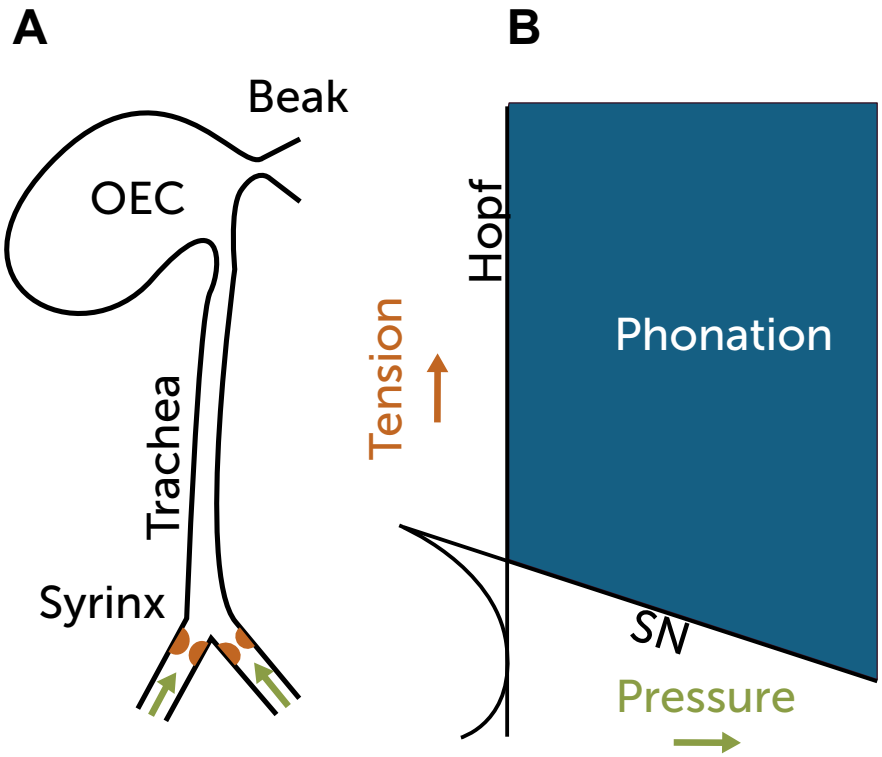

Figure 1.2. Schematic of sound source and filtering elements and the bifurcation diagram of the Mindlin model. A: The essential elements of zebra finch vocalization. Sound is generated by air sac pressure generating oscillations of the syrinx labia. This sound is filtered by the trachea and oropharyngeal cavity. B: Bifurcation diagram of the Mindlin model. As pressure and tension increase, they syrinx model moves to the phonating region, crossing either a saddle node or Hopf bifurcation. Low pitched notes <2 kHz lie near the saddle node bifurcation.

nerve, resulting in correctly timed, but very noisy syllables with a low fundamental frequency with a low fundamental frequency. This is consistent with predictions from the Mindlin model, where lower tension results in lower pitch and higher spectral entropy (Williams and McKibben, 1992).

The presence of a second sound source increases the complexity of modeling the syrinx and the sounds it produces. While some birds, such as cardinals, use two sources to produce simple pure tones, zebra finches can create complex sounds which have zero, one or two apparent fundamental frequencies. The two sources have interacting forces which may cause them to entrain (Amador and Mindlin, 2008). If they are entrained out of phase, frequency of the resulting sounds is doubled. The two sources can also create two independent oscillations and the resulting sound has two fundamental frequencies, confounding common pitch analysis techniques. Even a single source has intrinsic nonlinearities. *In vivo* preparation of an excised syrinx with one bronchial tube sealed shut shows period-doubling and aperiodic regimes as applied pressure increases, possibly due to nonlinear effects of the Bernoulli force as air flows through the bronchi and trachea (Fee et al., 1998). However, this result is disputed – this nonlinear effect does not arise artificially in the intact syrinx (Elemans et al., 2010), and specialized muscle tissue provides a mechanism for millisecond control of the labial tension (Elemans et al., 2008).

Though capturing all the acoustic complexity is an extremely challenging task, a single source model provides sufficient coverage of zebra finch vocal behavior to be a useful starting point. The low pressure/high tension region of parameter space produces pure tones, like those found in zebra finch high notes. The high pressure/low tension region of parameter space produces sounds with high harmonic content, similar to calls and low frequency sounds such as flat, harmonic notes. In theory, any zebra finch vocalization should have a unique pressure-tension trajectory (Perl et al., 2011), but in practice, the Mindlin syrinx model is very challenging to invert. The central issue is that a region of pressure-tension parameters

can produce nearly identical sounds, due to broad iso-frequency lines running parallel to the SNIC bifurcation, see Fig. 1.4 and (Lynch et al. (2016) supplemental figure 1). This challenge makes it currently impossible to automatically determine the unique pressure-tension trajectories for a given vocalization. Recently, the model was simplified by fixing the pressure parameter to a constant value during sound production which allows a tension trajectory to be automatically computed (Boari et al., 2015). That this simplification can be useful hints at the difficulty of finding the "true" pressure-tension trajectories of recorded zebra finch song. But it may also indicate that narrowing the inversion to a family of plausible solutions, of which the fixed pressure solution is a member, may still be a useful dimensionality reduction.

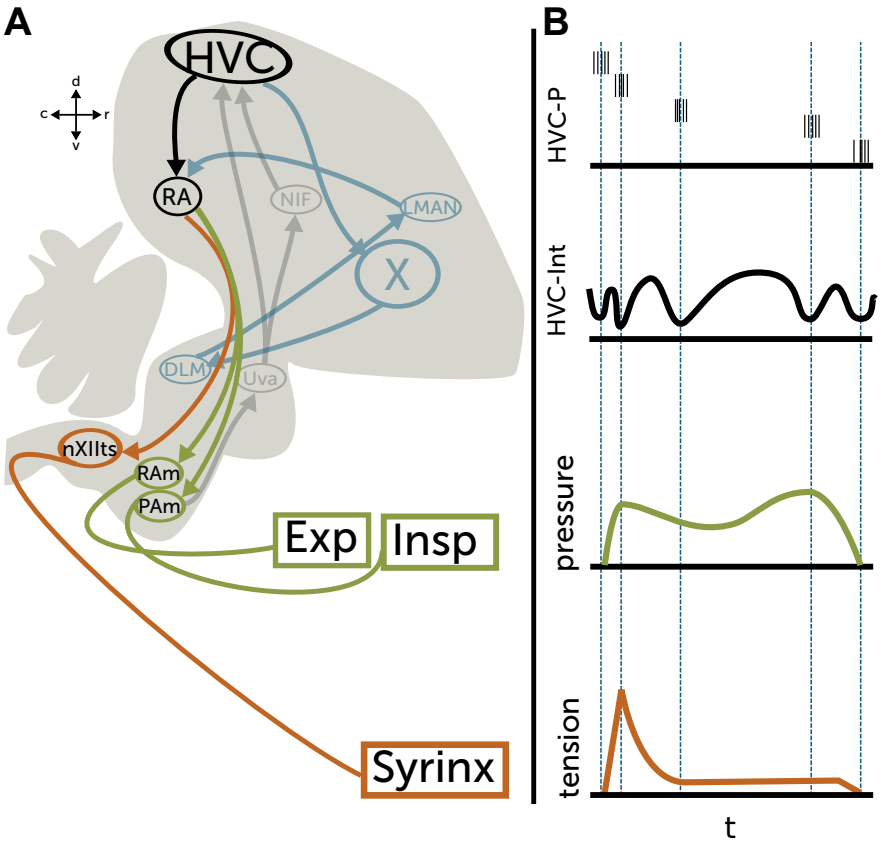

Figure 1.3. Anatomy of the song system and illustration of the GTE model. A: A simplified anatomical schematic of the zebra finch song system. B: An example sketch of the GTE model. HVC projection neurons burst and interneuron minima occur with zero time delay at moments of abstract gesture features of the syrinx tension and air sac pressure — onsets, offsets and local maxima.

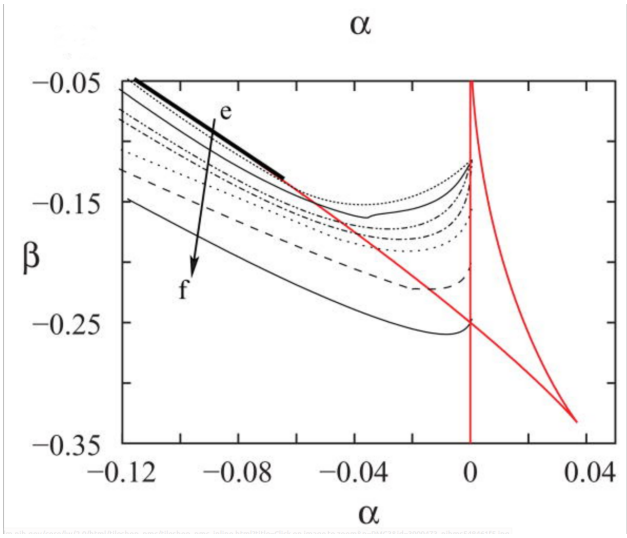

Figure 1.4. Iso-frequency curves in the output of the normal form of the Mindlin model. The bifurcation diagram of the Mindlin model of the syrinx. Red lines show bifurcation, the solid black bar indicates the SNILC bifurcation line. Dotted lines represent frequencies between 750 and 1900 Hz. The arrow from e to f denotes increasing frequency. With the exception of high notes, the fundamental frequency of most zebra finch syllables tend to be less than 2 kHz. Parameters α and β represent pressure and tension respectively (arbitrary units). For this figure the values of α and β have been inverted. The phonating region is where the forcing parameters are negative, rather than positive in Fig. 1.2. Figure from (Perl et al., 2011).

Syrinx tension is thought to be controlled by the six muscles associated with the syrinx, though the interactions of the syrinx muscles are poorly understood and individual contributions to pitch are a complex combination of agonistic and antagonistic (Elemans et al., 2006). The four intrinsic muscles are the ventral syringeal (vS), dorsal syringeal (dS), ventral tracheobronchial (vTB) and the dorsal tracheobronchial muscle (dTB). The two extrinsic muscles are the tracheolateral (TL) and sternotrachealis muscle (ST), Fig 1.5. In zebra finches, the intrinsic muscles are likely responsible for setting the position and tension of the syringeal membrane (Vicario, 1991). The muscle vS may be controlling the onset and offset of the syllable while dS may be involved in shaping the syllable during phonation (Vicario, 1991). During harmonic stacks and introductory notes, neither vS or dS are active, while more complex syllables have more complex interactions, Fig. 1.6. The clearest role for the intrinsic muscles (as defined by their time-varying activation over the course of song) is preparatory movement during gaps in phonation. Birdsong is a balance between vocalizing and respiration. The muscles of the syrinx must allow the bird to rapidly switch between a

vocalizing breath and a respiratory breath, and their essential role is hidden in the empty spaces of the spectrogram.

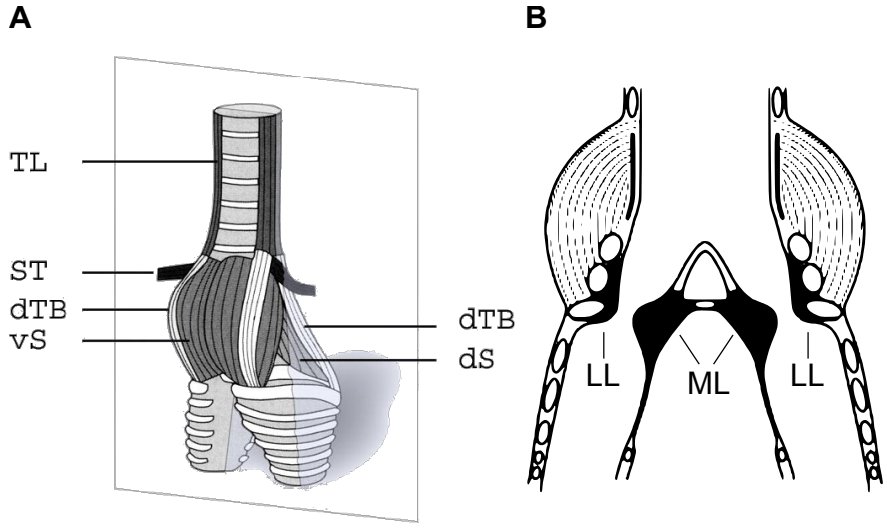

Figure 1.5. Anatomy of the syrinx. A: External view of the syrinx. B: Cross-section of the syrinx, showing the lateral labia (LL) and medial labia (ML). Figure adapted from (Riede and Goller, 2010b).

In some songbird species, such as the brown thrasher, there is a clear relationship between the EMG activation of the muscles and the fundamental frequency of the song. However zebra finch phonation appears to be more complex. Low frequency sounds in the 0.5-1.5 kHz range appear to have vS and dTB activity that correlate with fundamental frequency, through fundamental frequencies above that range are not correlated with either muscle (Goller and Riede, 2013; Sitt et al., 2008). Syllables with high pitches are generated by the right HVC exclusively, while lower pitched sounds are generated bilaterally (Jensen et al., 2007). The lack of EMG/pitch correlation in zebra finches may be due to unilateral phonation, but there is currently no understood mechanism behind this observation.

Air sac pressure also contributes to pitch (Amador and Margoliash, 2013). A minimum of 0.5 kPa of pressure is required for zebra finch phonation (Elemans et al., 2010). In canaries, oscillatory air sac pressure is essential for their characteristic sweeping notes (Gardner et al., 2001). In zebra finches, the effect is much more modest. Increased sub-syringeal pressure increases fundamental frequency slightly for low frequency notes but this linear relationship

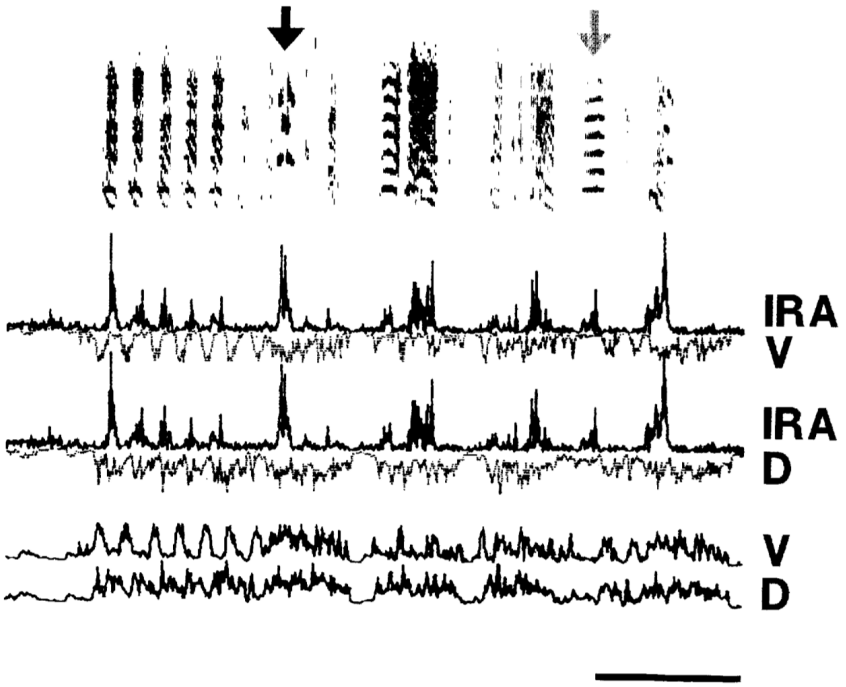

Figure 1.6. Muscle activity of vS and dS during zebra finch song. IRA: Rectified amplitude of the sound recording. V: vS EMG. D: dS EMG. Top: spectrogram of song with introductory notes. Middle/Bottom: Comparison of Acoustic amplitude and EMG activity in vS and dS. Note the muscle activity peaks during gaps in phonation. Gray arrow: harmonic stack, vS and dS are not active. Black arrow: complex note with a high fundamental, vS and dS show activation. Figure from Vicario (1991).

breaks down for high notes (Goller and Riede, 2013). Some high notes in zebra finch song occur during short inspiratory breaths have negative air sac pressure (unlike expiratory syllables) (Goller and Daley, 2001). In general, air sac pressure's major contribution to song is the temporal patterning of syllables. Expiration is driven by abdominal muscles, while thoracic muscles drive inspiration (Wild et al., 1998).

Zebra finch song is generated by a complex dance of the motor periphery, a combination of air pressure generated by the air sac system and the location and tension of the syrinx labia. The unknown interactions of muscles of the syrinx and their unknown contributions to both silent and phonating periods of song limit our understanding of the full configuration of the motor periphery during song. However, the Mindlin model provides an essential first step toward understanding the dynamics of the periphery during songbird vocalizations.

1.4 The zebra finch motor pathway

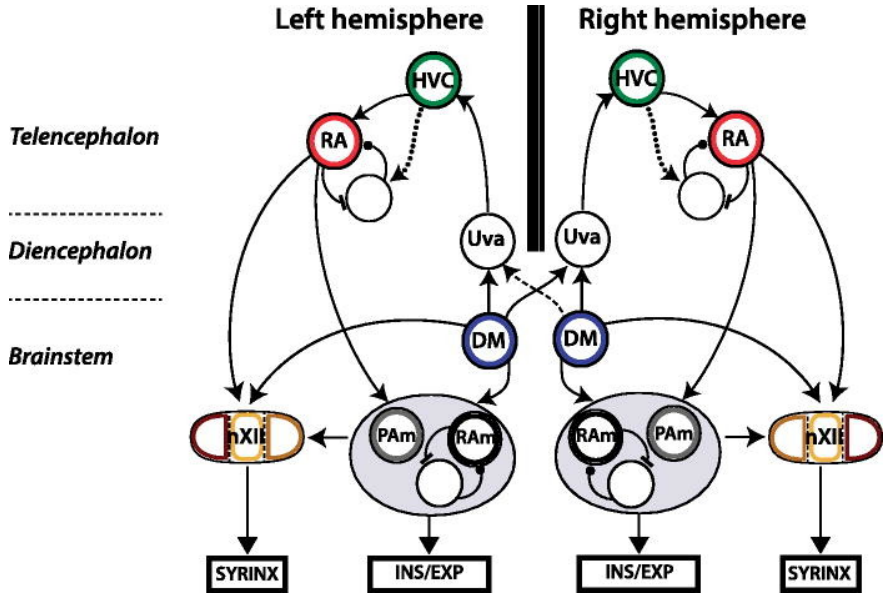

Figure 1.7. The zebra finch motor pathway. Figure from Mindlin (2017).

A schematic overview of the zebra finch motor pathway is shown in Fig. 1.7. All six muscles of the syrinx are innervated ipsilaterally, by the thracheosyringeal branch of the hypoglossal nerve (nXIIts). The nucleus nXIIts receives ipsilateral input from the forebrain central motor area robust arcopallium (RA), which in turn receives ipsilateral input from forebrain sensory-motor nucleus HVC. This strict lateralization allows for the independent control of the two sound sources in the syrinx, while also creating a synchronization problem: the independent control of the sound sources must be carefully aligned (Suthers et al., 1999). The hypoglossal nucleus also receives input from the dorsomedial nucleus of the intercollicular complex (DM) and possibly other respiratory-vocal nuclei of the pons and medulla (Wild, 1997). Stimulation of DM elicits calls, and DM is thought to be the principal output of the call pattern generation pathways in the midbrain (Simpson and Vicario, 1990). DM projects to many of the same brainstem sites as RA (Wild et al., 1997) and is active during both calls and song. DM is unique in that it has bilateral projections, possibly for call coordination.

The generation of air sac pressure required for phonation utilizes the same musculature as

breathing. The expiratory motorneurons controlling the abdominal muscles are innervated by nucleus retroambigualis (RAm) in the caudal medulla (Wild, 1993a). Inspiratory motorneurons are innervated by a rostral ventral respiratory group in the ventrolateral medulla called parambigualus (PAm) (Reinke and Wild, 1997; Ashmore et al., 2005).

The ventral region of RA projects to nXIIts, while the dorsal region projects to RAm and PAm. RA has major projections to nuclei of the pons and medula (Wild, 1997). There are a small number of bilateral projections from RA to the brainstem, though they are unlikely to be responsible for bilateral coordination (Wild, 1993b). There are also projections from RA back to HVC (Roberts et al., 2008).

HVC projects to RA, commonly thought of as the direct motor pathway, as well as to area X, a nucleus in the anterior forebrain pathway (AFP) implicated in song learning. Neurons projecting from HVC to RA (HVC_{RA} neurons) fire a single burst during a zebra finch motif (Hahnloser et al., 2002). HVC neurons projecting to area X (HVC_{X} neurons) also burst, but unlike HVC_{RA} neurons, a given HVC_{X} neuron can burst from one to four times per syllable. (A substantial population both of HVC_{RA} and HVC_{X} neurons are silent during singing.) In contrast, HVC interneurons have tonic activity, and have elevated firing throughout the motif (Yu and Margoliash, 1996). During sleep, the whole motor system can be engaged by replaying the bird's own song (BOS) to the bird. BOS stimulation activates the same song patterning seen during awake singing in HVC, RA and the muscles of the syrinx (Dave and Margoliash, 2000; Chi and Margoliash, 2001; Hahnloser et al., 2002; Young and Goller, 2012), though the respiratory muscles are not engaged, preventing BOS playback from eliciting song.

HVC receives input from thalamic nucleus uvaeformis (Uva), which in turn receives bilateral input from PAm and DM, a likely route for respiratory feedback and bilateral synchronization (Ashmore et al., 2005). Auditory information is routed from the primary auditory area field L through nucleus interfacialis of the nidopallium (NIf) and into HVC. HVC also

projects to Avalanche, a nucleus in the secondary auditory region caudal mesopallium (CM) (Roberts et al., 2017). Avalanche projects to both NIf and HVC, and may well be a critical source of auditory input to the forebrain song system.

There is significant sexual dimorphism in nXIIts, HVC and RA, with males having greater volume than their females (Clower et al., 1989), and lesions to HVC or RA eliminate song (Nottebohm et al., 1976). However, the system does not go unused in females: lesions to HVC prevent adaptive call timing between zebra finches of both sexes (Benichov et al., 2016).

Although not part of the motor pathway, the AFP receives input from HVC via HVC_X projection neurons. The AFP plays a role in developmental learning and adult song maintenance, though HVC_X neurons are not thought to be necessary for moment-to-moment control (Scharff et al., 2000).

1.5 The role of HVC in motor control

The most prominent model of HVC is a feedforward pre-motor model in which HVC_RA neurons are connected in a synfire chain that spans the duration of the motif. This theory requires an unbroken synaptic chain of HVC_RA neurons and an HVC_RA neuron burst associated with every moment of zebra finch motif. In this model, HVC has no association with the motor periphery, and only provides RA with timing information relative to the onset of the motif. This purely time encoding role for HVC is called the clock theory. In the strongest form, the clock theory predicts a uniform distribution of HVC_RA bursts during a motif (Hahnloser et al., 2002; Fee et al., 2004). Cooling experiments have shown that song tempo can be increased and decreased by changing the temperature of HVC, suggesting that HVC plays a major role in determining the timing of song (Long and Fee, 2008). The clock model specifies RA as the site of song motor learning, as synfire chains in HVC are extended to cover the song.

The clock model lacks parsimony with many studies of HVC. There is no clear mechanism

for interhemispheric coordination in the clock model, each HVC has independent synaptic chains of HVC_{RA} neurons and possibly independent HVC_X chains. Unilateral stimulation in HVC interrupts HVC patterning and song bilaterally (Vu et al., 1994), which suggests that there must be some bilateral propagation of information. Flashes of light will only interrupt song between syllables, implying that syllables are a fundamental unit of motor control (Cynx, 1990). In the clock model, there no distinction between times within syllables and times within gaps, and there is no clear reason for why syllables are not immediately interrupted. It has also been demonstrated that vocal exploration and variability is specific to individual vocal gestures (Ravbar et al., 2012). From the perspective of the clock model, learning and variability should be independent for each moment in time. During BOS stimulation, HVC neurons are selective to syllable combinations (Margoliash and Fortune, 1992). In the clock model, activation of a cell is completely dependent on the moment of time (and that moment's associated cell) for input, providing no explanation for combination sensitivity. There is evidence that learning a new gap duration between syllables causes changes in HVC activity (Ali et al., 2013), a phenomenon difficult to reconcile with a chain model, which would predict syllable timing changes to be learned in RA. Furthermore, the feedforward synaptic connectivity envisioned for synfire chains in HVC ignores extensive evidence for connectivity between the different classes of HVC neurons, and their physiological recruitment and interaction during singing.

All of this evidence against the clock model suggests there may be better models of HVC activity. Unlike primary motor areas responsible for fine motor control in primates, HVC does not directly innervate motoneurons, though ventral RA does. The midbrain call nucleus DM likely evolved to coordinate modulation of the already established pattern generating networks in areas like RAm and PAm with nXIIts. In songbirds, forebrain nuclei RA and HVC evolved subsequently to the call system to produce the syllable sequences and spectral elements characteristic of song. Song is impossible without respiration and many zebra

finch song syllables are taken directly from the call repertoire. This modulation cascade has parallels in mammalian thalamo-cortical circuits (Sherman, 2007), and like thalamo-cortical interactions, it is unlikely that brainstem nuclei simply relay commands from RA and HVC, as the clock model predicts. While a synfire chain is an appealing solution for controlling song tempo, memory traces can be used for interval timing, making a continuous synfire chain unnecessary for pattern generation (Staddon, 2005).

An alternate theory for HVC is the application of a gesture model to songbird electrophysiology (Amador et al., 2013), which makes two drastically different predictions about HVC's role in birdsong. This model of HVC is based on observations from inverting the Mindlin model of vocal production to recreated vocal gestures and analyzing the activity of HVC in relation to these gestures. The gesture trajectory extrema (GTE) model was developed by comparing HVC activity with zebra finch songs transformed into putative pressure and tension trajectories, representing the state of the motor periphery over time. The first prediction that arose from these observations is that $\mathrm{HVC_{RA}}$ and $\mathrm{HVC_{X}}$ bursts are not uniformly distributed over the motif, but are tightly synchronized to critical moments in the pressure and tension trajectories (4 ms σ). These moments are gesture onsets, offsets and maxima, termed gesture trajectory extrema. Each gesture can have up to four GTE, representing the times of onset, offset, maximum pressure, and maximum tension. The second prediction of the GTE model, and perhaps the most surprising, is that the $\mathrm{HVC_{RA}}$ bursts associated with a GTE do not precede the GTE in time, as would be predicted by a causal model of motor control, but are aligned to the exact moment of the GTE with zero time delay, forcing a completely new model of HVC's role in the song system, and presumably moving the critical causal role of the songbird forebrain to song motor control to downstream nucleus RA. See Fig. 1.3B for an illustration of the GTE model. The findings that during development protosyllables are first represented in HVC by bursts of projection neurons at onsets, and that syllable sequences are assembled so as to place projection neurons burst at

transitions within a sequence, are additional evidence of a gesture-like code in HVC during development (Okubo et al., 2015). The GTE model of HVC and the Mindlin model of the syrinx are somewhat independent. Improvements to the Mindlin model may cause updates to the GTE model and falsification of the GTE model does not invalidate the Mindlin model.

A limitation of the GTE model is that it does not fully specify a mechanism for song motor control (whereas the clock model does). Instead, the GTE model suggests that HVC contains a forward model of the song, which may be useful for comparing an internal prediction of the movement with auditory feedback (Amador et al., 2013). The forward model of HVC may be involved in "near-time" control of sound generation and can also contribute to song learning in addition to song control (Alonso et al., 2016). Yet, if HVC is primarily a forward model, it is unclear why temperature changes in HVC have such an immediate effect on song. It's also unclear why HVC stimulation interrupts song, or HVC lesions remove song entirely. If the clock model is incorrect, it is likely that song pattern generation is distributed across song motor system, but HVC is still an integral part of the circuit. Respiratory feedback from Uva, which might be critical for motif sequencing, must pass through HVC to reach RA, but it is uncertain how this signal is propagated to RA if motor activity is not preceded by HVC activity. One possible explanation is that synchronized oscillators are distributed across the motor system, allowing HVC signals to synchronize with zero-delay to the acoustic features. The possibility of synchronizing oscillators across the motor system has been discussed in (Mindlin, 2017).

Another major challenge facing the GTE model is the difficulty of inverting the Mindlin model to create pressure/tension trajectories. Once a plausible solution is found, it is experimentally challenging to validate the model by measuring zebra finch syrinx tension directly or evaluating the correct placement of GTE. The GTE model claims exact placement of GTE can be verified by recordings in HVC, though recent experiments cast doubt on that validation method (Picardo et al., 2016; Lynch et al., 2016). Possibly evaluating the response

of the song system to synthetic BOS could provide validation but future work is needed to assess the song systems response to the sensitivity of GTE placement. At present, it is unclear if a 60% response to synthetic song is evidence of correct GTE placement, or if the response is more sensitive to other features, such as the timing and rough spectral content of the syllables (Theunissen and Doupe, 1998). Improvements in song modeling may result in better stimuli to drive HVC neurons, which could help directly test the sensitivity to GTE placement.

Two recent papers sought to directly test the GTE model of HVC. Lynch et al. (2016) recorded from large numbers of projection neurons, including a bird with 136 projection neurons. The GTE model would predict that these bursts would densely cover the times of GTE and be extremely sparse in regions without GTE. However, the authors found that the bursts had 96% coverage of the motif, a finding difficult to reconcile with the GTE model. Picardo et al. (Picardo et al., 2016) used optical techniques to record from large populations of HVC projection neurons in singing head-fixed birds. In both papers, the authors found the projection neuron burst timing to be more likely taken from a uniform distribution than the tightly clustered distribution predicted by the GTE model. There are extensive problems with those two papers, however. One is the reliance of $\mathrm{HVC_X}$ neurons as the major source of data for evaluating the clock model. For example, of the 136 projection neurons in the example cited above, 107 were identified as being either $\mathrm{HVC_{RA}}$ or $\mathrm{HVC_X}$ neurons, but only 14 of those (13%) were $\mathrm{HVC_{RA}}$, contributing only 7% ($136 \times 0.13 = 18$) of the 249 total bursts for that bird. Given that $\mathrm{HVC_X}$ are not known to directly contribute to song production – neither real-time nor near-time (Scharff et al., 2000), this questions the validity of the conclusions as they relate to motor control. However, the original GTE model from Amador et al. (Amador et al., 2013) also did not distinguish between $\mathrm{HVC_{RA}}$ and $\mathrm{HVC_X}$ neurons. Currently, there is no clear hypothesis for which subpopulation of projection neurons are included in the GTE model.

While the broad distribution of HVC bursts found in these papers are unlikely to match the original distributions of the GTE model, Lynch did find fluctuations in the burst density that correlated with syllable onsets, and some variability based on syllable type. Though a uniform model provided a better fit to the data than the GTE model, it is possible that there may be a gesture-like model that provides a better fit to the data than uniformity.

1.6 The role of motor cortex in motor control

Fritsch and Hitzig (1870) performed the first experiments demonstrating electrical stimulation of canine cortex evoked contralateral movement. Their short burst galvanic stimulation protocol evoked muscle twitches, briefly activating muscle groups. The authors concluded that motor cortex is responsible for the control of small muscle groups. Shortly after, Ferrier performed similar experiments, but used long stimulation, with alternating current protocols. This increased activation period resulted in the evolution of complete motor patterns, such as chewing or raising a limb. From these results Ferrier concluded that motor cortex integrates many muscle groups into complex actions. Thus, the two sets of experiments resulted in two hypotheses for the role of motor cortex in motor control, and both groups attributed the alternate results to incompetence (Taylor and Gross, 2003). Leyton and Sherrington (1917) later described motor cortex as a system for integrating "fractional movements" into complete behavioral patterns, a view more consistent with Ferrier's theories. The degree to which motor cortex operates as fine motor control system versus an integrator of complex behavior remains a controversial topic.

1.6.1 Muscle force coding (kinetics)

Kinetic theories of motor control create a framework where cortical motor neurons encode muscle forces. Evarts pioneered the study of primate primary motor cortex (M1) with extracellular microelectrodes (Evarts, 1968). These experiments restricted the wrist movements

of a monkey (*Macaca mulatta*) to only 30 degrees of flexion. The monkey was trained to move a vertical rod between two positions under variable load. This experimental design restricted the motor output to only the wrist flexor and extensor muscles. Evarts found that M1 neurons fired preferentially for activation of a single muscle, with activity beginning 50-150 milliseconds before activation of the muscle, as measured by EMG. Thatch redesigned Evarts experiment to decorrelate force, position and direction of intended movement and found reasonably sized populations of neurons that best encoded each of these three variables individually. M1 neurons can encode static force, such as maintaining a position or posture (Werner et al., 1991). Morrow and Miller (2003) found that a simple linear decoder accurately predicted EMG activity during a precision grip task ($r^2 = 0.97$), though that accuracy dropped dramatically when the same decoder was applied to an alternate power grip task.

Correlations between muscle state and neural activity extend even beyond motor cortex. Santucci et al. (2005) found that recordings from frontal and parietal cortex can accurately reconstruct EMG activity, possibly due to EMG activity being correlated with many signals that are movement related even if not directly encoding muscle activation.

From these experiments and many others (see Ashe, 1997) it is clear evidence for a kinetic motor code in primary motor cortex. Kinetic experiments in the style of Evarts, however, reduce motor output to a single joint rotation. This reduction creates unnatural movement patterns. Realistic movement requires coordinated (and correlated) movement of the shoulder, elbow, wrist and fingers.

1.6.2 External coordinate coding (kinematics)

The next paradigm of cortical motor coding began with the experiments of Georgeopoulos and colleagues (Georgopoulos et al., 1982, 1983), who trained monkeys to perform straight-line hand movements from a center position to one of eight radial positions. They found that

a majority of neurons in M1 responded selectively to the angle of the reach direction. Unlike kinetic theories of motor control in which cortex encodes muscle states, these neurons respond to changes in position relative to an external coordinate system, or kinematic state. However, rather than uniquely coding a single movement direction, M1 neurons appeared to have bell-shaped tuning curves, firing at gradually lower firing rates as the movement direction became further from the neuron's preferred direction. A population vector code accurately decoded the movment targets. The population vector algorithm was also consistent with three-dimensional movement tasks (Georgopoulos et al., 1988). In tasks where a constant, "bias" force was applied to the hand during the task, the population vector matched the dynamic component of movement and not simple sum of the static and dynamic forces on the hand, as might be expected from a muscle-centric model, consistent with a cortical model of movement direction (Georgopoulos et al., 1992).

The population vector decoding algorithm also explains more complex trained movements. Finger trajectories could be reconstructed from monkeys trained to trace spiral patterns (Moran and Schwartz, 1999). The delay between cortical coded direction and actual direction varied linearly with the radius of the curve – tight curves resulted in longer prediction intervals. Unlike primary motor neurons, pre-motor neurons had a bi-modal distribution of delay times, with about half the cells centering around a zero millisecond delay. This is reminiscent of one prediction of the GTE model (above).

The kinetic view of motor coding has two fundamental differences from the kinematic view. The first is that cortical neurons encode abstracted dimensions of motor control: positions in an external coordinate system instead of the internal measures of activation forces of single muscles. Second, population codes change the role of individual neurons. Rather than small sub-populations taking complete control over a muscle, whole body movement is governed by a democracy of the whole motor cortex.

While there are experiments consistent with this theory of motor control, it's unclear

whether population vector codes are an organizing principle of motor control or a spurious result, due to the summation of many observed variables which happen to correlate to an external coordinate system. This theory was first proposed by Mussa-Ivaldi (1988), who argued that a simple kinetic code in motor cortex, where firing was related to muscle activation, resulted in population activity consistent with a population vector code, due to correlation between internal states and external coordinates. Similarly, Todorov (2000) argued that if biomechanical factors that modulate output force, muscle length and rate of lengthening, are modeled, then a purely kinetic model of cortex displays a code that appears to be a kinematic population vector, like that found by Georgopolous and colleagues.

Paninski et al. (2004) used a more complex behavioral task, in which monkeys followed a random visual stimulus, reducing some of the correlations in simpler designs (Paninski et al., 2004). They found a diversity of position- and velocity-tuned cells in M1. Velocity-tuned cells had a peak at 100 ms causal delay while position-tuned neurons had both positive and negative delays, possibly having both feed-forward and feedback inputs. While the external coordinate coding of position and velocity was significant, the instantaneous information was "surprisingly small", and the authors suggest that M1 may encode variables that are indirectly linked to the external coordinates of the hand. The authors also suggest that single electrode techniques may have biased previous studies by oversampling neurons that are highly modulated by hand position.

1.6.3 Internal coordinate coding (kinematics)

Caminiti et al. (1991) extended the 3D reaching task of Georgopolous by constructing three identical work spaces to the left, center and right of the animal. They found that the tuning curves of neurons rotated across workspaces in a manner consistent with a shoulder-centric coordinate system. Human errors in remembering points in 3D space are most consistent with a shoulder-centric coordinate system (Flanders et al., 1992). Wu and Hatsopoulos (2006)

found evidence for not only external coordinate representation, but shoulder coordinate and shoulder/elbow joint angle coordinate systems were all plausible M1 coding strategies, with no clear single representation within the neurons from which they recorded. The authors suggest that either the common coordinate system for motor cortex does not exists, or has yet to be discovered.

1.6.4 Gesture coding

In 1951, Karl Lashley posed the "problem of serial order in behavior" (Lashley, 1951; Rosenbaum et al., 2007). At that time, behaviorism dominated scientific dialogue in physiology and psychology, emphasizing the reduction of behavior to elemental reflexes. A behaviorist approach to creating sequential behavior is to chain reflexes together. For example, the first spoken word in a sentence might create an auditory stimulus that would in turn cause the second word to be pronounced, and so on. In this way, sequences may be built from chains of elemental reflexive components. Among other issues with this theory, Lashley gave examples of speech errors that are not consistent with a chain model, such as spoonerisms, where the first syllables are swapped in the words of a phrase, such as "blushing crow" instead of "crushing blow". These kinds of errors indicate some knowledge of the full sequence during its execution. Lashley proposed an alternative: sequences are built not from chains but from hierarchies or levels of abstraction.

The same problem exists in the control of motor sequences by cortex: are sequences generated by chaining kinetic (or kinematic) commands or by creating levels of abstraction?

One mechanism of extending time-independent kinematic commands to sequences is to allow preferred directions of movement to evolve over time, or become time dependent. Hatsopoulos et al. (2007) found that by allowing preferred direction to be time dependent, the instantaneous error of tracking estimation was significantly reduced. These evolving direction codes were deemed "pathlets", which encode arm gestures over 400 ms.

Griffin et al. (2015) recorded from primate corticomotoneuronal (CM) cells in M1 that project directly to motoneurons that innervate single skeletal muscles. The connectivity of CM neurons directly to motoneurons implies they are more likely to have a muscle activation code specific to a single muscle. One might expect these neurons to have a simple kinetic code. Surprisingly, the population of CM neurons for a single muscle are not a homogenous population. Less than half of CM neurons are maximally tuned to within 45 degrees of their target muscle's movement direction. A better model for CM neurons is that they fire selectively for the functional purpose of the muscle, such as when the muscle is a synergist or antagonist of the desired motor behavior. This functional motor code is an abstraction from pure muscle activation, suggesting a hierarchical code that places muscle activations within the context of a complete movement.

Hocherman and Wise (1991) studied arm trajectories in three conditions: a direct movement from the start to end point, a movement with clockwise curved trajectory and a movement with a counter-clockwise curved trajectory. A standard population vector model would predict that neurons should encode the instantaneous velocity of the hand. However the authors found populations of neurons in both motor and premotor areas that responded to the trajectory type. Not all neurons had gesture specificity, others encoded target location or initial direction (<20%), the latter being most consistent with external kinematic coding.

If gestures are being generated by motor cortex, are the instructions for the complete gesture or just an abstract description of the start and end points? Without visual feedback, monkeys can make correct movements to an end point even with perturbations to the trajectory (Polit and Bizzi, 1979). However, movement path curvature and speed appear to obey a 1/3 power law, suggesting the entire gesture is under cortical control and not just end points (Viviani and Stucchi, 1992), though some simulations are able to account for the power law without trajectory planning (Perrier and Fuchs, 2008).

Gesture coding neurons have been found in human supplementary motor cortex (Bouchard

et al., 2013), where there appears to be a phonetic map with motor precise timing.

1.6.5 Sequence coding

Another mechanism to extend kinetic and kinematic codes to sequence behavior might be circuits or sub-populations dedicated to sequence coding that could then instruct population vectors to evolve over time. This hypothesis is distinct from the gesture hypothesis because it creates a division between motor commands and timing, and does not rely on hierarchical abstraction. Carpenter et al. (1999) presented visual targets for center-out movement tasks, but displayed them in a random serial order, after which a single target was highlighted for the animal to move towards. During the movement period, 67.5% of cells encoded motor direction, but during the serial presentation of stimuli, 71.6% of neurons had selective response to either serial order or position of the targets. Serial order selective neurons had a phasic response and did not continue firing after the event they encoded, which suggests a distributed network for sequence encoding. A majority of motor direction neurons were also sensitive to serial order, which suggests that sequence coding is coupled with motor coding, rather than an independent circuit.

Another question is how highly trained sequences may be encoded differently than untrained sequences. Matsuzaka et al. (2007) trained monkeys for over two years on a visually guided sequencing task. Sequences were either random or one of two repeating patterns. These over-trained monkeys had neurons that responded specifically to sequence movements, but not to random untrained sub-sequences. This finding indicates that M1 neurons are capable of encoding abstract sequences instead of kinetics or kinematics.

Of course, the millisecond timing and sequencing of motor plans is unlikely to be restricted to primary motor cortex (Buhusi and Meck, 2005). The cerebellum has long been known as a location for millisecond timing of motor commands (Bracha, 2004). The pre-supplementary motor area (preSMA) and supplementary motor area (SMA) are involved in performing

motor sequencing (Shima and Tanji, 1998). Mita et al. (2009) recorded from preSMA and SMA neurons during a grasp and hold task requiring holds ranging from 2 to 8 seconds. In both areas, neurons exhibit duration selectivity. Most neurons displayed exponential firing rate decays or rises, which may be used to encode the passage of time. Phasic neurons (another category of neuron in preSMA) fired maximally at a fixed interval from the onset cue, possibly providing input to time-grading neurons. This motif may also be repeated in M2, where phasic neurons and integrator cells sensitive to the timing of behavior also are found (Murakami et al., 2014). Using phasic neurons as a clock has also been proposed in models of central pattern generation in the spinal cord (Grillner, 1981; Guertin, 2009).

1.6.6 Dynamic coding

So far, I have described two mechanisms for extending motor codes to sequences: chaining motor commands together with dedicated clock circuits and creating abstract gesture codes – neurons that do not directly encode movement parameters. A third hypothesis is dynamic coding (Shenoy et al., 2013; Gallego et al., 2017). Dynamic coding theory begins with a fundamentally different view from previous theories, especially kinematic and gesture coding: motor cortex is for *generating* movement not *representing* movement (Scott, 2008). With this perspective, it is not surprising that no single kinematic model captures the activity of all individual neurons. The output of the network is state dependent, evolving over time and determined by the underlying network connectivity, and while some neurons affect muscles directly, others represent the internal dynamics of the network. Internal dynamics are more important in modeling M1 dynamics than sensory areas such as V1 (Seely et al., 2016).

Scott (2008, 2004) argues that the multitude of kinetic and kinematic codes in motor cortex fail to provide a unifying framework for motor control. Neurons sensitive to movement direction do not uniformly distribute themselves among all directions, rather their preferred directions are biased towards directions of peak joint power of the elbow and shoulder (Scott

et al., 2001). Scott proposes a model based on optimal feedback control, a framework developed for robotics, where motor neurons do not encode output states, but rather provide input to the motor system, and control signals from cortex depend on the mechanics of the motor periphery.

The output motor neurons that do control the periphery may be operating under principles of optimal control rather than representing kinematics (Scott, 2004). For example, errors in population vector co-vary with peak joint power at the shoulder and elbow (Scott et al., 2001).

Population vectors treat each neuron as independent, and a wide variety of activity across neurons can result in the same population vector. However, individual neurons are not independent, and the space of measured activity states is far smaller than is theoretically possible with independent neurons. In fact, just a few dimensions are required for accurate reconstruction of motor cortex activity, suggesting that motor cortex dynamics live in a low dimensional manifold (Churchland et al., 2012; Cunningham and Yu, 2014). Commonly, principal component analysis (PCA) or factor analysis are used to find a reasonable approximation to the neural manifold. The dimensions of the neural manifold are termed latent variables. Within the manifold, variability is reduced as the animal prepares for movement (Churchland et al., 2010), possibly getting the system in the correct dynamical state to execute the desired movement patterns. Some latent variables are output-potent – movement along this dimension results in motor behavior, while others are output-null. Movement along output-null dimensions may be used to prepare for movement (Kaufman et al., 2014). Low dimensional neural ensembles have been identified in invertebrate motor patterns (Bruno et al., 2015). Finally, a low-dimensional manifold for motor control is also consistent with behavior constructed from motor primitives (Giszter, 2015; Graziano, 2006).

1.6.7 Limb versus vocal motor control

In human speech, the issue of motor control intrinsic or extrinsic coordinates is further confounded by the semiotic or symbolic nature of speech. The motor output of speech can be highly variable without stripping the action of its communicative intention (Grimme et al., 2011), which may mean that the cortical control of motor output is in the auditory domain (Blumstein and Stevens, 1979; Stevens, 1989). Using the acoustic domain or spectral features of speech for motor control is a type of external coordinate system. Alternately, speech control may be primarily motor, emphasizing the configuration of the speech articulators, as hypothesized by the motor theory (Liberman and Mattingly, 1985). The motor theory of speech production emphasizes cortical control in an intrinsic coordinate system. Within the speech community, the correct coordinate system has been the subject of decades of controversy, mirroring debates in primate reach and grasp control.

Limb and vocal control tasks also face different inertial regimes and timescales. The positional changes in speech primarily use soft tissue of relatively low mass, compared with the skeletal muscles used in limb movements. Limb movements must also incorporate external forces in tasks like lifting heavy objects, while speech is rarely modulated by external forces except gravity on the speech articulators. However, due to the fast acceleration of the articulators during speech, there are likely some inertial effects that require central control. The high rate of movement in speech may require different control paradigms, as the central mechanisms may not have time to incorporate real-time feedback. The fast articulatory gestures of human speech are a plausible basis for motor control because of their spatiotemporal nature and their development as pre-linguistic units in infants (Browman and Goldstein, 1986). Inter-gestural control is critically important for connecting phonemes but is a largely unexplored phenomenon in limb control.

Motor control in speech and limbs are both part of greater systems for interacting with the outside world, and their commonality may be best described by embedded cognition, where

models of cognition are understood by careful examination of the sensory signals and motor output of biological systems. Just as limb movement may be generated by low-dimensional neural manifolds, speech is likely generated under similar constraints (Fowler et al., 2008; Schneegans and Schöner, 2008).

In conclusion, there is considerable breadth in the approaches and emphasis of models of motor control in non-human primates and human vocal motor control, yet strikingly little overlap with the major theories of vocal motor control in songbirds. Perhaps this divergence is due to fundamentally different evolutionary constraints on the behaviors, which is then reflected in different activity patterns in mammalian motor cortex and motor control nuclei of songbirds. Whereas cortical neurons can be well described by probabilistic models such as Poisson processes, zebra finch HVC and RA projection neuron are practically deterministic in their firing behavior over time. Most striking is the single, time dependent burst of an $\mathrm{HVC_{RA}}$ neuron, where it is impossible to determine any meaningful relation to the panoply of time-varying motor features, at least at the level of the individual neuron. Though single neuron models from primate motor control may not inform our understanding of HVC, population models such as those proposed by dynamic coding, may have parallels in the song system.

1.7 Goals of this thesis

To date, a limitation of birdsong forebrain electrophysiology has been the use of the spectrogram to represent the motor output of the system. Both the clock and GTE models compare HVC activity to spectral features. In the GTE model, spectral features are transformed into pressure and tension estimates, but tension cannot be measured directly with current techniques and pressure measurements alone are insufficient as a large range of pressure/tension parameters produce the same fundamental frequency (see Fig. 1.4) which results in some uncertainty about the gesture trajectories. This uncertainty is particularly acute in zebra

finches, where the syrinx muscles have a complex and poorly understood effect on labial tension. Synthesized song can be presented to sleeping birds to validate gestures recovered from Mindlin model inversion, but it is unclear from a partial response (the best models result in 60% activation of HVC) whether the model is correct, or which part of the model remains incorrect. The largest variation in sleeping response to synthesized BOS is changes to the timing of whole syllables and the spectral distribution rather than fine-tuned changes to the pressure-tension trajectories (Amador et al., 2013; Theunissen and Doupe, 1998). The GTE model also only measures motor activity during phonation, missing essential preparatory motor behaviors where syringeal muscles are maximally activated (Vicario, 1991; Goller and Suthers, 1996b). A simplification of the biophysical model of the motor periphery to try to make it available to the larger birdsong community uses constant pressure during phonation and fits the tension parameter to the time-varying pitch of the vocalizations (Boari et al., 2015). It remains uncertain, however, how much utility will result from the gestures predicted by this simplified model – in the largest collection of HVC projection neuron data to date, the simplified model failed to provide any evidence for a GTE model in HVC (Lynch et al., 2016).

To address these outstanding issues in songbird motor control, I used two technological advances to provide a novel perspective. First, I used dense micro-electrode arrays to sample large populations of HVC neurons, as population recordings are an attractive approach for teasing apart the differences in predicted global HVC activity generated from the clock and gesture theories. Second, I recorded from the muscles of the syrinx directly. Potentially this presents some experimental difficulties in interpretation: the weak correlation between zebra finch song pitch and EMG activity may hide any relation between HVC and the motor output, particularly preparatory movement. Nevertheless, recording from syringeal muscles during song playback in acute (Chapter II) and chronic (Chapter III) preparations and during singing (Chapter IV) has several important features. It provides insight into the

state of the vocal periphery while eliminating the complexity of inverting the Mindlin model. Any such results would be more accessible to the larger scientific community. Second, this experimental design can directly test whether HVC population activity is divorced from the motor periphery, the central tenet of the clock model but one that has never been directly experimentally tested.

In the following chapters, I will demonstrate correlation between HVC and the motor output. My hope is that this work will move songbird motor control research towards better integration with the broader field of motor control.

CHAPTER 2

GLOBAL HVC ACTIVITY DURING AUDITORY PLAYBACK

2.1 Introduction

In songbirds, acoustic stimulation can drive the motor system. This was first reported for HVC multi-unit recordings in two awake canaries and one awake white-crowned sparrow (McCasland and Konishi, 1981). The first comprehensive single unit analyses were conducted in HVC of anesthetized (Margoliash, 1983) and later including awake white-crowned sparrows, (Margoliash and Konishi, 1985; Margoliash, 1986), awake starlings (Leppelsack et al., 1986) and anesthetized zebra finches (Schmidt and Konishi, 1998; Margoliash and Fortune, 1992), RA in sleeping zebra finches (Dave et al., 1998; Dave and Margoliash, 2000), and HVC of awake swamp sparrows and Bengalese finches (Prather et al., 2008). Whereas there is species variation in the effectiveness of auditory playback in awake birds, auditory playback presented to anesthetized or sleeping birds is an effective stimulus for all songbird species tested to date. The behavioral state dependent nature of sensory responses in the song system have been extensively studied (Dave et al., 1998; Schmidt and Konishi, 1998; Cardin and Schmidt, 2004). Here I focus on zebra finches.

When BOS is presented to a sleeping zebra finch, the motor pathway activates, entraining to BOS. Individual neurons in the song system motor nuclei are selectively activated for BOS, both in HVC (Rauske et al., 2003) and in RA (Dave and Margoliash, 2000). During sleeping replay, zebra finches also contract the muscles of the syrinx in the same pattern as awake song (Young and Goller, 2012). The anterior forebrain pathway, a pathway downstream of HVC implicated in song learning, is also activated by song replay (Solis and Doupe, 1997). Song replay activity is statistically similar to singing activity (Dave and Margoliash, 2000; Prather et al., 2008; Amador et al., 2013) suggesting that the same circuits are activated in singing and song replay. To date, there has been no direct comparison of individual HVC

neurons during both awake song and sleeping song replay in zebra finches, yet it is commonly assumed that measurements of HVC BOS response can be generalized to singing behavior, (see Amador et al., 2013; Hahnloser et al., 2002). Direct comparisons of awake individual HVC neurons during singing and in response to song playback in other species also supports this conclusion (Prather et al., 2008).

In a natural setting, a zebra finch never hears his own song while he sleeps. The selectivity of the response to BOS is sufficiently high that songs from conspecifics are extremely weak stimuli (Margoliash and Konishi, 1985; Margoliash and Fortune, 1992), raising the question of why the BOS response exists. During sleep, the motor system is spontaneously active for short periods of time, and the activity during these periods matches subsets of singing activation (Dave and Margoliash, 2000). It is likely this spontaneous replay is a type of off-line learning (Derégnaucourt et al., 2005). Song replay induced by BOS presentation may activate this sleep-related behavior. Rauske et al. (2010) found that RA neuron burst patterns change overnight, and these new burst patterns appear during song the following morning, providing strong support for this hypothesis. Given the apparent ubiquity of the replay phenomenon, and the common observation of auditory responses in awake birds of most species, perhaps the correct way to phrase this question is: "why are song system auditory responses so suppressed in zebra finches?" (Rauske et al., 2003). This remains unresolved.

Currently, there are two leading models which make predictions about the population activity in HVC. The GTE hypothesis predicts HVC activity to be modulated at times of GTEs – key points in the gestural trajectory of the peripheral motor state during song. In singing birds, HVC projection neurons were observed to burst at GTEs, while local minima in the firing rates of interneurons aligned with GTEs (Amador et al., 2013). Similar responses were observed in sleeping birds, considering well-isolated single-unit HVC neurons that were responsive to BOS playback. Another key point of the gesture hypothesis is that

HVC activity relates to GTEs with zero time delay, suggesting HVC cannot be a premotor structure in a simple feed-forward manner. More broadly, the GTE model predicts that HVC dynamics are dependent on syllable type and spectral content. The zero lag aspect of the GTE model can only arise if there are predictive mechanisms for adjusting the timing of HVC auditory responses (which in a "standard" model should lag) and singing related activity (which in a "standard" model should lead). There is extensive evidence that at least some song system auditory responses exhibit predictive response properties, integrating over multiple syllables (Margoliash, 1983; Margoliash and Fortune, 1992). The correspondence between times of auditory response and singing related activity in zebra finch RA (Dave and Margoliash, 2000), and in HVC in swamp sparrows and Bengalese finches (Prather et al., 2008) demonstrate that the internal dynamics driven by auditory stimulation are likely to be the same as motor driven dynamic activity during song. This correspondence suggests that the causal premotor system can align to auditory replay, which requires some degree of prediction.

Alternately, the clock model suggests that the activity of HVC projection neurons should be largely invariant to syllable type and structure, as the clock model does not allow for motor coding in HVC. Initially the clock model was proposed only for HVC_{RA} neurons. More recently it has been extended to include HVC_{X} neurons (Lynch et al., 2016; Picardo et al., 2016). Discussions and modeling studies of HVC synfire chains that arise from the clock model do not propose a role for HVC interneurons in structuring activity during singing, (but see Kosche et al. (2015)). Proponents of the clock model have found a slight increase in activity of projection neurons 29 ms before the onset of a syllable, but attribute this to a residual trace of song learning processes during development, and not directly to encoding any information about the state of the motor periphery at the onset of phonation (Okubo et al., 2015; Lynch et al., 2016; Troyer, 2016). In contrast to the GTE, any global fluctuations in HVC activity should provide essentially zero information about the motor state.

The gesture and clock models generate different predictions for global activity in HVC. Silicon micro-electrode arrays (MEAs) can sample from 16 to 64 different locations in a spatially extended grid than can span a large percentage of HVC. MEAs are an appealing approach for measuring global activity and have been used to track traveling LFP waves in other forebrain regions (Beckers et al., 2014). Dense probes – probes with distances between electrode pads of less than 100 microns – have pads with overlapping recording fields. They can help isolate single units and can even recover overlapping waveforms from neurons firing simultaneously (Yger et al., 2016).

In this chapter, we use MEAs to sample large populations of HVC neurons during song replay to test predictions generated by the gesture and clock models of HVC.

2.2 Methods

2.2.1 Song recording

Adult male zebra finches were placed in sound isolated chambers for one to three weeks. They were given unrestricted access to food and water and placed on a 9 p.m. to 9 a.m. reverse light cycle. Sound recordings were collected from Berhinger C-2 condenser microphones and saved to disk with custom software[1] running on the JACK audio framework. For each bird, a representative bout of song was selected and extracted from the song recordings (BOS), typically lasting 1-2 seconds and containing at least two full motifs and introductory notes. The maximum amplitude was normalized -10 dB RMS, and a reversed copy was created (REV). For each bird, a conspecific (CON) was also selected.

1. https://github.com/melizalab/jill

2.2.2 Surgical implantation

At least 24 hours before the first experiment, a brass head fixation pin was implanted on the skull. A custom brass pin was developed to prevent the MEA from contacting the pin during the experiment, Fig. 2.1. A 1-72 tap was used to allow the pin to be secured to a custom experimental stereotaxic apparatus.

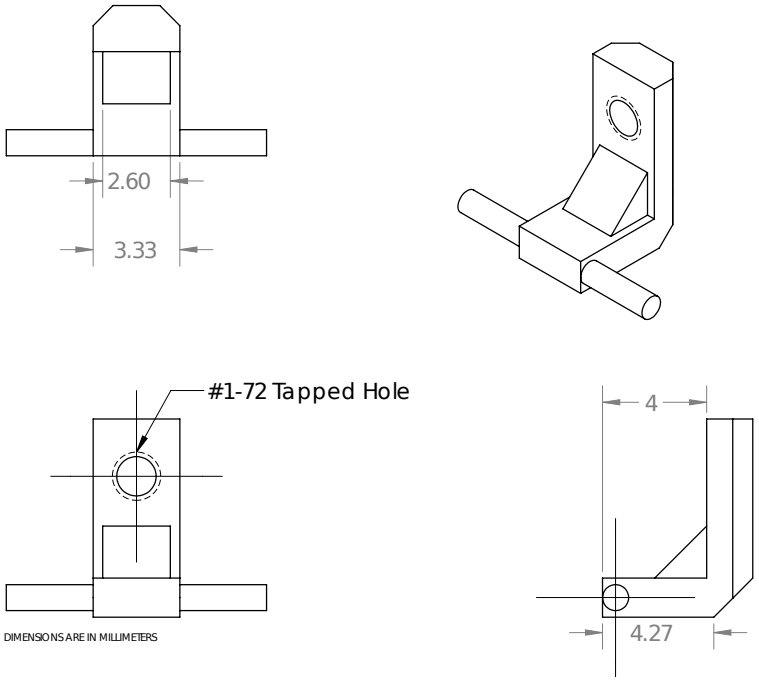

Figure 2.1. Custom brass head pin. Custom brass head fixing pin which prevents the stereotaxic apparatus from interfering with MEA placement.

Before the pin implantation surgery, the bird was first deprived of food for 1 hour, then placed under 2% isoflurane anesthesia at a rate of 2 liters per minute. Isoflurane concentration was adjusted such that the breathing rate remained between 1-2 Hz. The head position was aligned with a custom stereotaxic rig, fixing the beak at 45 degrees from horizontal. The custom 3D printed brass pin was placed 4 mm caudal to the midsagital sinus bifurcation and secured with dental acrylic to the skull. A small well of dental acrylic was built, centered around the mid sagital sinus and encompassing both HVCs. After the surgery, the bird was monitored for several hours to ensure a full recovery.

2.2.3 Experimental paradigm

All experiments were performed within the night cycle of the bird, during the period where the home cage lights were off. Within 1-2 hours of the start of the bird's night cycle, the bird was removed from the home cage and placed in a sound isolated experiment chamber. The bird was secured with the implanted pin, which stabilized the head, and wrapped in a cloth jacket, which restricted the body in an orientation comfortable for sleep. A craniotomy was prepared over HVC (0.3 mm rostral, 2.3 mm lateral), and the dura above HVC was removed. A second small craniotomy was placed on the same hemisphere 5 mm rostral to HVC and a silver grounding wire was implanted between the dura and skull. The acrylic well, which included both craniotomies, was filled with saline to protect the brain from exposure to air. A MEA probe was slowly lowered into HVC at a rate of 50 microns per minute, 30 minute period prior to the beginning of the recording session to allow the brain to relax from any deformation created by probe insertion. HVC was identified by increased activity from the MEA pads in response to song playback. The probe was lowered until the entire array was in HVC or until the lowest pads descended to the bottom of HVC. The models used were: A1x16-3mm-50-177-A16, A4x8-5mm-50-200-177-CM32, A4x2-tet-5mm-150-200-121-CM32, A1x32-Poly3-5mm-25s-177-CM32, A1x32-Poly3-6mm-50-177-CM32 from NeuroNexus Inc, and H2 from Cambridge Neurotech. MEAs were chosen such that the entire array could fit inside HVC, Fig. 2.2.

Data were recorded using an Intan RHD3212 amplifier connected to an Intan EVAL board and saved to a Linux workstation running the Intan Interface GUI. Data were recorded at a 16 bit depth and 30 kHz sampling rate. Stimulus data were generated by custom JACK software and presented to the zebra finch via a ribbon speaker (North Creek Music Systems). A copy of the stimulus and a timing pulse was sent to the EVAL board to synchronize the neural and stimulus data. Stimuli were 20 to 100 shuffled copies of BOS, REV and CON (conspecific song), with each stimulus preceded by 20 seconds of silence. Experiments lasted

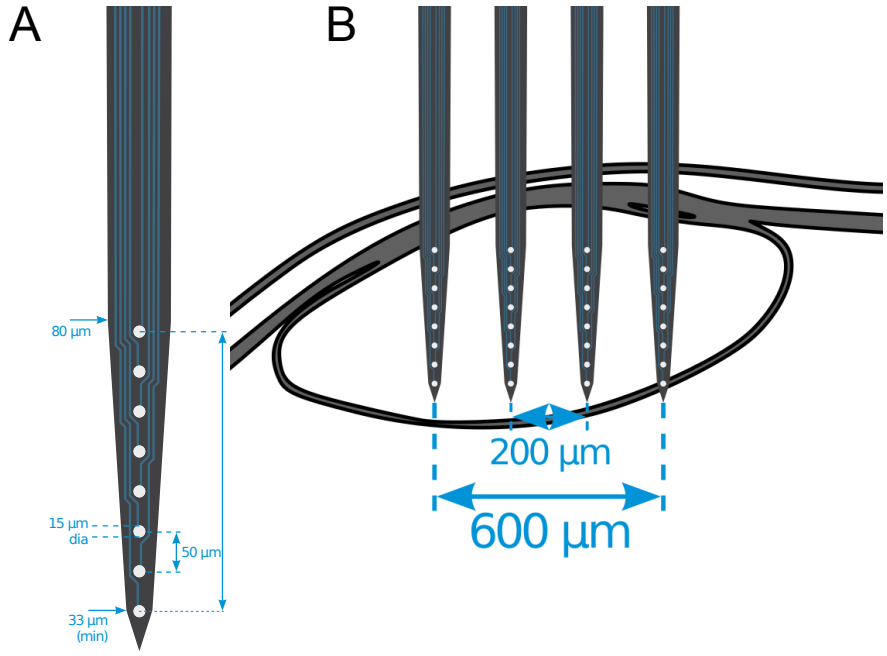

Figure 2.2. Probe geometry. A: geometry of a MEA single shank, Neuronexus A4x8-5mm-50-200-177-CM32. B. Comparison of the MEA geometry and a scaled sagittal section of HVC. Figure images adapted from www.zebrafinchatlas.org and www.neuronexus.com

no longer than four hours and the bird was given at least 48 hours to recover before any additional experiments. The vigilance (sleep) state of the animals was not monitored.

Upon the final experiment, the bird was perfused with heparinized saline and fixed with 10% neutral buffered formalin. Brain sections were stained with cresyl violet and the electrode tracts were identified to verify electrode placement within HVC, Fig. 2.3.

2.3 Data Analysis

Forty-one successful experiments were performed on ten birds.

The continuously recorded voltage data from each MEA pad was filtered and spike sorted using Spyking Circus (Yger et al., 2016). Spyking Circus takes advantage of the many to one mapping from MEA pads to individual neurons to determine spike waveform templates.

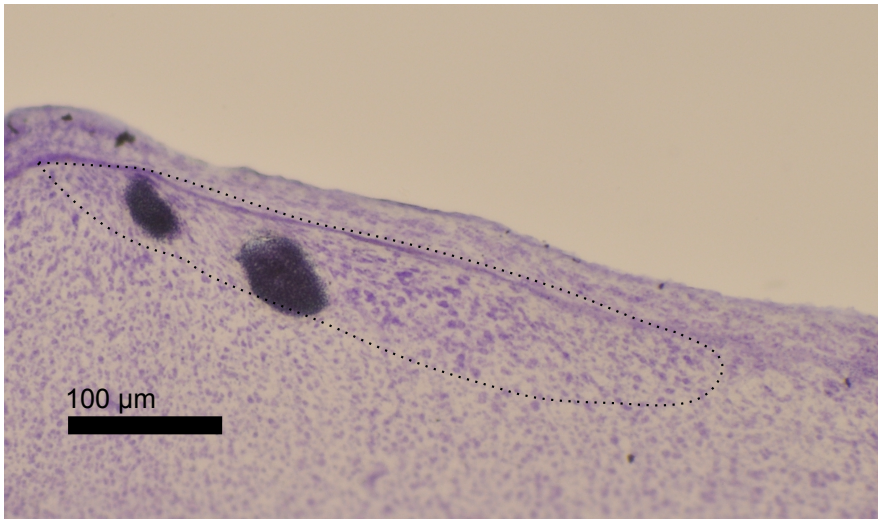

Figure 2.3. Electrode tract histology. A cresyl violet stained section containing two lesions in HVC created by MEA shanks. Dotted line: HVC border.

Manual review of putative spike clusters was performed with Phy[2]. Units were considered "good" if they met the following critera: a large amplitude, an asymmetrical waveform, and interspike-interval (ISI) histogram with few refractory period violations, see Fig. 2.4

Clusters failing these criteria were considered to be multi-unit activity (MUA). This is a conservative approach given that in HVC neurons tend to have highly synchronous activity in response to song playback (or singing; see Chapter 4), which will tend to flatten cross-correlograms.

Transient voltage artifacts due to bird movement were quite rare in these acute recordings but easily identified and discarded during spike sorting. Most recordings had no artifact clusters.

2. https://github.com/kwikteam/phy

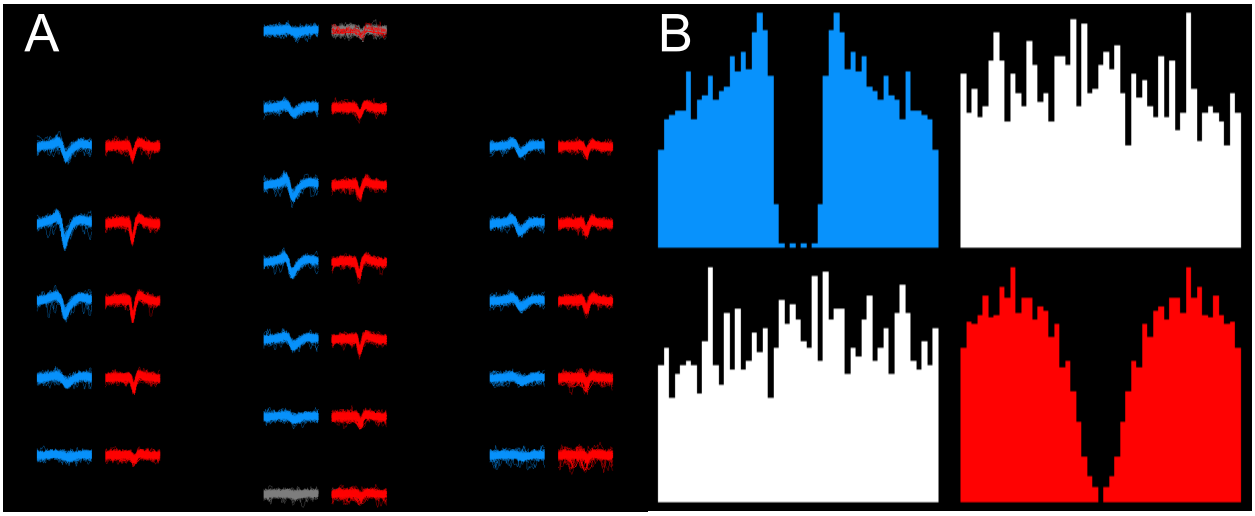

Figure 2.4. Spike sorting with Spyking Circus. Two "good" clusters of spikes representing two distinct units from HVC. This MEA had a single shank with three rows of electrodes spaced 25 μm apart (Neuronexus model A1x32-Poly3-25s). A: example waveforms of two units (2 ms in duration), in red and blue, spanning the top half of the array. B: ISI histograms of the two units in red and blue, and their cross-correlograms. Unique ISI distributions and flat cross-correlograms indicate that these are two distinct units. Each bar of the histograms represents 1 ms.

The firing rate, $r(t)$, was calculated as described in (Dayan and Abbott, 2001)

$$\rho(t) = \sum \delta(t - t_i)$$
$$r(t) = \frac{1}{\Delta t} \int_t^{t+\Delta t} d\tau h(\tau) \langle \rho(t - \tau) \rangle$$

Where $\rho(t)$ is the neural response function over a single trial, $\Delta t = 1$ ms, $h(t)$ is a smoothing kernel. Unless otherwise noted, all firing rate functions were smoothed with a Gaussian kernel with a 5 ms standard deviation.

For each bird, only the best experiment, as measured by total BOS response relative to baseline, was used for further analysis.

Spectral analysis was performed by custom software using multi-taper estimates of the short time Fourier transform. Multi-taper parameters were matched to the defaults of Sound Analysis Pro (Tchernichovski et al., 2000), (9.27 ms windows, 1.36 ms stride, 1024 point

DFT, 2 tapers and a NW parameter of 1.5). Spectrograms were visualized using the maximum spectral derivative, which helps visualize fine time-frequency variation in the spectrograms. For detailed description of multi-taper spectrograms and the maximum spectral derivative see Lusignan (2012).

For each stimulus, event labels were annotated with Praat[3]. These events include the syllable identities, their type, and the inter-syllabic transitions. Accepted syllable types were: noisy (syllables with broad power and little harmonic content), complex (syllables with distinct sub-elements and harmonic variation), flat (syllables with a single pitch, i.e. harmonic stacks), slide (syllables with a decreasing pitch, i.e. down-sweeps) and high (syllables with a fundamental frequency greater the 2 kHz). From the BOS stimulus the motif was selected for analysis, if multiple motifs were presented, the second motif was analyzed. All hand annotations of the spectral data were blind to the neural data.

Two acoustic features were generated from the stimuli, local maxima of amplitude speed (LMAS) and spectral edges (LMSE). AS was computed from the spectra by first computing the amplitude and then taking the absolute value of the derivative. The derivative was estimated with a Savitzky-Golay filter, which fits a third order polynomial to consecutive 60 ms windows of data. Both AS and SE were computed from the real-valued power spectra $P(t, f)$,

$$AS = |\frac{d \int df P(t, f)}{dt}|$$

SE was computed by convolving the spectrogram with an edge detection filter $k(t, f)$. A Gaussian gradient magnitude kernel was used, with a 2 ms standard deviation in time and

3. http://www.fon.hum.uva.nl/praat/

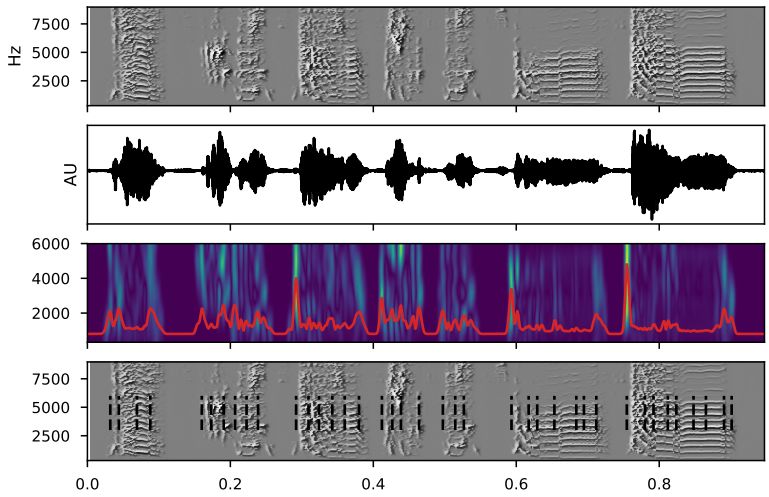

Figure 2.5. Spectral edges. Top: Spectrogram and oscillogram of BK196. Middle: Spectrogram convolved with an edge detection filter. Sum over frequencies is shown in red. Bottom: Spectrogram and the local maxima of the spectral edges (LMSE).

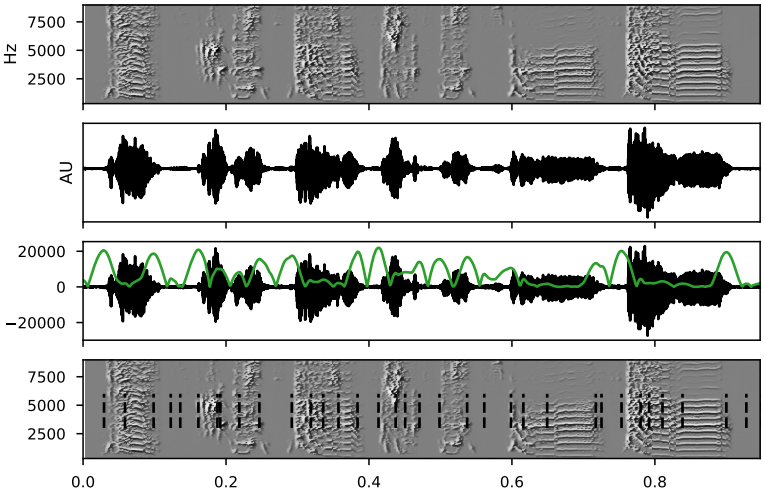

Figure 2.6. Amplitude speed. Top: Spectrogram and oscillogram of BK196. Middle: Oscillogram and the smoothed absolute value of the amplitude derivative in green (amplitude speed). Bottom: Spectrogram and the local maxima of the amplitude speed (LMAS).

a 500 Hz standard deviation in frequency. Fig. 2.5 gives an example of LMSE. and Fig. 2.6 gives an example of LMAS.

$$SE = \int df P(t, f) * k(t, f)$$

LMAS and LMSE were computed from the local maxima of AS and SE. If two maxima were found within 15 ms, the smaller maxima was discarded.

The distance between HVC events and spectral events was computed on a per-syllable basis. Only events within the syllable or within a 5 ms window on either side were included for analysis. For each HVC event, the distance to the closest spectral event was recorded. On average, each bird had 11.2 $\pm$ 2.9 local minima within these syllable windows (N=112). Synthetic minimum distance data were generated in the same method but with a random offset (range [0, L] where L is the length of the syllable window) added to all minima within a syllable. Synthetic HVC events that fell outside the syllable window after adding the random offset were shifted back into the syllable by subtracting the syllable window length L from the event time. For each syllable that had at least one HVC minima, this bootstrap procedure was run 10,000 times, creating 1,120,000 synthetic minimum distances.

For the analysis of continuous global activity, the global average firing rate was normalized by computing the z-score ($\mu = 0$, $\sigma = 1$) over the duration of the BOS stimulus, to allow the comparison of data across experiments. Bootstrapped confidence intervals of the average normalized rate were generated by sampling with replacement and computing bootstrap means. 95% confidence intervals represent the 2.5% to 97.5% range of the bootstrapped population for every time step.

Bird	Number of experiments
b12	5
bk196	13
k401	3
notag01	2
notag04	12
o122	2
o156	1
pink183	1
sae01	2
uae01	1

Table 2.1: Number of experiments by bird

2.4 Results

Forty-one experiments were performed on ten birds. Two birds, UAE01 and SAE01 were removed from the analysis for having only partial response to BOS. There were between one and 13 experiments in a given bird (Table 2.1). In each experiment, birds were presented with 40 repetitions of BOS, and 40 repetitions of either REV, CON or both. Stimulus presentation order was randomized, and 30 seconds of silence preceded each stimulus. This stimulation protocol lasted roughly 1 hour, and was typically performed one to four hours after the start of the bird's subjective night.

In the birds selected for analysis, BOS presentation elicited a strong population response in HVC, see Fig. 2.7. The population response, as measured by summing all action potentials, was similar for both all spikes detected on the array and only spikes from isolated units. This suggests that the population response reflects an underlying population distribution from which the isolated units are drawn. In contrast, the response to REV/CON was much weaker. Selectivity for BOS is a feature of HVC activity that is well established (Margoliash and Fortune, 1992; Theunissen and Doupe, 1998).

A similar strong response to BOS was observed in all of the data sets from each bird selected for analysis. A larger sample of examples are shown in Fig. 2.7. Features of the BOS response include substantial modulation of the firing activity (many peaks and

troughs) for most syllables. As previously reported, such modulation was greatly suppressed in response to CON/REV (data not shown). These features of response are quantitatively analyzed in what follows.

2.4.1 Alignment to times representing specific acoustic features of song

Three methods of determining spectral features were used, Fig. 2.8. Hand labels (HL) were manually placed at note transitions within the syllable. There is no well-established and broadly accepted standard for identifying such transitions by manual scoring of spectrographs. Thus, the other two methods, LMAS and LMSE, were automatically extracted from the stimulus to identify spectral features objectively. The utility of these methods for generating biological insight into neuronal activity patterns, however, has not previously been established. All three methods emphasized different spectral features, Fig. 2.8. Qualitatively, global firing rate minima appeared to align with a subset of times identified by these methods, and some apparent acoustic features of song that were not identified by these methods were aligned to other global firing rate minima. Qualitatively, other features of global firing rates such as maxima also appeared to align with a subset of the identified acoustic features of song.

To test the hypothesis that HVC minima align to one or more of these spectral features, we generated synthetic data sets by randomizing the HVC minima times during the syllable and comparing the true minimum distances with the synthetic bootstrapped random data. Combining the best day of data across all birds, none of the three measures resulted in a distribution significantly different from random data as measured by a one way F-test, HL features F=0.96, p=0.32, LMAS F=1.6×10^{-5}, p=0.99, LMSE F=0.35, p=0.55, Fig. 2.9.

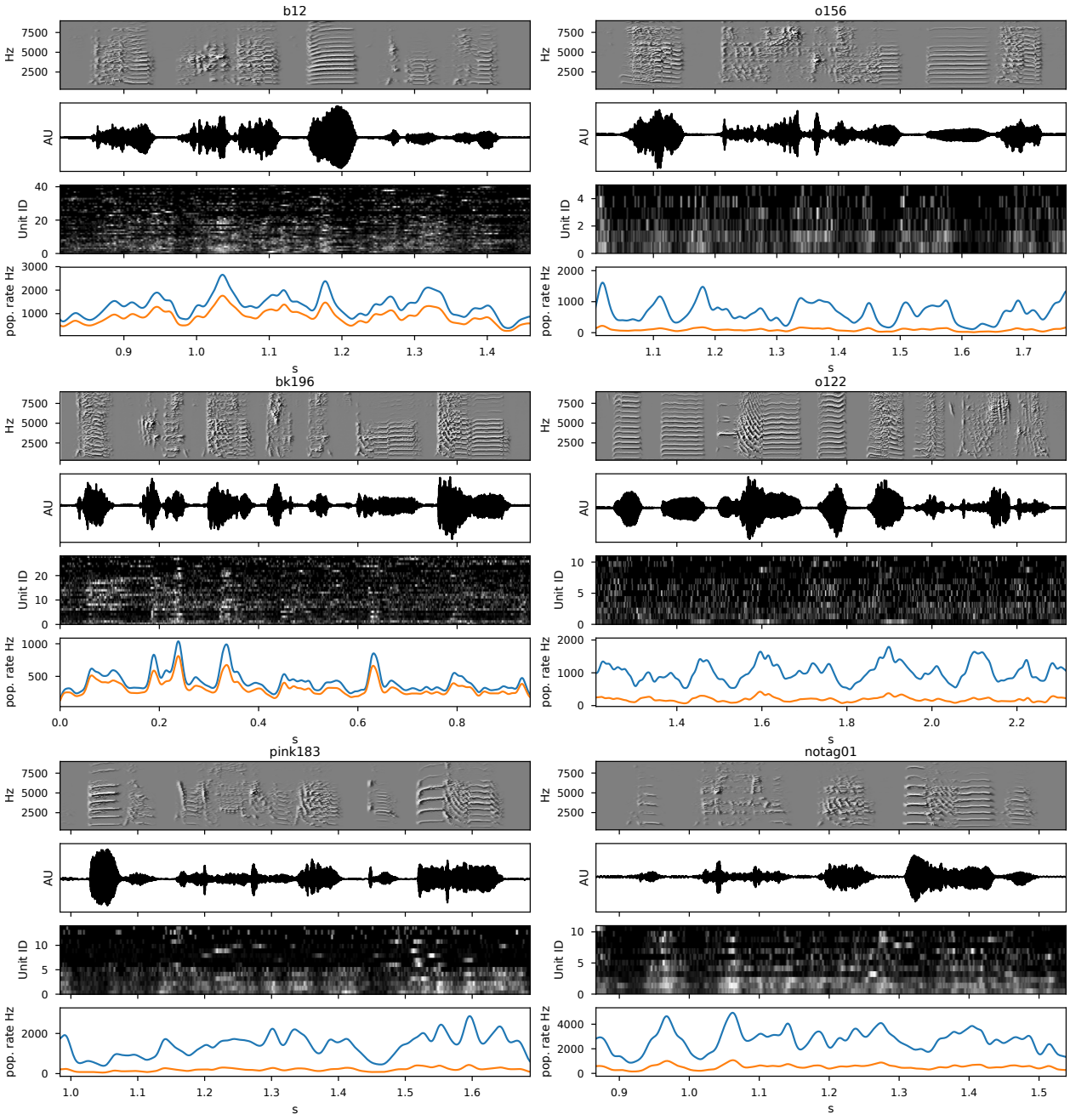

Figure 2.7. BOS stimulus and HVC population response in six birds Top panels: the spectrogram and oscillogram of the BOS stimulus (bird/exp). Third panel: the normalized firing rates of isolated HVC units. Bottom panel: global firing rates. The summed firing rate of all isolated units are in orange. In blue is the summed firing rate of all detected spikes. 40 BOS stimuli were presented.

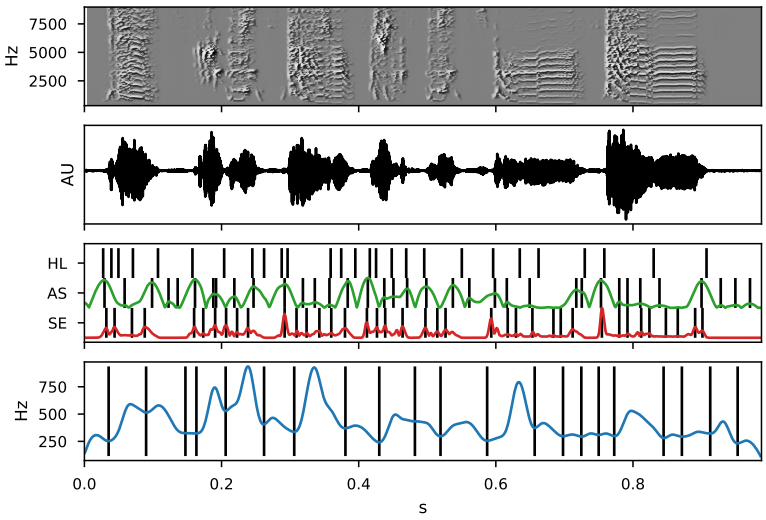

Figure 2.8. Spectral feature events. Top two panels: the spectrogram and oscillogram of a selected motif of BOS stimuli. Third panel: Three methods of approximating GTEs, hand labels (HL), and the local maxima of amplitude speed (AS) and spectral edges (SE). Bottom panel: local minima of the HVC population firing rate.

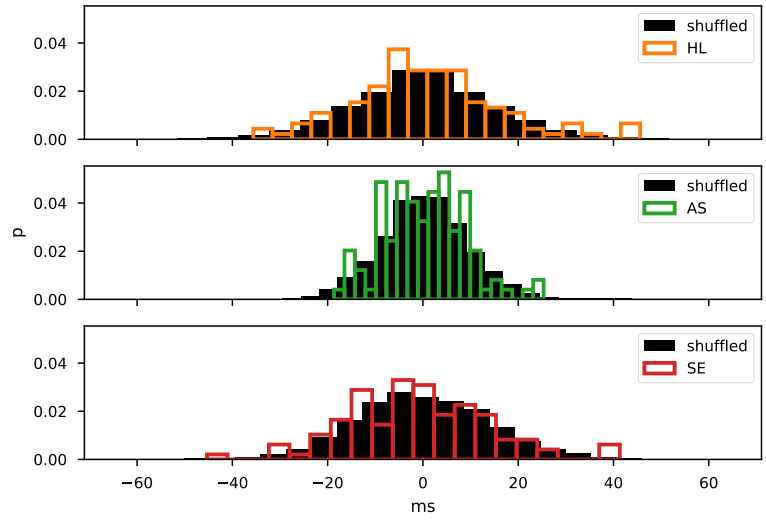

Figure 2.9. Zero delay histograms The distribution between the times of local minima in population firing rate and three spectral features. Top: hand-labeled song element edges (HL). Middle: Amplitude speed (AS). Bottom: Spectral edges (SE). True distributions are not significantly different from those generated with randomly shifted (shuffled) neural data.

2.4.2 HVC firing rates and features of singing

The previous sections of this chapter attempted to relate HVC firing and song features by comparing discrete data points – a list of HVC minima and a list of feature maxima. In this section I examine HVC population activity as continuous time-series data. With this shift in approach, local minima of the firing rate are not considered to be critical moments, and neither are the local maxima of spectral features. Instead, only the average firing rate as a time-series is analyzed.

First we examined the normalized average firing rates aligned to onsets and offsets of BOS syllables (N=139), Fig. 2.10. There was a pronounced increase in firing rate before the onset of syllable, which peaks at 22.6 ms before onset. The activity at this time is significantly greater than zero (one sample T-test, T=2.55, p=0.012). During syllable offsets, there is a significant reduction in normalized activity 21.3 ms after the end of the syllable (one sample T-test, T=-2.03, p=0.045).

The GTE model makes predictions about population minima, but we find significant peaks, not troughs in the population activity offset about 20 ms from the key times of interest (onsets and offset of syllables). It is possible that these peaks in activity are related to minima with zero lag relative to these time points, thus we examined the average distance between a local maxima and local minima across all birds during BOS playback (N=578). We found the average distance between HVC population minima and maxima to be 19.5 ms $\pm$ 1.0 ms, Fig. 2.12.

We next investigated whether syllable type affected HVC population activity. Syllables were manually classified into four broad categories: slides (a simple syllable with a decreasing pitch, similar to low volume calls or introductory notes) N=15; flat (syllables with constant pitch, similar to distance calls) N=9; noise (syllables with poorly defined pitch) N=5; and complex (syllables with multiple elements such as trills, noise and slides) N=17. From a motor control perspective, complex and flat syllables should represent two extremes. A gesture

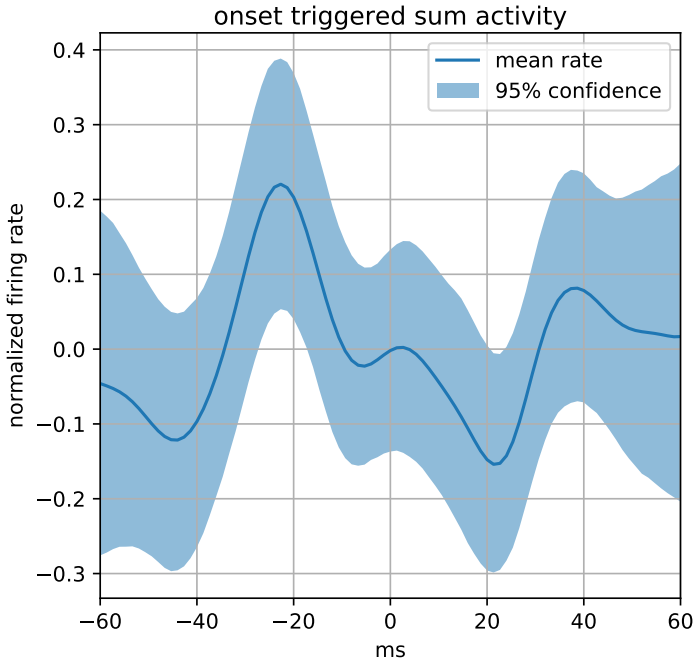

Figure 2.10. Onset triggered global activity. Average normalized population firing rate and bootstrapped 95% confidence intervals aligned to the onset of the syllables of a BOS stimulus. N=139. Syllable onset is at 0 ms.

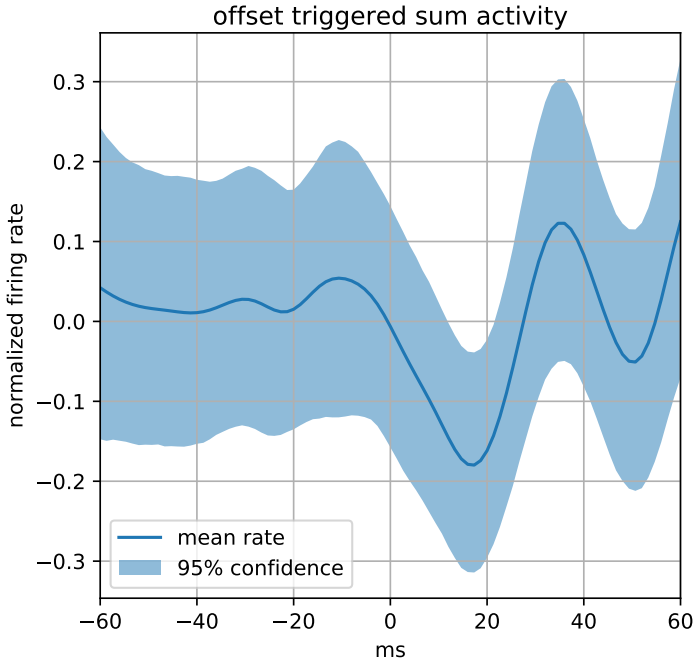

Figure 2.11. Offset triggered global activity. Average normalized population firing rate and bootstrapped 95% confidence intervals aligned to the offset of the syllables of a BOS stimulus. N=139. Syllable offset is at 0 ms.

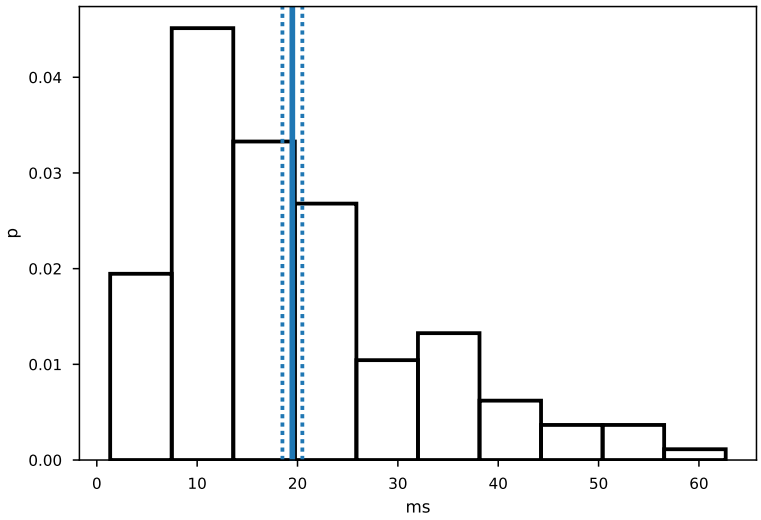

Figure 2.12. Distribution of distances between global firing rate peaks and troughs. Distribution of distances from a local maxima to a local minima in average HVC population rates across all birds, N=578. Blue line represents the mean of 19.5 ms, dotted line represents the 95% confidence interval

model would predict different HVC activity for these two types of behaviors. Consistent with this hypothesis, complex and flat syllables have different population activity shortly after syllable onsets, Fig. 2.13. The offset triggered population activity is shown in 2.14

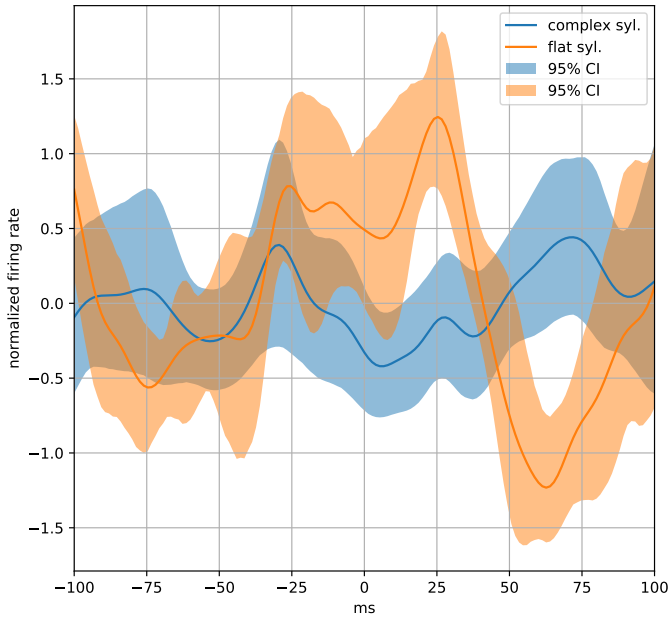

Figure 2.13. Onset firing rate by syllable type. Average normalized population firing rates for two syllable types: complex and flat (harmonic stack). Syllable onset is at 0 ms.

The GTE model predicts a lower rate of GTE in a flat syllable than a complex syllable, while the clock model does not differentiate between syllable types. To test this prediction, an average normalized firing rate was computed for each syllable in a single BOS motif for each bird. These averages were combined across birds to look for general changes in firing rate by syllable, Fig. 2.15. There was a significant difference between the average global activity during complex and flat syllables (Mann-Whitney U test, U=34.0, p=0.011).

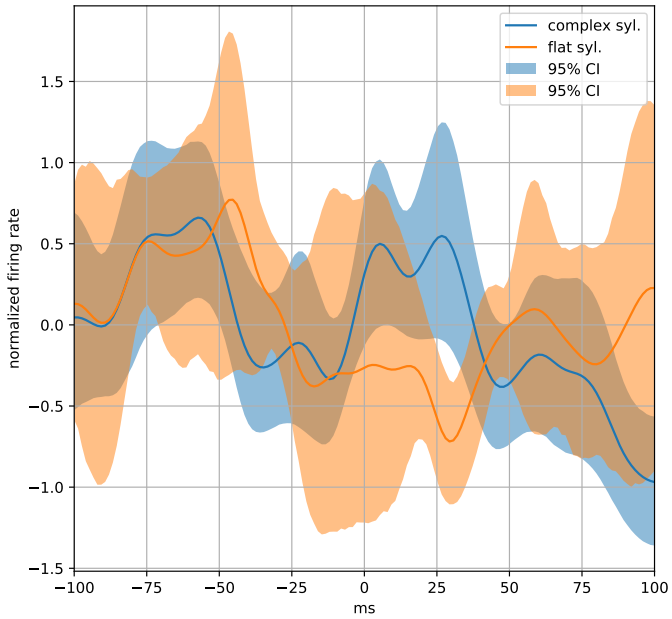

Figure 2.14. Offset firing rate by syllable type. Average normalized population firing rates for two syllable types: complex and flat (harmonic stack). Syllable offset is at 0 ms.

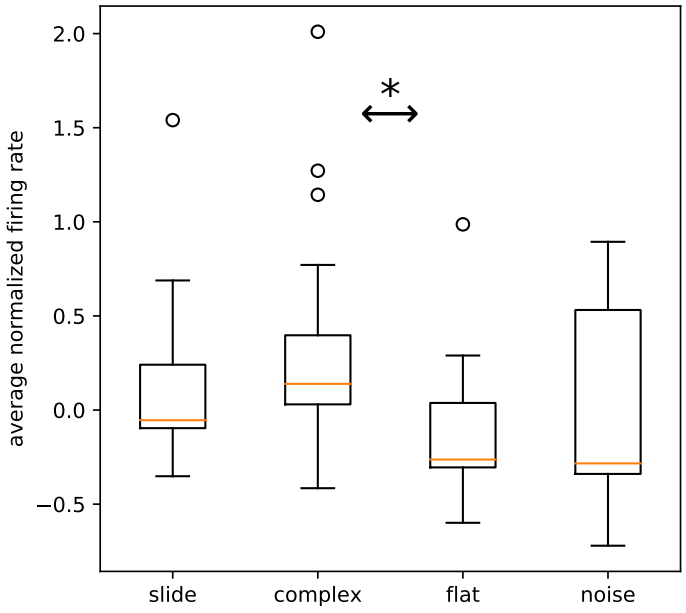

Figure 2.15. Offset firing rate by syllable type. Asterisk indicates $p < 0.05$.

2.5 Discussion

Here we provide evidence that HVC firing patterns may be unique to syllable types, shared across birds. The relation between population activity in HVC and syllable types is consistent with a gesture perspective of HVC motor coding. The GTE hypothesis predicts lower rate of bursts for HVC projection neurons during flat syllables and a higher rate of bursts for complex syllables, and we find flat syllables have a lower average population firing rate than complex syllables. It was previously known that HVC firing patterns are unique to each syllable within the motif of a single bird (Yu and Margoliash, 1996), what remained uncertain was whether these were unrelated to the motor periphery; as would be predicted by the clock model. The evidence for syllable identity specific firing in HVC, as described in this chapter, is not consistent with a gesture agnostic clock in HVC, though we are unable to comment on the method of sequence generation in HVC, which may still use a synfire-chain mechanism. While head-fixed sleeping experiments have technological advantages, our discussion of the clock model is also limited by not directly recording HVC activity in singing birds.

We were unable to replicate the findings in support of zero lag encoding of the GTEs using three simple spectral features. However we did not test the GTE model directly. Our hope was that using simple, easily calculated spectral features could replicate the findings of GTE coding in HVC, thereby allowing the greater songbird field to continue to refine gesture models of HVC without the challenges of inverting the Mindlin model to recover gestures. Recent papers that claim to disprove the GTE model, sought to compute GTE directly (Picardo et al., 2016; Lynch et al., 2016), however because there is no widely accepted algorithm for determining GTE, it is difficult to verify their reported GTE values. This underscores perhaps the greatest weakness of the GTE model: identifying the times of GTEs is not algorithmic and therefore challenging to independently replicate.

It is possible that the GTE model is correct, but very sensitive to accurate GTE place-

ment. This thesis and previous attempts at GTE model validation (Picardo et al., 2016; Lynch et al., 2016) may deviate sufficiently from the underlying GTE times, preventing model validation via the introduction of noise into the GTE times. Another possibility is that the original GTE model (gesture onsets, offsets and maxima) did not correctly identify the gesture elements HVC encodes. Perhaps HVC is modulated by onsets but not offsets, or some maxima but not others.

While we did not find significantly decreased population activity at syllable onsets and offsets, we did find significantly increased activity $\sim$ 20 ms before syllable onset and $\sim$ 20 ms after. The former observation matches previous reports that projection neuron burst density increases prior to syllable onset (Lynch et al., 2016). It is possible that the peak in activity preceding a syllable is evidence against the GTE model's zero time lag prediction. However since there is a $\sim$ 20 ms delay between an HVC population maxima and the next minima, the observation of a causal peak before a syllable and trough at the onset (Lynch et al., 2016; Amador et al., 2013) may be two descriptions of the same underlying oscillation. If there is cyclical activity in HVC with a period of approximately 40 ms (25 Hz), it may be fruitful to investigate behavior relative to the phase of this cycle. While these data are consistent with a premotor model of HVC with a $\sim$ 20 ms delay between HVC activation and the motor periphery, we cannot at this time rule out a zero lag synchronization of HVC and its downstream targets. The clearest way to validate the zero lag GTE model is by directly measuring the downstream targets of HVC (RA or the muscles of the syrinx), a test which will be explored over the next two chapters.

CHAPTER 3
ACTIVITY OF HVC AND THE SYRINX DURING SPONTANEOUS REPLAY

3.1 Introduction

Sleep is ubiquitous in birds and mammals, and is – at least – broadly distributed across vertebrates (Rattenborg and Amlaner, 2002). Sleep must provide some essential value to offset the risk of increased predation attendant to reduced sensory and motor function associated with sleep. While sleep is unlikely to have evolved for the purpose of memory consolidation, memory consolidation has been linked to sleep. In humans, sleep improves memory encoding and consolidation (Walker and Stickgold, 2006). In birds, sleep is an integral component of perceptual consolidation in adults, and sleep is critical for developmental song learning in juveniles (Brawn and Margoliash, 2015). Juvenile zebra finches display ultradian variation in their song structure (which can be experimentally disassociated from circadian variation), and the juveniles with the most ultradian variation produce the most accurate copies of the tutor song (Derégnaucourt et al., 2005). This suggests the involvement of an active process during sleep that modifies song system circuitry. Indeed, early in song development the firing statistics of RA neurons are significantly changed from one night to the next (Shank and Margoliash, 2009). RA neurons have increased bursting that precede changes in a juvenile's ultradian singing patterns, and RA neurons alter their firing when juveniles are experimentally induced to switch song tutor models (Shank and Margoliash, 2009).

Neuronal replay in zebra finches has been observed in RA, where subsets of precisely timed bursts during song are replayed during sleep (Dave and Margoliash, 2000). Nighttime replay may help consolidate sensory-motor memories formed during the day (Shank and Margoliash, 2009). Changes in neuronal activity in basal ganglia circuits over periods of sleep help to minimize experimentally induced vocal errors, which supports this perspective

(Andalman and Fee, 2009). Continuous recordings of RA neurons before and after a night of sleep show subtle changes to burst structure after sleep (Rauske et al., 2010). Replay-like bursting has been seen across the song system, including in HVC, and bursts in RA are likely driven by activity in HVC (Hahnloser and Fee, 2006). There is preliminary evidence of song replay in the song muscles of the syrinx (Young and Goller, 2012). The entire descending motor pathway of the song system appears to be active during nighttime replay, with the exception of the respiratory nuclei (hence preventing sound output during sleep).

If night-time replay in the song system is representative of singing activity, then replay can be used as a window into HVC's role in singing. According to the clock model, neural activity in HVC should not correlate with activity of the muscles during song. Alternately, the GTE hypothesis predicts a temporary precise relationship between activity in HVC and syrinx muscle activation. In this chapter, I record from HVC and the muscles of the syrinx simultaneously during sleep, to examine the coordination of sleep replay across the motor system.

3.2 Methods

Adult male zebra finches with previously recorded song were selected and placed on a reversed light cycle (9 p.m. to 9 a.m.). The three birds that contributed usable data (see below) were allowed to adjust to the reversed light cycle for at least 14 days prior to any surgeries.

3.2.1 Surgical preparation for syringeal recordings

All syringeal implants were performed by Franz Goller or performed under his direct instruction. Birds were first food deprived one hour before surgery or until no seeds were visible in the crop. An anesthesia mixture of 92 mg/mL ketamine and 8 mg/mL xylazine was prepared and 80 μL of the ketamine/xylazine mixture was injected into the tissue of the breast muscle. An additional 0-5% isoflurane anesthesia at 2 liters per minute was continuously

administered such that a 1–2 Hz respiratory rate was maintained. An incision to the skin was made over the interclavicular air sac and any fatty tissue was carefully retracted to expose the air sac. An incision of the air sac exposed the syrinx. Every bird received two EMG electrodes, one inserted into the ventral syringeal (vS) and another inserted into the dorsal tracheobronchial (dTB) syringeal muscles. The EMG electrodes consisted of pairs of 0.001" Teflon coated stainless steel wire (California Fine Wire) with 500 microns of Teflon manually removed with a scalpel, forming the conductive tip of the electrode. The stainless steel tips of the wires were implanted in the muscular tissue and secured to the outermost fascia with Vetbond tissue adhesive (3M), precisely applied with a 30 gauge hypodermic needle with the bevel removed. The EMG wires were guided subcutaneously to the back with a #26 tapestry needle. Two loops of wires in the thoracic cavity provided strain relief. The air sac was closed with tissue adhesive and the skin was closed with 6/0 silk sutures. Conductive epoxy was used to connect the EMG wires to 1 mm grid socket connectors (Mill-Max 861 series). The bird was fitted with a custom cotton/elastic backpack to house the connectors. A head pin was implanted following the procedure outlined in the previous chapter. After the surgery, birds were monitored for 48 hours or full recovery before any experiments.

3.2.2 Experimental paradigm

Six experiments were performed on the three birds k403, o133, and k405. A seventh experiment (k405 experiment 2) was discarded upon analysis of the data, as the MEA probe did not appear to be placed within HVC. Thirty minutes before the start of the experiment, an array was inserted in HVC, as described in Chapter 2. Recordings from HVC were made with Neuronexus A4x8-5mm-50-200-177-CM32 arrays, which were placed on the same side of the body as the syrinx EMG implants. The stimulus protocol and data collection system was similar to the description in Chapter 2 of this thesis, but with an additional period of 30 minutes without stimulation at the start of the experiment. The 32 MEA channels, the two

channels of EMG activity, and the stimulus signal were recorded to disk at 30 kHz sampling rate.

3.2.3 Analysis

Spike sorting was performed with KlustaKwik/Phy (Rossant et al., 2016). Cluster quality was evaluated based on the inter-spike interval, amplitude and shape of the spike waveform, Fig. 3.1. (We have since moved to using SpyKing Circus software for isolating single neurons from MEA recordings, see Chapter 2.) Clusters were classified as either good units or multi-unit activity (MUA).

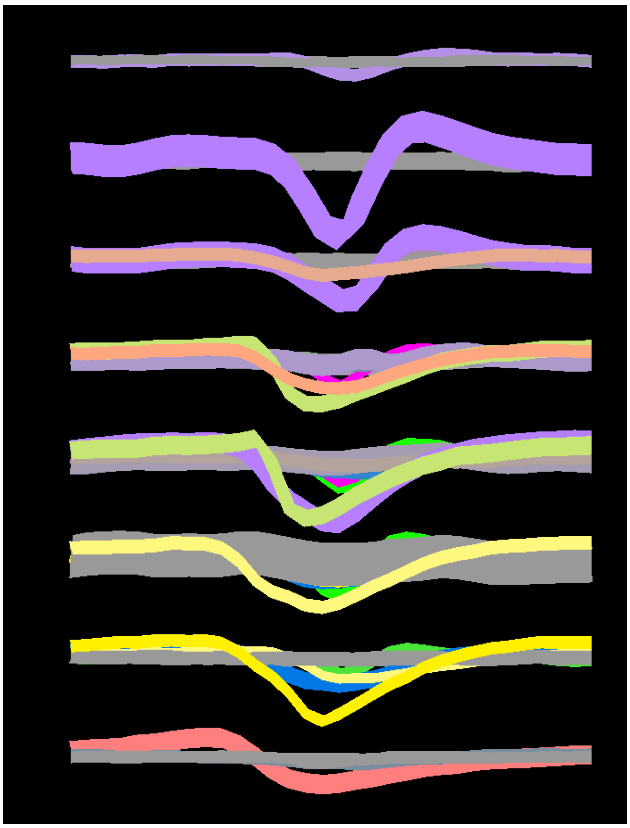

Figure 3.1. KlustaKwik MEA waveforms. Average spike waveforms from clusters identified on a single shank of a MEA. Width indicates standard deviation of the cluster waveforms. Each pad (recording surface) on the MEA is separated by 50 μm.

The raw EMG signals were recorded using bipolar amplifiers, which recorded the differential voltage between EMG electrode pairs. The EMG amplitude was estimated by filtering

the EMG signal with a 30 Hz cut-off high pass filter, taking the absolute value, and applying a 100 Hz cut-off low pass filter, Fig. 3.2. Both filters were implemented using 3rd order Butterworth, applied forward-backward to eliminate phase shift.

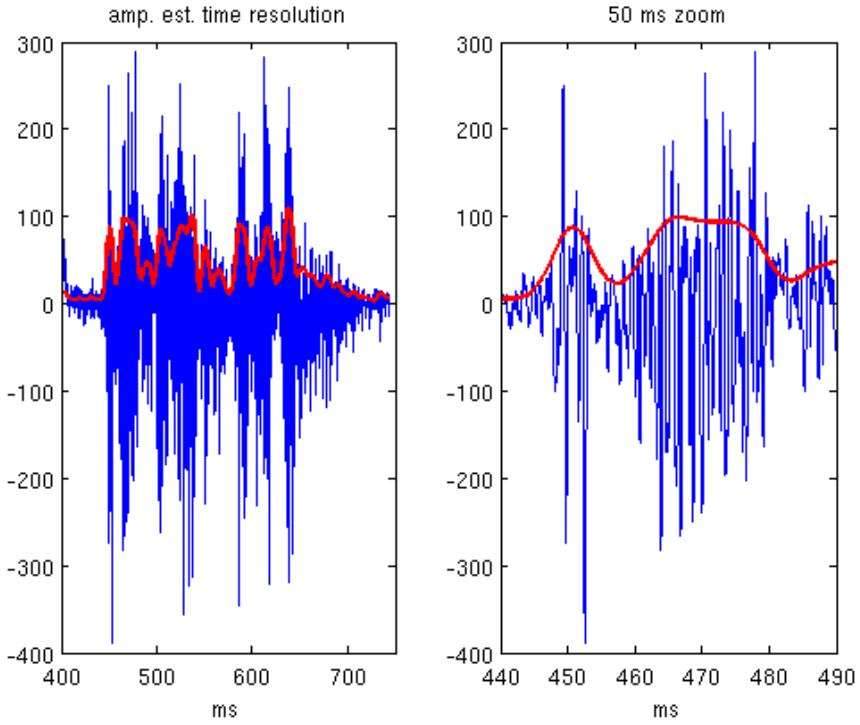

Figure 3.2. EMG amplitude estimation. VS EMG activation during a period of spontaneous activation. The raw EMG signal is in blue, and the amplitude estimate is in red. Left: an episode of a spontaneous EMG activation. Right: EMG activity in a 50 ms window during the episode, showing the difference between the raw signal and the amplitude estimate.

A spike triggered average (STA) was used to compare HVC firing with EMG activity. The STA was computed for each isolated HVC unit as well as MUA clusters. STA were computed by

$$STA(\tau) = \langle \frac{1}{n} \sum_{i=1}^{n} a(t_i - \tau) \rangle$$

where $t_{1,2,3...n}$ are the spike times of a unit, and τ is a time lag. $STA(0)$ represents the average activity of signal a at zero delay.

A generalized linear model was constructed with a Poisson link function to predict EMG

amplitude from the activity in HVC.

$$E(Y_t|\mathbf{x}_{t-15}) = e^{\theta \mathbf{x}_{t-15}}$$

where $\mathbf{x}$ is a vector of five parameters: a constant and the activity of each shank 15 ms before time t, and Y_t is the target EMG activity. The model target was EMG activation 15 ms after the input HVC activity, and was fit by taking two minutes of spontaneous activity and tested on a second segment of two minutes.

After the final experiments, the birds were euthanized and perfused with 10% neutral buffered formalin. Electrodes tracts were recovered via cresyl violet staining as described in Chapter 2.

3.3 Results

3.3.1 Responses to auditory stimuli

BOS response elicited replay activity in the motor system for some birds and some recording sites e.g. Fig. 3.3, but not others Fig. 3.4. A summary of the BOS response data is presented in 3.5. The population firing rate for two of the experiments (k405_1, k403_3) showed selectivity of the response to BOS as compared to the response to REV or CON.

In both recordings where BOS selective auditory responses were observed, there were also good EMG recordings. Good recordings were obtained from both muscles for k405_1 and for vS (but not vTB) for k403_3. The overall activation in the muscles followed the overall selectivity for BOS, as would be expected if BOS was driving activity in the descending motor pathway (Williams and Nottebohm, 1985). In another experiment (k403_1), the syringeal recordings demonstrated selective activation for BOS although the HVC recordings did not. It is speculative how this arose (e.g. an unidentified technical issue), but nevertheless this indicates that the BOS-selective auditory responses were intact. In the other two experiments

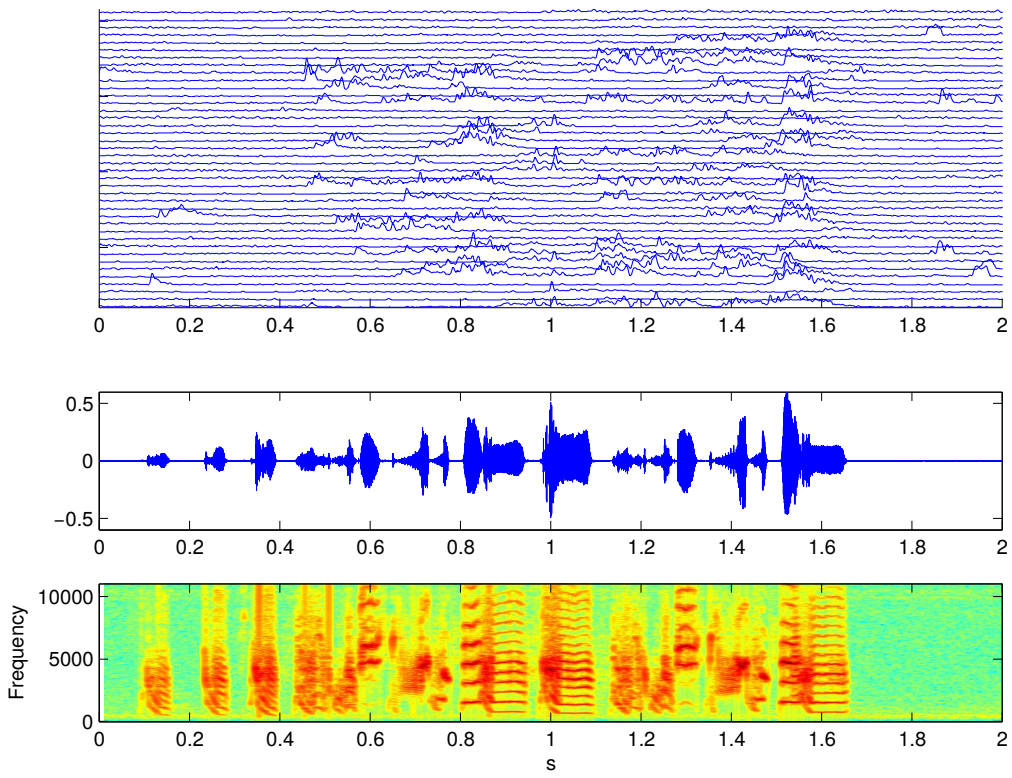

Figure 3.3. K403 vTB response to BOS stimuli. Top: amplitude of vTB, over trials of BOS stimulation. Bottom: Oscillogram and spectrogram of BOS stimulus. Bird: K403.

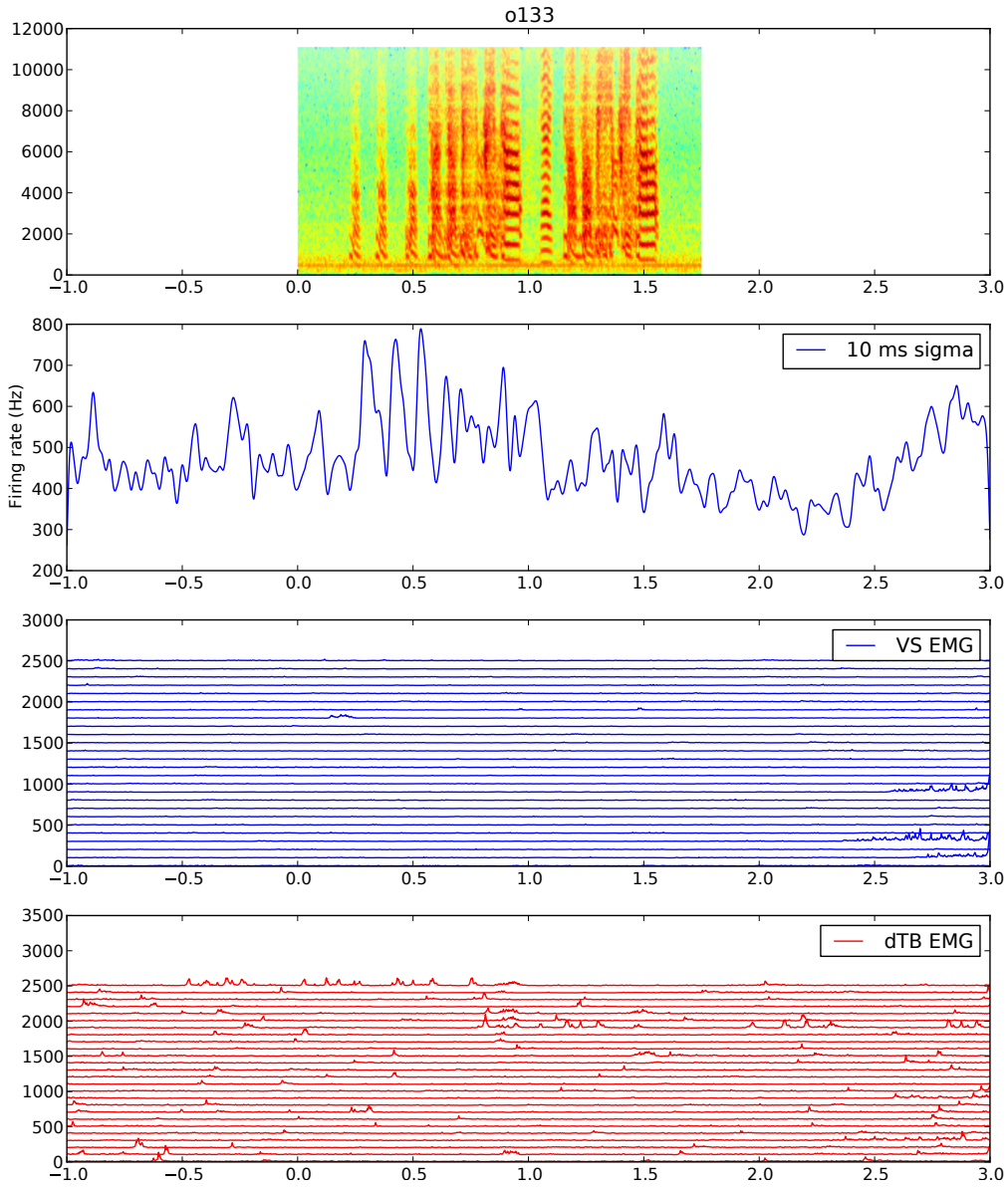

Figure 3.4. O133 vTB and vS response to BOS stimuli. Top: spectrogram of O133 BOS stimulus, and the HVC population response, smoothed with a 10 ms σ Gaussian window. Bottom: Amplitude of vS and dTB over each BOS presentation.

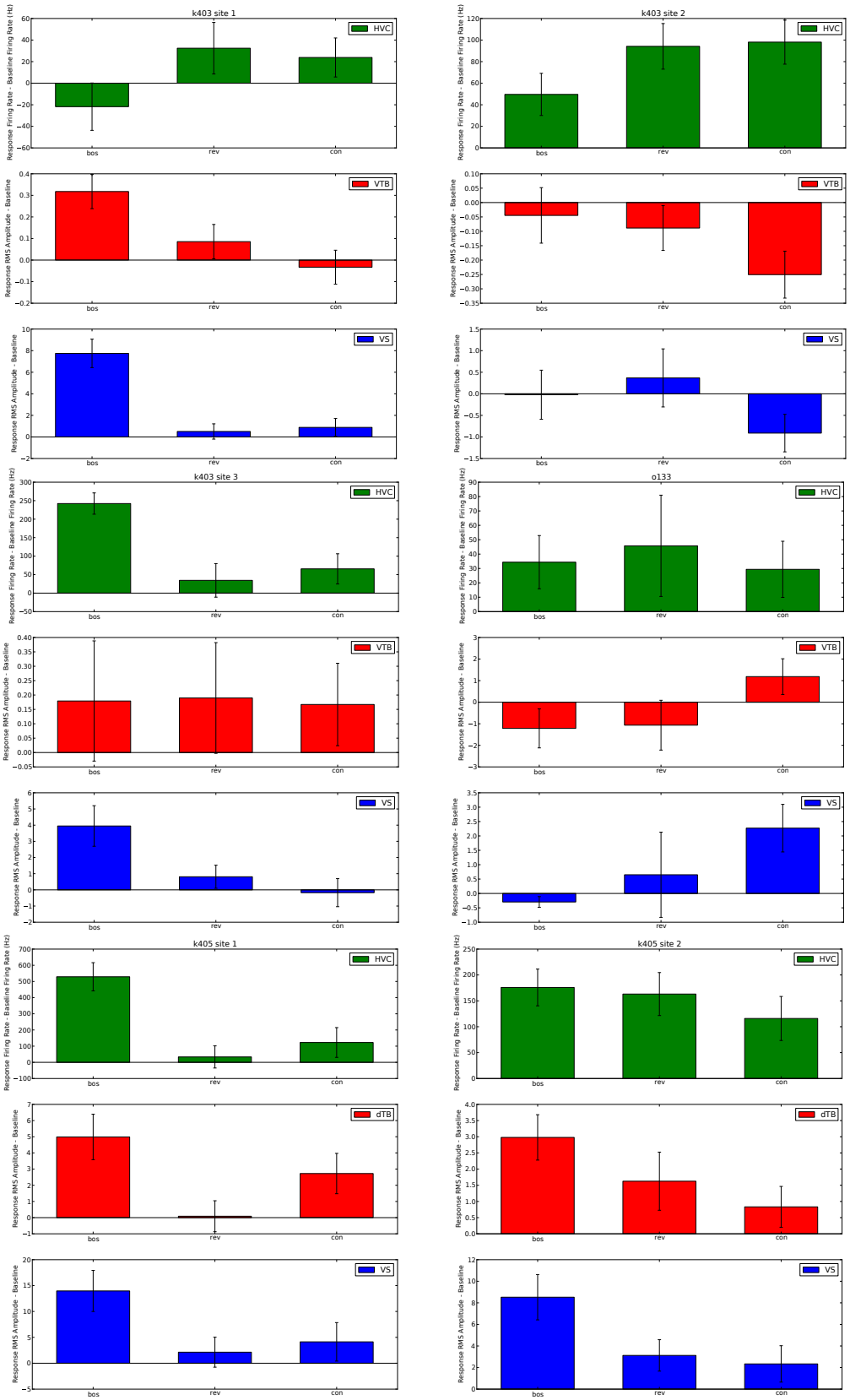

Figure 3.5. EMG and HVC BOS selectivity Each set of three panels represents the overall BOS response from an experiment.

(k403_2, o133) the syringeal recording were of low quality, and BOS-selective activation was not observed. The others are shown in Figures 3.6, 3.7 and 3.4 (note k405 site 2 was discarded because the array was not in HVC).

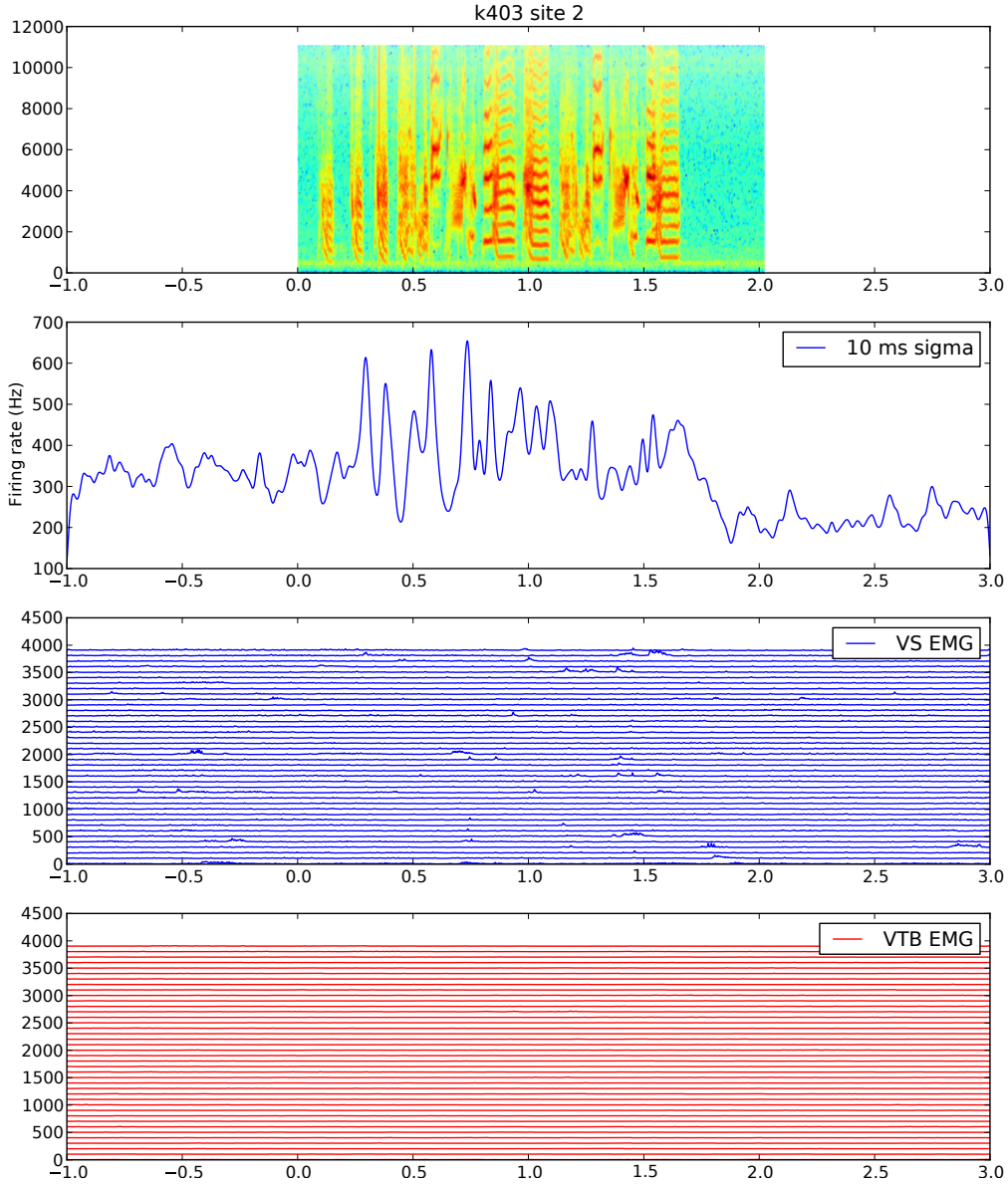

Figure 3.6. K403 experiment 2 BOS response Top: spectrogram of K403 BOS stimulus, and the HVC population response, smoothed with a 10 ms σ Gaussian window. Bottom: Amplitude of vS and dTB over each BOS presentation.

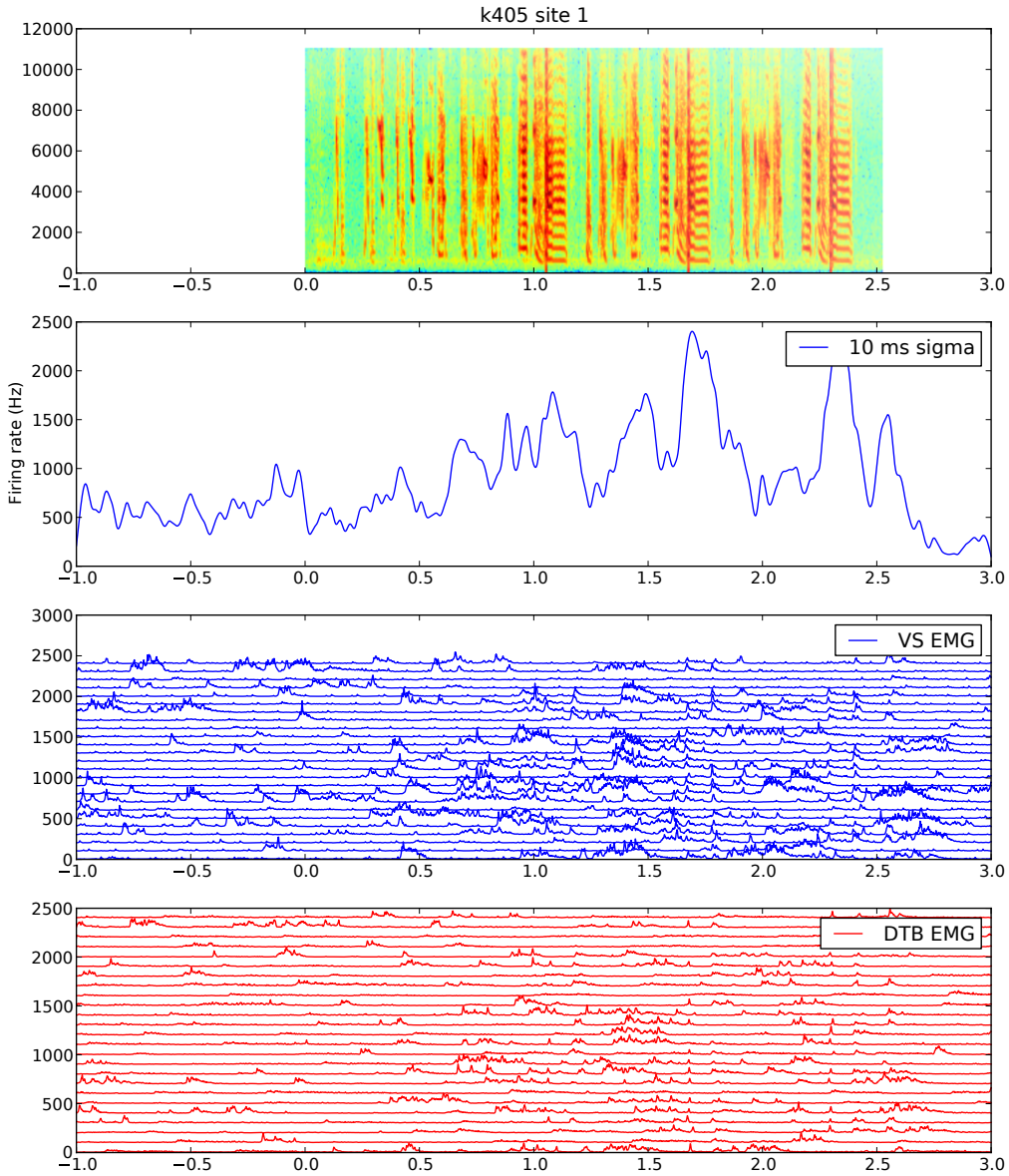

Figure 3.7. K405 experiment 1 BOS response Top: spectrogram of K405 BOS stimulus, and the HVC population response, smoothed with a 10 ms σ Gaussian window. Bottom: Amplitude of vS and dTB over each BOS presentation.

3.3.2 Spontaneous activity

Additional analysis were restricted to regions of spontaneous activity during periods of recording without auditory stimulation. Spontaneous activation in HVC occurred during periods of elevated EMG activity. This was apparent for both isolated units, Figs. 3.8, 3.9, and multi-units, Figs. 3.10, 3.11.

Across the four shanks of the MEA, spaced at 200 μm intervals, periods of activation were synchronized, both with each other and the EMG amplitude, Fig. 3.8.

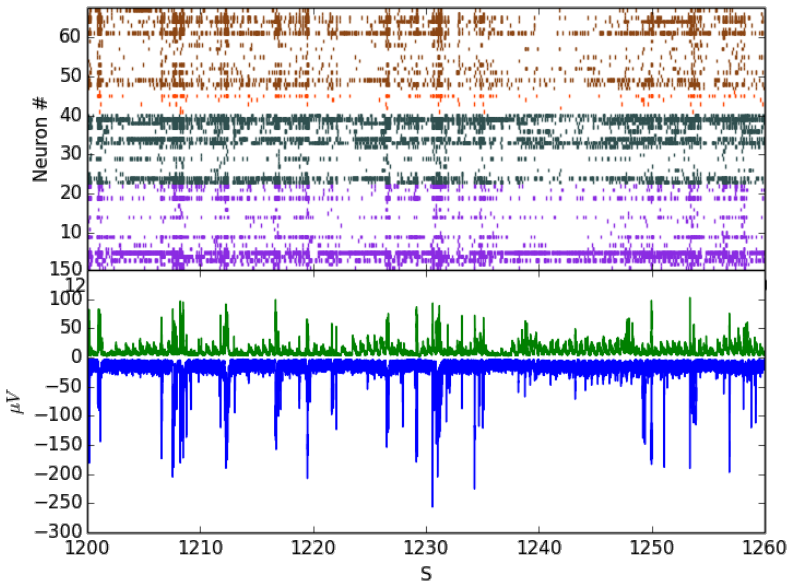

Figure 3.8. K405 spontaneous activity over 60 seconds. Top: spontaneous activity from single units across all four shanks of the array, each color represents a shank. Bottom: Spontaneous amplitude of vS (blue) and dTB (green), vS is inverted to aid visual comparison with dTB.

For roughly half of the clusters with more than 2000 spikes (253/489), computing a spike triggered average between the neuronal activity and the raw EMG signals yielded a sharp bipolar curve with a peak between 10 and 30 ms lag, Figs. 3.12 and 3.13. Computing a STA with the EMG amplitude estimate instead of the raw signal identified peaks at approximately 15 ms and 18 ms lag, Fig. 3.14 this was also true of some MUAs, Fig. 3.15. A summary of the peak location and peak amplitude is shown in Fig. 3.16. Considering, for example, the

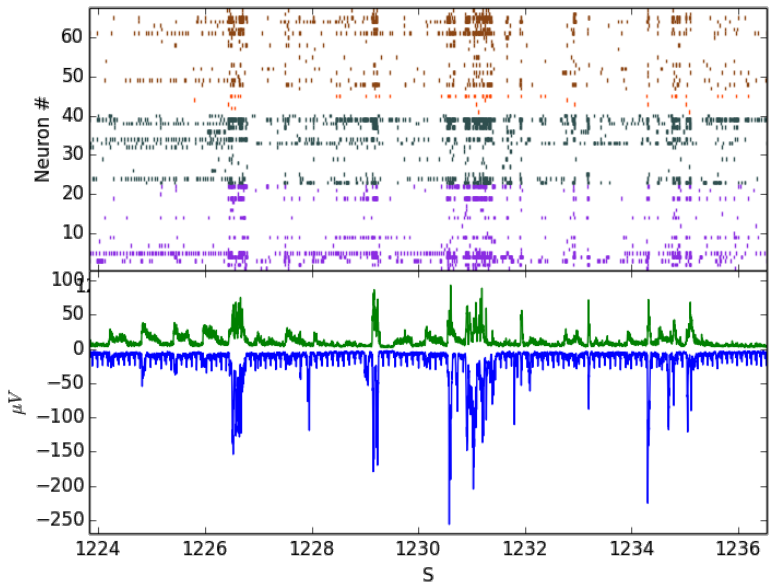

Figure 3.9. K405 spontaneous activity over two seconds. top: spontaneous activity from single units across all four shanks of the array, each color represents a shank. Bottom: Spontaneous amplitude of vS (blue) and dTB (green), vS is inverted to aid visual comparison with dTB.

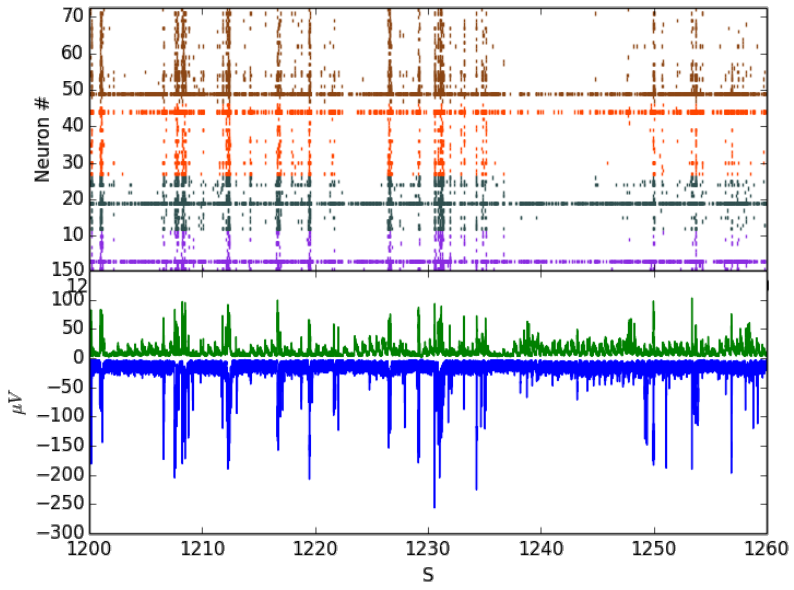

Figure 3.10. K405 multi-unit spontaneous activity over 60 seconds. Top: spontaneous activity from multi-units across all four shanks of the array, each color represents a shank. Bottom: Spontaneous amplitude of vS (blue) and dTB (green), vS is inverted to aid visual comparison with dTB.

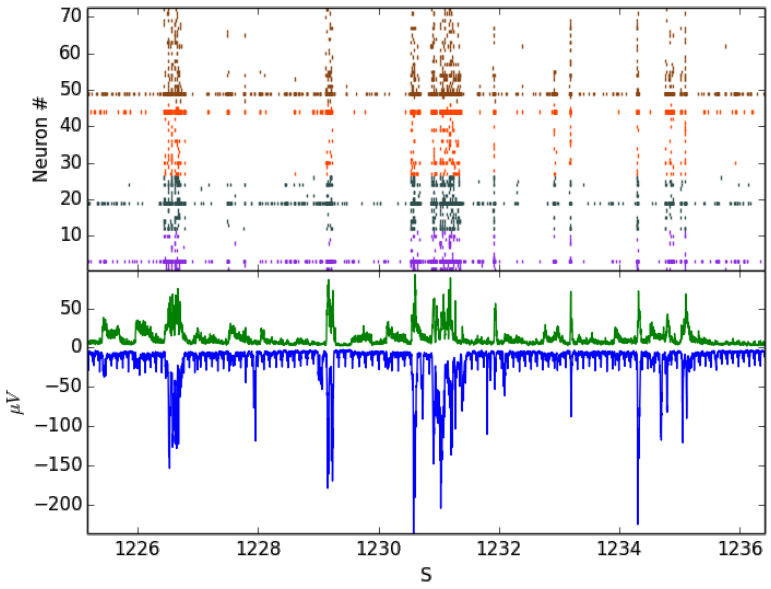

Figure 3.11. K405 multi-unit spontaneous activity, 2 seconds Top: spontaneous activity from multi-units across all four shanks of the array, each color represents a shank. Bottom: Spontaneous amplitude of vS (blue) and dTB (green), vS is inverted to aid visual comparison with dTB.

raw signal STAs, the two experiments with strong BOS selectivity also show a clear peak in the distribution between HVC unit and syringeal muscle timing. For k403_3, 44/72 neurons had peaks between 10 and 25 ms; for k405_1, 80/80 neurons had peaks between 10 and 25 ms. In contrast, for two of the experiments with weak BOS selectivity (k403_2 and o133), the distributions are flat. Interestingly, for k403_1, there is a strong peak in the distribution of the timings of HVC units and syringeal activation. Of the three experiments with weak BOS selectivity, it is noteworthy that this experiment was the only one with strong selective syringeal activation. This indicates that synchronization down the motor pathway can be preserved in the absence of HVC BOS selectivity.

For subsequent figures, only units with peaks greater than 30 μV were included – an arbitrary cut-off value that eliminated much of the low amplitude peaks with high timing variability. In the two experiments with dTB recordings, the average STA peak was 13.0 ms and 7.5 ms, Fig. 3.17. Across the experiments with vS recordings, the average time of the STA peak was 14.6 ms, Fig. 3.18. The values of the vS STA peak lag are shown in table

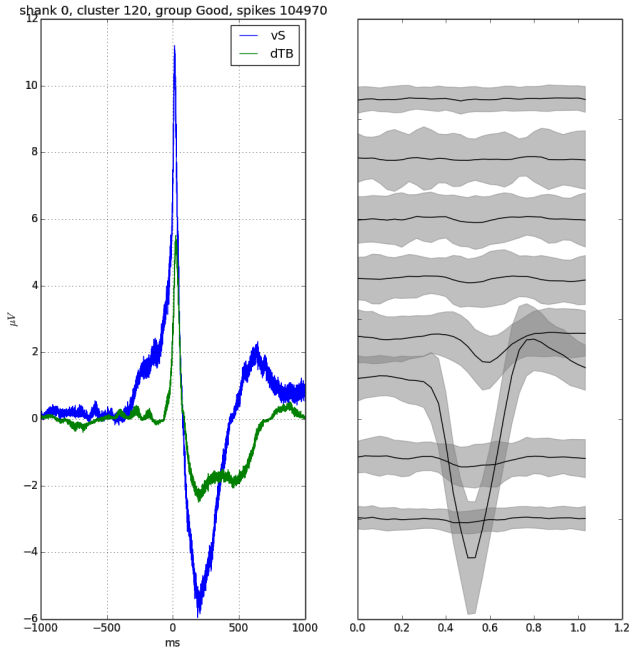

Figure 3.12. K405 single unit #120 STA for vS and vTB. Left: spike triggered average for vS (blue) and vTB (green). Right: average waveform ± standard deviation for the single unit.

3.1.

The simple relationship between HVC activity and EMG activity can be measured by looking at correlation. For each shank, the total activity was correlated with EMG activity at a 15 ms lag. A simple GLM model was also created. The correlations of shank activity and the GLMs are shown in Fig. 3.19 and example data are shown in Fig. 3.20

Experiment	mean vS STA peak (ms)
k403_1	22.4
k403_2_bottom	14.6
k403_2_top	11.2
k403_3	15.2
k405_1	13.8
o133	10.8

Table 3.1: Mean vS STA peak across experiments

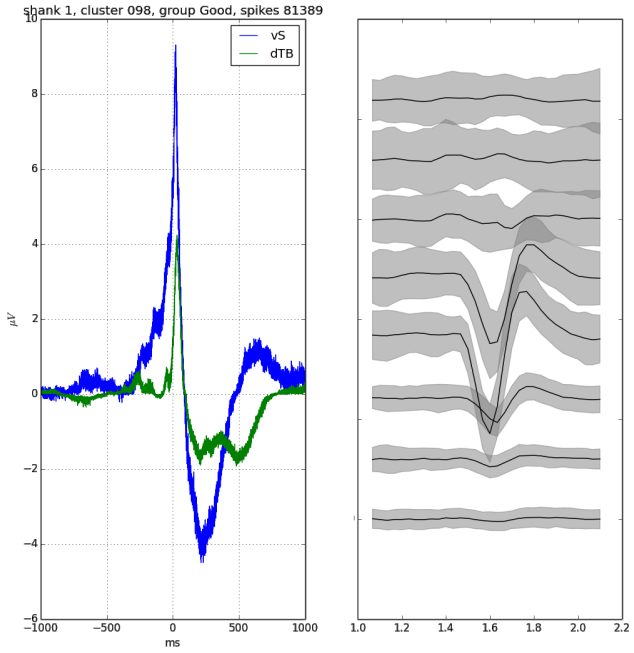

Figure 3.13. K405 single unit #98 STA for vS and vTB. Left: spike triggered average for vS (blue) and vTB (green). Right: average waveform ± standard deviation for the single unit.

3.4 Discussion

We have simultaneously recorded neurons from HVC and the muscles of the syrinx throughout the night in a head-fixed zebra finch preparation. For many neurons, activity was correlated with syrinx muscle activity at remarkably short latencies, and was highly correlated with EMG amplitude, as measured by the STA of single units and global activity across the shanks of the MEA. These data suggest that during spontaneous replay, HVC encodes the motor periphery at a lag consistent with motor control. If spontaneous replay reactivates the neural circuits responsible for song, these results cast serious doubts on the clock model, which predicts motor-related activity RA but not HVC. It is likely that syringeal activity is also strongly correlated with HVC activity when birds are presented with BOS. This too would narrow the potential domain of the clock model (limited at best to describing activity in HVC during singing).

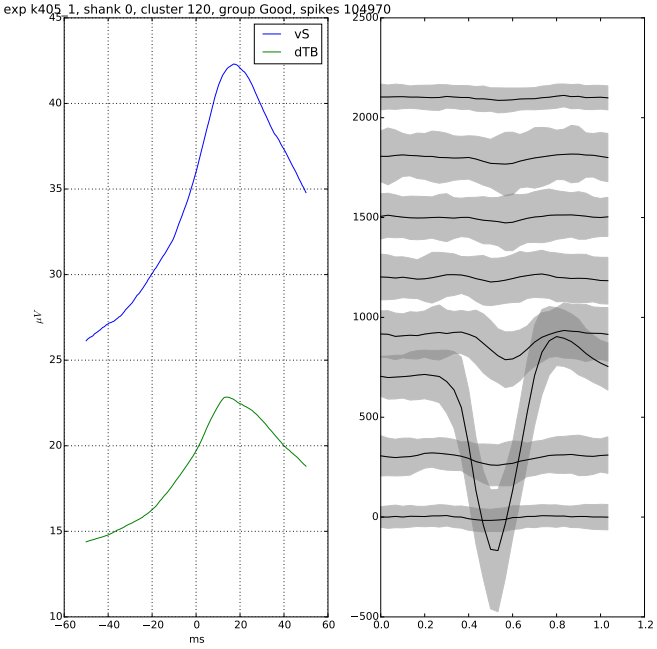

Figure 3.14. K405 single unit #120 STA for vS and vTB. Left: spike triggered average for vS (blue) and vTB (green), using amplitude estimates instead of the raw EMG voltages. Right: average waveform ± standard deviation for the single unit.

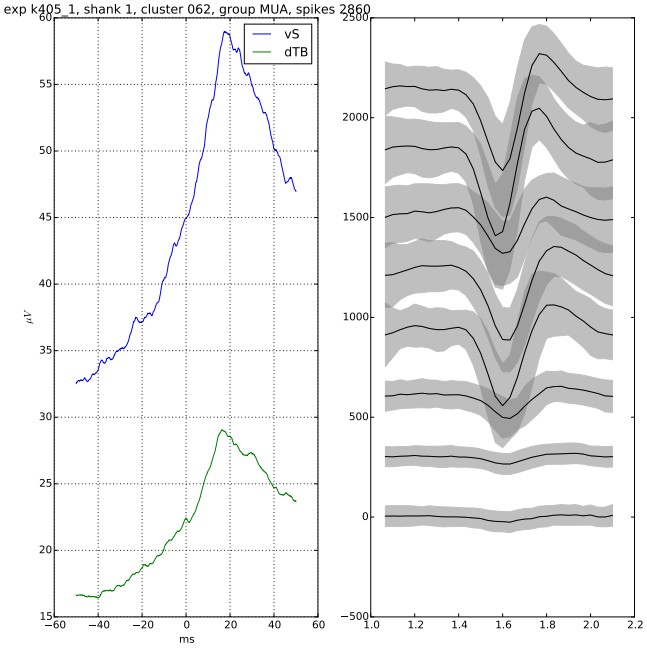

Figure 3.15. K405 single unit #62 STA for vS and vTB. Left: spike triggered average for vS (blue) and vTB (green), using amplitude estimates instead of the raw EMG voltages. Right: average waveform ± standard deviation for the single unit.

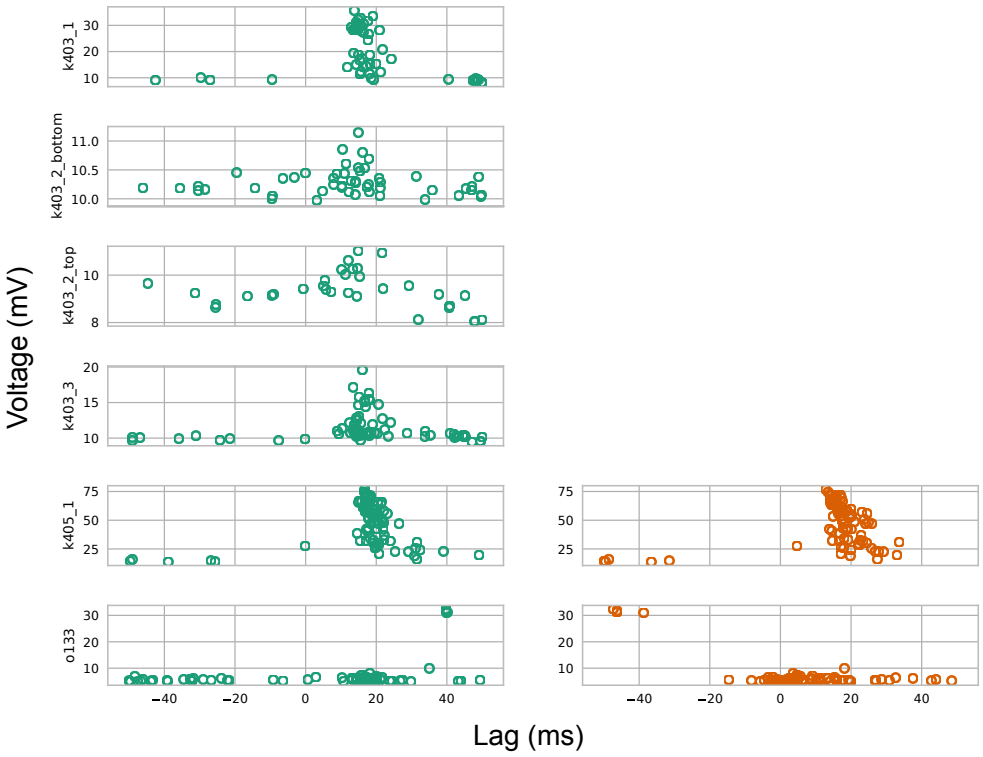

Figure 3.16. STA peak voltage and lag by experiment For each experiment, the peak lag time (Y axis) and voltage (X axis) is plotted for every unit. Left: vS. Right: vTB.

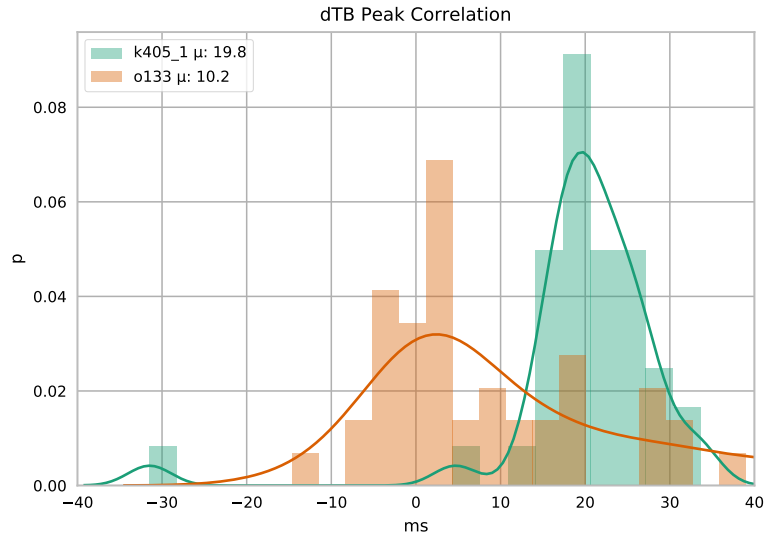

Figure 3.17. Distribution of dTB STA peak times. For the two experiments with dTB recordings, the average lag time of the STA peak was 19.8 ms and 10.2 ms for K405_1 and o133 respectively.

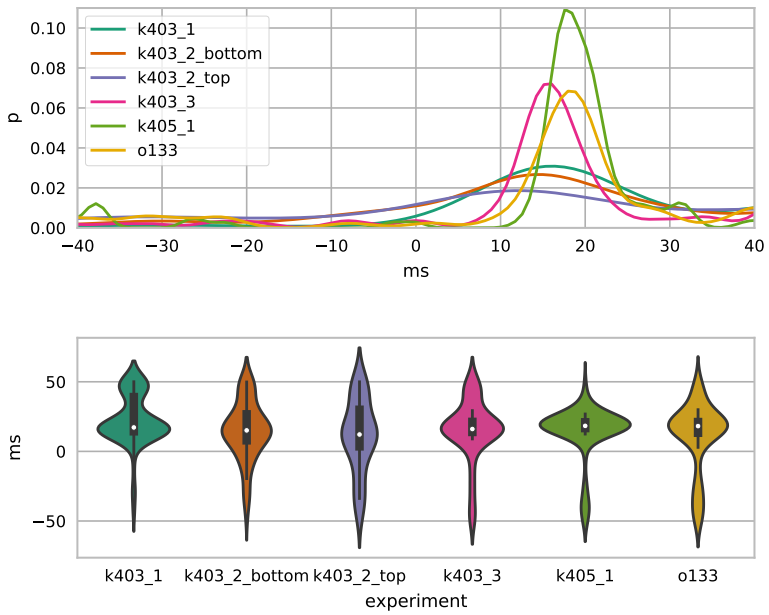

Figure 3.18. Distribution of vS peak times. Top: distribution of vS STA peaks by experiment. Bottom: same data, but with experiment ID as the X axis.

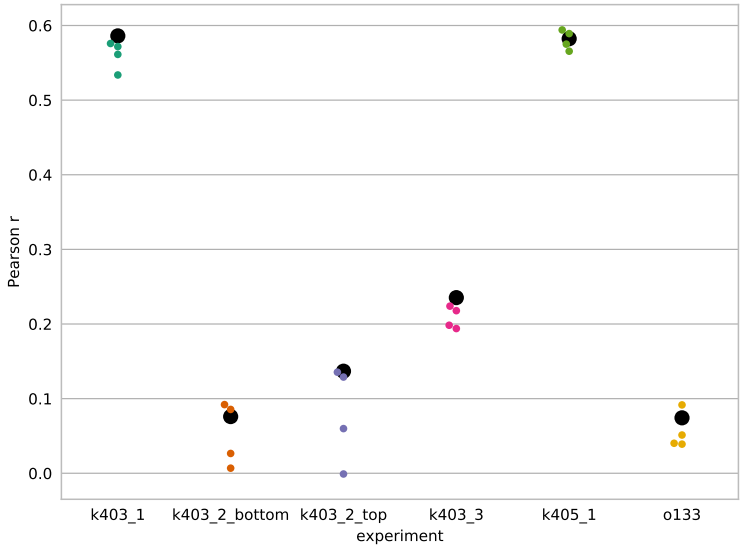

Figure 3.19. Correlation between total shank activity and vS. Black dot indicates GLM fit to the four shank activity rates.

Broadly, these results are consistent with a premotor kinetic code in HVC, a novel hypothesis for HVC. While the population dynamics of HVC have yet to be fully explored, it's possible that the role of HVC is to perform an inversion: converting a desired sequence into a motor code. Many neurons recorded in this chapter did not have a significant EMG STA. This population of unresponsive neurons may encode other muscles involved in singing or may be used to generate dynamic sequences, like populations seen in primate motor cortex (Scott, 2004; Kaufman et al., 2014; Gallego et al., 2017).

A major limitation of this study is that sleeping replay, rather than song, was studied. Sleeping replay is known to be driven by input from NIf to HVC (Hahnloser and Fee, 2006), though NIf input to HVC is not required for song (Cardin et al., 2005). It is possible that bursts from NIf may create cascades of activation in the song system that are not representative of song network reactivation. If this hypothesis is correct, spontaneous sleep replay cannot elucidate the relationship between HVC and syrinx muscles during song. However, the results from RA recordings suggest that sleep replay matches the temporally precise

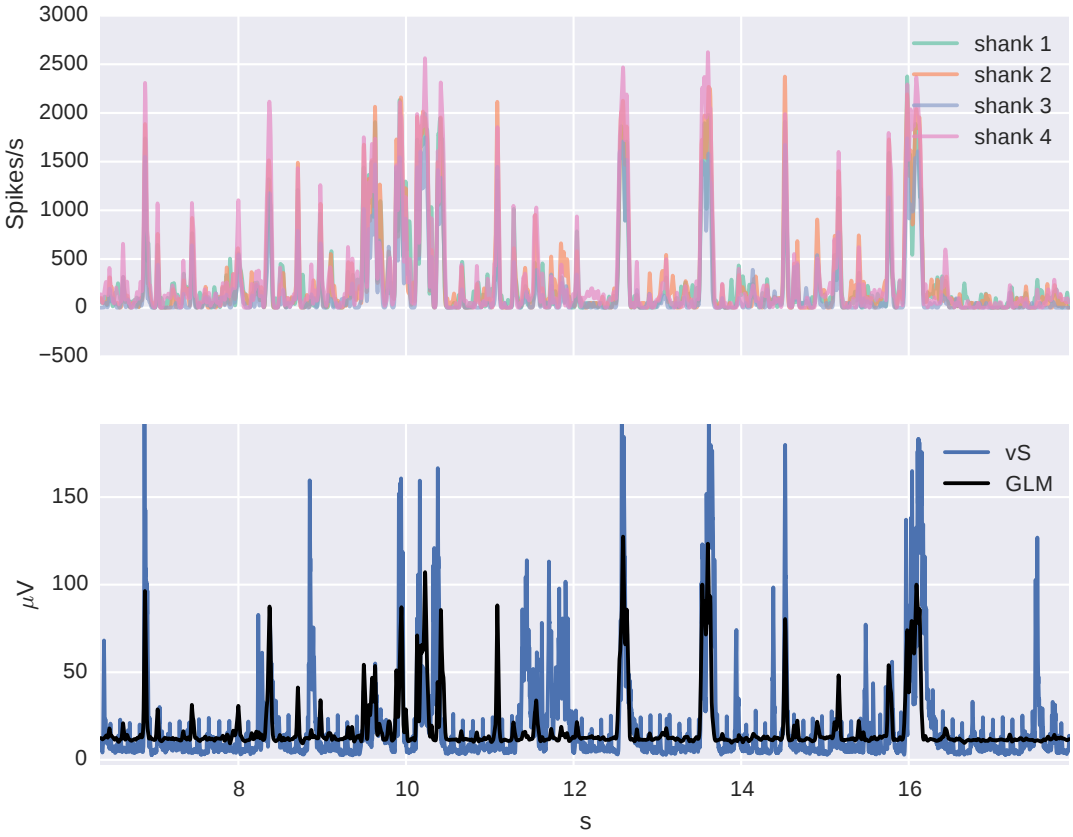

Figure 3.20. GLM prediction of vS amplitude Top: spontaneous HVC activity by electrode shank. Bottom: the spontaneous activity of vS and a prediction from a GLM model, using the four values of shank activity as input with a 15 ms lag.

activation seen during song. Similarly, EMG activity during sleep can match multi-syllable segments of pattern seen during song (G. Mindlin personal communication). Based on these results, I suspect the synchronized activation of HVC and the syrinx observed in this chapter should generalize to singing behavior.

The GTE hypothesis predicts that HVC activity should be correlated with syringeal activity. Our results strongly support that prediction.

The GTE hypothesis also predicts that HVC activity should reflect changes in the periphery with zero lag, however in examining spontaneous discharge we found that nearly all neurons with significant muscle correlations had activity that lead the muscles by 7.5–25 ms. The zero lag observations were made in singing birds, but also in head-fixed birds listening to BOS playback. Further analysis of the present recordings during BOS playback should prove to be valuable.

Another hypothesis of the GTE model is that interneurons should have minima during transitions of the song. It's not clear what the exact prediction is for the muscles of the syrinx, but simple correlation shown in this chapter is unexpected. Given this correlation, if the GTE hypothesis is correct, we would expect the muscles to have minima during GTE in singing birds, i.e. the syrinx muscles should be minimally active at the onset, offset and maxima of gestures. Another possibility is that sleep replay does not have the 0 ms lag alignment of HVC and the syrinx, but singing activity does.

In this chapter I've demonstrated the first evidence of kinematic coding in HVC during spontaneous replay. In the next chapter I will investigate whether this relationship is maintained during singing.

CHAPTER 4

ACTIVITY OF HVC AND THE SYRINX DURING SONG

4.1 Introduction

Over the years, a large body of evidence has indicated that acoustic responses in HVC are linked to motor (singing) behavior (Margoliash, 1983; Margoliash and Konishi, 1985; Sutter and Margoliash, 1994). Auditory driven activity in HVC depends on features of the individual bird's song and integrates over multiple syllables of song. Similar results have also been observed in RA responses (Dave and Margoliash, 2000), which depend on the neuromodulatory tone in HVC (Dave et al., 1998; Shea and Margoliash, 2003).

However, there is only a relatively small number of studies examining the representation of motor activity in HVC during singing (Troyer, 2016). Amador et al. (2013) found evidence for a gesture-based zero-lag code in HVC, (the GTE model, see previous chapters). Ali et al. (2013) found that changes in song timing alter the activity in HVC in a way that is consistent with motor coding. Despite these results, the dominant view of HVC coding during song is the clock model (Picardo et al., 2016; Lynch et al., 2016), which claims the activation state of HVC is independent of features of singing behavior (song). In the narrow version of the clock model, only the activity of $\mathrm{HVC_{RA}}$ during singing is modeled (Hahnloser et al., 2002). It is asserted that the timing of each $\mathrm{HVC_{RA}}$ is related to a particular moment in time, and not to any other feature of singing behavior. In a broader version of the clock model, a similar claim is made that includes $\mathrm{HVC_X}$ neurons (Picardo et al., 2016; Lynch et al., 2016).

In the last two chapters I demonstrated evidence of activity in HVC during auditory and spontaneous replay relating to song structure and to syringeal activity, behaviors which are likely intimately related to singing (Dave and Margoliash, 2000; **?**; Amador et al., 2013). Nonetheless, the strongest test of the clock model is to investigate potential motor codes in HVC during singing. In this chapter I develop techniques for making recordings with

a lightweight chronic implant with HVC electrode arrays coupled with syringeal EMGs. I use this system to test whether the activity in HVC is independent of the motor periphery during singing. This experiment tests a fundamental prediction of the clock model, and can test predictions of the GTE model as well.

4.2 Methods

4.2.1 Chronic microdrive

Custom manual screwdrives were used to move chronically implanted electrodes. The screwdrive is based on a design from Vandecasteele et al. (2012). Constructed from inexpensive parts, the drive modifies a three pin header (DigiKey S1012E-36-ND), by swapping the middle pin with a 0-90 screw, converting the plastic header block into an electrode shuttle, Fig. 4.1. Required materials: brass strip, 12"x0.5"x0.025", 3 pin connector header, #65 drill bit, 0-90 tap, 0-90 machine screw, 0-90 hex nut.

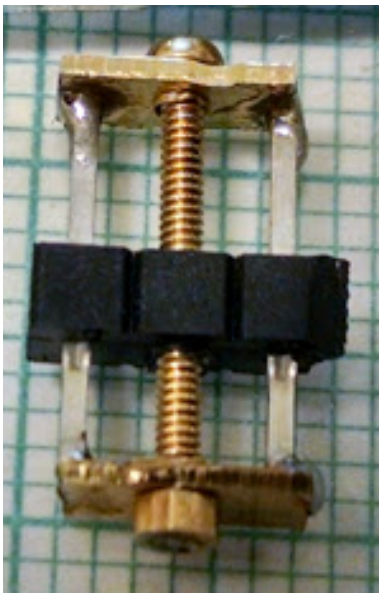

Figure 4.1. Custom screwdrive This lightweight screwdrive is easy and inexpensive to build. Design from (Vandecasteele et al., 2012).

Assembly instructions:

1. Remove middle header pin
2. Tap the middle header pin hole with the 0-90 tap
3. Cut two pieces of brass plate to the length and width of the header
4. Use a #65 bit to drill a hole into the center of both brass pieces. Steps 3-4 can be performed on a CNC mill with a 0.052" square end mill, see Fig. 4.2.
5. Pass the screw through one brass plate, the central header pin hole, the second brass plate and the nut. Tighten the nut until the brass plates apply light pressure to the two outer header pins.
6. Solder the header pins to brass plates. Also solder the nut to the screw. DO NOT solder the nut to the brass plate.
7. Test the full range of shuttle travel by turning the screw.

Figure 4.2. Milled brass tabs for screwdrive. Brass tabs milled on a ShopBot CNC router using a 0.052" square end mill.

After the screwdrive is assembled, the electrode guides and connector must be attached. Materials: Omnetics NPD-14-WD-12.0-C-GS or similar dual-row Nano pin connector, 27 gauge hypodermic tubing, 5 mil Pt/Ir wire, small sockets (TE Connectivity & Connectors 322-HCS5P2-100).

Assembly instructions:

1. Cut the connector leads to 30 mm, stripping about 2 mm of shielding from the ends
2. Cut 6 mm of hypodermic tubing, grinding off the crimped edges
3. Thread the wire from a connector lead into the tube, securing with a crimp. If not all the wire can fit inside the tube, secure the remaining wires to the outside of the tube with a layer of conductive paint followed by a layer of no-sag epoxy.
4. Repeat steps 2-3 for four total hypodermic tubes. Optionally have one tube connected to the reference channel.
5. Using no-sag epoxy, glue the hypodermic tubes, one at time, to the screwdrive shuttle, using the epoxy to electrically insulate the outside of each tube.
6. To make the ground wire, flatten a 3 mm segment of 5 mil Pt/Ir wire with pliers and solder to the ground channel connector lead.
7. Repeat step 6 for EEG/EOG channels
8. For EMG channels, solder four small sockets to four connector leads

Electrodes are installed by removing insulation at the base of the electrode and inserting the base into a hypodermic tube. A small bend in the electrode shaft will ensure an electrical connection to the barrel. A complete screwdrive array is shown in Fig. 4.3.

4.2.2 Custom lightweight headstage

The headstage used in these experiments was designed by Hamish Mehaffey, which in turn was modified from the 32 channel Omnetics minature headstage design from open-ephys.net. The headstage achieves a small form factor of 15 mm by 15 mm and a weight of only 0.74 g. Custom two layer 0.020" PCBs, Fig. 4.4 (pcbuniverse.com). I modified the Mehaffey design to include impedance measurement.

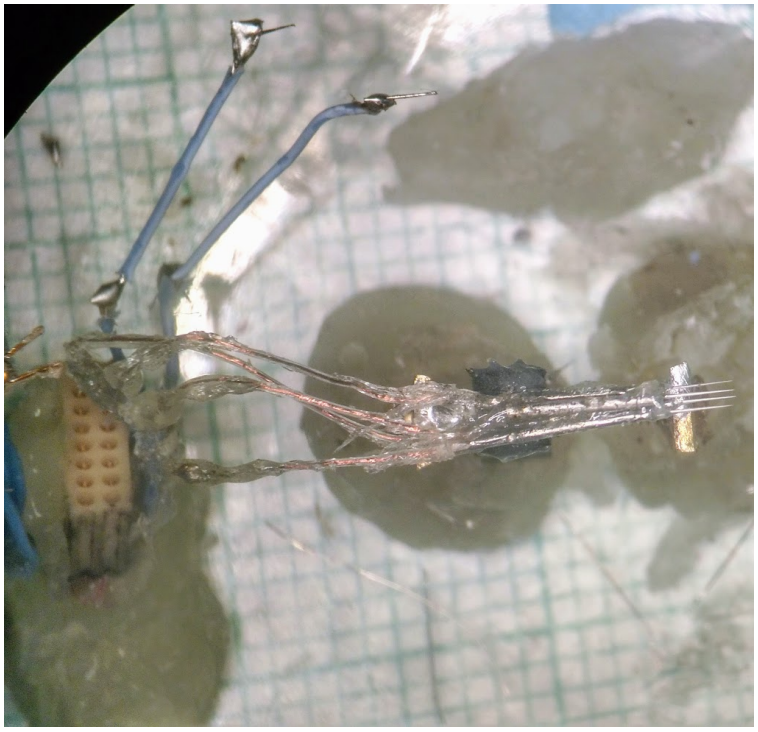

Figure 4.3. Assembled screwdrive A fully assembled screwdrive loaded with four Pt/Ir electrodes. Also visible is the Omnetics Nano pin connector and Pt/Ir EEG electrodes.

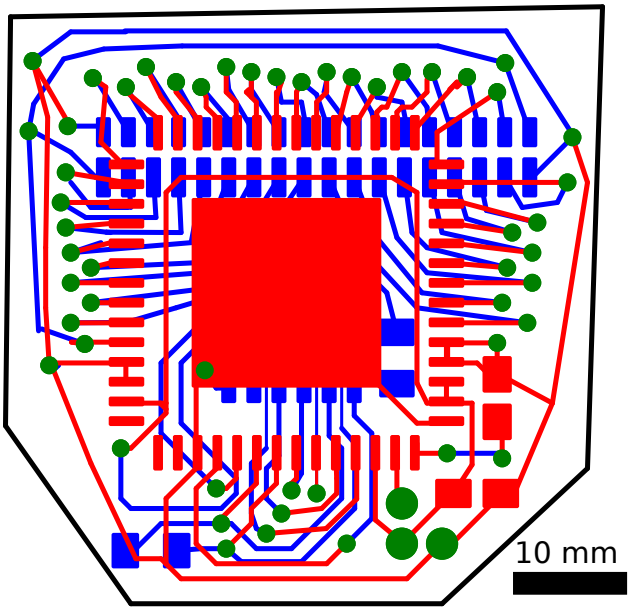

Figure 4.4. Headstage design. Two-sided PCB headstage design by Hamish Mehaffey. Red: top layer, designed to fit an Intan RHD2132 amplifier chip. Blue: bottom layer, with pads for two Omnetics connectors. Some connection points (green circles) connect the two sides of the PCB by means of vias. Full specifications are located at https://github.com/kylerbrown/headstage_nano.

4.2.3 Lightweight tether

The Intan RHD2132 amplifier connects to the Intan EVAL board using the electrical specification of the SPI protocol. This requires power, ground, and four control signals. In this application, the control signals are low-voltage differential signals (LVDS), each requiring two wires. This gives a total of 10 wires required to connect the headstage to the EVAL board.

I built custom lightweight tethers to extend from the headstage connected to the screwdrive to the top of the acoustic chamber. Materials: surgical tubing (0.080" O.D. 0.060" I.D.), 2x 12 channel bipolar Omnetics Nano connector, straight tails 2x, 8x 12" Cooner 1103 40 AWG wire, 2x 12" 32 AWG, 7 strand wire. The thicker 32 AWG wire is used for power and ground, while the ultra lightweight Cooner wire is used for the remaining channels. Each LVDS pair are twisted together, as is power and ground, resulting in 5 twisted pairs. The twisted pairs are then braided and passed through surgical tubing, which provides stiffness to help translate rotations of freely moving bird into torque on the commutator. The total length is trimmed to provide a bird full mobility in the cage. The connectors are then soldered to the cable.

4.2.4 Low torque slip-ring commutator

During chronic recordings, the bird must be able to freely move through the cage without getting tangled. A slip-ring commutator connects with the tether at the top of cage and permits continuous rotation of the tether while maintaining electrical connection, and allows the bird to freely rotate. I collaborated with Daniel Baleckaitis to design the commutator based on a design described in (Micco Jr., 1977), which specified a low torque slip-ring currently available from Moog, Inc. as products AC2690 and AC259, Fig. 4.5A. We later discovered these slip-rings to be drop-in replacements for commutators from DragonFly Inc.. Baleckaitis designed a 3D printable custom housing for the commutator, developed

the procedures for assembling a commutator, and assembled all the devices, Fig. 4.5B.

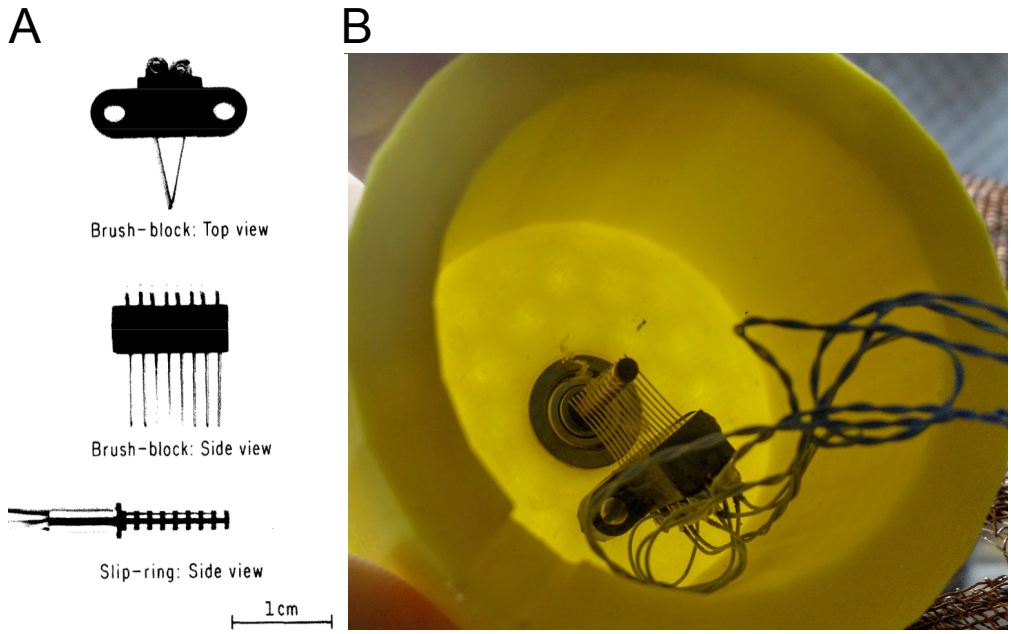

Figure 4.5. Custom slip-ring commutator A: The components of the slip-ring commutator, reproduced from (Micco Jr., 1977). B: Assembled slip-ring commutator in a 3D printed housing.

4.2.5 Microphone analog filter

Microphone signals were routed to the ADC channels of the EVAL board. Before digitization, the signals required a constant DC offset to allow the unipolar ADC channels to accept bipolar microphone signals. The microphone signals also require low-pass filtering to remove anti-aliasing artifacts. I designed a custom analog circuit to prepare four channels of acoustic data for digitization, Fig. 4.6

4.2.6 Surgery

Adult male zebra finches weighing at least 13 g were placed in sound isolation home cages one week prior to surgery. During the surgery, both the EMG electrodes and HVC screwdrive was installed, requiring about four hours of surgery time. To prevent dehydration, 50 μL saline solution was injected subcutaneously at the start of the surgery and after two hours.

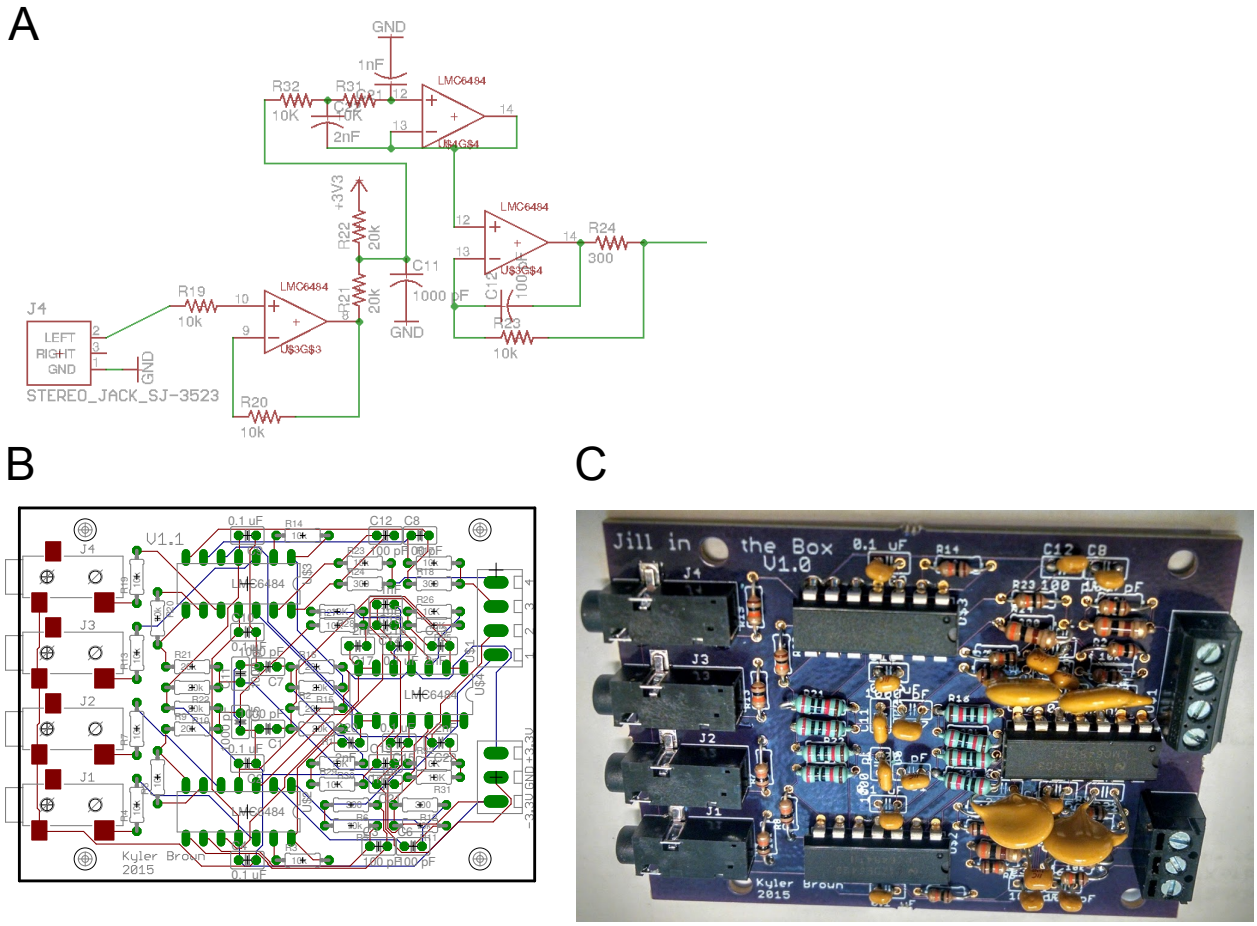

Figure 4.6. Microphone level shifting and anti-aliasing circuit. A: circuit schematic for a single channel. The first (leftmost) op-amp provides a unity-gain buffer that shifts the levels to 0–3.3 V. Resistors R21, R22 and C11 provide a first order low-pass filter with a 15 kHz cut-off. The second (center) op-amp provides a second order Butterworth low pass filter with a 10 kHz cut-off. The final op-amp (right) forms a unity-gain buffer to drive the input to the ADC. B: four channel PCB schematic, C: assembled filter board. Adapted from an Intan level-shifter schematic.

EMG electrodes were implanted using protocol outlined in chapter 3 of this thesis, but with the wires routed subcutaneously to the top of the head. This challenging surgery lead to a full recovery in 9 out of 35 birds.

After the surgery, 0.2 mg/kg of meloxicam was administered via intramuscular injection. Baytril (5 mg/kg) was also administered, once a day for 3–5 days, to prevent infection. The home cage was enriched with millet sprays, cuddle bone, mirrors and grit, in addition to food and water ad libitum. A second cage partition housed a female zebra finch to promote singing. After full recovery, birds were connected to the headstage and tether for data collection.

4.2.7 Electrodes advancement

During every day of recording, the electrodes were advanced 70 μm (a 1/4 turn of the screwdrive). This typically was done in the morning, with data collection commencing no less than 5 minutes after the array was advanced.

Eventually all the electrodes would descended past the ventral edge of HVC. This was identified by a qualitative change in broad multi-unit background activation typical of electrodes in HVC during song. The days after such occurrences, the trajectory was reversed. In a few instances good recordings were obtained while the array was being moved dorsally. The probability of successful recordings while retracting the electrodes was increased in later experiments by fixing the shaft of the electrode to the hypodermic barrel of the array with epoxy, preventing electrodes from slipping out of the barrel during retraction.

4.2.8 Processing continuously sampled data

Three streams of continuously sampled data were streamed to disk, microphone data, HVC data and EMG data. Microphone data were digitally filtered with a forward-backward highpass 3$^{\text{rd}}$ order Bessel filter with a cut-off of 300 Hz to remove fan noise.

HVC data were band-pass filtered between 300–1200 Hz with a forward-backward band-pass 3$^{\text{rd}}$ order Butterworth filter. While a Bessel filter provided better preservation of the spike shape, a Butterworth filter resulted in better spike detection, see (Wiltschko et al., 2008). To help remove movement artifacts, the median signal across all HVC electrodes was used as a digital reference. However, because each electrode had a variable amount of movement contamination, possibly due to variable impedance, a coefficient was found to optimally scale the reference signal before it was subtracted from the electrode. For each channel i, the resulting signal is

$$\bar{a}_i(t) = a_i(t) - c_i * median(a_{n \neq i}(t))$$

where c_i is a reference scaling coefficient, such that the RMS value of $\bar{a}_i(t)$ was minimized, and the reference signal is the median value across all remaining electrodes.

EMG data were band-pass filtered between 150–600 Hz with a forward-backward band-pass 3$^{\text{rd}}$ order Butterworth filter, and digitally referenced using the method described for HVC electrodes.

4.2.9 Spike sorting

The Plexon OfflineSorter software package was used to detect and sort spikes from the HVC data. OfflineSorter provides a rich environment for analyzing single electrode recordings. The electrodes in the arrays used for these experiments were separated by approximately 250 μm, more than the $\sim$ 200 μm maximum separation for seeing a given unit on multiple electrodes. Thus, the signal from each electrode was analyzed independently of the signals from the other electrodes. In previous chapters, other methods (Spyking Circus, KlustaKwik/Phy) were used to take advantage of the dense electrodes arrays, but these spike sorting methods were not applicable for the array described in this chapter.

The following procedure was used to spike sort the data with Plexon OfflineSorter. After

the raw data were loaded, I selected a spike detection threshold, based on the distribution of peaks in waveform and examples of elevated background activity indicative of song (I did not have access to the acoustic data during spike sorting). The spike detection threshold was placed at 6 standard deviations, though this threshold was increased for traces with larger spike amplitudes. Then a positive artifact threshold was set, to just above the amplitude of the largest spikes in the dataset (identified by shape). Any waveforms with amplitudes larger than than the artifact threshold were eliminated, as were any spikes within 10 ms after a spike crossing the artifact threshold. This step was repeated for a negative artifact threshold. I used the principal components to identify clusters of spikes likely generated by a single source. Waveforms at the center of the cluster were used to create a spike template. Any waveforms within two mean absolute deviations were assigned to the spike cluster, though this threshold was manually adjusted to preserve the correspondence between the template waveform and the mean waveform, and to provide a qualitative match to size of the cluster.

After sorting spikes into separate clusters, I evaluated the quality of the sorting procedures on a cluster-by-cluster basis. Clusters were labeled as "good" (meaning well isolated single units) or multi-unit activity (MUA) based on the size and shape of the average spike waveform, the percentage of refractory period violations and the shape of the inter-spike interval (ISI) histogram, Fig. 4.7. Plexon OfflineSorter does not allow clusters to be compared with acoustic data, which made it impossible to sort spikes based on their response to song. This limitation was particularly acute with projection neuron spikes, which have very sparse firing at precisely timed moments in song. It is possible that a spike sorting approach which incorporates extra-channel data, especially the acoustic data, could greatly improve the spike sorting and in particular help isolate HVC projection neurons. After spike sorting was complete, each unit was individually examined. Multi-units were classified as either responsive, if they exhibited a stereotyped response across renditions of the motif,

or non-responsive. Good units were classified into non-responsive, tonic or phasic. Tonic neurons displayed a stereotyped firing pattern across renditions of the motif, while phasic neurons fired primarily as a single location. I considered tonic neurons as putative HVC interneurons, and phasic neurons as putative HVC projection neurons, however given the difficultly of accurately spike sorting HVC projection neurons, this category of cells should be considered to have heavy multi-unit contamination.

Plexon OfflineSorter was also used for spike detection of the EMG channels with a threshold of -9 standard deviations. As in the treatment of brain recordings, times of movement artifact were detected and removed. No spike sorting was attempted on the EMG data, though it may have been possible. It may be of interest to future experimenters to spike sort dense EMG arrays.

4.2.10 Syllable detection and labeling

The large number of vocalizations recorded prohibited a complete manual review and analysis of the acoustic data (song recordings). Instead, a semi-automated approach blended hand labeled data with machine learning. For each bird, vocalizations were detected using an approach outlined in Koumura and Okanoya (2016). First, segments that surpassed a manually adjusted amplitude threshold were labeled. Second, adjacent segments with silent intervals shorter than 20 ms were joined. Third, any segment shorter than 20 ms was removed. Next the segments were hand labeled with custom software, typically 200–500 motifs. These labels were then used to train a deep convolutional neural network (CNN). The input dimensions were 81×66, representing 66 frequency values over 81 timepoints spanning 691.2 ms and frequencies between 300–8000 Hz. The output is the probabilities of each syllable type (or silence) at the central time of the window. The input is then shifted one timestep to the right, and a new prediction is generated for the new central time. The model is similar in approach to (Koumura and Okanoya, 2016), but with additional layers. In particular, the

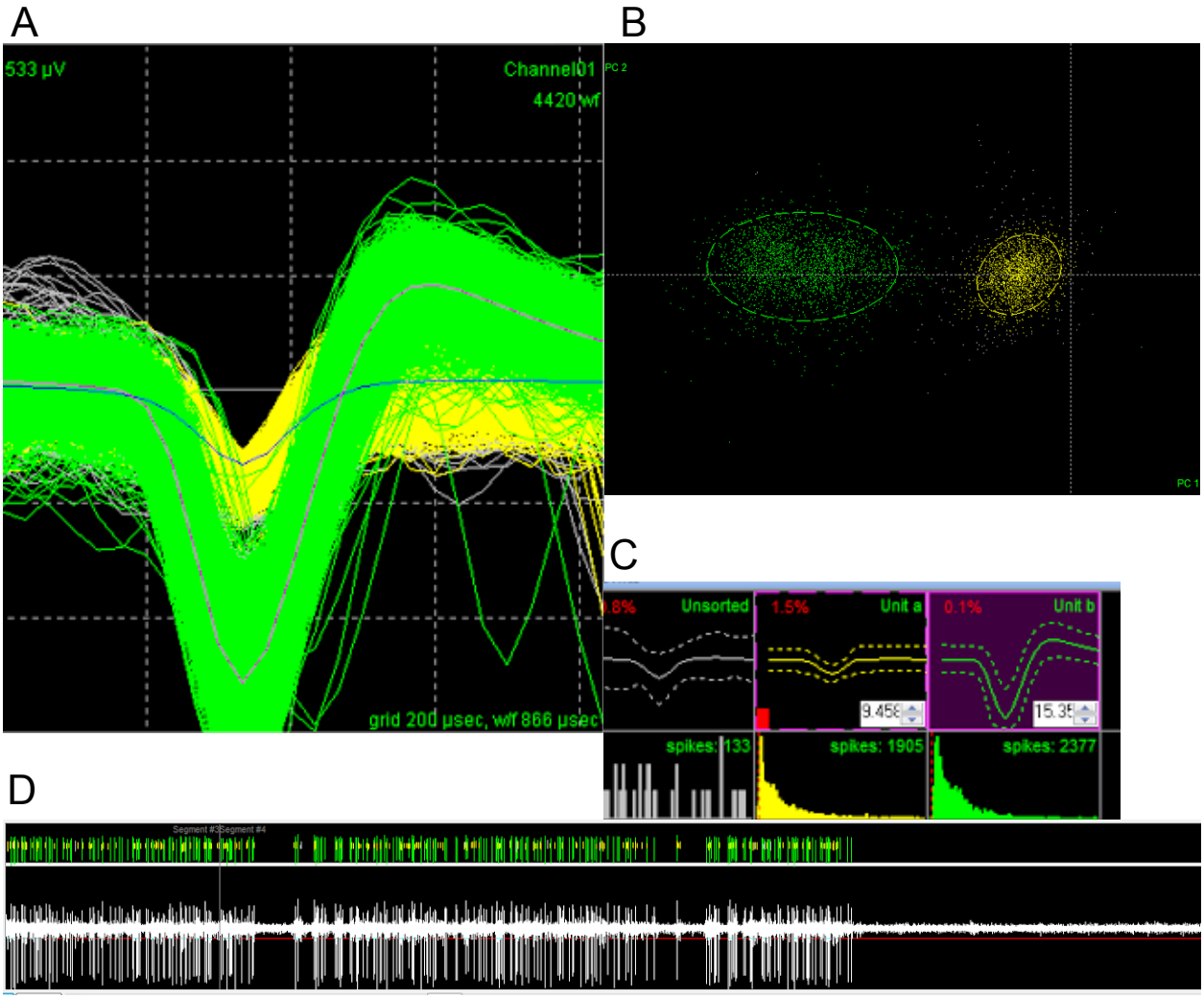

Figure 4.7. Plexon OffLine Sorter example. Plexon Offline sorter user interface windows from a spike sorting session from bird orange184. A song selective interneuron is shown in green, while multi-unit activity is shown in yellow. A: Cluster waveforms. B: Spike waveforms projected onto the first two principal components. C. ISI distributions. D. A segment of raw voltage data over time.

addition of batch normalization layers drastically decreased the time to convergence (Ioffe and Szegedy, 2015).

The structure of the network is shown in Table 4.1. There are three types of layers, convolution, max pooling and batch normalization. The convolution layers perform a convolution on the data, while the pooling layers reduce the dimensionality by taking only the maximum value from a local region of the data matrix. Final processing is done by flattening the data matrix and using a traditional multi-layer perceptron architecture (dense layers) to classify the data into one of the syllable categories. The drop-out layer randomly removes a fraction of the dense units to help prevent over-fitting. For a detailed description of convolutional neural networks see Goodfellow et al. (2016).

The performance of the model exceeded 99% accuracy for a holdout dataset. For one bird, (Black206) a syllable was systematically missed at the end of the motif, requiring additional training data for the model to perform correctly.

The model was implemented in Keras, a neural network library built on top of TensorFlow. TensorFlow uses a flexible architecture that allows GPU accelerated computation. We ran our models on a NVidia GTX 1080 ti graphics card. The amount of time required to train the network for an individual bird was 20 minutes, and the remaining acoustic data were classified by the network in 2–7 hours depending on the amount of data. In total about 500,000 syllables were classified over roughly 24 hours. The primary computational bottleneck was computing short fast Fourier transform windows to generate the input arrays for the model, which ran on a single CPU. A simple solution to this bottleneck was to run 2+ jobs simultaneously, using 2+ CPUs, and splitting the GPU memory between jobs.

Using this song parser, all acoustic data without manual labels were analyzed.

Layer	Output Shape	Param #
Conv2D	81, 62, 16	96
MaxPooling2	81, 31, 16	0
Batch	81, 31, 16	64
Conv2D	81, 28, 16	1040
MaxPooling2	81, 14, 16	0
Batch	81, 14, 16	64
Conv2D	81, 11, 32	2080
MaxPooling2	81, 5, 32	0
Batch	81, 5, 32	128
Conv2D	81, 2, 32	4128
MaxPooling2	81, 1, 32	0
Batch	81, 1, 32	128
Conv2D	78, 1, 64	8256
MaxPooling2	39, 1, 64	0
Batch	39, 1, 64	256
Conv2D	36, 1, 64	16448
MaxPooling2	18, 1, 64	0
Batch	18, 1, 64	256
Conv2D	15, 1, 128	32896
MaxPooling2	7, 1, 128	0
Batch	7, 1, 128	512
Conv2D	4, 1, 128	65664
MaxPooling2	2, 1, 128	0
Batch	2, 1, 128	512
Flatten	256	0
Dense	256	65792
Dropout	256	0
Dense	8	2056
Total params		200,376
Trainable params		199,416
Non-trainable params		960

Table 4.1: CNN syllable classifier model architecture.

4.2.11 Time warping song motifs

While song is highly stereotyped, there is some natural variation in the duration of motif of 5–10% (Aronov and Fee, 2012). To align all repetitions to a single motif, I used dynamic time warping (DTW) (Anderson et al., 1996), Fig. 4.8. DTW tries to find the most parsimonious mapping from one rendition of a song to another by shrinking or stretching time over small segments, such that the two spectrograms are maximally aligned. The DTW procedure creates a warp trace, which maps time from one motif to time in another, Fig. 4.9. This allowed all motif data to be compared against a single canonical motif.

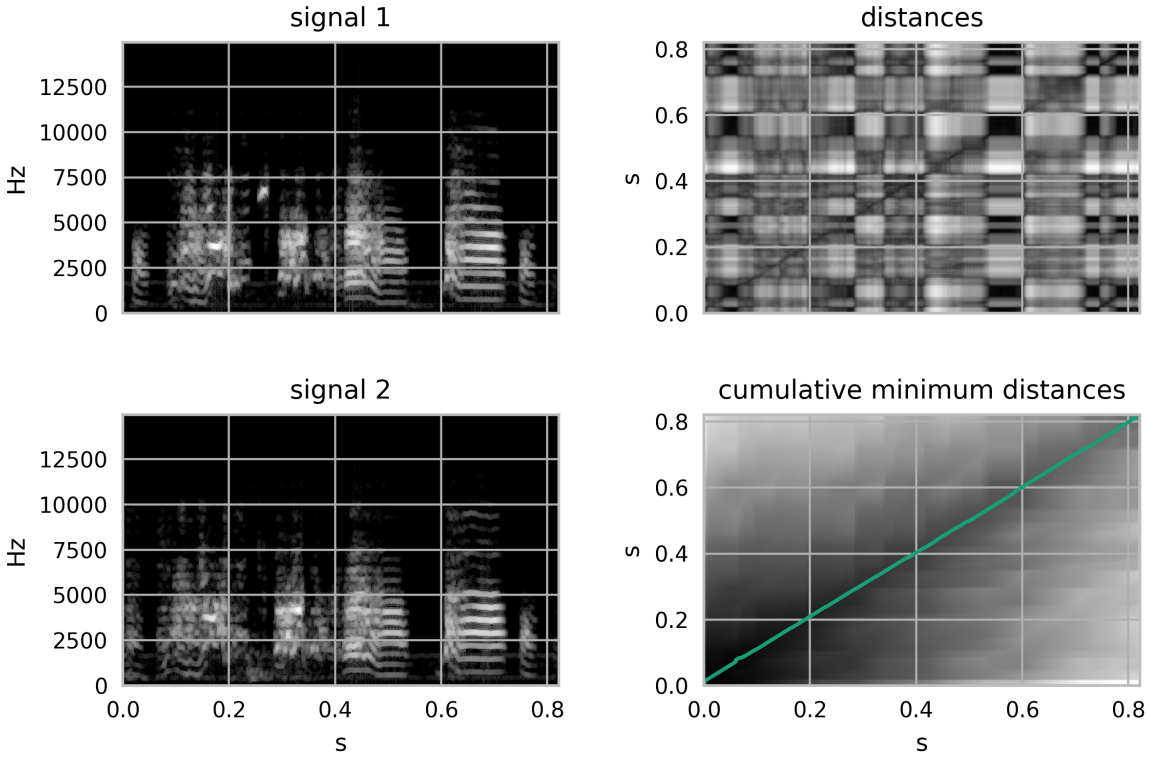

Figure 4.8. DTW algorithm. This figure demonstrates the components of the DTW algorithm. Top right: The pairwise absolute difference in frequencies across all times between the two signal. Bottom right: The cummulative minimum path from the beginning to the end of signal 2. In Green: the warp path that maps times in signal 2 to times in signal 1. The early introductory note in signal 2 results in a slightly positive Y intercept, mapping the first milliseconds of signal 2 onto the beginning of signal 1.

A pad of 100 ms was added to each side of the motif, to ensure the warp found the correct

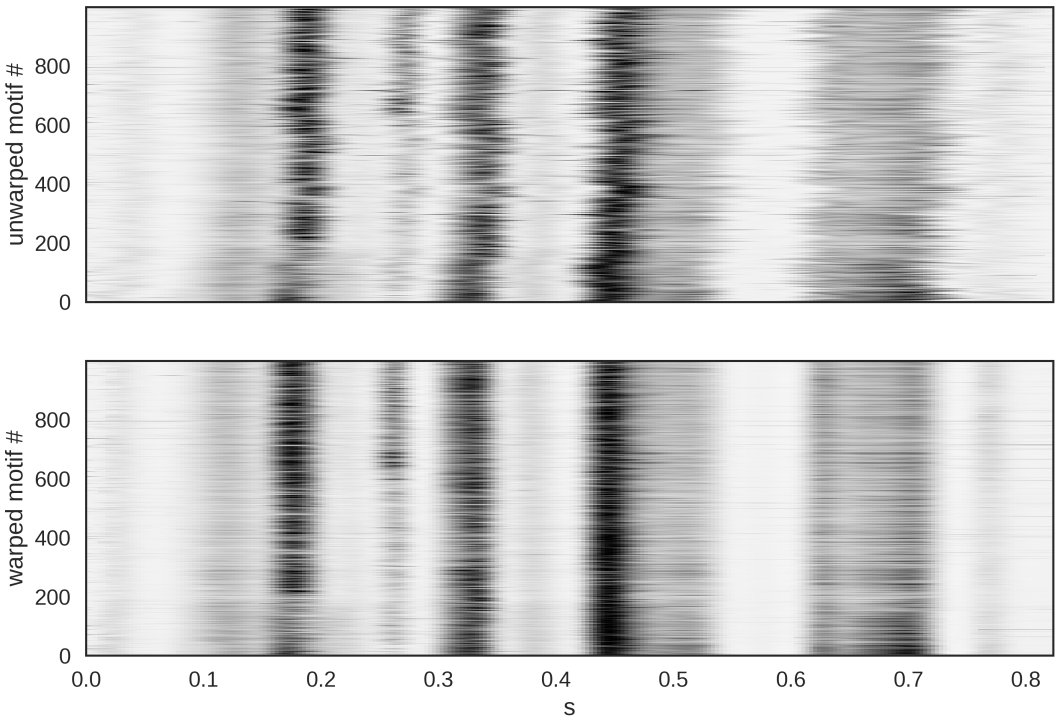

Figure 4.9. DTW aligned motifs Top: Amplitude over time of the first 1000 motifs of one bird, Black33. Bottom: The amplitude of the same motifs after dynamic time warping.

onset and offset if the labels had some error, and to allow for analysis of time lags near the onset and offset of the motif.

4.2.12 Spike triggered averaging

To compute spike triggered averages (STAs), the enriched dataset was used, increasing the density of song in the dataset, however also including introductory notes and some calls. For each spike of an HVC unit, an interval of 200 ms, centered on the spike was collected and averaged over all spikes of that unit, see Chapter 3 and (Dayan and Abbott, 2001) for details on computing STAs.

4.2.13 Linear model relating HVC and syringeal activity

A normalized firing rate was computed over the motif of length T from time-warped spikes from all responsive N units and M multi-units, and smoothed with a 5 ms standard deviation Gaussian kernel (see Chapter 2). The dimensionality of the HVC response was reduced from $(N + M) \times T$ to $3 \times T$ via principal component analysis.

The array of PCA values over time formed the input matrix to the linear model, $\mathbf{x}(t)$. The target output, $y(t)$ was found by computing the average time warped firing rate across the EMG channels for a day with excellent EMG recording quality.

A linear model was created to predict the EMG activity at time t,

$$\hat{y}(t) = \mathbf{a}\mathbf{x}(t + k) + b.$$

The coefficients $\mathbf{a}$ and b were fit using ordinary least squares.

The parameter k is a constant offset, allowing the model to predict muscle activity with zero delay ($k = 0$), a premotor delay ($0 < k < 25$ms) or other delays. I used the ranges (-300 ms $< k <$ -50 ms) and (50 ms $< k <$ 300 ms) to generate a null distribution of predictions

for a given experiment. The logic of this choice is that even in any cases where HVC might exhibit simple linear premotor encoding in the time range -50 ms $< k <$ 50 ms, it is likely to exhibit weaker or no simple linear premotor encoding at more distant intervals.

To prevent over-fitting, K-fold cross-validation was used to generate $\hat{y}$. The song was divided into 3 equal partitions, and the model coefficients were fit to the data in the first two partitions and tested on the last partition. This model fitting and prediction procedure was repeated for all 3 combinations of possible test/train splits, resulting in a complete prediction of the motif by appending the predictions of the three models.

4.3 Results

Data were collected from 9 birds. One bird, Orange240, was removed from the analysis because the electrode array was not placed in HVC, another bird, Orange247, had very little neuronal data but considerable singing data, and accordingly was excluded from neuronal but not behavioral analyses. Recordings ended after the HVC electrodes failed. On average, 138.1 $\pm$ 31.7 motifs per day were detected in the seven birds providing HVC data (N = 7 birds, 138 days, 19046 motifs) (see Table 4.2). The relatively low number of average motifs per day were the result of the burden of carrying the headgear, which reduced a bird's proclivity to sing.

For computing spike triggered averages, all spike sorted data for each unit were used. For analyses of motif data, only the time-warped data from each rendition were used, and each motif was considered a trial.

From the seven birds, a total of 356 MUAs and 205 units were recorded. I first assessed a stereotyped response to song by inspection. I rejected 102 MUAs and 73 units for failing to display a stereotyped response to song. Of the remaining units, 49 displayed phasic firing patterns, indicative of HVC projection neurons (though partially contaminated with MUA in all cases), and the remaining 103 units had tonic activity during song, and are considered

Bird	# of motifs	# of HVC recording days
black206	832	10
black33	6946	31
black5	33	5
orange184	3818	26
orange240	3431	16
orange247	246	1
pink119	8	3
pink121	640	24
white17	6660	38

Table 4.2: Number of detected motifs per bird for all seven birds providing usable HVC recordings

Bird	MUA	Interneuron	Projection	non-responsive MUA	non-responsive units
Black206	37	8	0	1	0
Black33	90	36	23	5	6
Black5	14	7	2	1	2
Orange184	47	14	17	18	17
Orange240	0	0	0	8	6
Orange247	1	0	1	0	0
Pink119	8	4	0	0	1
Pink121	39	16	5	4	3
White17	19	18	2	65	38

Table 4.3: Breakdown of neural data by bird. Numbers of responsive sites with multiunit activity (MUA), tonic neurons (Interneuron), and phasic neurons (Projection), and numbers of MUAs and units that were not unresponsive during singing.

putative HVC interneurons. This is the commonly used definition in the field, e.g. Amador et al. (2013); Lynch et al. (2016). In what follows, projection neurons and interneurons are used as descriptive terms, without any direct evidence of the projection targets of the neurons.

While each bird had a canonical motif of two or more syllables, there was some variation in the syllable transitions. The amount of variation was bird specific, see Fig. 4.10. Only data collected during canonical motifs were selected for analysis completed to date.

For each bird, the data from a microphone, the HVC electrodes and the EMG electrodes were continuously recorded, including during singing. Digital filtering and referencing helped reduce movement artifacts and single units from HVC and the EMG from vS could be reliably

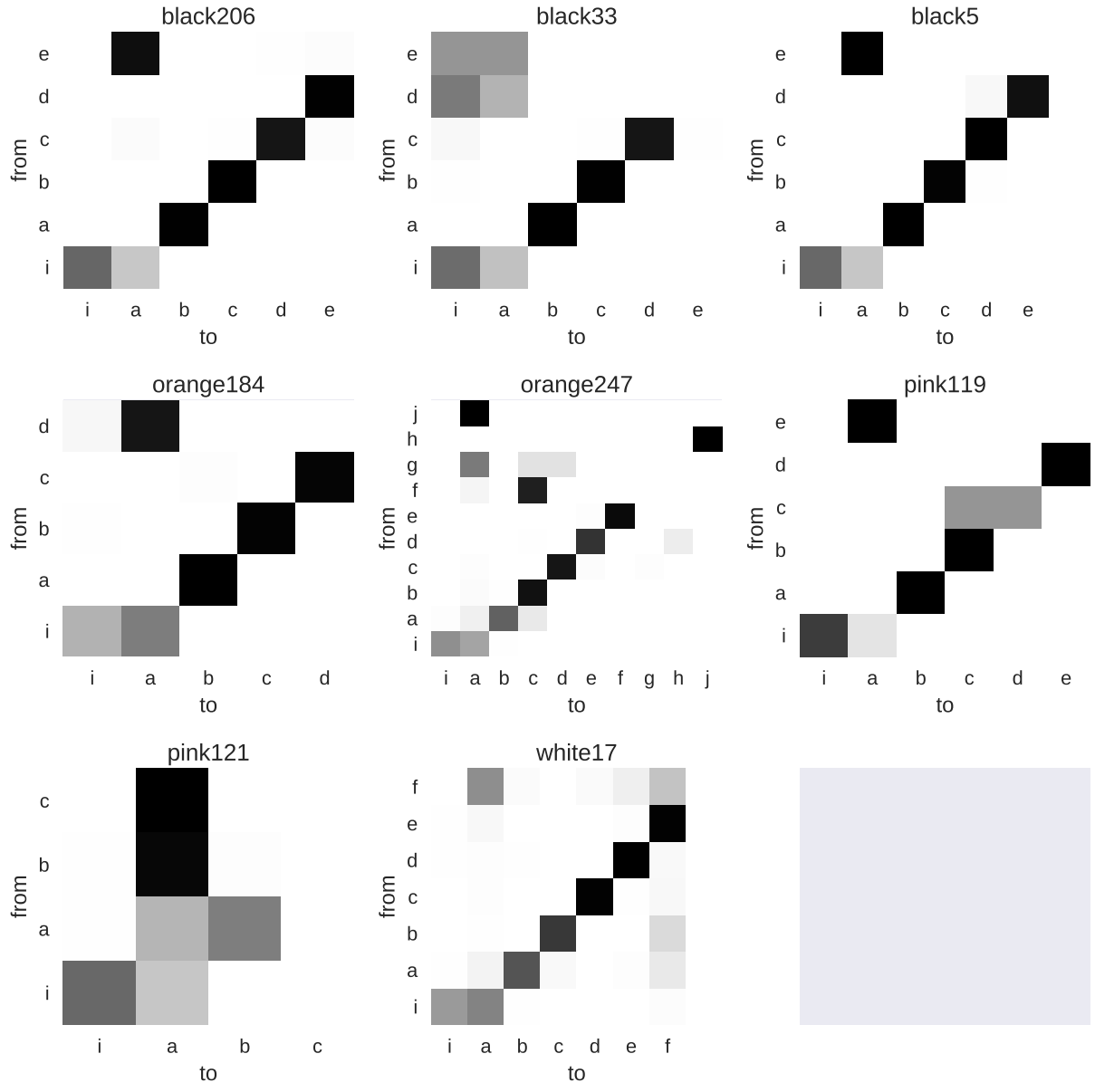

Figure 4.10. Syllable transition probabilities for each bird. White indicates a transition probability of zero while black indicates a probability of one. Note: Pink121's cannoical motif was AABA.

recorded, Figs. 4.11, 4.12, 4.13, 4.14, 4.16, 4.17 and 4.18 These figures depict raw data examples from which single units were extracted.

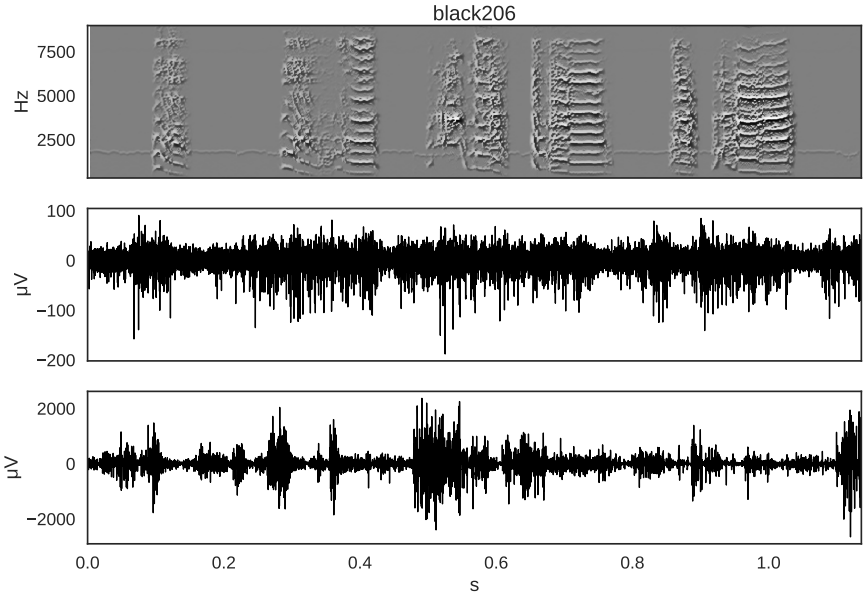

Figure 4.11. Raw voltage data for bird Black206. Top: spectrogram of one motif. Middle: voltage trace from an HVC electrode. Bottom: voltage trace from an EMG electrode.

After spike sorting, units were classified as tonic, phasic or MUA (see Table 4.3). Representative samples of the spike times for these units are displayed in Figs. 4.19, 4.20, 4.21, 4.22, 4.23, 4.24, 4.25, and 4.26. Due to limitations of the spike-sorting process (see methods), projection neurons are contaminated with background multi-unit activity. Also, trial-to-trial variability in individual interneurons rates is likely due to changing signal-to-noise ratios as the bird moves in the cage, which cause small changes in the location of electrodes in the brain.

Similar to the sleeping data in chapter 3 of this thesis, MUAs and interneurons appear to co-activate with syringeal activity. That is, regions of dense neural activity appear to coincide with vS muscle activation. To investigate this phenomenon, I performed spike triggered averaging (STA) of the vS EMG signal and the microphone amplitude, Figs. 4.27, 4.28, 4.29, 4.30, 4.31, 4.32 and 4.33. See the methods of Chapter 3 for more details on the

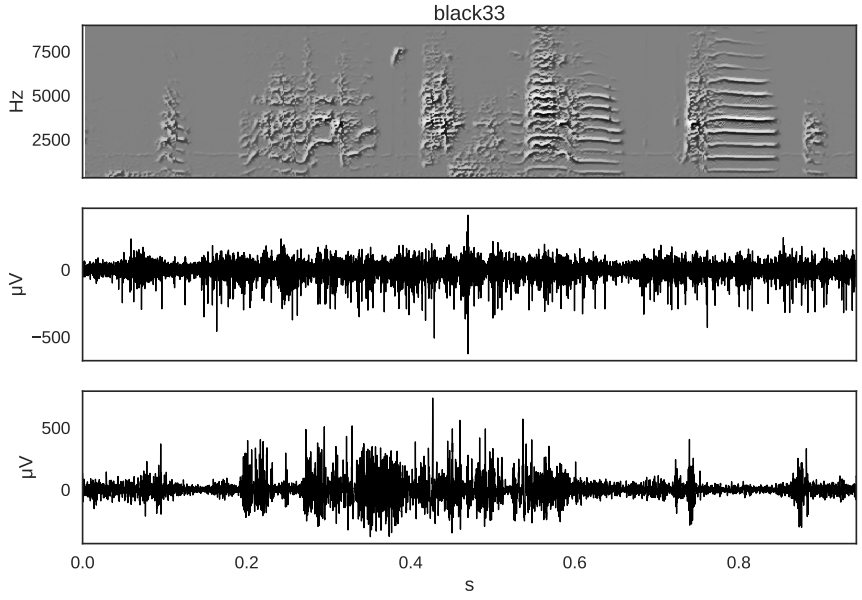

Figure 4.12. Raw voltage data for bird Black33. Top: spectrogram of one motif. Middle: voltage trace from an HVC electrode. Bottom: voltage trace from an EMG electrode.

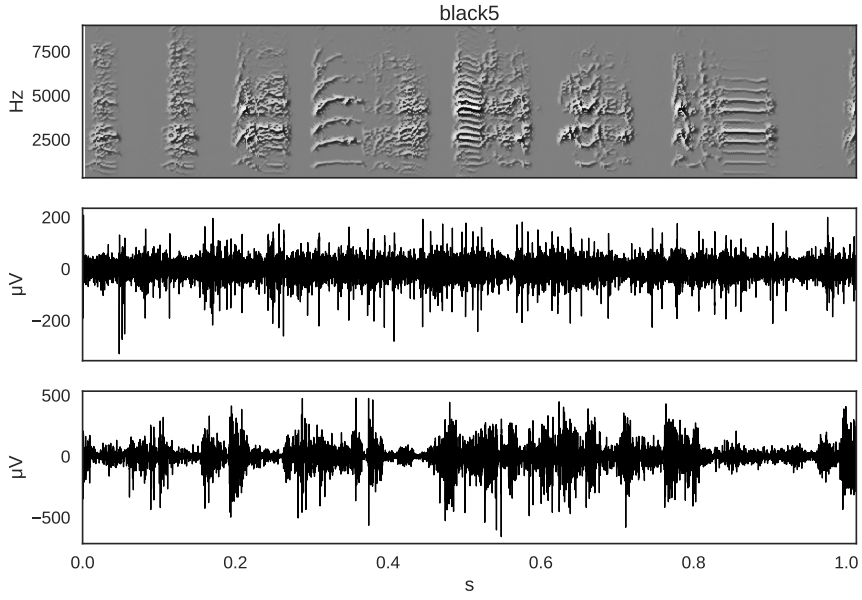

Figure 4.13. Raw voltage data for bird Black5. Top: spectrogram of one motif. Middle: voltage trace from an HVC electrode. Bottom: voltage trace from an EMG electrode.

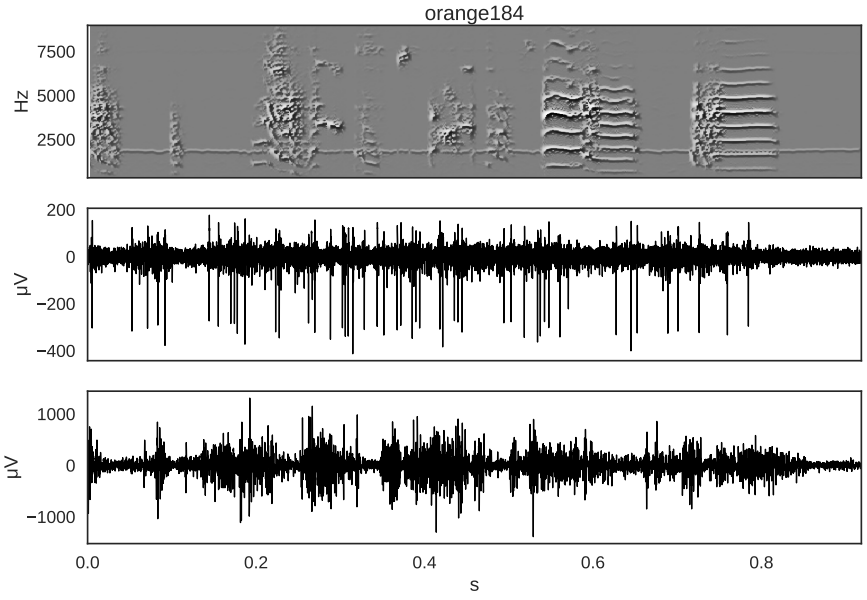

Figure 4.14. Raw voltage data for bird Orange184. Top: spectrogram of one motif. Middle: voltage trace from an HVC electrode. Bottom: voltage trace from an EMG electrode.

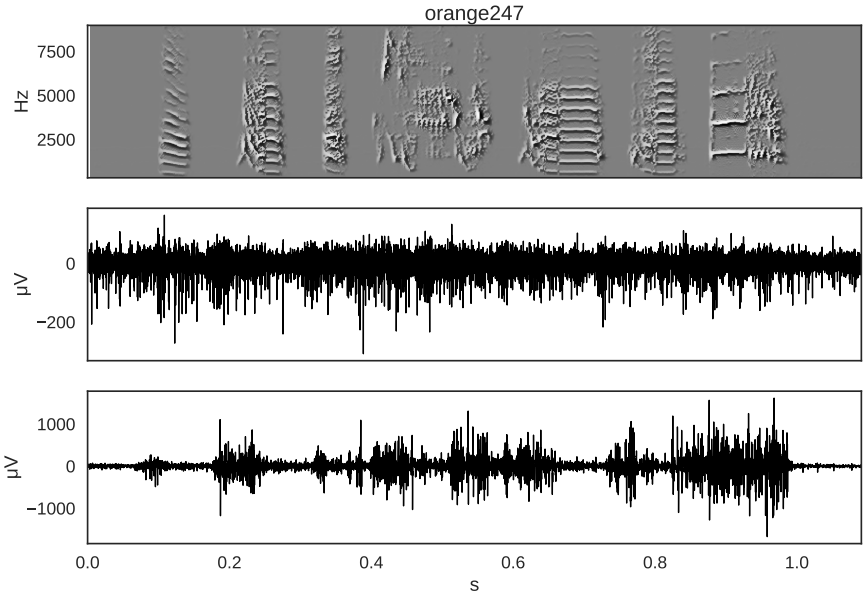

Figure 4.15. Raw voltage data for bird Orange247. Top: spectrogram of one motif. Middle: voltage trace from an HVC electrode. Bottom: voltage trace from an EMG electrode.

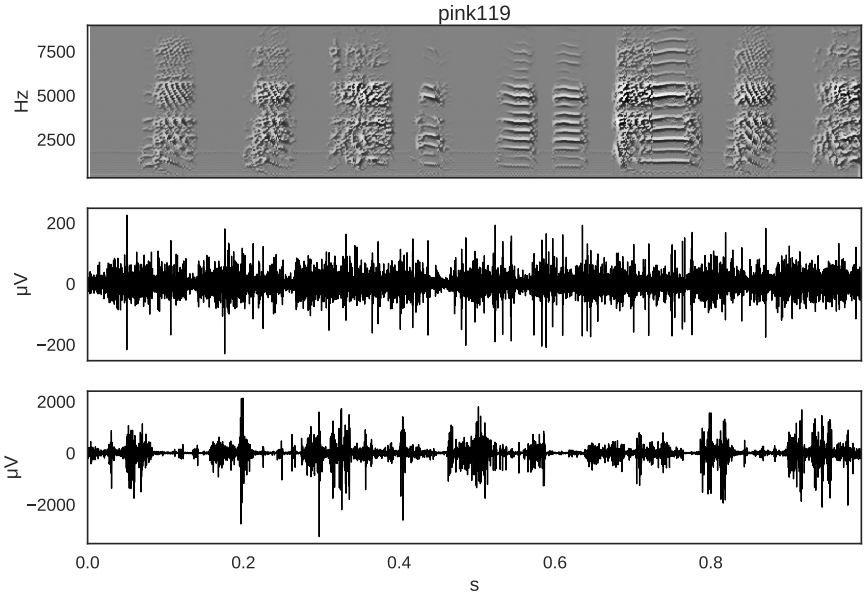

Figure 4.16. Raw voltage data for bird Pink119. Top: spectrogram of one motif. Middle: voltage trace from an HVC electrode. Bottom: voltage trace from an EMG electrode.

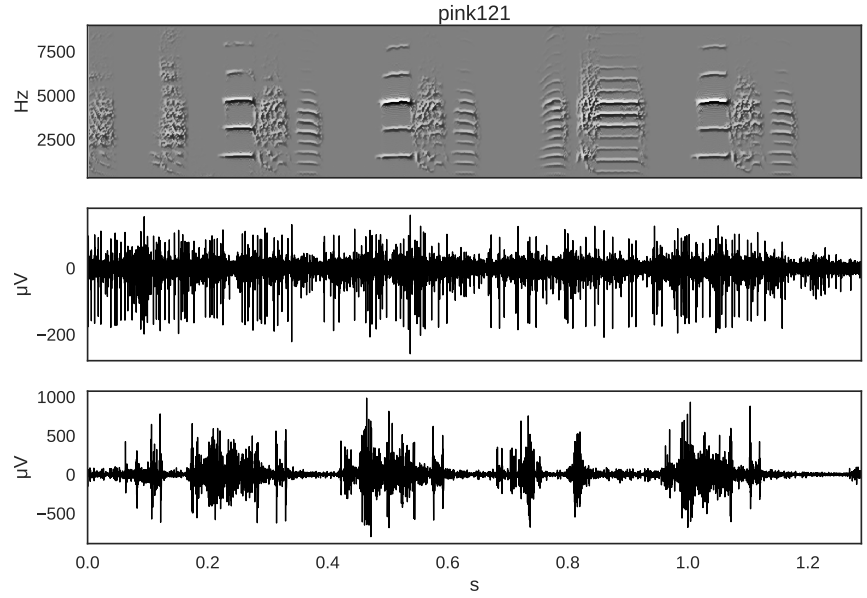

Figure 4.17. Raw voltage data for bird Pink121. Top: spectrogram of one motif. Middle: voltage trace from an HVC electrode. Bottom: voltage trace from an EMG electrode.

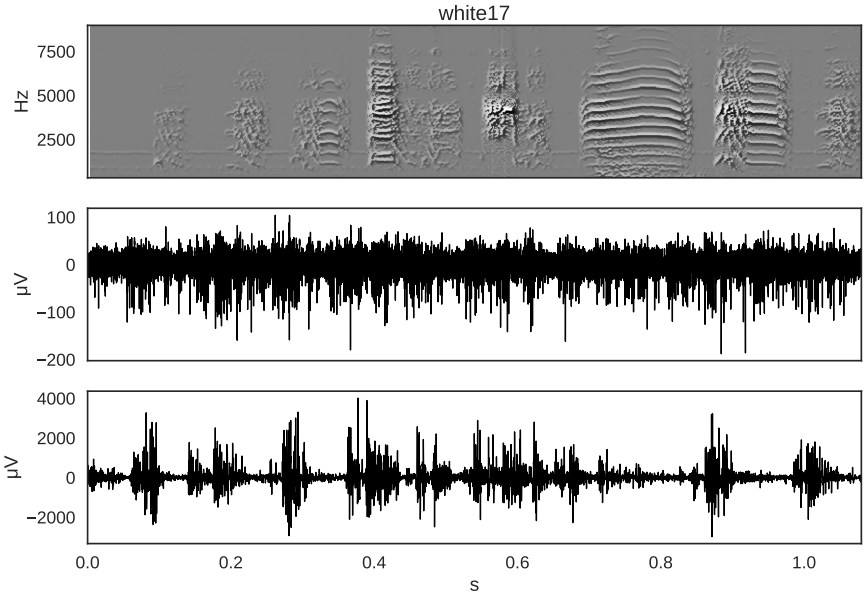

Figure 4.18. Raw voltage data for bird White17. Top: spectrogram of one motif. Middle: voltage trace from an HVC electrode. Bottom: voltage trace from an EMG electrode.

STA calculation. For these STAs, any spike within 10 seconds of the beginning or end of a motif was included (the enriched dataset). Thus each calculated STA uses all the data spike sorted for each unit.

The average time of the EMG STA peak (across units for all birds) for putative interneurons was 16.0 ms (7.54–24.1 ms 95% confidence interval). This is significantly different from 0 (T-test, T=3.8, $p<0.001$). For multi-units, this delay was a slightly shorter 12.5 ms (8.4–16.4 ms 95% confidence interval), and the mean was also significantly different from 0 (T-test, T=5.9, $p<10^{-8}$). While the mean of the distribution is significantly non-zero, there appear to be zero-lag multi-units, though these may represent movement artifacts as the contamination of a common noise source in will result in artificially higher correlation between the two signals at zero time lag.

Slightly larger delays were found in comparison with the microphone amplitude, as would be expected in a causal chain, Figs. 4.34, 4.35, 4.36, 4.37, 4.38 and 4.39. Interneurons had an average peak microphone amplitude STA lag of 30.0 ms (21.1–38.3 ms 95% CI) and multiunits

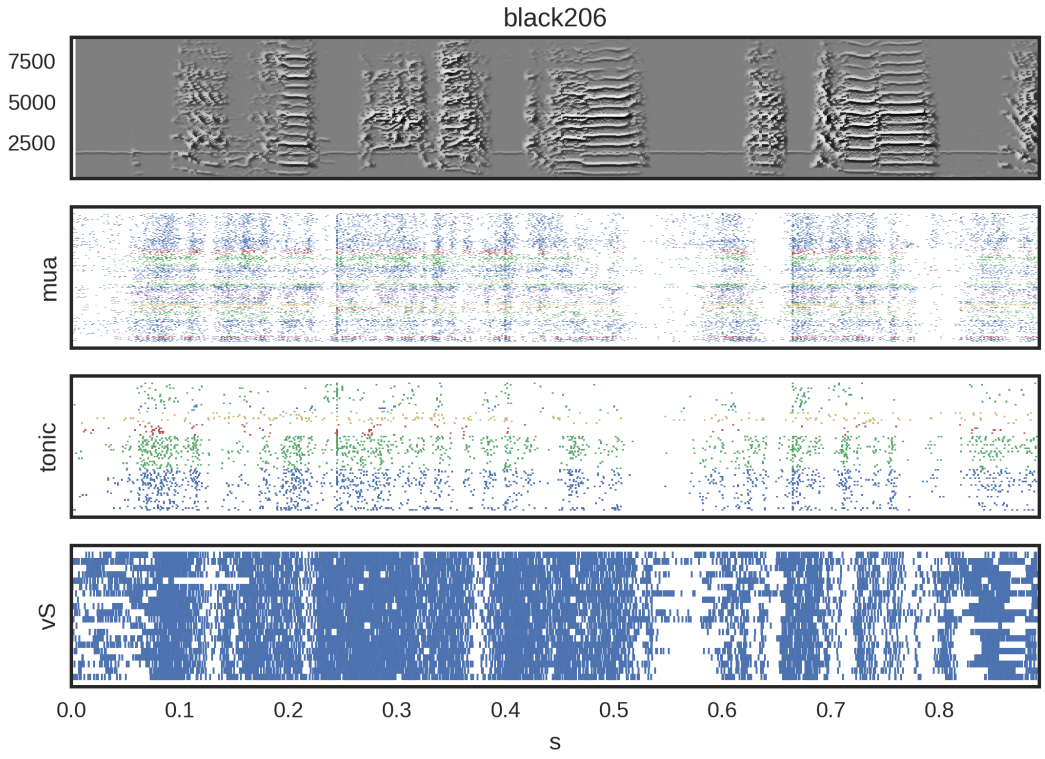

Figure 4.19. Warped motif rasters for Black206. Firing of HVC neurons relative to the cannonical motif. Top: spectrogram of one motif. Middle: representative rasters for HVC units. Bottom: spikes detected from the vS EMG channel.

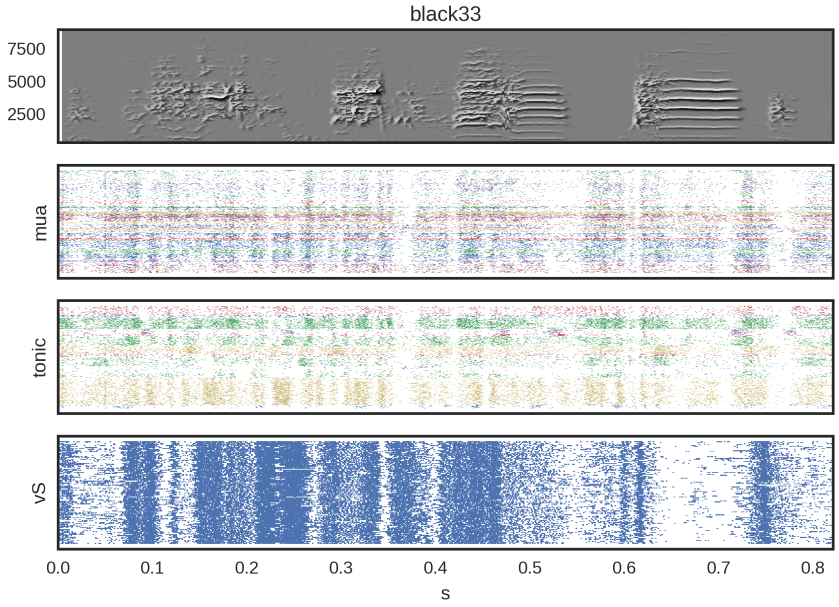

Figure 4.20. Warped motif rasters for Black33. Firing of HVC neurons relative to the cannonical motif. Top: spectrogram of one motif. Middle: representative rasters for HVC units. Bottom: spikes detected from the vS EMG channel.

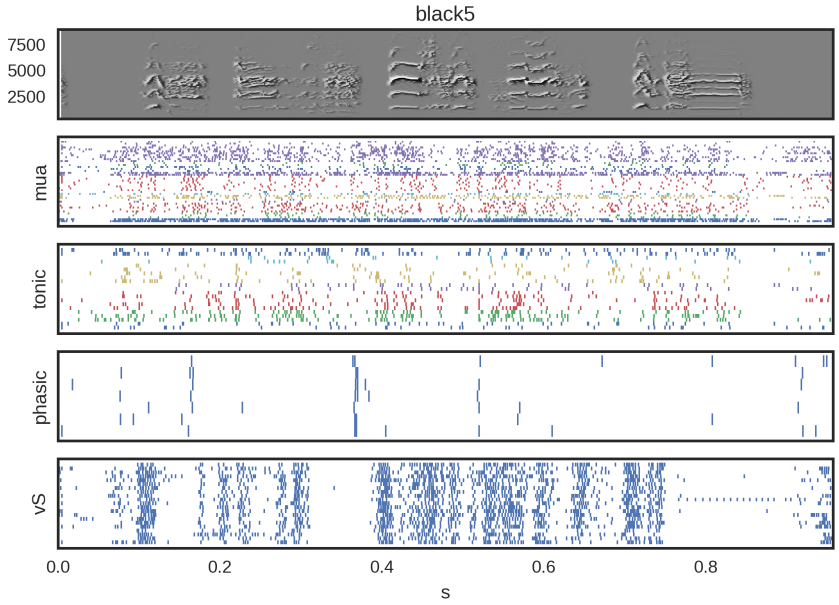

Figure 4.21. Warped motif rasters for Black5. Firing of HVC neurons relative to the cannonical motif. Top: spectrogram of one motif. Middle: representative rasters for HVC units. Bottom: spikes detected from the vS EMG channel.

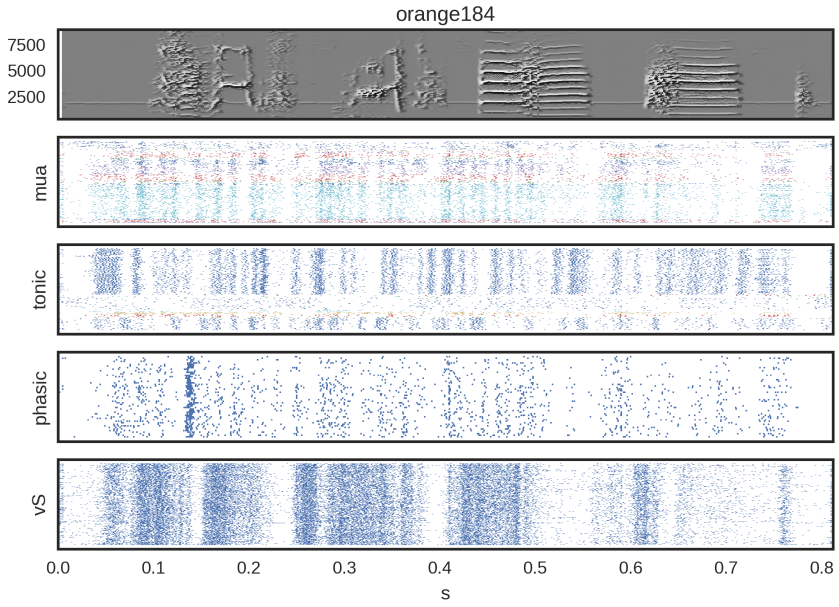

Figure 4.22. Warped motif rasters for Orange184. Firing of HVC neurons relative to the cannonical motif. Top: spectrogram of one motif. Middle: representative rasters for HVC units. Bottom: spikes detected from the vS EMG channel.

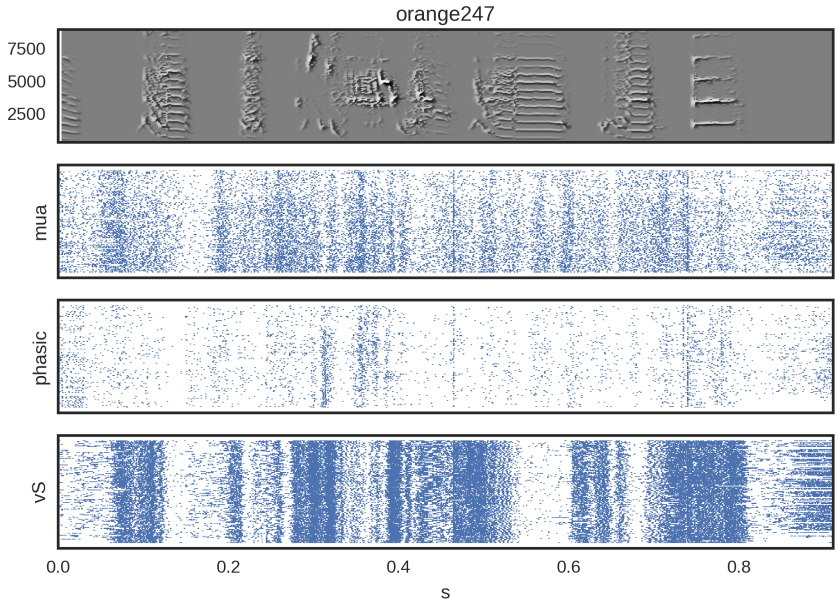

Figure 4.23. Warped motif rasters for Orange247. Firing of HVC neurons relative to the cannonical motif. Top: spectrogram of one motif. Middle: representative rasters for HVC units. Bottom: spikes detected from the vS EMG channel.

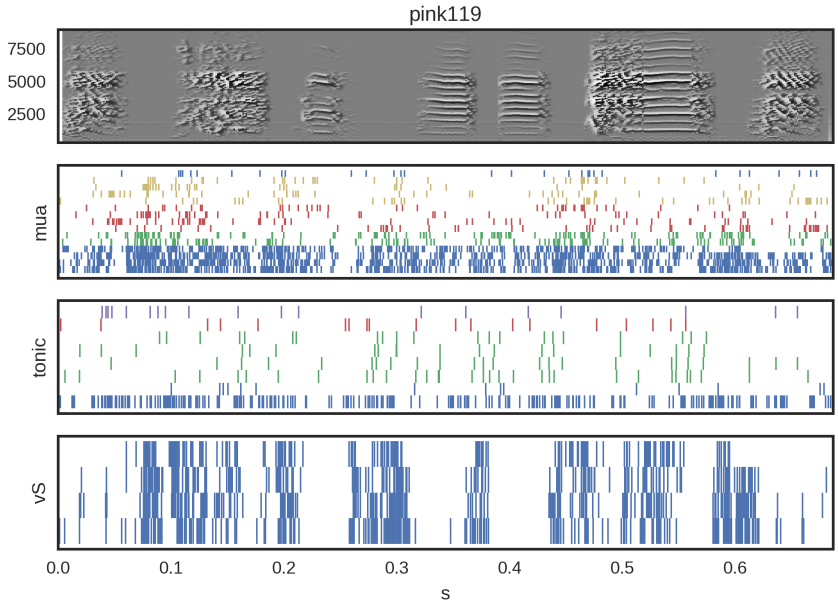

Figure 4.24. Warped motif rasters for Pink119. Firing of HVC neurons relative to the cannonical motif. Top: spectrogram of one motif. Middle: representative rasters for HVC units. Bottom: spikes detected from the vS EMG channel.

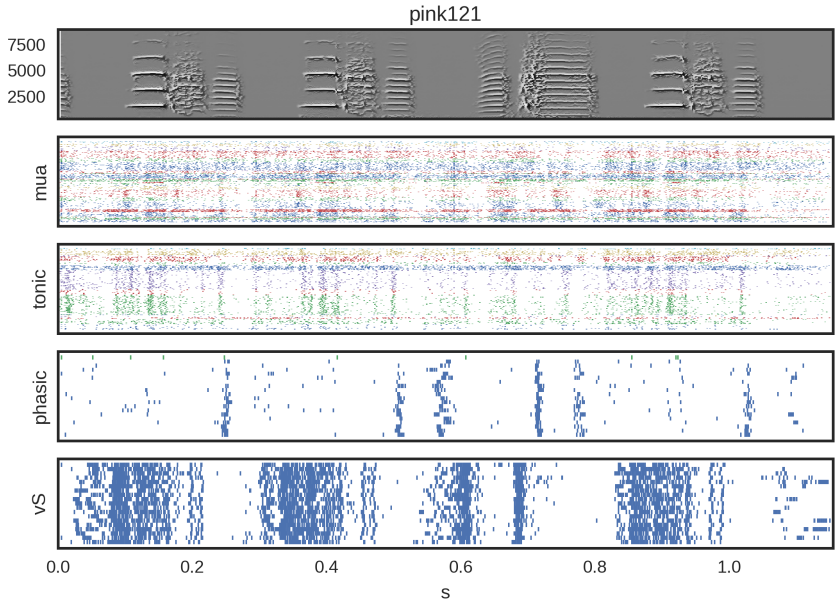

Figure 4.25. Warped motif rasters for Pink121. Firing of HVC neurons relative to the cannonical motif. Top: spectrogram of one motif. Middle: representative rasters for HVC units. Bottom: spikes detected from the vS EMG channel.

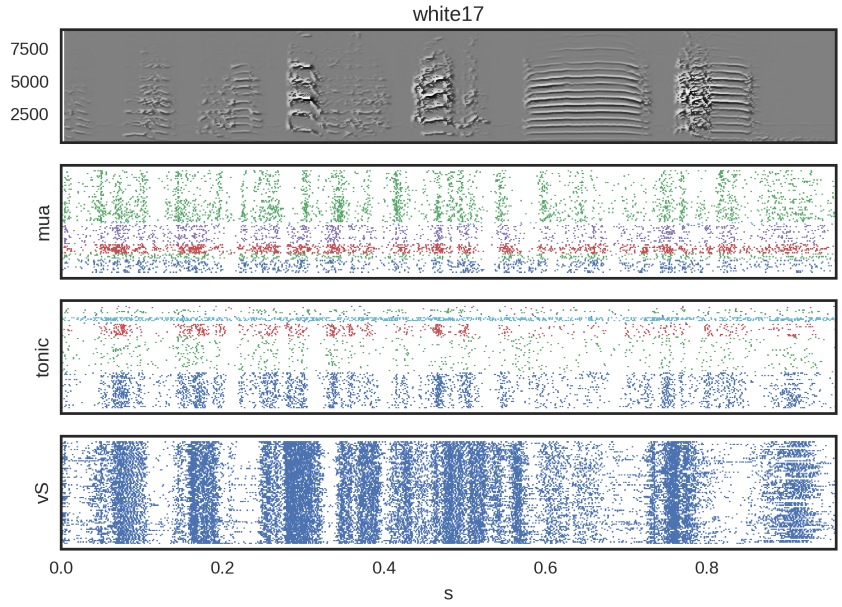

Figure 4.26. Warped motif rasters for White17. Firing of HVC neurons relative to the cannonical motif. Top: spectrogram of one motif. Middle: representative rasters for HVC units. Bottom: spikes detected from the vS EMG channel.

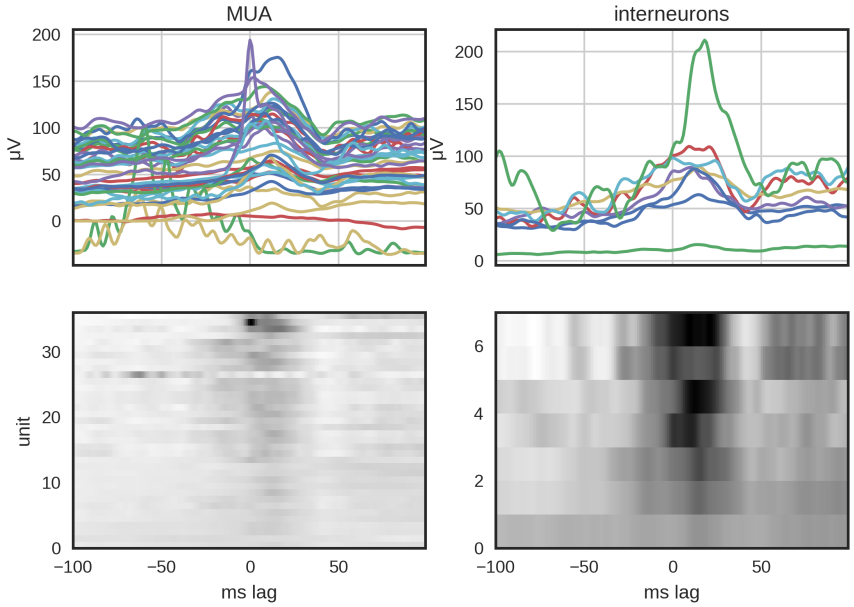

Figure 4.27. Black206 EMG STA. The spike triggered average of the EMG signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

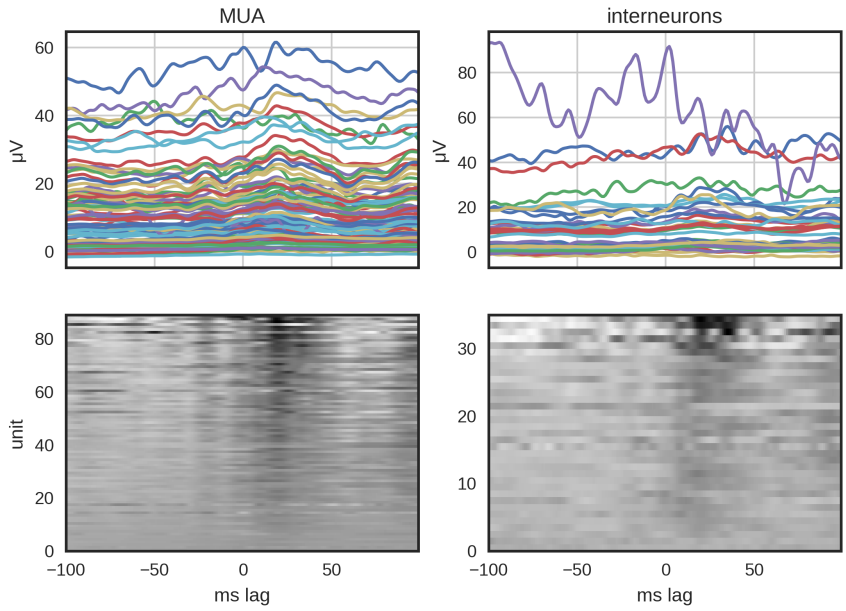

Figure 4.28. Black33 EMG STA. The spike triggered average of the EMG signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

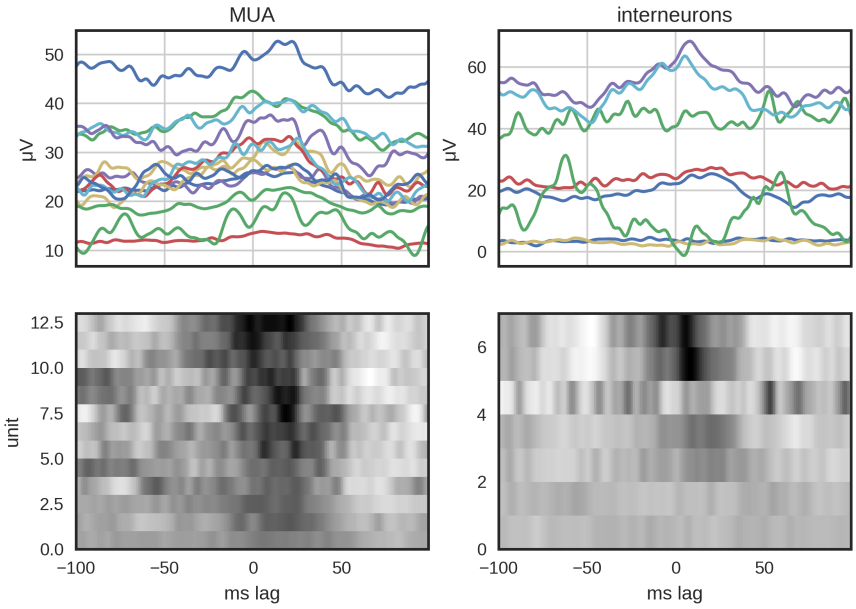

Figure 4.29. Black5 EMG STA. The spike triggered average of the EMG signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

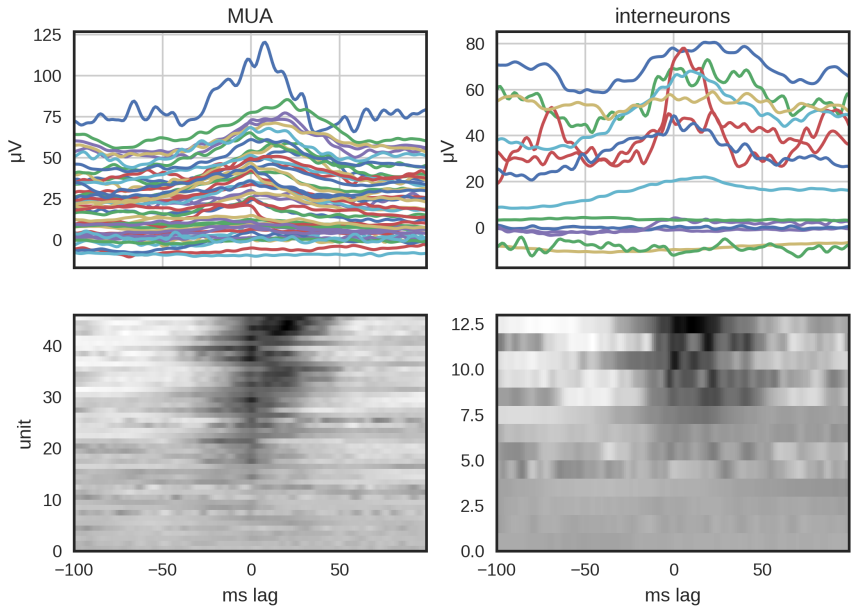

Figure 4.30. Orange184 EMG STA. The spike triggered average of the EMG signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

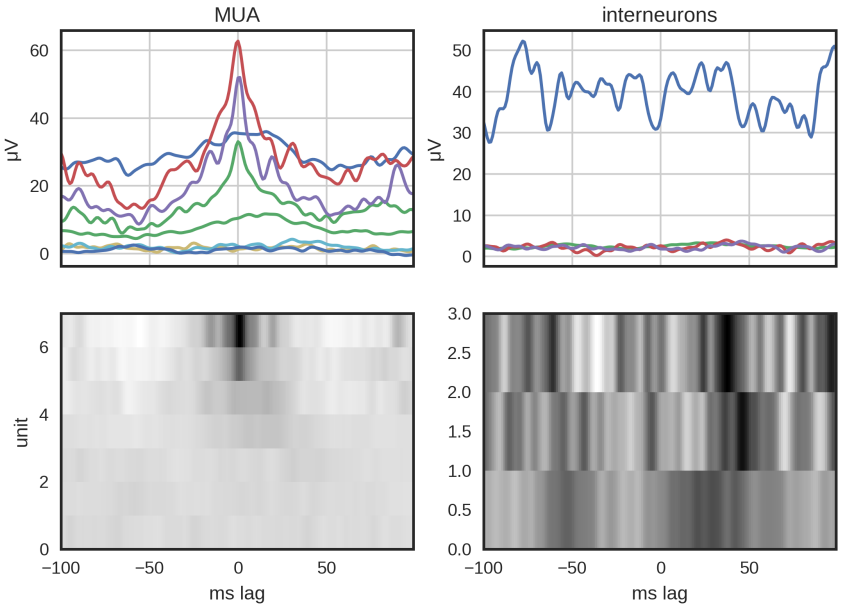

Figure 4.31. Pink119 EMG STA. The spike triggered average of the EMG signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

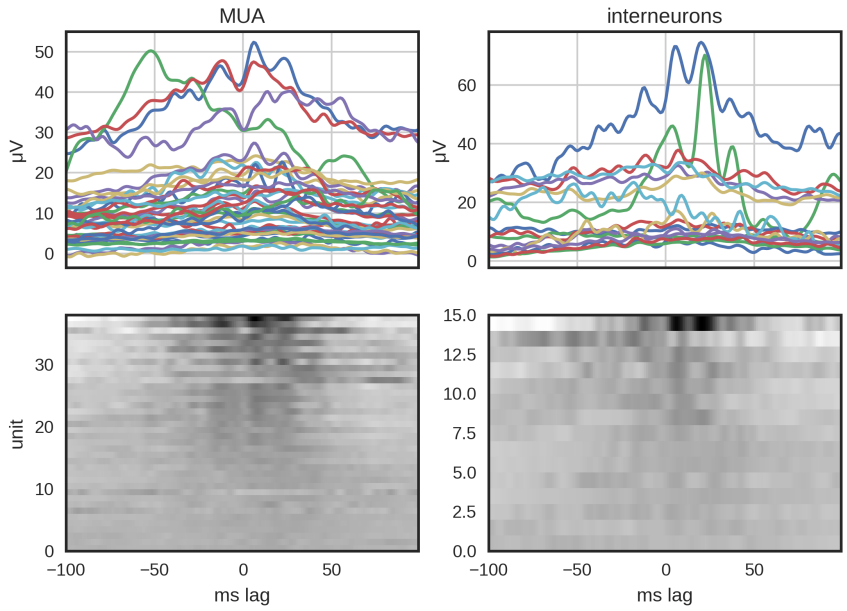

Figure 4.32. Pink121 EMG STA. The spike triggered average of the EMG signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

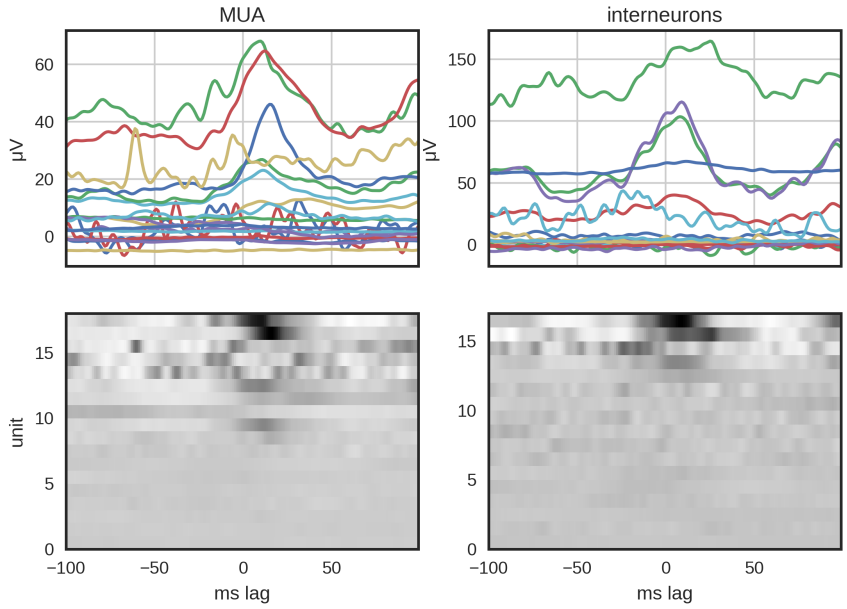

Figure 4.33. White17 EMG STA. The spike triggered average of the EMG signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

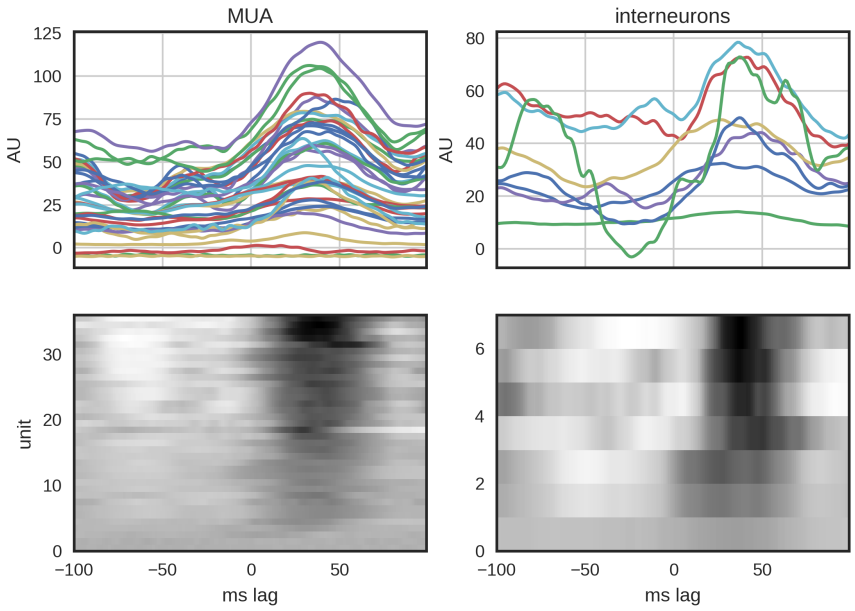

Figure 4.34. Black206 acoustic STA. The spike triggered average of the microphone amplitude signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

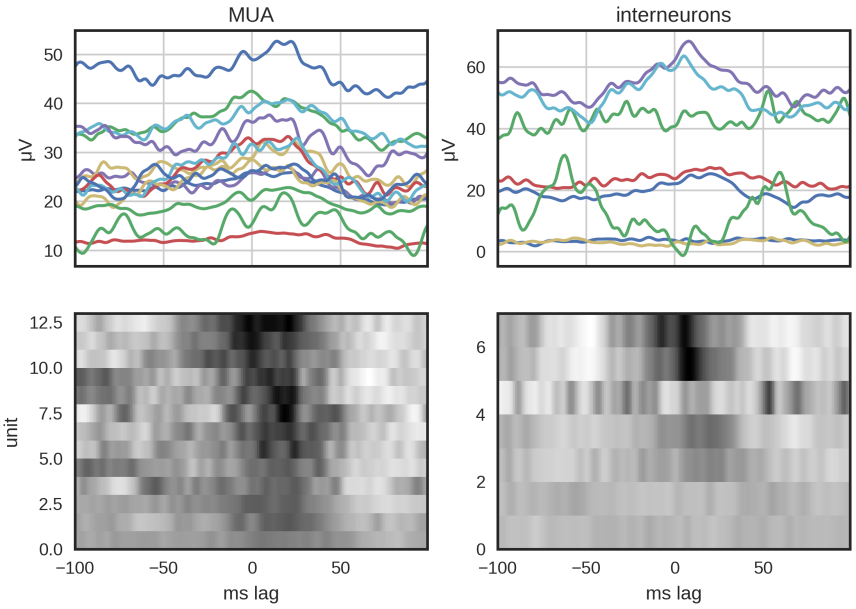

Figure 4.35. Black5 acoustic STA. The spike triggered average of the acoustic amplitude signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

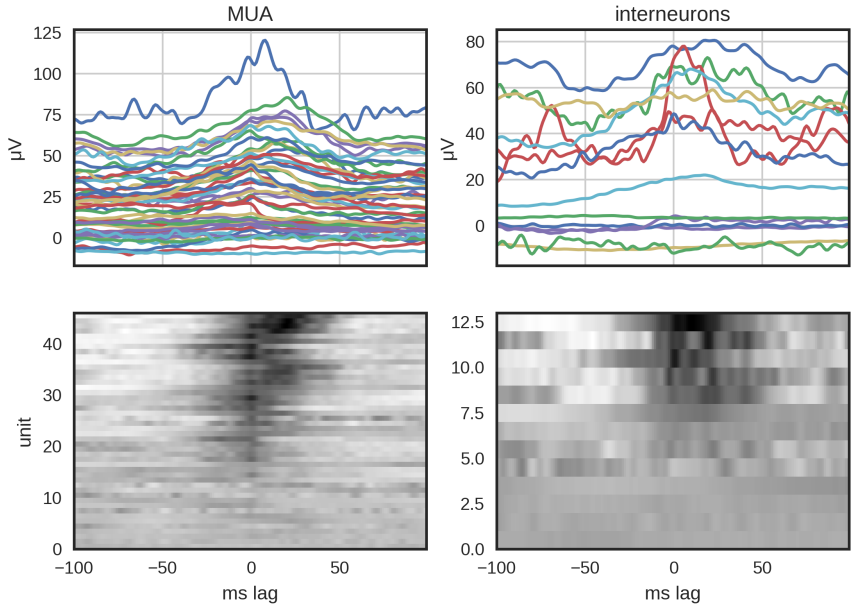

Figure 4.36. Orange184 acoustic STA. The spike triggered average of the acoustic amplitude signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

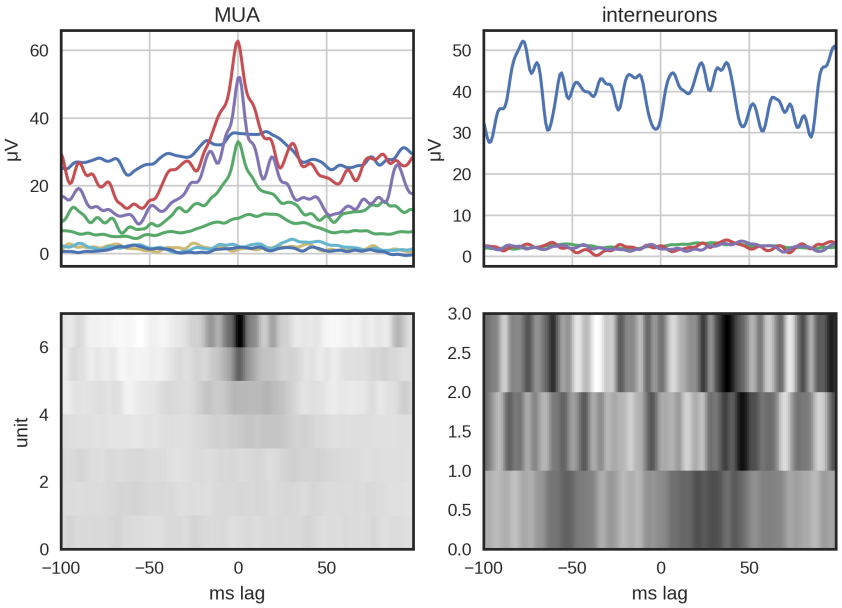

Figure 4.37. Pink119 acoustic STA. The spike triggered average of the acoustic amplitude signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

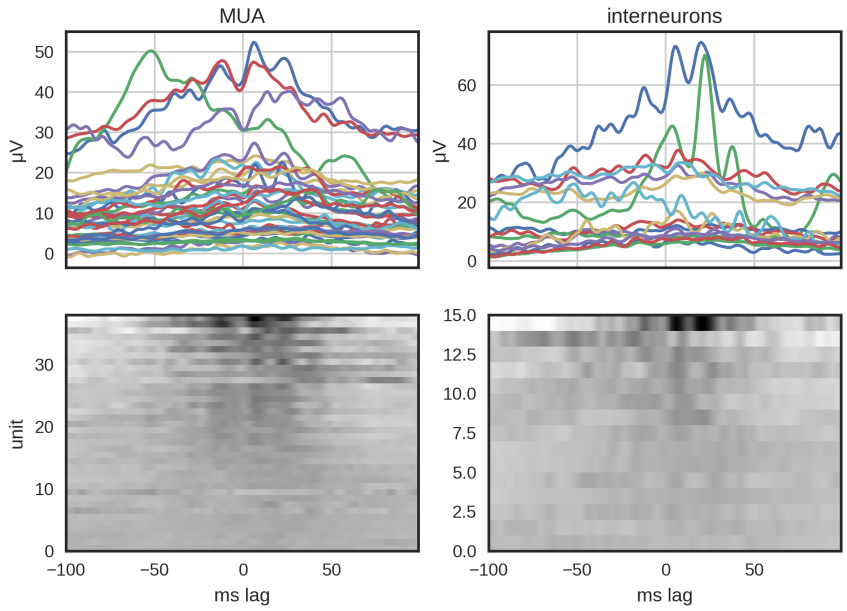

Figure 4.38. Pink121 acoustic STA. The spike triggered average of the acoustic amplitude signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

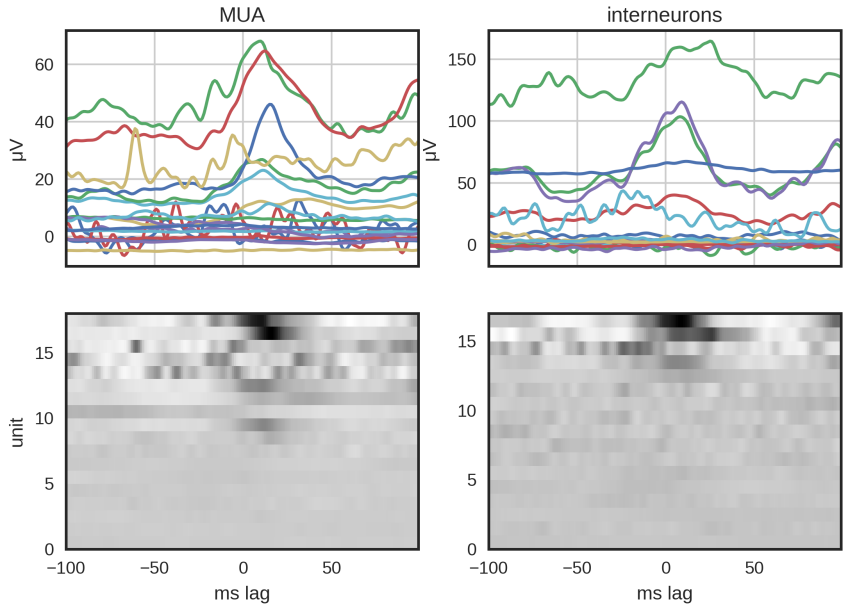

Figure 4.39. White17 acoustic STA. The spike triggered average of the acoustic amplitude signal from MUA (left) and putative interneurons (right). The data are shown as individual traces (top) and rows sorted by RMS value (bottom).

had a peak lag of 32.5 ms (28.2–36.5 ms 95% CI). For both interneurons and multiunits, the mean of the peak microphone amplitude lags were significantly different from 0 (T=6.85, $p<10^{-10}$ and T=15.2, $p<10^{-30}$ respectively). There are noticeable differences in peak lag time between multiunits and interneurons, Fig. 4.40.

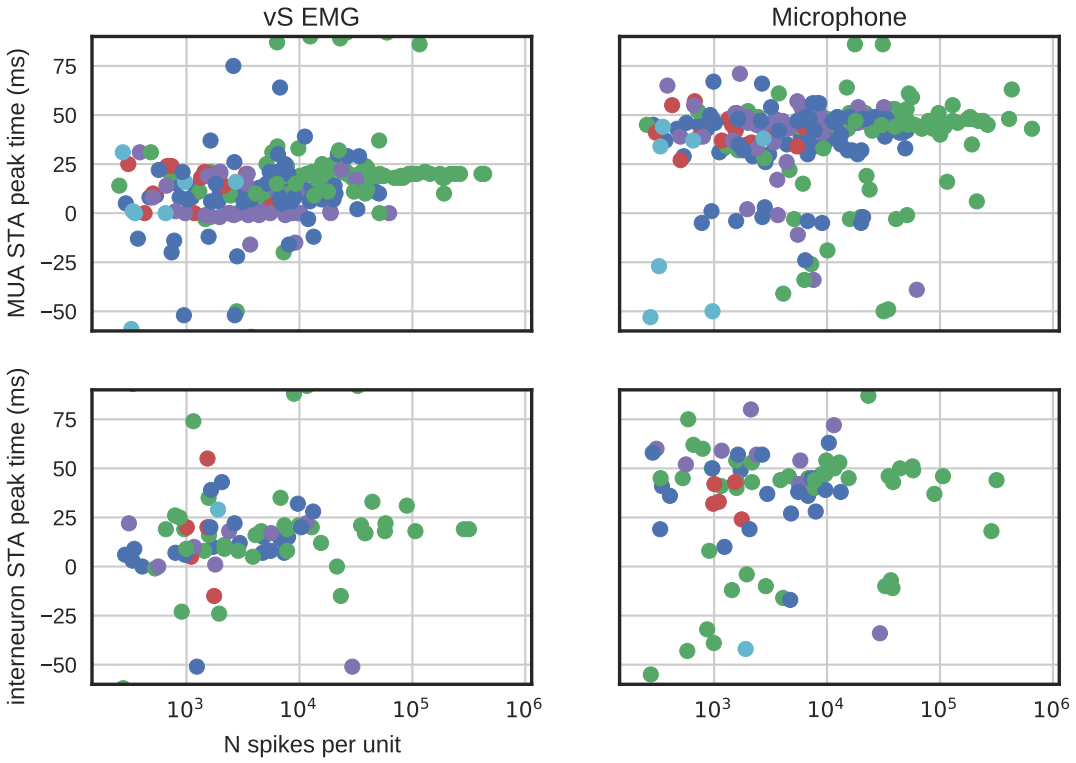

Figure 4.40. Distribution of peak STA lags. First column: STA from vS EMG amplitude, second column STA from microphone amplitude. First row: HVC multiunits, second row: HVC interneurons.

For each bird, the peak STA value for the vS EMG was correlated with the peak STA value of the microphone amplitude, both MUA and interneurons have been combined for Fig. 4.41, correlation values in Table 4.4.

The STA results suggest that HVC activity has a linear relationship with muscle activation during zebra finch vocalizations that include song.

To investigate the linear relationship between HVC activity and the vS, I developed a linear model to predict muscle activation. Time-warped firing rates were created for all

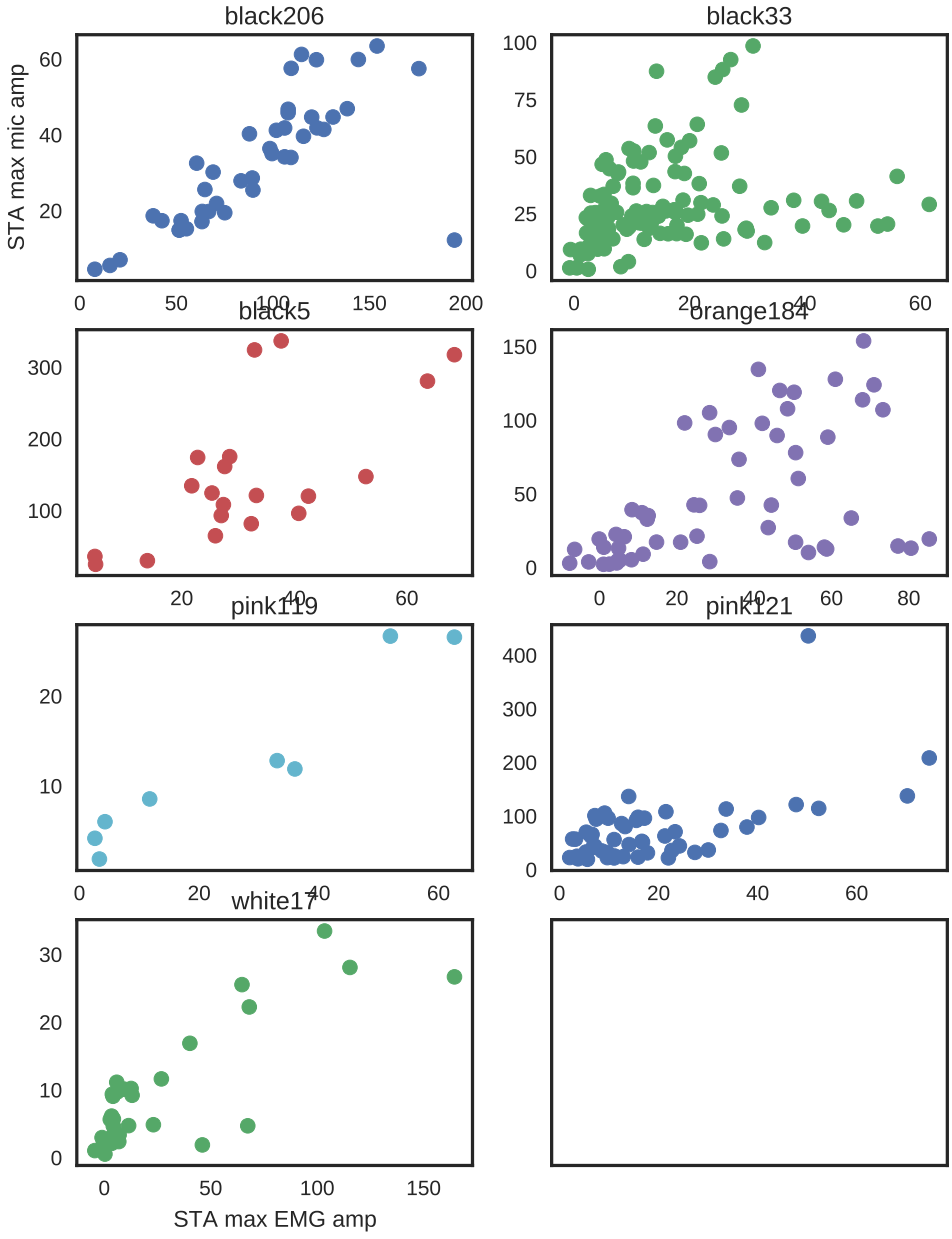

Figure 4.41. Maximum EMG STA peak over maximum acoustic amplitude STA peak.

Bird	N	r	p
Black206	35	0.724697	8.536543e-07
Black33	90	0.268931	1.037592e-02
Black5	14	0.272079	3.466872e-01
Orange184	44	0.477020	1.060825e-03
Pink119	7	0.951355	9.764389e-04
Pink121	39	0.595236	6.416449e-05
White17	17	0.821040	5.354239e-05

Table 4.4: Pearson correlation values between peak EMG STA and peak microphone amplitude STA, combining multiunits and interneurons.

Bird	U	p
Black206	105.0	3.635264e-16
Black33	2013.0	3.781124e-08
Black5	88.0	2.994340e-16
Orange184	4.0	1.139596e-16
Pink119	1205.0	3.384738e-11
Pink121	719.0	2.824561e-13
White17	1583.0	1.044051e-09

Table 4.5: Mann-Whitney U test between the MSE of the pre-motor lags and the MSE of the null lags. P values are Bonferroni corrected

responsive HVC interneurons and multi-units. PCA reduced HVC activity to three time varying parameters. A simple linear model with three coefficients was used to predict vS activity at timelags from -300 ms to 300 ms. Models with an absolute lag greater than 50 ms were considered Null models. Lags between 0 ms and 25 ms were considered plausible pre-motor delays, with 0 ms predicted by the GTE model. For all birds, the model performance exceeded 95% of the null performance distribution as measured by variance explained (VE) and mean squared error (MSE) during all or part of the plausible pre-motor lag range, Figs. 4.42 – 4.48. Model performance was significantly better in the pre-motor lag range than the null lag range, see Table 4.5 and Fig. 4.49.

The average performance peak within the premotor range was 12 ms. In Figs. 4.51 – 4.57 the PCA inputs, motor target, and model prediction for a 12 ms delay is shown.

One potential caveat is whether movement artifacts may have contaminated the electrode signals in the array recordings, which created a spurious correlation. Movement artifacts were

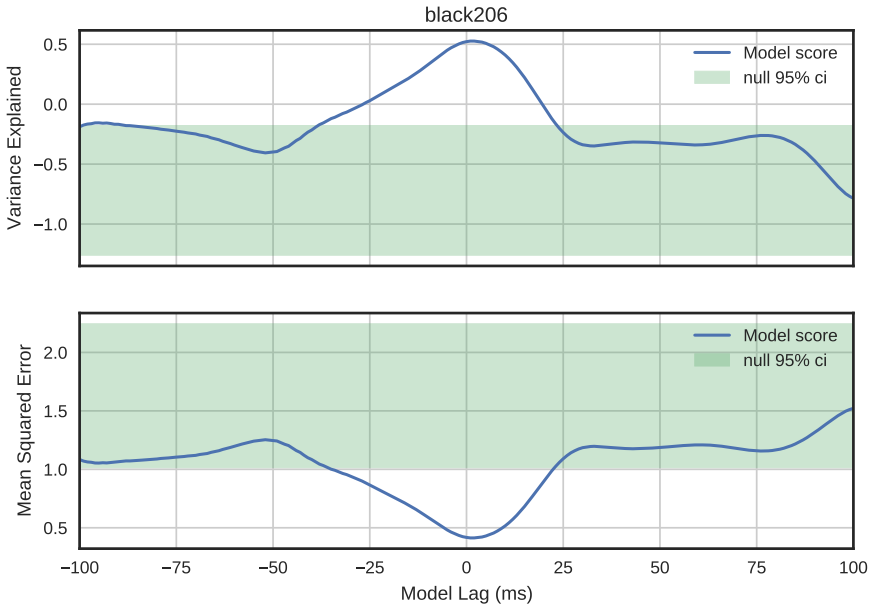

Figure 4.42. Linear EMG model performance for Black206. Top: normalized average firing rate for all responsive multi-units and interneurons. Middle: The three principal components with the largest eigenvalues. Bottom: The normalized EMG activity during song and the model prediction.

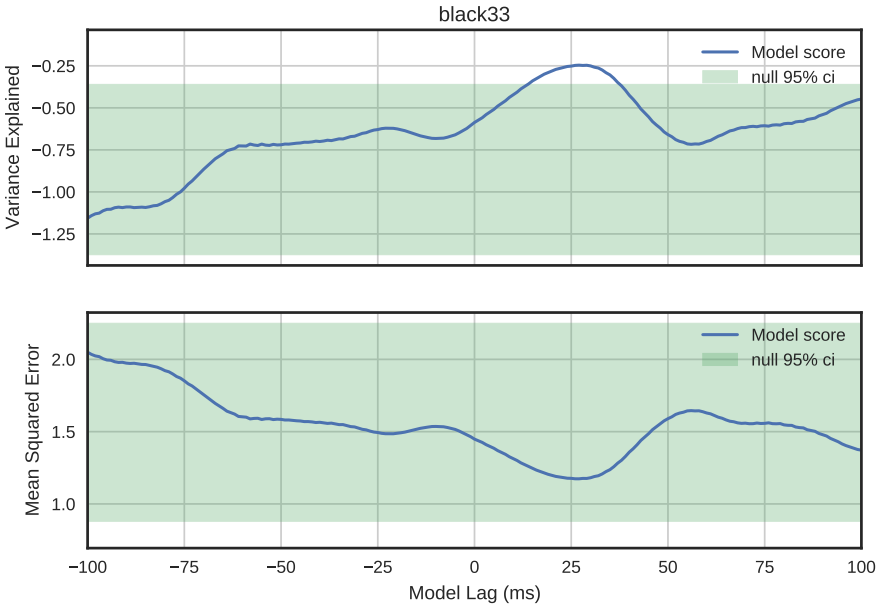

Figure 4.43. Linear EMG model performance for Black33. Top: model performance as measured by Variance Explained over varying model lags (k).

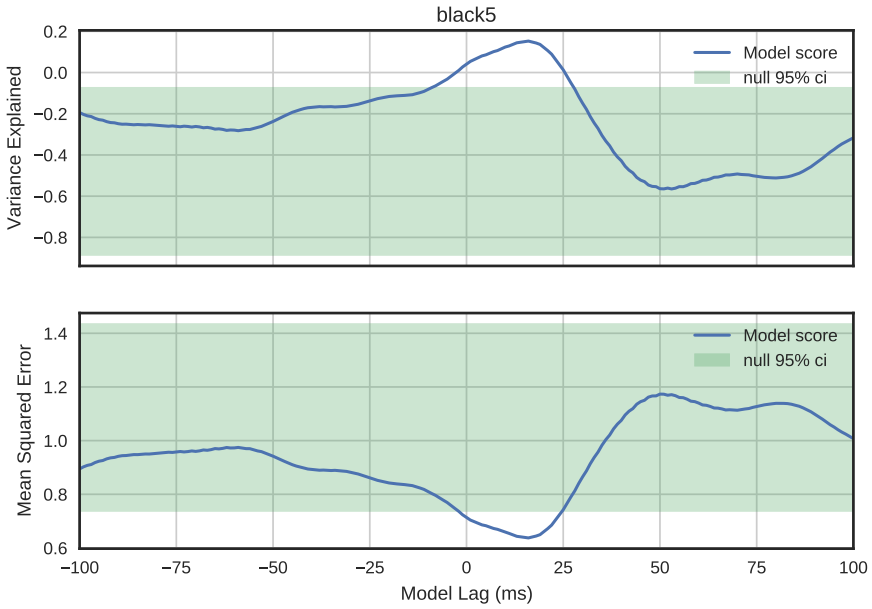

Figure 4.44. Linear EMG model performance for Black5. Top: model performance as measured by Variance Explained over varying model lags (k).

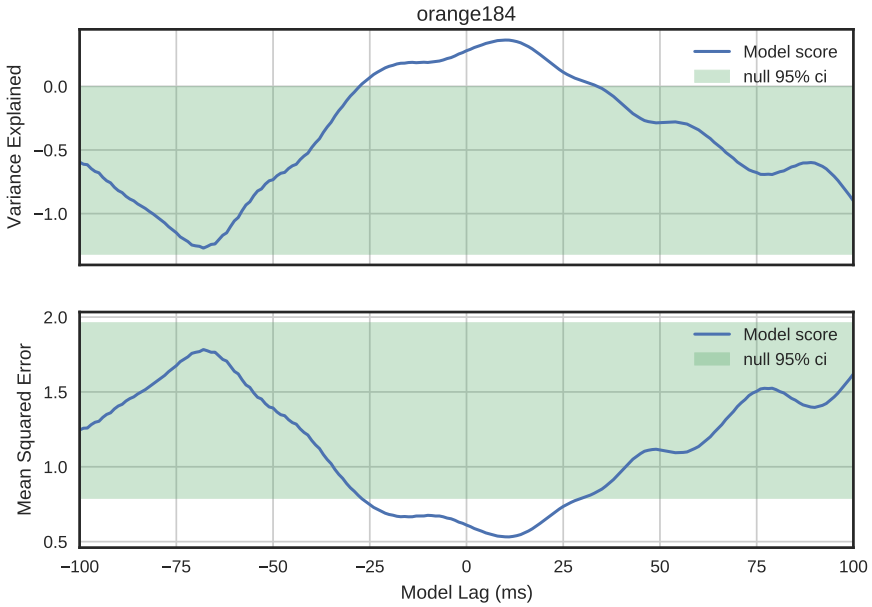

Figure 4.45. Linear EMG model performance for Orange184. Top: model performance as measured by Variance Explained over varying model lags (k).

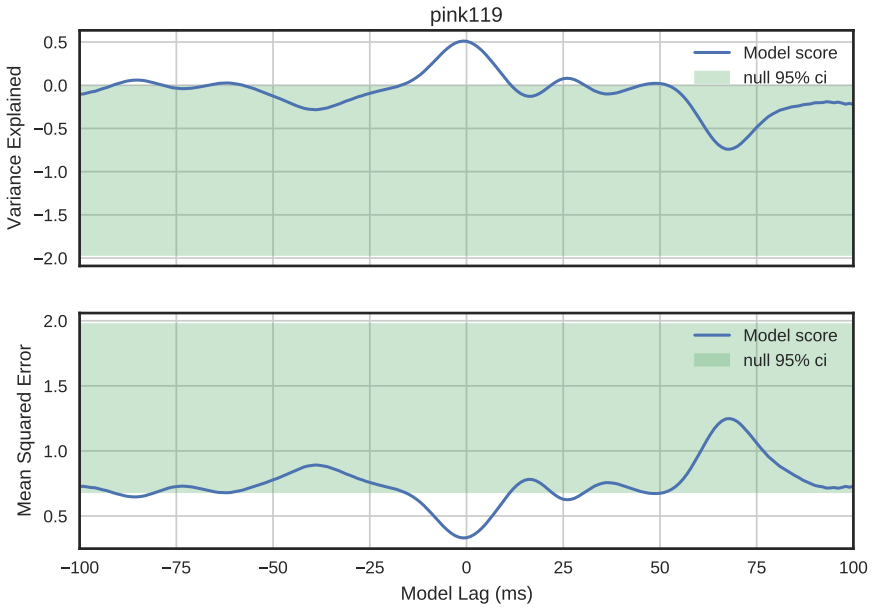

Figure 4.46. Linear EMG model performance for Pink119. Top: model performance as measured by Variance Explained over varying model lags (k).

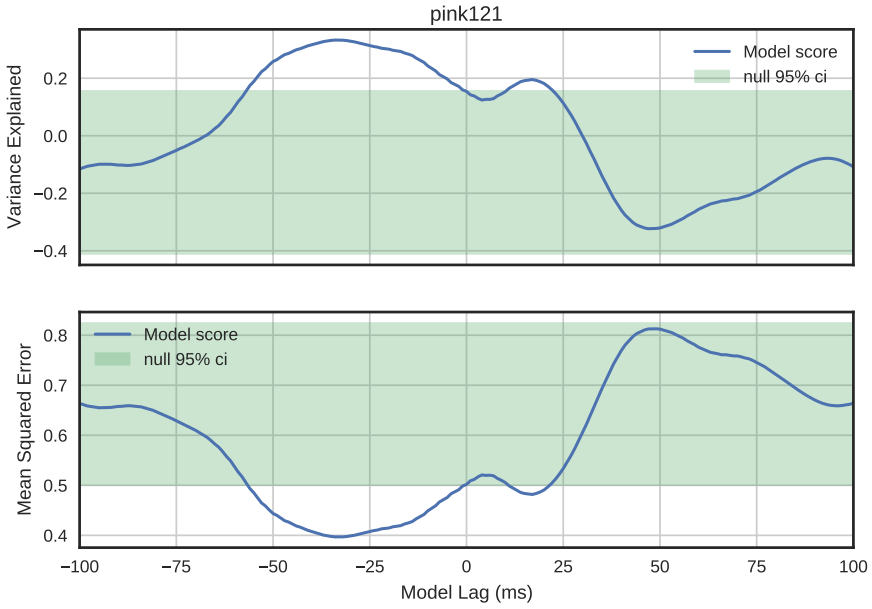

Figure 4.47. Linear EMG model performance for Pink121. Top: model performance as measured by Variance Explained over varying model lags (k).

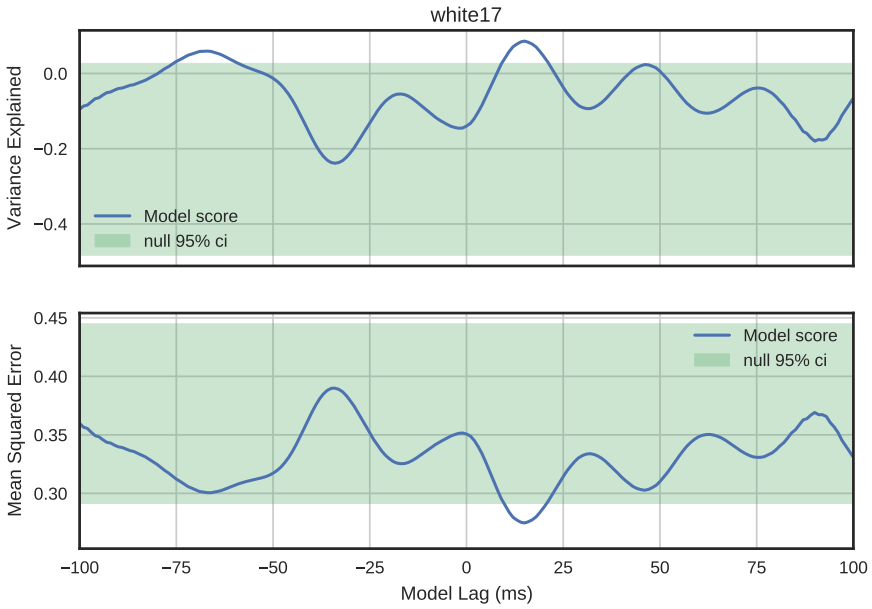

Figure 4.48. Linear EMG model performance for White17. Top: model performance as measured by Variance Explained over varying model lags (k).

treated particularly carefully. First, digital referencing eliminated most of the common signal in the electrode channels. Second, during spike-sorting with Plexon OfflineSorter, movement artifact can be identified by their symmetrical waveforms. Third, for the linear model, we used a muscle activity estimate based on spikes as detected from Plexon OfflineSorter, which should have further eliminated movement artifacts from being included in estimate of the EMG activity. Finally, the conclusions are the same when the analysis is restricted to well isolated units only, Fig. 4.50.

4.4 Discussion

In this chapter I demonstrated the feasibility of simultaneous recordings of syrinx EMG and single neurons in HVC. These simultaneously recorded areas appear to show a linear relationship, which is consistent with a kinetic code in HVC. Previous findings have anticipated this result – HVC activation prior to vocalization was reported in (McCasland and Konishi, 1981;

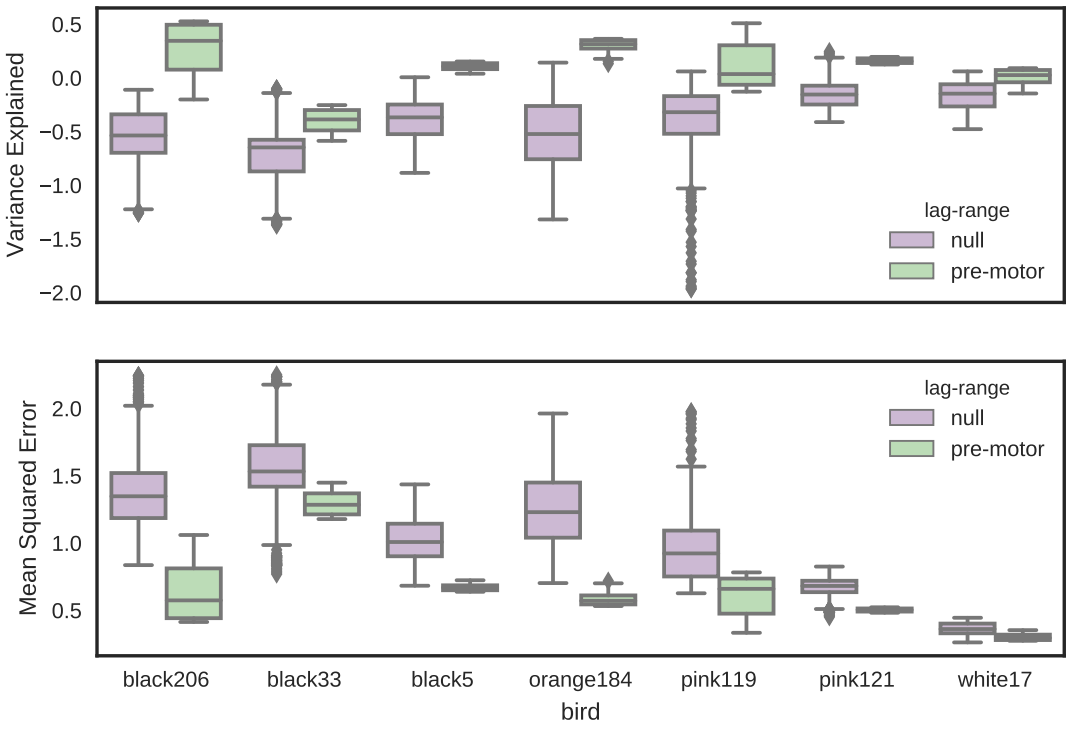

Figure 4.49. Performance of linear model across birds. For each bird, the distribution of performances with the premotor range 0–25 ms was compared against lags ranging from -300 ms to -50 and 50 ms to 300 ms.

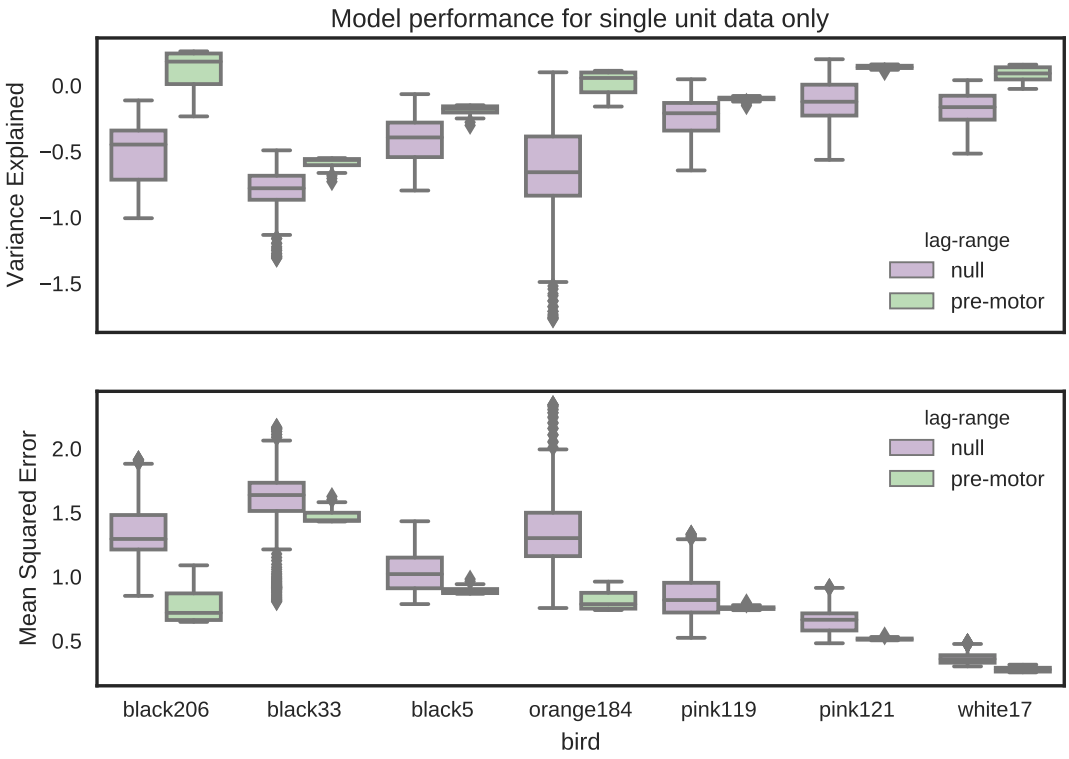

Figure 4.50. Performance of linear model across birds, restricted to interneurons. For each bird, the distribution of performances with the premotor range 0–25 ms was compared against lags ranging from -300 ms to -50 and 50 ms to 300 ms. Multi-units were excluded from the model inputs.

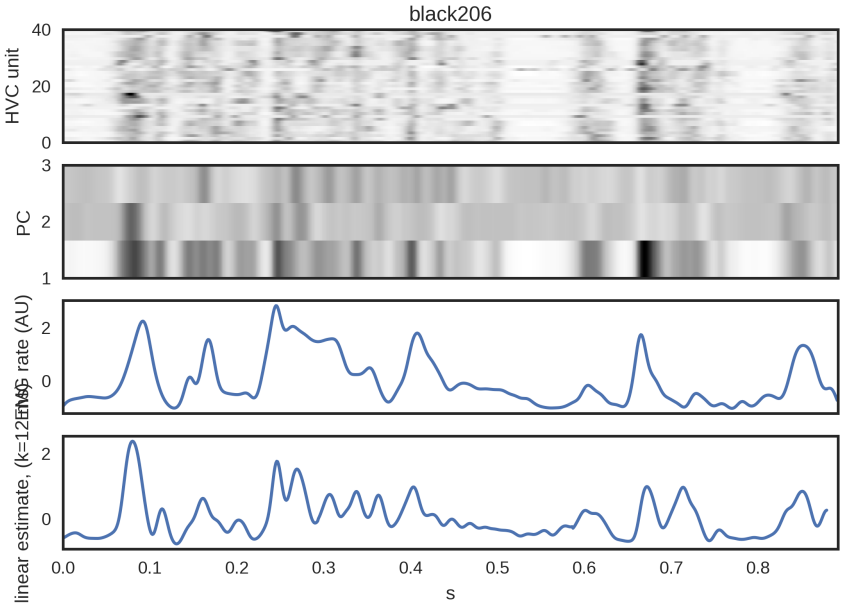

Figure 4.51. Linear EMG model for Black206. Top: normalized average firing rate for all responsive multi-units and interneurons. Middle: the three principal components with the largest eigenvalues. Bottom: the normalized EMG activity during song and the model prediction.

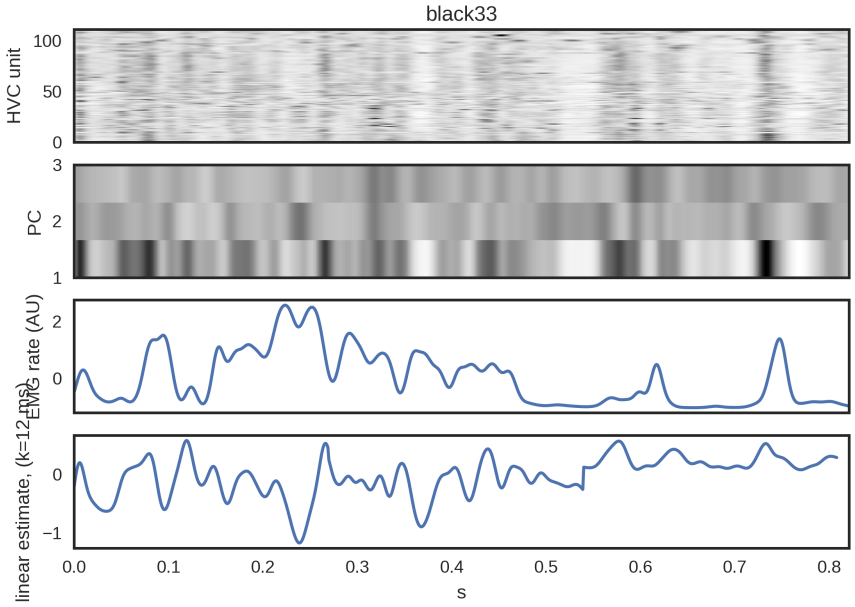

Figure 4.52. Linear EMG model for Black33. Top: normalized average firing rate for all responsive multi-units and interneurons. Middle: the three principal components with the largest eigenvalues. Bottom: the normalized EMG activity during song and the model prediction.

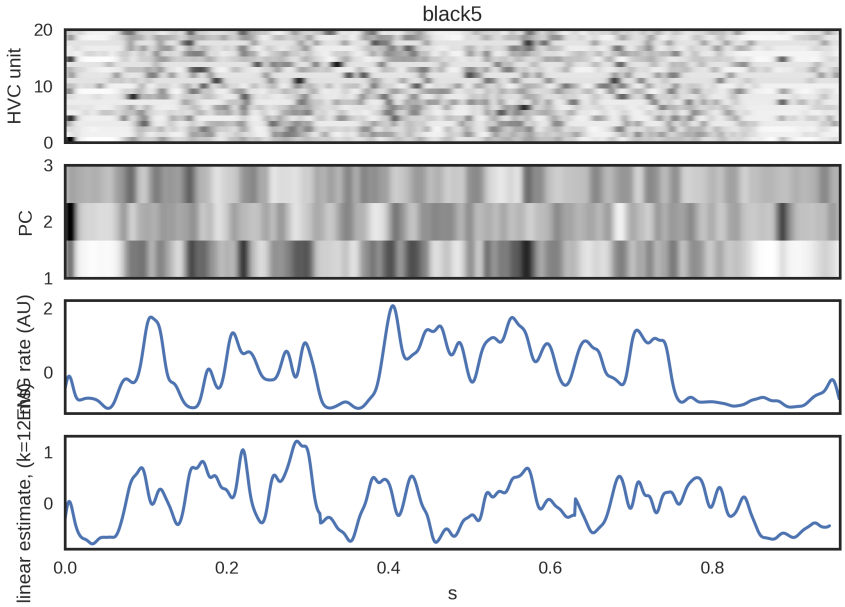

Figure 4.53. Linear EMG model for Black5. Top: normalized average firing rate for all responsive multi-units and interneurons. Middle: the three principal components with the largest eigenvalues. Bottom: the normalized EMG activity during song and the model prediction.

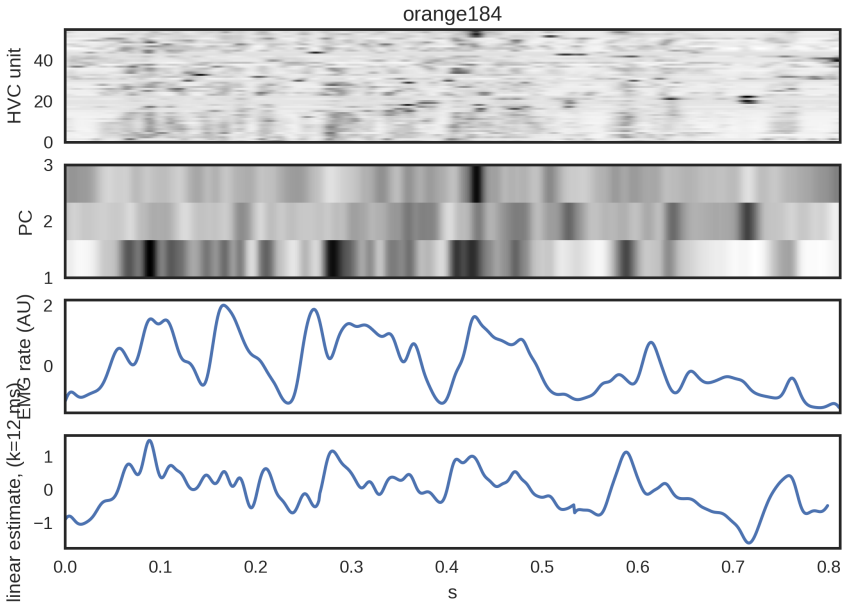

Figure 4.54. Linear EMG model for Orange184. Top: normalized average firing rate for all responsive multi-units and interneurons. Middle: the three principal components with the largest eigenvalues. Bottom: the normalized EMG activity during song and the model prediction.

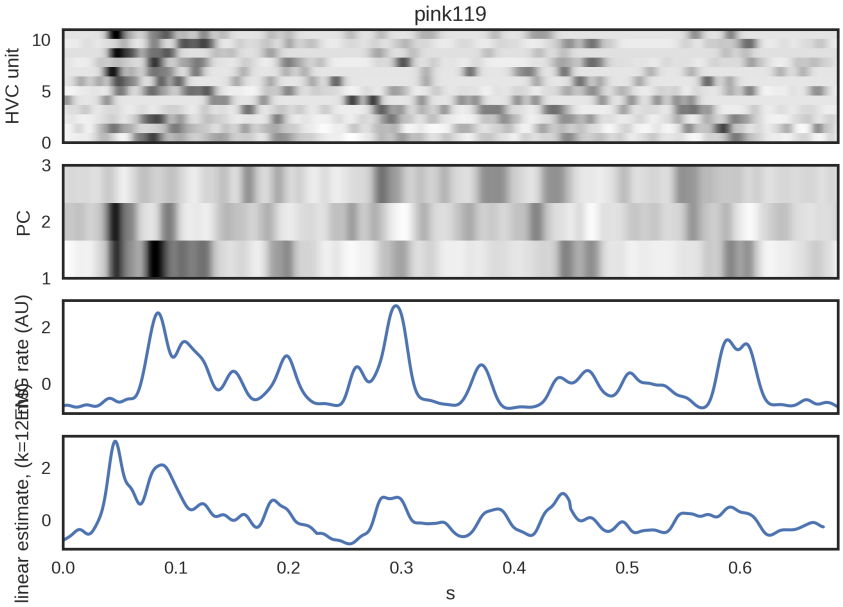

Figure 4.55. Linear EMG model for Pink119. Top: normalized average firing rate for all responsive multi-units and interneurons. Middle: the three principal components with the largest eigenvalues. Bottom: the normalized EMG activity during song and the model prediction.

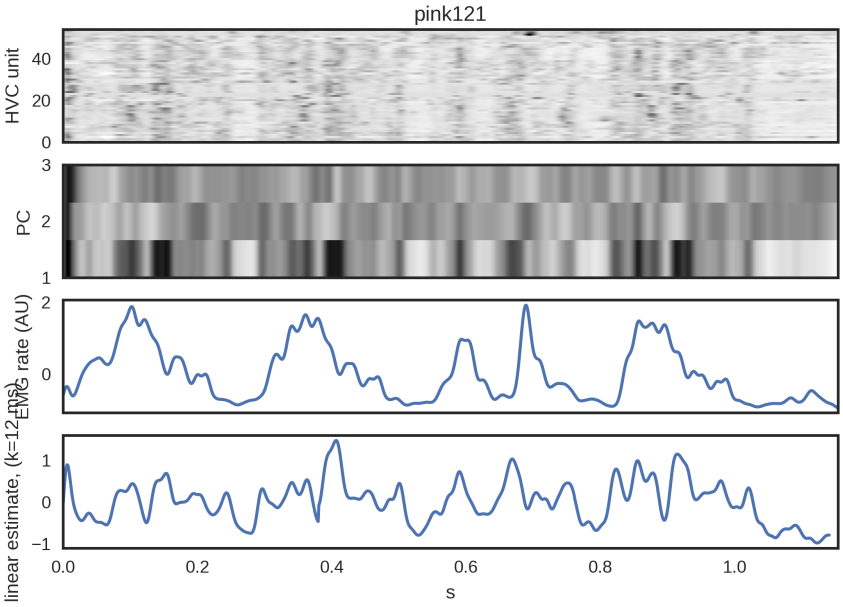

Figure 4.56. Linear EMG model for Pink121. Top: normalized average firing rate for all responsive multi-units and interneurons. Middle: the three principal components with the largest eigenvalues. Bottom: the normalized EMG activity during song and the model prediction.

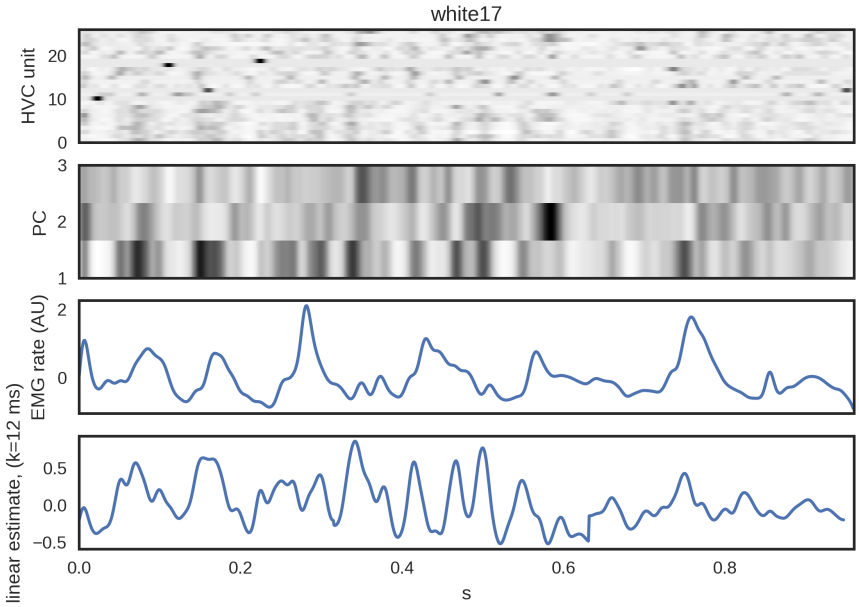

Figure 4.57. Linear EMG model for White17. Top: normalized average firing rate for all responsive multi-units and interneurons. Middle: the three principal components with the largest eigenvalues. Bottom: the normalized EMG activity during song and the model prediction.

Yu and Margoliash, 1996) and stimulation of HVC results in contraction of the vocal muscles (for a brief discussion see Katz and Gurney, 1981) and the respiratory muscles (Ashmore et al., 2005; Méndez et al., 2012). During song, a simple linear model with a premotor delay accurately predicted muscle activation, which is in line with the results from the previous chapters, where the data were collected during replay. In addition, there is some evidence for zero-lag encoding of the motor periphery.

The data in this chapter suggests that HVC is not exclusively a "clock", though we cannot rule out the existence of a clock in the HVC projection neurons. The existence of motor codes in HVC calls into question the utility of a motor-agnositic clock in HVC. If a motor code exists in the interneurons, it is reasonable to suspect a similar code for the projection neurons given the micro-circuitry of HVC (Mooney and Prather, 2005; Kosche et al., 2015; Vallentin et al., 2016).

The GTE model does not make predictions of transformations along motor pathway

– it makes predictions of onset, offsets and maxima of gestures, with zero time lag. The interneurons identified in this study with a causal pre-motor delay may cause a subset of GTE encoding projection neurons to fire at a near-zero delay due to release from inhibition (Kosche et al., 2015).

There is a striking similarity between the first principal component of HVC activity and vS activity. This kinetic code in HVC may be a single dimension of a low-dimensional HVC manifold that may also include respiratory dynamics, which mirrors contemporary ideas of control in motor cortex (Gallego et al., 2017). An avenue for future research would be to investigate the existence of a respiratory dimension in HVC. This hypothesis has many parallels to the GTE hypothesis, which maps activity in HVC to the low-dimensional Mindlin model. If a low-dimensional manifold exists in HVC, the internal dynamics within this manifold may be the sequencing mechanism of HVC, which is an alternative to the clock model consistent with the data presented in this thesis.

CHAPTER 5

GENERAL DISCUSSION

5.1 Implications for the activity of HVC projection neurons

The primary focus of this thesis is HVC interneurons and their relation to the syrinx muscles. I found evidence for an interneuron population that co-activates with the muscles, at a plausible pre-motor delay of 10-20 ms. HVC interneurons represent only 10% of the population of cells within HVC, and do not directly project outside HVC (Scotto-Lomassese et al., 2007). Therefore, HVC interneurons cannot be directly driving motor outputs of the vocal system unless they influence the activity of the projection neuron population. Some models of HVC do not require any patterned interneuron activity for the sequencing of song (Li and Greenside, 2006; Jin et al., 2007; Long et al., 2010). If these models are correct, the coactivation of interneurons and the muscles does not inform our understanding of songbird motor control.

HVC interneurons form synaptic connections with both projection neuron types, and receive inputs from both as well. These bi-directional connections allow the two projection neuron classes to influence each other via disynaptic inhibition (Mooney and Prather, 2005; Kosche et al., 2015), Fig. 5.1.

Some computational models of HVC suggest interneurons are important for song sequencing by shaping HVC_{RA} activity (Gibb et al., 2009; Yildiz and Kiebel, 2011). This hypothesis was validated by the infusion of gabazine, a GABA receptor antagonist into HVC, which eliminated normal singing behavior (Kosche et al., 2015). HVC_{RA} bursts cause disynaptic inhibition to both HVC_{RA} (Kosche et al., 2015) and HVC_{X} (Mooney and Prather, 2005) populations. Bursts from local projection neurons require temporary gaps in activity from the local inhibitory pool (Kosche et al., 2015), and the time varying activity of HVC interneurons is likely driven by excitatory input from projection neurons. However, motor feedback

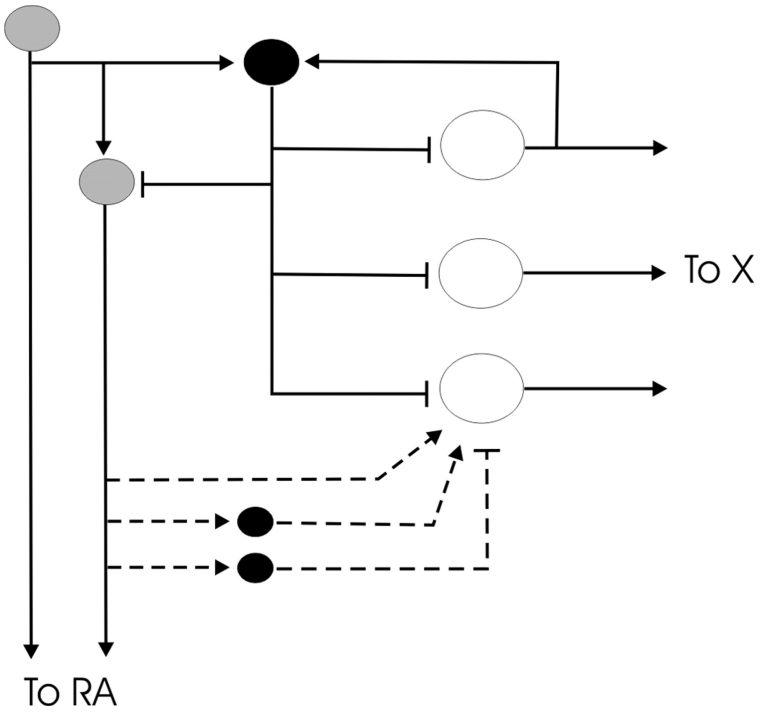

Figure 5.1. The HVC microcircuit. HVC interneurons (black) from inhibitory synaptic connections onto HVC_{RA} (gray) and HVC_{X} (white) projection neurons. Both projection neuron classes form excitatory synaptic connections onto HVC interneurons. Figure from (Mooney and Prather, 2005).

from Uva and auditory feedback from NIf and field L may also contribute to interneuron activity.

If projection neurons drive the interneuron populations from which I recorded, there may be a projection neuron population that also encodes the muscle state with a premotor delay. A second population of projection neurons may fire once the resulting inhibitory burst has subsided. Based on population activity recorded in Chapter 2, these moments of local minima are on average 20 ms after peak activity, placing the bursts of the second population of projection neurons at zero time delay relative to the motor periphery. It may be the activity of this second population of projection neurons that were reported in (Amador et al., 2013). Broadly, the evidence for motor coding presented in this thesis are more consistent with the GTE model than the clock model, though a detailed model incorporating both the present results with the GTE model will require additional data, particularly the simultaneous recording of the EMG and a population of projection neurons,

possibly by using the optical technique developed in (Picardo et al., 2016).

5.2 Implications for song learning

As syllables are learned, interneuron patterning stabilizes. These stabilized patterns are dependent on the acquisition of individual syllables (Vallentin et al., 2016). This observation is consistent with the acquisition of an internal model in HVC of the motor periphery.

HVC both receives auditory inputs and projects to the anterior forebrain pathway, making it a likely candidate for evaluating sensory feedback and propagating an error signal to the AFP. I speculate that HVC_X neurons may be the site of this error signal.

As mentioned earlier, the microcircuitry of HVC suggests that a subpopulation of projection neurons is also premotor. This subpopulation of premotor projections neurons will cause disynaptic inhibition in HVC_X cells. HVC_X cells show rebound firing after inhibition (Daou et al., 2013). Auditory projections to HVC synapse on these HVC_X neurons (Prather et al., 2008). Thus one hypothesis for motor learning in HVC is coincidence detection in HVC_X cells from a rebound premotor disynaptic inhibition and auditory input.

5.3 Motor coding in HVC

In section 1.5 of this thesis, I outlined the major theories of cortical motor control. Based on the evidence presented in Chapter 3 and Chapter 4, there may be a representation of muscle activation in HVC (a kinetic code). However, there are many confounding variables. First, as demonstrated in Chapter 3 and described elsewhere (Goller and Suthers, 1996a; Vicario, 1991), the muscles of the syrinx have highly correlated activity. The limitations of these experiments prevent us from distinguishing whether neurons in HVC encode specific, individual muscles or a synergistic muscle state, or another parameter correlated with muscle activation, such as labial tension (an internal kinematic code). Though the specific motor

output cannot be determined, the results indicate some motor representation in HVC.

In both Chapter 3 and Chapter 4, simple population models provided predictions of vS activation that are significantly better than random, suggesting there may be latent variables in HVC that represent the motor periphery. The Mindlin model provides a powerful dimensionality reduction for the sound source: pressure and tension. The sum activity (Chapter 3) and first principal component (Chapter 4) predict vS muscle activation, which is related to labial tension. I speculate that air sac pressure is another latent variable represented in HVC. If true, HVC may not only encode GTE, but also represent the gesture trajectories. Though more experiments are needed to verify this hypothesis, I speculate that HVC population activity is best described by the dynamic model of motor control, where external gesture trajectories are mirrored by an internal model, as described in (Gallego et al., 2017).

5.4 The choice of model system

Zebra finches are the most widely studied model of sing and has many advantages as a model system. They adapt well to laboratory conditions, and singing readily in the presence of humans. The stereotyped song makes it easy to assess variability during song learning. However, the zebra finch model also has significant limitations.

First, complex, noisy syllables common in zebra finch song are difficult to analyze with common spectral analysis tools. During reinforcement learning experiments, the modulated syllable tends to be flat notes (harmonic stacks), due to having a well defined fundamental frequency (Ali et al., 2013). However flat notes have atypical syrinx muscle activity (Vicario, 1991). (The findings in this thesis suggest that previous studies of learning in zebra finches may be biased towards syllable types that may not be representative of HVC activity for all syllable types.) The complex, non-harmonic sounds of zebra finch song greatly increase the complexity of modeling the periphery. While synthetic song estimates can be approximated

by a single sound source, the fitting process may be greatly simplified by choosing a model organism without the spectral complexity of the zebra finch syllable repertoire.

Second, the lack of behavioral variability arising from the stereotyped motif of the zebra finch poses unique challenges for testing models of motor control. A dynamical system is by definition time dependent. Any moment in time can be mapped to a single state in the dynamical system, and it may not be meaningful to try and distinguish between the two.

Brown thrashers (*Toxostoma rufum*) seem like an appealing model. They are a common bird throughout eastern North America, and may have the largest repertoire of any North American bird, with at least 1,100 song variants (Boughey and Thompson, 1981; Stap, 2006). All measured syrinx muscle EMGs (vS, vTB and dTB) are correlated with fundamental frequency (Goller and Suthers, 1996b,a), Fig. 5.2. Another advantage is the brown thrasher's song sequence. Each syllables is repeated twice. This provides an excellent test of the clock mode, which would predict that a projection neuron should fire during one of the two repeated syllables but not both. If individual projection neurons burst at the same moment in both syllables, either the clock is limited to single syllables or the burst has some relation to the motor periphery. Given the extremely broad repertoire of the brown thrasher, $\mathrm{HVC_{RA}}$ neurons may burst in more than a single syllable. If $\mathrm{HVC_{RA}}$ neurons burst during similar spectral features or similar times from the syllable onset, they could provide evidence for the GTE or clock models respectively.

Another promising model is the domestic canary (*Serinus canaria*). Like the brown thrasher, canary song consists of repeated syllables with simple tonal notes. The biomechanics of canary song have been extensively studied (Gardner et al., 2001; Mindlin, 2017; Suthers et al., 2004; Hartley and Suthers, 1990). There is also a preexisting literature of canary neurophysiology (Halle et al., 2003; Nottebohm et al., 1976; Goldin and Mindlin, 2017), a distinct advantage over the brown thrasher.

The simpler tonal notes of the brown thrasher and the canary aid in acoustic analysis

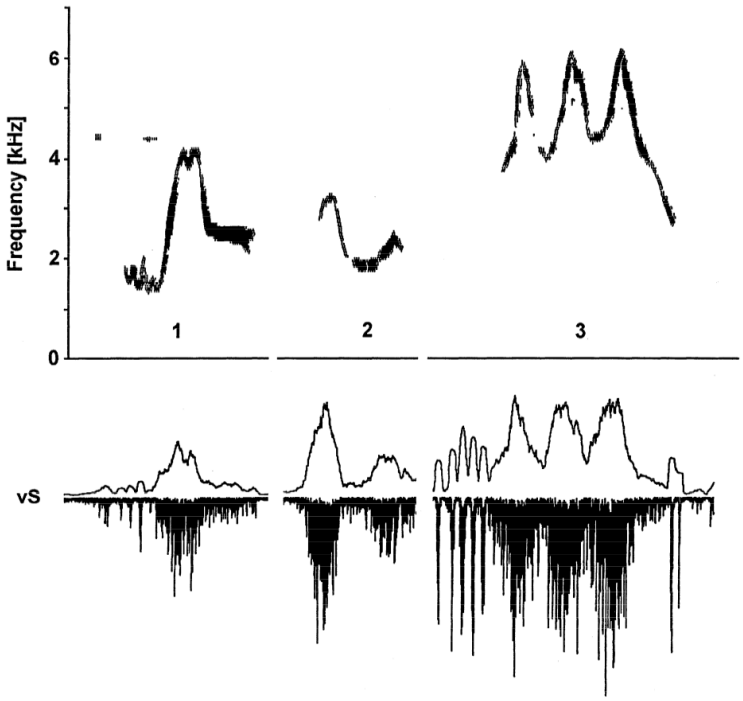

Figure 5.2. EMG activity predicts pitch in brown thrashers. Three sound segments and the associated vS activation. Pitch and vS activity are better correlated in brown thrashers than zebra finches, see Fig. 1.6. From (Goller and Suthers, 1996a).

and modeling, and repeating syllables, which helps to decouple motor state from a specific moment in time. I speculate that these models, or models with these advantages, will provide better insight into song motor control than the zebra finch.

5.5 Future directions

A major question is whether air sac pressure is encoded in HVC, along with syrinx muscle activation. A sensible next experiment would be recording air sac pressure with simultaneous array recordings in HVC.

It's possible that common syllable types have a stereotyped muscle activation across birds, allowing an estimate of muscle state without measuring the muscles directly. Just as inverting the Mindlin model converts song into pressure and tension trajectories, it would be useful to have a model that could predict muscle activation from recorded song, particularly

during the gaps in phonation, where the Mindlin model cannot make predictions (beyond placing the pressure and tension values in the non-phonating region).

A model for estimating muscle activity by syllable would also benefit from a complete description of zebra finch notes (the spectral elements from which syllables are created). To date, no systematic description of all zebra finch notes has been published, likely due to the challenge of objectively identifying note boundaries within syllables. A complete description of zebra finch notes and their underlying muscle activity could provide a powerful tool for interpreting neural activity in the song system.

REFERENCES

Adret, P., C. D. Meliza, and D. Margoliash (2012, October). Song tutoring in presinging zebra finch juveniles biases a small population of higher-order song-selective neurons toward the tutor song. *J. Neurophysiol. 108*(7), 1977–1987.

Ali, F., T. M. Otchy, C. Pehlevan, A. L. Fantana, Y. Burak, and B. P. Ölveczky (2013, October). The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong. *Neuron 80*(2), 494–506.

Alonso, R. G., A. Amador, and G. B. Mindlin (2016, October). An integrated model for motor control of song in Serinus canaria. *J. Physiol. Paris 110*(3 Pt A), 127–139.

Amador, A. and D. Margoliash (2013, July). A Mechanism for Frequency Modulation in Songbirds Shared with Humans. *J. Neurosci. 33*(27), 11136–11144.

Amador, A. and G. B. Mindlin (2008, December). Beyond harmonic sounds in a simple model for birdsong production. *Chaos 18*(4), 043123.

Amador, A., Y. S. Perl, G. B. Mindlin, and D. Margoliash (2013, March). Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature 495*(7439), 59–64.

Andalman, A. S. and M. S. Fee (2009, July). A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *PNAS 106*(30), 12518–12523.

Anderson, S. E., A. S. Dave, and D. Margoliash (1996, August). Template-based automatic recognition of birdsong syllables from continuous recordings. *J. Acoust. Soc. Am. 100*(2 Pt 1), 1209–1219.

Arneodo, E. M., Y. S. Perl, F. Goller, and G. B. Mindlin (2012). Prosthetic avian vocal organ controlled by a freely behaving bird based on a low dimensional model of the biomechanical periphery. *PLoS Comput. Biol. 8*(6), e1002546.

Aronov, D. and M. S. Fee (2012, October). Natural Changes in Brain Temperature Underlie Variations in Song Tempo during a Mating Behavior. *PLOS ONE 7*(10), e47856.

Ashe, J. (1997, September). Force and the motor cortex. *Behav. Brain Res. 87*(2), 255–269.

Ashmore, R. C., J. M. Wild, and M. F. Schmidt (2005, September). Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J. Neurosci. 25*(37), 8543–8554.

Ballintijn, n. and n. Cate (1998, May). Sound production in the collared dove: a test of the 'whistle' hypothesis. *J. Exp. Biol. 201 (Pt 10)*, 1637–1649.

Beckers, G. J., J. v. d. Meij, J. A. Lesku, and N. C. Rattenborg (2014, February). Plumes of neuronal activity propagate in three dimensions through the nuclear avian brain. *BMC Biology 12*(1), 16.

Benichov, J. I., S. E. Benezra, D. Vallentin, E. Globerson, M. A. Long, and O. Tchernichovski (2016, February). The Forebrain Song System Mediates Predictive Call Timing in Female and Male Zebra Finches. *Current Biology 26*(3), 309–318.

Blumstein, S. E. and K. N. Stevens (1979, October). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am. 66*(4), 1001–1017.

Boari, S., Y. S. Perl, A. Amador, D. Margoliash, and G. B. Mindlin (2015, November). Automatic reconstruction of physiological gestures used in a model of birdsong production. *J. Neurophysiol. 114*(5), 2912–2922.

Bolhuis, J. J., M. Everaert, R. C. Berwick, and N. Chomsky (2013, March). *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*. MIT Press. Google-Books-ID: BufxCwAAQBAJ.

Bouchard, K. E., N. Mesgarani, K. Johnson, and E. F. Chang (2013, March). Functional Organization of Human Sensorimotor Cortex for Speech Articulation. *Nature 495*(7441), 327–332.

Boughey, M. J. and N. S. Thompson (1981, January). Song Variety in the Brown Thrasher (Toxostoma rufum). *Zeitschrift für Tierpsychologie 56*(1), 47–58.

Bracha, V. (2004). Role of the cerebellum in eyeblink conditioning. *Prog. Brain Res. 143*, 331–339.

Brawn, T. P. and D. Margoliash (2015). A Bird's Eye View of Sleep-Dependent Memory Consolidation. *Curr Top Behav Neurosci 25*, 207–237.

Browman, C. P. and L. M. Goldstein (1986). Towards an articulatory phonology. *Phonology 3*(01), 219–252.

Bruno, A. M., W. N. Frost, and M. D. Humphries (2015, April). Modular deconstruction reveals the dynamical and physical building blocks of a locomotion motor program. *Neuron 86*(1), 304–318.

Buhusi, C. V. and W. H. Meck (2005, October). What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci 6*(10), 755–765.

Caminiti, R., P. B. Johnson, C. Galli, S. Ferraina, and Y. Burnod (1991, May). Making arm movements within different parts of space: the premotor and motor cortical representation of a coordinate system for reaching to visual targets. *J. Neurosci. 11*(5), 1182–1197.

Cardin, J. A., J. N. Raksin, and M. F. Schmidt (2005, April). Sensorimotor Nucleus NIf Is Necessary for Auditory Processing But Not Vocal Motor Output in the Avian Song System. *Journal of Neurophysiology 93*(4), 2157–2166.

Cardin, J. A. and M. F. Schmidt (2004, September). Noradrenergic inputs mediate state dependence of auditory responses in the avian song system. *J. Neurosci. 24*(35), 7745–7753.

Carpenter, A. F., A. P. Georgopoulos, and G. Pellizzer (1999, March). Motor cortical encoding of serial order in a context-recall task. *Science 283*(5408), 1752–1757.

Catchpole, C. K. (1987, April). Bird song, sexual selection and female choice. *Trends in Ecology & Evolution 2*(4), 94–97.

Chi, Z. and D. Margoliash (2001, December). Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron 32*(5), 899–910.

Churchland, M. M., J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy (2012, July). Neural population dynamics during reaching. *Nature 487*(7405), 51–56.

Churchland, M. M., B. M. Yu, J. P. Cunningham, L. P. Sugrue, M. R. Cohen, G. S. Corrado, W. T. Newsome, A. M. Clark, P. Hosseini, B. B. Scott, D. C. Bradley, M. A. Smith, A. Kohn, J. A. Movshon, K. M. Armstrong, T. Moore, S. W. Chang, L. H. Snyder, S. G. Lisberger, N. J. Priebe, I. M. Finn, D. Ferster, S. I. Ryu, G. Santhanam, M. Sahani, and K. V. Shenoy (2010, March). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci 13*(3), 369–378.

Clower, R. P., B. E. Nixdorf, and T. J. DeVoogd (1989, July). Synaptic plasticity in the hypoglossal nucleus of female canaries: structural correlates of season, hemisphere, and testosterone treatment. *Behav. Neural Biol. 52*(1), 63–77.

Cunningham, J. P. and B. M. Yu (2014, November). Dimensionality reduction for large-scale neural recordings. *Nat Neurosci 17*(11), 1500–1509.

Cynx, J. (1990). Experimental determination of a unit of song production in the zebra finch (Taeniopygia guttata). *Journal of Comparative Psychology 104*(1), 3–10.

Daou, A., M. T. Ross, F. Johnson, R. L. Hyson, and R. Bertram (2013, September). Electrophysiological characterization and computational models of HVC neurons in the zebra finch. *J. Neurophysiol. 110*(5), 1227–1245.

Dave, A. S. and D. Margoliash (2000, October). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science 290*(5492), 812–816.

Dave, A. S., A. C. Yu, and D. Margoliash (1998, December). Behavioral State Modulation of Auditory Activity in a Vocal Motor System. *Science 282*(5397), 2250–2254.

Dayan, P. and L. F. Abbott (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Computational neuroscience. Cambridge, Mass: Massachusetts Institute of Technology Press.

Derégnaucourt, S., P. P. Mitra, O. Fehér, C. Pytte, and O. Tchernichovski (2005, February). How sleep affects the developmental learning of bird song. *Nature 433*(7027), 710–716.

Eales, L. A. (1987, October). Song learning in female-raised zebra finches: another look at the sensitive phase. *Animal Behaviour 35*(5), 1356–1365.

Elemans, C. P. H., R. Laje, G. B. Mindlin, and F. Goller (2010, October). Smooth operator: avoidance of subharmonic bifurcations through mechanical mechanisms simplifies song motor control in adult zebra finches. *J. Neurosci. 30*(40), 13246–13253.

Elemans, C. P. H., A. F. Mead, L. C. Rome, and F. Goller (2008). Superfast vocal muscles control song production in songbirds. *PLoS ONE 3*(7), e2581.

Elemans, C. P. H., I. L. Y. Spierts, M. Hendriks, H. Schipper, U. K. Müller, and J. L. van Leeuwen (2006, March). Syringeal muscles fit the trill in ring doves (Streptopelia risoria L.). *J. Exp. Biol. 209*(Pt 5), 965–977.

Evarts, E. V. (1968, January). Relation of pyramidal tract activity to force exerted during voluntary movement. *Journal of Neurophysiology 31*(1), 14–27.

Fee, M. S., A. A. Kozhevnikov, and R. H. R. Hahnloser (2004, June). Neural mechanisms of vocal sequence generation in the songbird. *Ann. N. Y. Acad. Sci. 1016*, 153–170.

Fee, M. S., B. Shraiman, B. Pesaran, and P. P. Mitra (1998, September). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature 395*(6697), 67–71.

Feher, O., H. Wang, S. Saar, P. P. Mitra, and O. Tchernichovski (2009, May). De novo establishment of wild-type song culture in the zebra finch. *Nature 459*(7246), 564–568.

Flanders, M., S. I. H. Tillery, and J. F. Soechting (1992, June). Early stages in a sensorimotor transformation. *Behavioral and Brain Sciences 15*(2), 309–320.

Fletcher, N. H., T. Riede, and R. A. Suthers (2006, February). Model for vocalization by a bird with distensible vocal cavity and open beak. *J. Acoust. Soc. Am. 119*(2), 1005–1011.

Fowler, C., M. Richardson, K. Marsh, and K. Shockley (2008). Language use, coordination, and the emergence of cooperative action. In *Coordination: Neural, behavioral and social dynamics*. Elsevier.

Fritsch, G. and E. Hitzig (1870). Über die elektrische Erregbarkeit des Grosshirns (2009 reprinted translation). *Epilepsy Behav 15*(2), 123–130.

Gallego, J. A., M. G. Perich, L. E. Miller, and S. A. Solla (2017, June). Neural Manifolds for the Control of Movement. *Neuron 94*(5), 978–984.

Gardner, T., G. Cecchi, M. Magnasco, R. Laje, and G. B. Mindlin (2001, November). Simple motor gestures for birdsongs. *Phys. Rev. Lett. 87*(20), 208101.

Georgopoulos, A. P., J. Ashe, N. Smyrnis, and M. Taira (1992, June). The motor cortex and the coding of force. *Science 256*(5064), 1692–1695.

Georgopoulos, A. P., R. Caminiti, J. F. Kalaska, and J. T. Massey (1983). Spatial coding of movement: a hypothesis concerning the coding of movement direction by motor cortical populations. In *Neural Coding of Motor Performance*, Volume 7 of *Exp Brain Res Suppl*, pp. 327–336. Springer-Verlag.

Georgopoulos, A. P., J. F. Kalaska, R. Caminiti, and J. T. Massey (1982, November). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci. 2*(11), 1527–1537.

Georgopoulos, A. P., R. E. Kettner, and A. B. Schwartz (1988, August). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *J. Neurosci. 8*(8), 2928–2937.

Gibb, L., T. Q. Gentner, and H. D. I. Abarbanel (2009, September). Inhibition and recurrent excitation in a computational model of sparse bursting in song nucleus HVC. *J. Neurophysiol. 102*(3), 1748–1762.

Giszter, S. F. (2015, August). Motor primitives—new data and future questions. *Current Opinion in Neurobiology 33*, 156–165.

Goldin, M. A. and G. B. Mindlin (2017, August). Temperature manipulation of neuronal dynamics in a forebrain motor control nucleus. *PLoS Comput. Biol. 13*(8), e1005699.

Goller, F. and B. G. Cooper (2004, June). Peripheral motor dynamics of song production in the zebra finch. *Ann. N. Y. Acad. Sci. 1016*, 130–152.

Goller, F. and M. A. Daley (2001, November). Novel motor gestures for phonation during inspiration enhance the acoustic complexity of birdsong. *Proc. Biol. Sci. 268*(1483), 2301–2305.

Goller, F. and O. N. Larsen (1997, December). A new mechanism of sound generation in songbirds. *PNAS 94*(26), 14787–14791.

Goller, F. and T. Riede (2013, June). Integrative physiology of fundamental frequency control in birds. *J Physiol Paris 107*(3), 230–242.

Goller, F. and R. A. Suthers (1996a, July). Role of syringeal muscles in controlling the phonology of bird song. *J. Neurophysiol. 76*(1), 287–300.

Goller, F. and R. A. Suthers (1996b, February). Role of syringeal muscles in gating airflow and sound production in singing brown thrashers. *J. Neurophysiol. 75*(2), 867–876.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep learning*. Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press.

Graziano, M. (2006). The Organization of Behavioral Repertoire in Motor Cortex. *Annual Review of Neuroscience 29*(1), 105–134.

Grieves, R. M. and K. J. Jeffery (2017, February). The representation of space in the brain. *Behav. Processes 135*, 113–131.

Griffin, D. M., D. S. Hoffman, and P. L. Strick (2015, November). Corticomotoneuronal cells are "functionally tuned". *Science 350*(6261), 667–670.

Grillner, S. (1981). Control of Locomotion in Bipeds, Tetrapods, and Fish. In *Comprehensive Physiology*. John Wiley & Sons, Inc. DOI: 10.1002/cphy.cp010226.

Grimme, B., S. Fuchs, P. Perrier, and G. Schöner (2011, January). Limb versus speech motor control: a conceptual review. *Motor Control 15*(1), 5–33.

Guertin, P. A. (2009, December). The mammalian central pattern generator for locomotion. *Brain Res Rev 62*(1), 45–56.

Hahnloser, R. H. and M. S. Fee (2006, September). Sleep-related spike bursts in HVC are driven by the nucleus interface of the nidopallium. *Journal of Neurophysiology*.

Hahnloser, R. H. R., A. A. Kozhevnikov, and M. S. Fee (2002, September). An ultra-sparse code underliesthe generation of neural sequences in a songbird. *Nature 419*(6902), 65–70.

Halle, F., M. Gahr, and M. Kreutzer (2003, September). Effects of unilateral lesions of HVC on song patterns of male domesticated canaries. *J. Neurobiol. 56*(4), 303–314.

Hartley, R. S. and R. A. Suthers (1990, December). Lateralization of syringeal function during song production in the canary. *J. Neurobiol. 21*(8), 1236–1248.

Hatsopoulos, N. G., Q. Xu, and Y. Amit (2007, May). Encoding of Movement Fragments in the Motor Cortex. *J. Neurosci. 27*(19), 5105–5114.

Hocherman, S. and S. P. Wise (1991). Effects of hand movement path on motor cortical activity in awake, behaving rhesus monkeys. *Exp Brain Res 83*(2), 285–302.

Hubel, D. H. and T. N. Wiesel (1959, October). Receptive fields of single neurones in the cat's striate cortex. *J Physiol 148*(3), 574–591.

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. In *Bird vocalizations (Hinde RA, ed)*, pp. 61–74. London: Cambridge University Press.

Ioffe, S. and C. Szegedy (2015, February). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]*.

Jensen, K. K., B. G. Cooper, O. N. Larsen, and F. Goller (2007, November). Songbirds use pulse tone register in two voices to generate low-frequency sound. *Proc Biol Sci 274*(1626), 2703–2710.

Jin, D. Z., F. M. Ramazanoglu, and H. S. Seung (2007, December). Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. *J Comput Neurosci 23*(3), 283–299.

Katz, L. C. and M. E. Gurney (1981, September). Auditory responses in the zebra finch's motor system for song. *Brain Res. 221*(1), 192–197.

Kaufman, M. T., M. M. Churchland, S. I. Ryu, and K. V. Shenoy (2014, March). Cortical activity in the null space: permitting preparation without movement. *Nat Neurosci 17*(3), 440–448.

Kawai, R., T. Markman, R. Poddar, R. Ko, A. Fantana, A. Dhawale, A. Kampff, and B. Ölveczky (2015, May). Motor Cortex Is Required for Learning but Not for Executing a Motor Skill. *Neuron 86*(3), 800–812.

Knudsen, E. I. (1987). Neural derivation of sound source location in the barn owl. An example of a computational map. *Ann. N. Y. Acad. Sci. 510*, 33–38.

Konishi, M. (1965a, January). Effects of Deafening on Song Development in American Robins and Black-headed Grosbeaks. *Zeitschrift für Tierpsychologie 22*(5), 584–599.

Konishi, M. (1965b, January). The Role of Auditory Feedback in the Control of Vocalization in the White-Crowned Sparrow. *Zeitschrift für Tierpsychologie 22*(7), 770–783.

Kosche, G., D. Vallentin, and M. A. Long (2015, January). Interplay of Inhibition and Excitation Shapes a Premotor Neural Sequence. *J. Neurosci. 35*(3), 1217–1227.

Koumura, T. and K. Okanoya (2016, July). Automatic Recognition of Element Classes and Boundaries in the Birdsong with Variable Sequences. *PLOS ONE 11*(7), e0159188.

Lachlan, R. F., C. A. A. van Heijningen, S. M. Ter Haar, and C. Ten Cate (2016). Zebra Finch Song Phonology and Syntactical Structure across Populations and Continents-A Computational Comparison. *Front Psychol 7*, 980.

Lashley, K. S. (1951). The problem of serial order in behavior. In *Cerebral mechanisms in behavior; the Hixon Symposium*, pp. 112–146. Oxford, England: Wiley.

Leppelsack, H.-J., D. Margoliash, and U. Häusler (1986). Responses to acoustic stimulation in the vocal control center HVc of the starling. In *Verhandlungen der Deutschen Zool Ges*, pp. 224–225. New York: Stuttgart.

Leyton, A. S. F. and C. S. Sherrington (1917, July). Observations on the Excitable Cortex of the Chimpanzee, Orang-Utan, and Gorilla. *Exp Physiol 11*(2), 135–222.

Li, M. and H. Greenside (2006, July). Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus HVC. *Phys Rev E Stat Nonlin Soft Matter Phys 74*(1 Pt 1), 011918.

Liberman, A. M. and I. G. Mattingly (1985, October). The motor theory of speech perception revised. *Cognition 21*(1), 1–36.

Long, M. A. and M. S. Fee (2008, November). Using temperature to analyze temporal dynamics in the songbird motor pathway. *Nature 456*(7219), 189–194.

Long, M. A., D. Z. Jin, and M. S. Fee (2010, November). Support for a synaptic chain model of neuronal sequence generation. *Nature 468*(7322), 394–399.

Lusignan, M. E. (2012). *Growing up singing: Behavioral and physiological perspectives on song acquisition*. Ph. D. thesis, University of Chicago, Chicago, IL. ISBN: 9781267602015.

Lynch, G. F., T. S. Okubo, A. Hanuschkin, R. H. R. Hahnloser, and M. S. Fee (2016, May). Rhythmic Continuous-Time Coding in the Songbird Analog of Vocal Motor Cortex. *Neuron 90*(4), 877–892.

Margoliash, D. (1983, May). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci. 3*(5), 1039–1057.

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

Margoliash, D. and E. S. Fortune (1992, November). Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *J. Neurosci. 12*(11), 4309–4326.

Margoliash, D. and M. Konishi (1985, September). Auditory representation of autogenous song in the song system of white-crowned sparrows. *Proc. Natl. Acad. Sci. U.S.A. 82*(17), 5997–6000.

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology 71*(2, Pt.2), 1–25.

Marler, P., M. Konishi, A. Lutjen, and M. S. Waser (1973, May). Effects of Continuous Noise on Avian Hearing and Vocal Development. *Proc Natl Acad Sci U S A 70*(5), 1393–1396.

Marler, P. and M. Tamura (1962, September). Song "Dialects" in Three Populations of White-Crowned Sparrows. *The Condor 64*(5), 368–377.

Matsuzaka, Y., N. Picard, and P. Strick (2007). Skill representation in the primary motor cortex after long-term practice. *Journal of Neurophysiology 97*(2), 1819–1832.

McCasland, J. S. and M. Konishi (1981, December). Interaction between auditory and motor activities in an avian song control nucleus. *Proc. Natl. Acad. Sci. U.S.A. 78*(12), 7815–7819.

Micco Jr., D. J. (1977, November). Lightweight, multi-contact, slip-ring commutator for recording and stimulation with small animals. *Brain Research Bulletin 2*(6), 499–502.

Mindlin, G. B. (2017, September). Nonlinear dynamics in the study of birdsong. *Chaos 27*(9), 092101.

Mooney, R. and J. F. Prather (2005, February). The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *J. Neurosci. 25*(8), 1952–1964.

Moran, D. W. and A. B. Schwartz (1999, November). Motor Cortical Activity During Drawing Movements: Population Representation During Spiral Tracing. *Journal of Neurophysiology 82*(5), 2693–2704.

Morris, D. (1954, January). The Reproductive Behaviour of the Zebra Finch (Poephila Guttata), With Special Reference To Pseudofemale Behaviour and Displacement Activities. *Behaviour 6*(1), 271–322.

Morrow, M. M. and L. E. Miller (2003, April). Prediction of muscle activity by populations of sequentially recorded primary motor cortex neurons. *J. Neurophysiol. 89*(4), 2279–2288.

Murakami, M., M. I. Vicente, G. M. Costa, and Z. F. Mainen (2014, November). Neural antecedents of self-initiated actions in secondary motor cortex. *Nat. Neurosci. 17*(11), 1574–1582.

Mussa-Ivaldi, F. A. (1988, August). Do neurons in the motor cortex encode movement direction? An alternative hypothesis. *Neurosci. Lett. 91*(1), 106–111.

Méndez, J. M., G. B. Mindlin, and F. Goller (2012, June). Interaction between telencephalic signals and respiratory dynamics in songbirds. *J. Neurophysiol. 107*(11), 2971–2983.

Nordeen, K. W. and E. J. Nordeen (1992, January). Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol. 57*(1), 58–66.

Nottebohm, F., T. M. Stokes, and C. M. Leonard (1976, February). Central control of song in the canary, Serinus canarius. *J. Comp. Neurol. 165*(4), 457–486.

Nowicki, S. (1987, January). Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature 325*(6099), 53–55.

Ohms, V. R., P. C. Snelderwaard, C. ten Cate, and G. J. L. Beckers (2010, July). Vocal Tract Articulation in Zebra Finches. *PLoS One 5*(7).

Okubo, T. S., E. L. Mackevicius, H. L. Payne, G. F. Lynch, and M. S. Fee (2015, November). Growth and splitting of neural sequences in songbird vocal development. *Nature 528*(7582), 352–357.

Paninski, L., M. R. Fellows, N. G. Hatsopoulos, and J. P. Donoghue (2004, January). Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J. Neurophysiol. 91*(1), 515–532.

Payne, R. B., W. L. Thompson, K. L. Fiala, and L. L. Sweany (1981). Local Song Traditions in Indigo Buntings: Cultural Transmission of Behavior Patterns across Generations. *Behaviour 77*(4), 199–221.

Perl, Y. S., E. M. Arneodo, A. Amador, F. Goller, and G. B. Mindlin (2011, November). Reconstruction of physiological instructions from Zebra finch song. *Phys Rev E Stat Nonlin Soft Matter Phys 84*(5 Pt 1), 051909.

Perrier, P. and S. Fuchs (2008, September). Speed-curvature relations in speech production challenge the 1/3 power law. *J. Neurophysiol. 100*(3), 1171–1183.

Picardo, M. A., J. Merel, K. A. Katlowitz, D. Vallentin, D. E. Okobi, S. E. Benezra, R. C. Clary, E. A. Pnevmatikakis, L. Paninski, and M. A. Long (2016, May). Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch. *Neuron 90*(4), 866–876.

Polit, A. and E. Bizzi (1979, January). Characteristics of motor programs underlying arm movements in monkeys. *J. Neurophysiol. 42*(1 Pt 1), 183–194.

Prather, J. F., S. Peters, S. Nowicki, and R. Mooney (2008, January). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature 451*(7176), 305–310.

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *Journal of Comparative and Physiological Psychology 93*(2), 260–277.

Rattenborg, N. and C. Amlaner (2002). Phylogeny of sleep. In *Sleep Medicine*, pp. 7–22. Hanley and Belfus, Inc.

Rauske, P. L., Z. Chi, A. S. Dave, and D. Margoliash (2010, February). Neuronal stability and drift across periods of sleep: premotor activity patterns in a vocal control nucleus of adult zebra finches. *J. Neurosci. 30*(7), 2783–2794.

Rauske, P. L., S. D. Shea, and D. Margoliash (2003, March). State and neuronal class-dependent reconfiguration in the avian song system. *J. Neurophysiol. 89*(3), 1688–1701.

Ravbar, P., D. Lipkind, L. C. Parra, and O. Tchernichovski (2012, March). Vocal exploration is locally regulated during song learning. *J. Neurosci. 32*(10), 3422–3432.

Read, A. F. and D. M. Weary (1992, October). The Evolution of Bird Song: Comparative Analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences 338*(1284), 165–187.

Reinke, H. and J. M. Wild (1997, March). Distribution and connections of inspiratory premotor neurons in the brainstem of the pigeon (Columba livia). *J. Comp. Neurol. 379*(3), 347–362.

Riede, T. and F. Goller (2010a, January). Functional morphology of the sound-generating labia in the syrinx of two songbird species. *J. Anat. 216*(1), 23–36.

Riede, T. and F. Goller (2010b, October). Peripheral mechanisms for vocal production in birds - differences and similarities to human speech and singing. *Brain Lang 115*(1), 69–80.

Roberts, T. F., E. Hisey, M. Tanaka, M. G. Kearney, G. Chattree, C. F. Yang, N. M. Shah, and R. Mooney (2017, July). Identification of a motor-to-auditory pathway important for vocal learning. *Nat Neurosci 20*(7), 978–986.

Roberts, T. F., M. E. Klein, M. F. Kubke, J. M. Wild, and R. Mooney (2008, March). Telencephalic Neurons Monosynaptically Link Brainstem and Forebrain Premotor Networks Necessary for Song. *J. Neurosci. 28*(13), 3479–3489.

Roper, A. and R. Zann (2006, May). The Onset of Song Learning and Song Tutor Selection in Fledgling Zebra Finches. *Ethology 112*(5), 458–470.

Rosenbaum, D. A., R. G. Cohen, S. A. Jax, D. J. Weiss, and R. van der Wel (2007, August). The problem of serial order in behavior: Lashley's legacy. *Hum Mov Sci 26*(4), 525–554.

Rossant, C., S. N. Kadir, D. F. M. Goodman, J. Schulman, M. L. D. Hunter, A. B. Saleem, A. Grosmark, M. Belluscio, G. H. Denfield, A. S. Ecker, A. S. Tolias, S. Solomon, G. Buzsáki, M. Carandini, and K. D. Harris (2016, April). Spike sorting for large, dense electrode arrays. *Nat Neurosci 19*(4), 634–641.

Santucci, D. M., J. D. Kralik, M. A. Lebedev, and M. A. L. Nicolelis (2005, September). Frontal and parietal cortical ensembles predict single-trial muscle activity during reaching movements in primates. *European Journal of Neuroscience 22*(6), 1529–1540.

Scharff, C., J. R. Kirn, M. Grossman, J. D. Macklis, and F. Nottebohm (2000, February). Targeted Neuronal Death Affects Neuronal Replacement and Vocal Behavior in Adult Songbirds. *Neuron 25*(2), 481–492.

Schmidt, M. F. and M. Konishi (1998, October). Gating of auditory responses in the vocal control system of awake songbirds. *Nat. Neurosci. 1*(6), 513–518.

Schneegans, S. and G. Schöner (2008). Dynamic field theory as a framework for understanding embodied cognition. In *Handbook of cognitive science: An embodied approach*, pp. 241–271. Elsevier.

Scott, S. H. (2004, July). Optimal feedback control and the neural basis of volitional motor control. *Nat. Rev. Neurosci. 5*(7), 532–546.

Scott, S. H. (2008, March). Inconvenient truths about neural processing in primary motor cortex. *J. Physiol. (Lond.) 586*(5), 1217–1224.

Scott, S. H., P. L. Gribble, K. M. Graham, and D. W. Cabel (2001, September). Dissociation between hand motion and population vectors from neural activity in motor cortex. *Nature 413*(6852), 161–165.

Scotto-Lomassese, S., C. Rochefort, A. Nshdejan, and C. Scharff (2007, March). HVC interneurons are not renewed in adult male zebra finches. *Eur. J. Neurosci. 25*(6), 1663–1668.

Searcy, W. A. (1992, February). Song Repertoire and Mate Choice in Birds. *Integr Comp Biol 32*(1), 71–80.

Searcy, W. A. and M. Andersson (1986). Sexual Selection and the Evolution of Song. *Annual Review of Ecology and Systematics 17*(1), 507–533.

Seely, J. S., M. T. Kaufman, S. I. Ryu, K. V. Shenoy, J. P. Cunningham, and M. M. Churchland (2016, November). Tensor Analysis Reveals Distinct Population Structure that Parallels the Different Computational Roles of Areas M1 and V1. *PLoS Comput. Biol. 12*(11), e1005164.

Shank, S. S. and D. Margoliash (2009, March). Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature 458*(7234), 73–77.

Shea, S. D. and D. Margoliash (2003, December). Basal forebrain cholinergic modulation of auditory activity in the zebra finch song system. *Neuron 40*(6), 1213–1226.

Shenoy, K. V., M. Sahani, and M. M. Churchland (2013, July). Cortical Control of Arm Movements: A Dynamical Systems Perspective.

Sherman, S. M. (2007, August). The thalamus is more than just a relay. *Curr Opin Neurobiol 17*(4), 417–422.

Sherrington, C. S. (1910, April). Flexion-reflex of the limb, crossed extension-reflex, and reflex stepping and standing. *The Journal of Physiology 40*(1-2), 28–121.

Shima, K. and J. Tanji (1998, December). Both supplementary and presupplementary motor areas are crucial for the temporal organization of multiple movements. *J. Neurophysiol. 80*(6), 3247–3260.

Simpson, H. B. and D. S. Vicario (1990, May). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci. 10*(5), 1541–1556.

Sitt, J. D., A. Amador, F. Goller, and G. B. Mindlin (2008, July). Dynamical origin of spectrally rich vocalizations in birdsong. *Phys Rev E Stat Nonlin Soft Matter Phys 78*(1 Pt 1), 011905.

Solis, M. M. and A. J. Doupe (1997, August). Anterior forebrain neurons develop selectivity by an intermediate stage of birdsong learning. *J. Neurosci. 17*(16), 6447–6462.

Staddon, J. E. R. (2005, July). Interval timing: memory, not a clock. *Trends in Cognitive Sciences 9*(7), 312–314.

Stap, D. (2006). *Birdsong: a natural history*. Oxford ; New York: Oxford University Press. OCLC: ocm63472807.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics 17*, 3–46.

Sturdy, C. B., L. S. Phillmore, and R. G. Weisman (1999). Note types, harmonic structure, and note order in the songs of zebra finches (Taeniopygia guttata). *Journal of Comparative Psychology 113*(2), 194–203.

Suthers, R. A., F. Goller, and C. Pytte (1999, May). The neuromuscular control of birdsong. *Philos. Trans. R. Soc. Lond., B, Biol. Sci. 354*(1385), 927–939.

Suthers, R. A., E. Vallet, A. Tanvez, and M. Kreutzer (2004, September). Bilateral song production in domestic canaries. *J. Neurobiol. 60*(3), 381–393.

Suthers, R. A. and S. A. Zollinger (2004, June). Producing Song: The Vocal Apparatus. *Annals of the New York Academy of Sciences 1016*(1), 109–129.

Sutter, M. L. and D. Margoliash (1994, November). Global synchronous response to autogenous song in zebra finch HVc. *J Neurophysiol 72*(5), 2105–2123.

Takahashi, D. Y., A. R. Fenley, Y. Teramoto, D. Z. Narayanan, J. I. Borjon, P. Holmes, and A. A. Ghazanfar (2015, August). The developmental dynamics of marmoset monkey vocal production. *Science 349*(6249), 734–738.

Taylor, C. S. R. and C. G. Gross (2003, October). Twitches versus movements: a story of motor cortex. *Neuroscientist 9*(5), 332–342.

Tchernichovski, O. and P. P. Mitra (2002, December). Towards quantification of vocal imitation in the zebra finch. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol. 188*(11-12), 867–878.

Tchernichovski, O., P. P. Mitra, T. Lints, and F. Nottebohm (2001, March). Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science 291*(5513), 2564–2569.

Tchernichovski, O., F. Nottebohm, C. E. Ho, B. Pesaran, and P. P. Mitra (2000, June). A procedure for an automated measurement of song similarity. *Animal Behaviour 59*(6), 1167–1176.

Theunissen, F. E. and A. J. Doupe (1998, May). Temporal and Spectral Sensitivity of Complex Auditory Neurons in the Nucleus HVc of Male Zebra Finches. *J. Neurosci. 18*(10), 3786–3802.

Thorpe, W. H. (1958, October). The Learning of Song Patterns by Birds, with Especial Reference to the Song of the Chaffinch Fringilla Coelebs. *Ibis 100*(4), 535–570.

Titze, I. R. (1988, April). The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am. 83*(4), 1536–1552.

Todorov, E. (2000, April). Direct cortical control of muscle activation in voluntary arm movements: a model. *Nat. Neurosci. 3*(4), 391–398.

Troyer, T. W. (2016, May). Continuous Time Representations of Song in Zebra Finches. *Neuron 90*(4), 672–674.

Vallentin, D., G. Kosche, D. Lipkind, and M. A. Long (2016, January). Inhibition protects acquired song segments during vocal learning in zebra finches. *Science 351*(6270), 267–271.

Vandecasteele, M., S. M, S. Royer, M. Belluscio, A. Berényi, K. Diba, S. Fujisawa, A. Grosmark, D. Mao, K. Mizuseki, J. Patel, E. Stark, D. Sullivan, B. Watson, and G. Buzsáki (2012, March). Large-scale Recording of Neurons by Movable Silicon Probes in Behaving Rodents. *JoVE (Journal of Visualized Experiments)* (61), e3568–e3568.

Vicario, D. S. (1991, January). Contributions of syringeal muscles to respiration and vocalization in the zebra finch. *J. Neurobiol. 22*(1), 63–73.

Viviani, P. and N. Stucchi (1992, August). Biological movements look uniform: evidence of motor-perceptual interactions. *J Exp Psychol Hum Percept Perform 18*(3), 603–623.

Vu, E. T., M. E. Mazurek, and Y. C. Kuo (1994, November). Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci. 14*(11), 6924–6934.

Walker, M. P. and R. Stickgold (2006). Sleep, Memory, and Plasticity. *Annual Review of Psychology 57*(1), 139–166.

Watson, S., S. Townsend, A. Schel, C. Wilke, E. Wallace, L. Cheng, V. West, and K. Slocombe (2015, February). Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology 25*(4), 495–499.

Werner, W., E. Bauswein, and C. Fromm (1991, September). Static firing rates of premotor and primary motor cortical neurons associated with torque and joint position. *Exp Brain Res 86*(2), 293–302.

Westneat, M. W., J. H. Long, W. Hoese, and S. Nowicki (1993, September). Kinematics of birdsong: functional correlation of cranial movements and acoustic features in sparrows. *J. Exp. Biol. 182*, 147–171.

Wild, J. M. (1993a, March). The avian nucleus retroambigualis: a nucleus for breathing, singing and calling. *Brain Res. 606*(2), 319–324.

Wild, J. M. (1993b, December). Descending projections of the songbird nucleus robustus archistriatalis. *J. Comp. Neurol. 338*(2), 225–241.

Wild, J. M. (1997, November). Neural pathways for the control of birdsong production. *J. Neurobiol. 33*(5), 653–670.

Wild, J. M., F. Goller, and R. A. Suthers (1998, September). Inspiratory muscle activity during bird song. *J. Neurobiol. 36*(3), 441–453.

Wild, J. M., D. Li, and C. Eagleton (1997, January). 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*(3), 392–413.

Williams, H. and J. R. McKibben (1992, January). Changes in stereotyped central motor patterns controlling vocalization are induced by peripheral nerve injury. *Behav. Neural Biol. 57*(1), 67–78.

Williams, H. and F. Nottebohm (1985, July). Auditory responses in avian vocal motor neurons: a motor theory for song perception in birds. *Science 229*(4710), 279–282.

Wiltschko, A. B., G. J. Gage, and J. D. Berke (2008, August). Wavelet filtering before spike detection preserves waveform shape and enhances single-unit discrimination. *J. Neurosci. Methods 173*(1), 34–40.

Woodgate, J. L., M. M. Mariette, A. T. D. Bennett, S. C. Griffith, and K. L. Buchanan (2012, March). Male song structure predicts reproductive success in a wild zebra finch population. *Animal Behaviour 83*(3), 773–781.

Wu, W. and N. Hatsopoulos (2006, November). Evidence against a single coordinate system representation in the motor cortex. *Exp Brain Res 175*(2), 197–210.

Yger, P., G. L. B. Spampinato, E. Esposito, B. Lefebvre, S. Deny, C. Gardella, M. Stimberg, F. Jetter, G. Zeck, S. Picaud, J. Duebel, and O. Marre (2016, August). Fast and accurate spike sorting in vitro and in vivo for up to thousands of electrodes. *bioRxiv*.

Yildiz, I. B. and S. J. Kiebel (2011, December). A Hierarchical Neuronal Model for Generation and Online Recognition of Birdsongs. *PLOS Computational Biology 7*(12), e1002303.

Young, B. K. and F. Goller (2012). Playback-induced syringeal motor rehearsal of song without concurrent respiratory movements. Program No. 205.02. New Orleans, LA.

Yu, A. C. and D. Margoliash (1996, September). Temporal Hierarchical Control of Singing in Birds. *Science 273*(5283), 1871–1875.

Zann, R. (1993). Variation in song structure within and among populations of Australian Zebra Finches. *The Auk 110*(4), 716–726.

Zann, R. A. (1996). *The Zebra Finch: A Synthesis of Field and Laboratory Studies*. Oxford University Press. Google-Books-ID: 5KO6cZH0WbEC.