Title
Motor Planning for Syllable Sequence and Phonology in Bridsong

Permalink
https://escholarship.org/uc/item/1x7457pg

Author
Wohlgemuth, Melville Joseph

Publication Date
2008

Peer reviewed|Thesis/dissertation

Motor Planning for Syllable Sequence and Phonology in Birdsong

by

Melville Joseph Wohlgemuth III

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

GRADUATE DIVISION

of the

This work is dedicated to my family: Mel, Ann and Allaire.

Acknowledgements

There are many people who were integral to the completion of the following dissertation. First and foremost, I have to thank Michael Brainard, my thesis advisor, for all of the guidance he has provided me with throughout my graduate career. He demonstrated a lot of patience with me, and was a tremendous source of scientific support for me. I would also like to thank my thesis committee: Allison Doupe, Jonathan Horton and Loren Frank. They expected a lot from me, and as result, my thesis is something in which I take great pride. The interaction I had with my committee taught my how to be a rigorous scientist, and made me a much more critical thinker about the design and implementation of scientific research.

There are also two specific people in the Brainard Laboratory I wish to thank: Sam Sober and Evren Tumer. Much of the following work was done in collaboration with Sam Sober. He and I worked very well together, and I think we both benefited from our collaboration. Evren Tumer was integral to many of the concepts and analyses contained in the following dissertation. He was always willing to help, and was a great friend throughout my time at UCSF.

I also have to thank my friends and family. Science is a long, somewhat lonesome path. Without the love and support of the people in my life outside of science, I would have never made it through graduate school. My father and sister were particularly instrumental in my mental well being. They are always supportive and always have good advice, and their role as my psychologists was very important at times throughout the last six years.

Lastly, I want to thank my mother. Although she always wanted me to go to medical school and become a 'real' doctor, I think she would be very proud of me if she was still alive today. I almost think she would be more surprised than proud, actually.

<u>Contributions</u>

The work presented below was done in collaboration with Samuel J. Sober.

Abstract:

       Song production in birds requires planning on two different time-frames: the immediate muscle program for the vocal effectors (vocal and respiratory musculature), as well as the more temporally extended control of sequencing the shorter vocal elements (syllables) into longer vocalizations. The songs of the Bengalese finch exhibit variability in both dimensions: syllables are produced with a small amount of variability from rendition to rendition, and the syllables themselves can be used in a variety of sequences. The following research examines three question in regards to variability in Bengalese finch song.

1. Does a syllable's sequence affect its phonology?
2. Are changes in phonology are a result of differences in brain activity?
3. Is there a correlation between changes in brain activity and changes in phonology?

       In analyzing the songs of Bengalese finches, we find that syllable structure is altered by sequence in two different ways: syllable production is modified by the immediate history of the system, as well as long time-scale patterns in the sequencing of syllables. These sequence effects upon syllable production are not the result of purely peripheral influences, however. There are multiple ways in which changes to central neural activity correlate with changes in syllable phonology. Changes in the activity of the robust nucleus of the arcopallium (RA) neurons were correlated with mean changes in syllable production, as well as with changes to syllable production around the mean. The

correlation between mean changes in RA activity and syllable structure implies that the RA motor program is phonological in nature. We also find that RA activity correlates specifically with changes in syllable features such as pitch, amplitude and entropy. In taking these results together, a more complete model of syllable control by RA can be constructed. The data suggest that general patterns in RA activity relate to the identity of individual syllables, and that changes in syllable structure are made by increasing or decreasing that activity of specific pools of RA neurons. The results of this research are the first demonstration of phonological encoding by RA neurons.

Table of Contents

List of Figures

Introduction

Chapter 1

Chapter 2

Chapter 3

Appendix 1

GENERAL INTRODUCTION

Maintaining a social system requires communication between its constituents. The methods of communication are quite variable, with many different motor and sensory modalities being employed to convey and receive information (Bradbury and Vehrencamp, 1998). Although there is much to be learned from each method of communication, we, as humans, are particularly interested in those organisms that use learned vocalizations in maintaining their social relationships. Understanding how humans learn and maintain language is of great interest, considering the importance of vocal behaviors in our social interactions. As a result, the learned vocalizations of passerine birds have been studied extensively as a model for human language acquisition and maintenance (Doupe and Kuhl, 1999).

Young birds learn their songs from adult tutors (Catchpole and Slater, 1995). The song development process is very similar to the ontogeny of language in humans (Doupe and Kuhl, 1999). Young birds first go through a sensory acquisition phase where most of their efforts are spent listening to their tutor. It is during this stage that they memorize their tutor's song to serve as a 'template' (Catchpole and Slater, 1995). This template serves as a target vocalization throughout the process of song learning (Catchpole and Slater, 1995). The next phase of song learning is typically termed sensory-motor integration. As the name suggests, this stage involves refining the motor control necessary to produce an accurate copy of the tutor song (Catchpole and Slater, 1995). Auditory feedback is the birds' primary way of judging the similarity of the acoustic

output to their memorized template (Nottebohm 1968), with proprioception used to a lesser degree (Wild et al, 2002). Exactly how audition and proprioception drive song development is still an unresolved issue. The final phase of song learning, termed crystallization, is essentially the endpoint of song learning. Once song crystallization commences, young birds have typically mastered the production of their own songs, and little to no changes are made to songs afterwards (Catchpole and Slater, 1995). There are some birds, called open-learners, that retain the ability to learn new vocalizations throughout life (e.g. parrots and mockingbirds); however, the following research is focused on birds that are closed learners, and make few changes to their songs after crystallization (Catchpole and Slater, 1995). The three stages of song learning described here mimic the process of language acquisition in humans. In human language development, babies first listen to their parents, followed by a phase dominated by babbling as they develop the ability to vocalize; and eventually as we age, adjustments to vocalizations (e.g. like learning a new language) become increasingly more difficult (Bishop, 1999). It is because of these similarities between birdsong and human language that the songbird system is so beneficial in furthering our understanding of speech development, and motor learning in general.

The process of vocal motor learning follows a trajectory similar to other motor behaviors: early attempts are often quite variable, but over time, and with practice, motor control is refined and competency is reached (Mitra et al, 1998). When a bird is learning its song, it must master motor control on two temporal levels: the phonation of individual syllables, as well as the ordering of those syllables into songs. Both syllable phonation

and sequencing are honed over time, but even in the adult behavior, a small amount of variability remains (Sakata et al, 2008; Kao and Brainard, 2006). An interesting question is whether that residual variability has any purpose, or is just an indication of the limits of motor control. One way to address this question is to determine if the source of this variability is outside of the central nervous system (CNS), or if it is in fact under the control of the CNS. If this variability is being generated in the CNS, the brain may be able to use it to make changes to song.

Control of song in the avian CNS is accomplished by a specialized set of nuclei that are responsible for the acquisition, production and maintenance of song (Figure 1.1c). These nuclei can be loosely divided into two, interconnected circuits: the motor pathway and the anterior forebrain pathway (AFP). Constituting the motor pathway are two main nuclei, HVC (used as a proper name), and the robust nucleus of the arcopallium (RA) (Figure I.1).
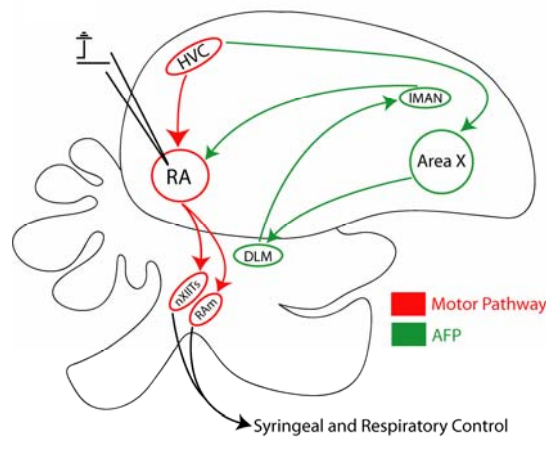
Figure I.1: Saggital view of the avian song system. In red is the motor pathway, in green is the anterior forebrain pathway (AFP). Neural recordings for the following study were performed in RA.

Because these brain areas are responsible for the motor commands for song, lesions of either structure in young or adult birds will almost completely ablate normal song behavior (Catchpole and Slater, 1995). In terms of basic anatomy, RA sits 'downstream' from HVC, with most neuronal projections originating in HVC and synapsing onto RA neurons (Wild, 2004). Recently, however, a projection back from RA to HVC has been identified (Roberts et al, 2008), opening up the possibility that RA can directly influence ongoing motor commands in HVC. Previous research on HVC and RA has alluded to different functional roles for these two nuclei in song production. Recordings in HVC have demonstrated that neurons in this structure tend to fire at only one time point during song production (Hahnloser et al, 2002; Yu and Margoliash, 1996). The nature of this encoding seems to be a temporal one, with different neurons in HVC controlling the timing of song production. RA neurons on the other hand, fire at multiple time points during song (Leonardo and Fee, 2005; Yu and Margoliash, 1996), and synapse directly onto brainstem motoneurons controlling the syrinx and respiration during song production (Wild, 1993). The anatomical connections and firing properties of RA neurons suggest that this structure is responsible for taking timing commands from HVC and translating them into continuous motor commands for song. In this way, the song sequence is imposed upon RA by HVC, and RA then takes this information and produces muscle activation patterns for syllable phonation.

Unlike the motor pathway, the AFP is not directly required for adult song production, but it is necessary for song learning and maintenance. The main nuclei of the AFP include Area X, the dorsolateral thalamus (DLM), and the lateral magnocellular

nucleus (lMAN) (Figure 1.1c). HVC sends projections to both RA and Area X, and the

AFP reconnects to the motor pathway with lMAN sending projections to RA (Figure

1.1c). Lesions to AFP structures disrupt the normal song learning process in young birds

(Bottjer, 1994; Scharff, 1991; Sohrabji, 1990), but removal of AFP input in adults only

has modest effects upon song (Bottjer, 1994). The role of the AFP in adulthood is

believed to be more important for song maintenance than song production. As a result,

lesions to AFP nuclei inhibit changes to song (Brainard and Doupe, 2000), but not

singing in general (Bottjer, 1994). The anatomy of the connections between the AFP and

motor pathway suggests that these two pathways work in concert for song production and

maintenance. More specifically, the interaction of these two circuits may play a critical

role in how the birdbrain utilizes auditory information in order to develop and adapt vocal

motor behaviors.

Neurons in lMAN and HVC both synapse onto neurons in RA. This site of

convergence is thought to mediate the integration of auditory and motor information

(Konishi, 2004). lMAN's projection onto RA is mostly mediated by NMDA receptors,

whereas those from HVC are composed of a mixture of NMDA and AMPA receptors

(Mooney and Konishi, 1991; Stark and Perkel, 1999). Because of this difference in

receptors, the influence of lMAN upon RA is quite different than that from HVC. In the

adult bird, HVC can directly excite neurons in RA, but the cells in RA must already be

depolarized in order for lMAN neurons to excite RA. In this way, the NMDA receptor,

and the neurons in lMAN, act as "coincidence detectors," because their influence is

dependent on the occurrence of prior events (Duquid and Sjostrom, 2006; Yuste et al,

1999). One property of connected neurons is that the more they are activated, the stronger the connection between them becomes (Voronin, 1983). As a result, coincidence detection in the lMAN-RA synapse can drive a potentiation in the connections between neurons in lMAN and RA, as well as between neurons in HVC and RA. One plausible model of the process of sensory-motor integration for song involves the strengthening and weakening of synapses between lMAN, RA and HVC; those patterns of activity in RA that result in accurate copies of song are strengthened, and those that result in more errors are weakened.

The direct and immediate influence of lMAN upon song behavior has been assessed in several previous reports. Many of these studies have found a relationship between lMAN activity and the variability of song. In juvenile birds, temporary inactivation of lMAN increases the stereotypy of the song (Olveczky et al, 2005); and lesions lead to premature crystallization of song (Bottjer, 1994; Scharff, 1991). Similar results have been found in adult birds, with lMAN removal resulting in increases in song stereotypy (Kao and Brainard, 2006; Brainard, 2004). Chronic neural recordings in lMAN have also demonstrated a direct relationship between variability in the firing of lMAN neurons and variability in song (Hessler and Doupe, 1999). In terms of specific effects of lMAN upon song, an interesting result was found when lMAN was electrically stimulated during singing. Exogenous activation of neurons in lMAN caused reliable changes in the pitch and amplitude of the ongoing syllables (Kao et al, 2005). This experiment shows that lMAN can induce specific changes, on a moment-by-moment basis to song structure. Considering the role the lMAN has in both song variability and

spectral parameters, its influence upon RA most likely includes an instructive role. It is unclear if the variability impressed upon RA by lMAN can bias the song in one particular direction, but it is clear that lMAN can influence song structure in reliable ways. In view of the anatomy of the song circuit, the influence of lMAN upon song production is most likely mediated through its affects upon the firing properties of RA neurons.

The research discussed below addresses the question of how changes in the patterns of brain activity drive changes in song. The basic anatomy of the song circuit, as well as previous research on the motor pathway, suggests that RA is the most likely structure in which changes in neural activity correlate with changes in the behavior. As a result, the experiments discussed are dedicated to elucidating the role that RA may have in changing song features.

CHAPTER 1 HYPOTHESIS

We have studied the songs of the Bengalese finch because, much like human language, there is variability in both phonology and sequence. A variably sequenced song is advantageous when attempting to disambiguate control signals for phonology and sequence. Bengalese finch song has two different types of sequence varieties: convergence points and divergence points. Convergent syllables are syllables that can be preceded by more than one other syllable, and divergent syllables are those syllables that are followed by more than one syllable. There are analogous features of human language, and the production of the phonemes involved in these vocal sequences is reliably altered depending upon context (Daniloff and Hammarberg, 1973). It is believed

that the differences between phonemes found in different sequences result from the influence of different motor transitions (Öhman, 1966). The first chapter deals with the following hypothesis about the interactions of syllables and sequence in birds:

The same syllable found in different sequences will have consistent changes in production across those different sequences.

The above hypothesis was tested by making very precise measurements of spectro-temporal properties of syllables, and then determining how these properties change as a syllable's sequence changes. This can be accomplished at two levels of sequence changes. At convergence and divergence points, we can relate changes in the syllable structure to the structures of the immediately preceding and following syllables, respectively. Any reliable changes as a result of the immediately adjacent syllable are most likely a result of the carry over of, or the preparation for, different motor acts. We can also assess the effects of sequence across broader portions of song. There are long-range differences in the incidence of certain syllable sequences. Because the local motor contexts are the same for these broad changes, any modifications in syllable structure must be through a different mechanism than those observed as a result of adjacent syllable differences. This mechanism may instead be the result of differences in motor planning, rather than compensations for changing motor contexts.

CHAPTER 2 HYPOTHESIS

An open area of research in the field of birdsong is the level of interaction between phonology and sequence during song learning and production. Of specific interest is how RA encodes the same syllable found in different sequences. Previous research on RA encoding of repeated syllables in zebra finch song suggested that these syllables might be encoded identically in RA, but the data set was not extensive enough for a thorough analysis. Because the Bengalese finch often uses the same syllable in multiple sequences, there is a greater opportunity to look more closely at the issue of syllable versus sequence encoding in RA. A complicating layer on top of this issue may be any significant differences in phonology found for the same syllable found in different sequences (Chapter 2). If sequence does reliably alter phonology, how will RA encode each syllable? The second chapter will explore this question by addressing the following hypothesis about syllable encoding in RA:

Differences in syllable phonology as a result of sequencing are proportional to differences in neural activity in RA.

There are multiple ways in which RA may encode the same syllable found in different contexts, but we predict that syllable structure has the largest influence upon patterns of activity in RA. This hypothesis will be addressed by comparing the similarity of syllables to the similarity of the underlying neural activity in RA.

CHAPTER 3 HYPOTHESIS

A prevailing hypothesis about the role of motor variability in adult behaviors is that it serves to make fine scale adjustments to a behavior. There is measurable variability in the adult songs of birds, and it is a likely that a portion of that variability is driven by the avian CNS. The overall amount of variability in lMAN has been correlated with song variability, but there has yet to be a direct relationship identified between specific changes in brain activity and specific changes in the song. As mentioned earlier, RA neurons are the most likely cells in the avian CNS to be correlated with song structure. The third chapter is focused on the following hypothesis about RA:

> The measurable variability in pitch, amplitude, and entropy of syllables is correlated with specific changes in the firing properties of RA neurons.

In order to address this issue, we made chronic neural recordings in the RA of adult Bengalese finches during singing. On each successive iteration of a syllable, the overall amount of activity in RA was compared to specific changes in the spectral properties of syllables. Through this experiment, we will attempt to identify a direct relationship between changes in brain activity and changes in song.

Through the exploration of the three hypotheses outlined above, we hope to elucidate the role RA plays in generating song structure. These results will fill be informative on how lMAN might influence patterns of activity in RA to drive specific

changes in song. Although these conclusions will be primarily applicable to the

maintenance of adult song, they will also be informative about how the song system can

generally make changes to song throughout a bird's life.

CHAPTER 1

THE EFFECTS OF VARIABLE SEQUENCING UPON SYLLABLE STRUCTURE

Abstract

For both human language and the songs of passerine birds, vocal complexity is generated through variable sequencing of shorter, more elemental vocalizations. The motor planning for this task involves controlling the muscles for each elemental unit (phonemes for humans, and syllables for birds), as well as the program for sequencing these elemental units in multiple ways. There are two general varieties of vocal sequence variability in bird song: convergence points, where at least two different syllables precede the same syllable; and divergence points, where at least two different syllables follow the same syllable. For both convergence points and divergence points, the birds learn how to vocalize the same syllable in multiple motor contexts. We are interested in the effects of variable sequencing upon syllable structure, and what these effects might tell us about the control of syllable production. We found that syllable phonology was modified on two different time-scales: at convergence points, the production of the common syllables were modified by the structure of the immediately preceding syllables; and for both convergence and divergence points, we found that syllable structure covaried with song structure on an extended time-scale. The short-term effects upon convergent syllables may be a result of constraints of the song system, whereas the more temporally extended correlations between syllable phonology and sequence suggest that these song features are not completely independent during song production.

Introduction

The muscle control for complex motor behaviors occurs on two different time-scales: the immediate control of muscle position for each discrete movement, as well as the sequencing of these elemental acts into longer, more complicated motor patterns. A classic example of such a behavior is human speech. Our languages are composed of individual motor units, or phonemes, that are variably sequenced in order to produce complicated arrangements of vocalizations. The production of human speech requires planning for each individual phoneme (the basic unit of speech), as well as the sequencing of phonemes into words and sentences (Bishop, 1999). In an effort to understand how variably sequenced vocal behaviors are generated, we have analyzed the songs of the Bengalese finch (*Lonchura domestica*). In a fashion similar to human language, Bengalese finches learn their songs from other individuals (Catchpole and Slater, 1995), and these learned vocalizations involved shorter vocal elements, or syllables, arranged into variable sequences (Okanoya 2004).

There are two different varieties of sequence variability in Bengalese finch song: convergence points and divergence points. These are places in song that can be considered transition points because syllable sequencing can follow multiple trajectories. A convergence point (or convergent syllable) is defined as a syllable that can be preceded by at least two different syllables. The term convergence is used because the vocal musculature is transitioning to the same position from multiple, different positions. Conversely, divergence points (or divergent syllables) are song syllables that can be followed by at least two different syllables, reflecting a change in muscle position starting from a single, common position. For human language, similarly sequenced vocalizations

14

exhibit differences in phonation. There are left-to-right effects, where past vocalizations alter the current one (convergence); and right-to-left effects, where future vocalizations influence the ongoing utterance (divergence) (Daniloff and Hammarberg, 1973). We are interested in whether convergent and divergent syllables in Bengalese finch song are similarly modified by adjacent vocalizations.

Not only are we interested in the differences adjacent syllables may cause, it is also important to understand how patterns in the sequencing of syllables affect their production. Throughout Bengalese finch song, some syllable sequences are more common than others at various time points (Figure 1.1a). This results in a bias for the production of one particular convergent or divergent sequence over another at different time points in song. An unanswered question is whether these biases in sequencing affect the phonology of syllables. If differences in sequencing do drive consistent and extended (across multiple syllables) effects upon syllable structure, it suggests that the production of these sequences may involve different motor plans. In order to provide a more complete description of song control, it is important to disambiguate any effects of adjacent syllables from those that may result from temporally extended biases in syllable sequencing.

The songs of the Bengalese finch exhibit variable sequencing of shorter, more elementary vocalizations. The songs of this species therefore require planning on two different time-scales: for phonology and sequence. Knowledge about the interaction of phonology and sequence during song production may provide insight into the hierarchical

structuring of a Bengalese finch song. In previous work on the songs of Bengalese finches, sequence and phonology were assumed to be independent entities (Nakamura and Okanoya, 2008; Sakata and Brainard, 2006). Sequencing information is generated in HVC (Hahnloser et al, 2002), while phonology is thought to be controlled by the robust nucleus of the arcopallium (RA, Leonard and Fee, 2005). These data suggest that the same syllable found in different sequences would have very similar patterns of RA activity, but different HVC activity. If convergent or divergent sequences exhibit differences in production, the source of those differences may be peripheral, central, or both. Understanding the nature of sequence effects upon phonology on both short and long time-scales can serve to appropriately orient research on the neural control of variably sequenced vocalizations.

Methods

The songs of 14 adult male Bengalese finches (*Lonchura domestica*), ages 4 months to 2 years old, were recorded (labeled BF1 through BF14 in the text). Omni-directional, lavalier microphones (Countryman™) were used to record the songs, after which they were bandpass filtered between 50 and 10000 Hz before being digitized at 32 kHz (National Instruments™ BNC-2090 A/D board). Customized acquisition software (LabView™) was used for identifying and saving songs. Many songs over the course of many weeks were recorded, but a random selection of 40 songs from each bird was used for the final analysis. These songs were taken from multiple days and always from morning hours in order to avoid any diurnal effects upon syllable phonology and sequencing (before 12:00 PM, which is 5 hours after the lights are turned on in a 14 hour day).

Once the 40 songs were randomly selected from the entire data set, they were imported into Matlab™ 7.1 for analysis. Songs were segmented into syllables using both amplitude and temporal thresholds. Syllables were then visually labeled using a different letter for each unique syllable in the bird's songs, and eight different spectro-temporal measurements were made for each syllable. The eight measurements used were: duration, pitch, time to half-peak amplitude, frequency slope, amplitude slope, entropy of the spectral density, entropy of loudness versus time, and entropy of the full spectrum.

Absolute measurements were made for duration (milliseconds), pitch (hertz), time to half-peak amplitude (milliseconds), and frequency slope (hertz/millisecond). Pitch was measured by performing an autocorrelation of the analog amplitude trace. The peak offset of the autocorrelation vector was then divided by the sampling rate in order to calculate the fundamental frequency. For syllables with flat frequency profiles, the pitch measurement was made over a 16 millisecond window centered on the middle of the syllable. For frequency modulated syllables, the pitch measurement was made across a 16 millisecond window starting 5 milliseconds after onset of the syllable. The fundamental frequency (equivalent to the spacing of harmonics) was used as the pitch measurement. The time to half-peak amplitude measurement was made in milliseconds by determining when the volume had reached half its maximum over the course of its duration. Frequency slope was calculated through a number of steps. First, a fast-fourier transformation was performed on the syllable oscillogram, producing a spectrogram of the syllable. A cross-correlation was successively performed between 1 millisecond bins, separated by 3 milliseconds, across the time axis of the spectrogram. The slope of a regression line was calculated for the cumulative sum of the peaks of the cross correlation across measurement time. This slope was then used as the frequency slope measurement.

Unlike the above measurements, amplitude slope and the three entropy features, measurements were normalized across all birds to values between 0 and 100. For amplitude slope, a value of 0 equates to a completely flat amplitude profile, while a value of 100 assigned to a syllable with maximal temporal modulation. Amplitude slope was first calculated by dividing the spectrogram of each syllable into two halves (across time).

The sum of the absolute value of each half was then calculated. Using the sum of each

half of the syllable, amplitude slope was calculated as follows:

Amplitude Slope = 50 + 50*((sum1-sum2)/(sum1+sum2))

The entropy of the spectral density measurement was also scaled between 0 and 100.

Higher entropy values represent noisy syllables, and lower values are those syllables with

only one defined peak. Entropy of the spectral density is calculated as follows:

Entropy S.D. = 100*sum(spec*log2(spec))/(log2(length(spec))

where *spec* represents the absolute value of the power at each frequency (below 8000 Hz)

collapsed across all time bins. Entropy of loudness over time describes how many peaks

the oscillogram contains, with low values for one very sharp peak, and high values for

flat amplitude profiles. Loudness was calculated by first smoothing the oscillogram of a

syllable with a 5 millisecond Gaussian window. These values were then normalized by

dividing by the maximum of the smoothed waveform. The entropy of loudness over time

was then calculated as follows:

Entropy L v. T = -100*sum(loud.*log2(loud))/log2(length(loud))

where *loud* is the previously calculated loudness waveform. Lastly, the entropy of the

full spectrum measurement was used to estimate the consistency of spectro-temporal

features of a syllable from one time bin to the next. Tonal and harmonic syllables have

values closer to zero, and noisiest syllables have values closer to 100. Entropy of the full

spectrum was calculated by first measuring the absolute value of every point in the

spectrogram, and then vectorizing the values. The values in this vector were sorted from

lowest to highest, and then normalized by dividing by the maximum value. The entropy

of the full spectrum was then calculated by the following equation:

$$\text{Entropy F.S.} = 100*(-\text{sum}(sort\_spec*\log2(sort\_spec))/$$

$$\log2(\text{length}(sort\_spec))$$

where *sort_spec* is the vectored, sorted and normalized spectrogram.

These eight spectro-temporal measurements were made for every syllable from the 40 songs of each bird. There were 113 different syllables identified in the songs of the birds recorded, with 41995 total syllables used for analysis. Bengalese finches usually start their songs with introductory syllables that are low amplitude, noisy syllables. Introductory syllables are also some of the more variable syllables, and skewed the results of the principal components analysis discussed below (analysis not shown). Because of our interest specifically in the sequencing of *song* syllables, as well as the variable nature of introductory syllables, introductory syllables were excluded from our analysis.

Once these measurements were made for all song syllables, a z-score transformation was performed across all syllables, for all birds, within each different measurement. This was done to normalize the values for all measurements in order to weigh relative differences in each parameter more heavily than absolute differences. A principal components analysis (PCA) was then performed on all z-score values from all birds. The PCA was performed on the entire data set for two reasons: first, we wanted to ensure that any differences in recording technique within each bird were not a factor in differentiating syllables; and second, we wanted to summarize the results from all birds, and performing a PCA within each bird would prevent us from combining data across

individual birds. After the PCA was calculated, we plotted the variance explained versus number of principal components in order to determine how many components to use for our final analysis. We found that the first six principal components explained 95% of the variability in the data set (Appendix 1, Figure 2b), and therefore used only these six principal components for subsequent analyses.

The distance between syllables in PCA space was used as a measure of similarity. As mentioned previously, only the first six principal components were used. We used the Euclidean distance between the center of mass of two syllables in six-dimensional space as our similarity measurement (referred to as COM distance in the rest of the paper). Syllables that are more spectrally different have larger COM values than those that are more similar to each other. We are most interested in how, on average, syllable structure changes depending on sequence. Because of this, we looked at the mean changes in syllable production, and therefore used COM distance rather than a measure like Mahalanobis distance, which also takes into account the variability of the clusters. We used the COM distance in two ways: first, it was used as a method to quantitatively justify our syllable-labeling scheme; and secondly, it was used as a way to judge the similarity of two different syllables.

To confirm the accuracy of our syllable labeling with respect to spectro-temporal feature calculations, we measured the COM distances between every syllable in a bird's repertoire. Distances were calculated between different iterations of the same syllable in the same sequence, the same syllable within different sequences, and completely different

syllables.   Plotting all syllable distances from all birds yielded a bimodal, but continuous distribution of values (Supplemental Figure 1.3a). If the values did in fact fall into two, separate distributions, it would be evidence that COM distances for syllables with the same label were non-overlapping with those between differently labeled syllables. In order to determine the extent of overlap, values for syllables in each category were plotted separately (Appendix 1, Figure 1.3b; same syllable (SAME), different sequence (SEQ), and different label (DIF) in green, blue and black, respectively). The values for the same syllable in the same sequence and differently labeled syllables are non-overlapping, but there is a small amount of overlap in the COM values for the same syllable in different sequences and differently labeled syllables (red highlighted area in Appendix 1, Figure 1.3b). This overlap is most likely caused by our labeling methodology. Syllables are labeled within each bird, and then the values from all birds were combined for this analysis. As a result, two different syllables labeled within one bird may be more similar to each other than a highly variable single syllable in another bird.

In order to determine if combining data across birds was contributing to the overlap of SEQ and DIF syllable values, we developed a system of ranking syllable similarity within each bird. The question we wished to answer was how often two SEQ syllables within a single bird's repertoire are quantitatively more similar to each other than they are to any other syllable. If they were more similar to each other than to any other labeled syllable, they would receive a rank of one (1); but, if one other syllable was more similar to either sequentially unique syllable, it would be ranked as a two (2). A

rank of three (3) results when two other syllables are more similar to the SEQ syllable, and then a rank of four (4), five (5), and so on for each additional syllable that is more similar. This analysis was performed for all syllables across all birds, and the results are summarized in Appendix 1, Figure 1.3c. Approximately 90% of the same syllables found in multiple sequences were more similar to each other than to any other syllable, but there were some sequentially different syllables that received ranks of two (10), and four (1). We wanted to determine if our original labeling was inaccurate, or if our PCA analysis of syllable similarity was imprecise. In the majority of greater than one-ranked syllables (7 out of 11), the quantification of syllable features fell short of our visual labeling method. The syllables where our PCA approach failed were generally low amplitude, noisy syllables. As a result, we took the conservative approach and excluded these syllables from the analyses discussed below.

As mentioned above, we also used COM distances as a way to measure the similarity of two syllables. This measurement was made for syllables with the same label in the same sequence, the same syllable in different sequences, and for two differently labeled syllables. For the same syllable found in the same sequence, the entire data set was randomly split into two groups, and the distance between each half was calculated. This was repeated 1000 times, and the mean of all 1000 comparisons was used for analysis. The differences in COM value distributions for these three syllable types were then tested for significance using a Kolmogorov-Smirnov test. COM distances were also calculated between syllables adjacent to transition points in song. For convergent sequences, these are the syllables before the convergent syllable, and for divergent

sequences, these are the syllables after divergence. Correlation coefficients were then

calculated between the convergent/divergent syllable and the adjacent syllables in order

to determine if any significant relationship existed between them.

In order to determine the temporal relationship of syllable similarity and song

sequencing, we calculated the similarities of the syllables after convergence and before

divergence and measured any trends between syllable similarity and the distance from the

transition point. We also wanted to investigate how syllable production varied over even

longer time-scales. We looked at this issue by measuring changes to syllable structure

depending upon the *next* transition. In the example shown in Figure 1.7, consider a

syllable A that can precede either syllable B or syllable F. In this bird's song, there are

instances where an A-B transition is followed by another A-B, but there are also

instances where it is followed by an A-F transition. We compared the distance between

the A's in the sequence A-B, when the *next* syllable A transition is A-B or A-F (**A**b-ab

versus **A**b-af). These syllables are termed successive transition syllables, and the analysis

was performed for both convergence and divergence points. Two control analyses were

used to put upper and lower boundaries on the COM distances for successive transition

syllables. The lower bound was calculated by randomly selecting two groups of syllable

A iterations in sequence A-B and calculating the distance between them (SS control).

Because of the random selection, the groups being compared will contain syllable A in

both **A**b-ab and **A**b-af sequences. This was done 1000 times, and a mean distance was

derived from those trials. This analysis estimates how likely the successive transition

syllables distances are within the pool of all syllable A's in sequence A-B. The upper

boundary control was defined as the distance between syllable A's in each convergent or divergent sequence (i.e., **A**-b versus **A**-f). All of these comparisons were paired, and a Wilcoxon signed-rank test was therefore used to test for significance between each category of syllables.

We also looked at the differences in convergent point similarity before and after deafening. The songs of four birds from previous experiments were analyzed. Forty songs before deafening were compared to forty songs after deafening. Convergence points were identified in the pre-deafening songs, and only those convergence points were analyzed in the post-deafening songs. All of the post-deafening songs used were recorded within one week of the deafening surgery to limit any long-term effects of deafening. These again were paired comparisons, and a Wilcoxon signed-rank test was used to test for any significant effects of deafening.

Results

Bengalese finch song is composed of acoustically continuous segments (50 to 100 milliseconds) that are termed 'syllables', surrounded by short (~10 milliseconds) periods of silence (see song of BF2 in Figure 1.1). Song syllables are labeled with a different letter for each discreet syllable in the bird's song.
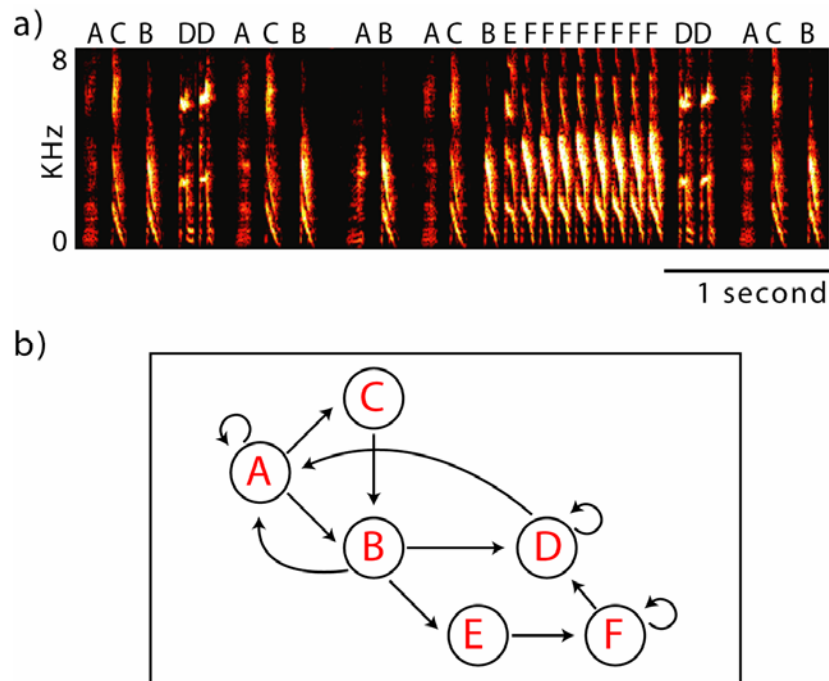
Figure 1.1: (a) Spectrogram of the song of BF1. Syllables are visually labeled with unique letters. Note that some syllables are found embedded in multiple sequences.(b) syllable transition diagram for the song shown in A and B. The directions of the arrows represent possible transitions from one syllable to another. For instance, syllable A may follow syllables C, B, or A. These transitions can be seen in the labeled song in Figure 1B.

The song shown in Figure 1.1 contains six (6) different syllables, labeled: A, B, C, D, E, and F. Syllable A can be found in multiple sequences of syllables: syllable A is repeated several times in a row, as well as sung before syllables C and B. The complex sequencing of Bengalese finch song can be more readily appreciated through the

construction of a "transition diagram" (Figure 1.1b). In the transition diagram for the song shown, it is apparent that the sequencing of Bengalese finch song is not linear: the same syllable is sung in multiple sequences of syllables. The example song has four different transition points (syllables A, B, D and F), or places in song where the sequencing is probabilistic rather than deterministic. This behavioral feature is ubiquitous across Bengalese finches, with all recorded birds (14 out of 14) having at least one syllable used in multiple sequences.

JUSTIFICATION OF SYLLABLE SIMILARITY MEASUREMENTS

Unique syllables were initially identified visually, but it is important to show that our visual labeling scheme can be justified quantitatively. We measured eight different acoustic features for each syllable as a way of assessing similarity. A PCA was performed on these eight spectral measurements for each syllable sung in forty songs collected from each bird (41995 syllables in 479 songs from 14 birds). For three example syllables from BF11, average spectrograms and smoothed, rectified amplitude waveforms are shown in Figures 1.2a through 1.2c. Examples of the clustering of these syllables by PCA are shown in Figures 1.2d through 1.2e; shown are ellipsoids, centered at the mean of the first three principal components, with radii that are one standard deviation. By plotting the first three principal components of each syllable, we can compare the clustering of syllables by PCA to our initial visual labeling of syllables. The ten syllable clouds segregate in PCA space, supporting the original categorization of syllables by visual labeling.
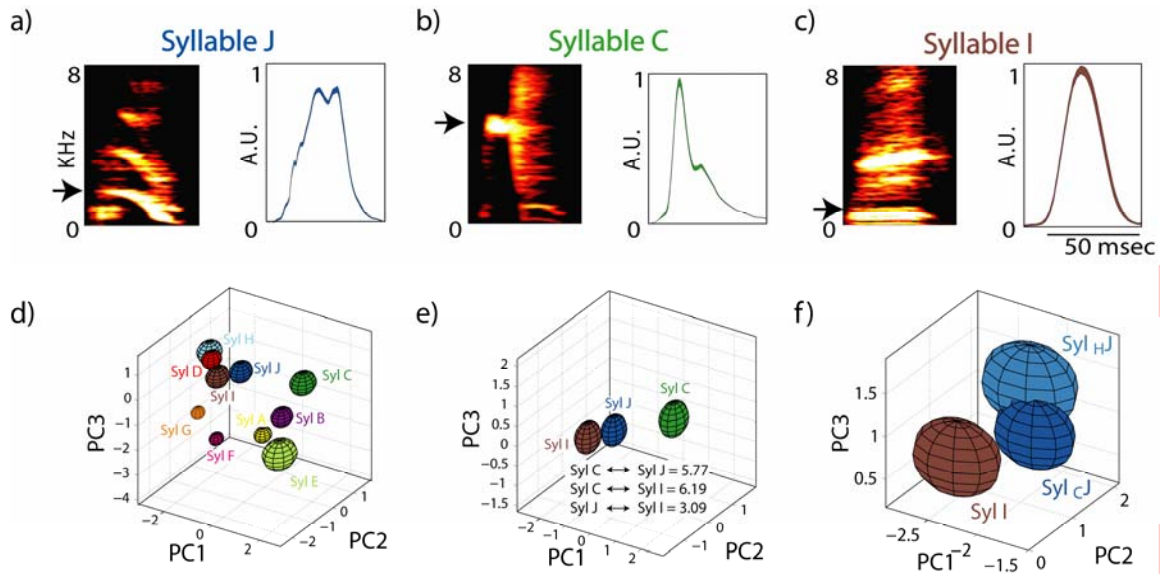
Figure 1.2 (a) spectrogram and smoothed, rectified amplitude waveforms for syllable J of BF12. Black arrow marks measured pitch of syllable. (b) spectrogram and smoothed, rectified amplitude waveforms for syllable C of BF12. Black arrow marks measured pitch of syllable. (c) spectrogram and smoothed, rectified amplitude waveforms for syllable I of BF12. Black arrow marks measured pitch of syllable. (d) example of syllable clustering after PCA. Potted are mean +/- one standard deviation ellipsoids for the first three principal components of ten differently labeled syllables in the song of BF12. (e) center of mass distances (COM) between syllable clouds. Syllables J, C, and I are all distinct syllables from the song of BF12. The COM distance between syllable J and C is larger than it is from syllable J to I (COM = 5.77 and 3.09, respectively). These COM distances are then used as a measure of syllable similarity, with lower values for more similar syllables. (f) clusters for the same syllable in different sequences are closer than for differently labeled syllables. Syllable J of BF12 is found after syllables C (dark blue ellipsoid) and H (lighter blue ellipsoid). The clusters for syllable J in each sequence are overlapping, whereas clusters for syllables with different labels are non-overlapping (e.g. syllable I in brown)

Calculating the distances between clusters in PCA space provides a quantitative measure of syllable identity. We have used the center-of-mass distance (COM distance) between the means as a measure of syllable similarity. In our data, the distances are smaller for 'spectrally similar' syllables, whereas more disparate syllable types have greater distance measures. Shown in Figure 1.2e are the syllable clusters and COM distances for the three example syllables of BF11 shown previously (Figure 1.2a, b, and c). Our PCA of syllable features places syllable J further from syllable C than it does from syllable I (J-to-I COM distance is 3.09, and J-to-C distance is 5.77).

28

We systematically analyzed each labeled syllable and determined quantitatively whether it was the most appropriate label for that particular syllable (Appendix 1, Figure 1.3 and accompanying text). In ninety percent of the cases, the original visual labeling system was substantiated by our quantification of spectral structure. For the other 10%, there was an unresolved discrepancy between the visual and quantitative labeling of syllables. Because of the inherent variability of syllable production, it is difficult to find a perfect measure of syllable similarity. In order to avoid any shortcomings of our quantification, syllables where visual and quantitative labeling did not match were not included in the final analysis.

SIMILARITY OF CONVERGENT AND DIVERGENT SYLLABLES

There are two main varieties of sequence variation in the songs of Bengalese finches: sequence convergence and sequence divergence. Although we have labeled syllables with the same letter when found in different sequences, it is important that we quantitatively demonstrate that these syllables are indeed the same acoustic output. In Figure 1.2f is an example of the clusters for a syllable found in two different sequences. Shown are the syllable clouds for syllable J in the sequences H-J (in light blue) and C-J (in dark blue), as well as the cloud for the next most similar syllable, syllable I (brown). The two ellipsoids (mean +/- one S.D.) for syllable J are overlapping, but they are both distinct from syllable I. The COM distances provide quantitative evidence for this outcome: the distance between $_H$J and $_C$J is only 0.75, whereas the COM distance between either example of syllable J and syllable I is approximately 3.10. This is just one

29

of many examples demonstrating the similarity of convergent and divergent syllables as compared to differently labeled syllables. While there are subtle differences in the production of syllables found in different sequences, these differences are not large enough for them to be classified as categorically unique.

As was shown above, when the same syllable is found in different sequences, there are subtle but significant differences in its production. In Figure 1.3 are two example syllables that demonstrate this phenomenon.
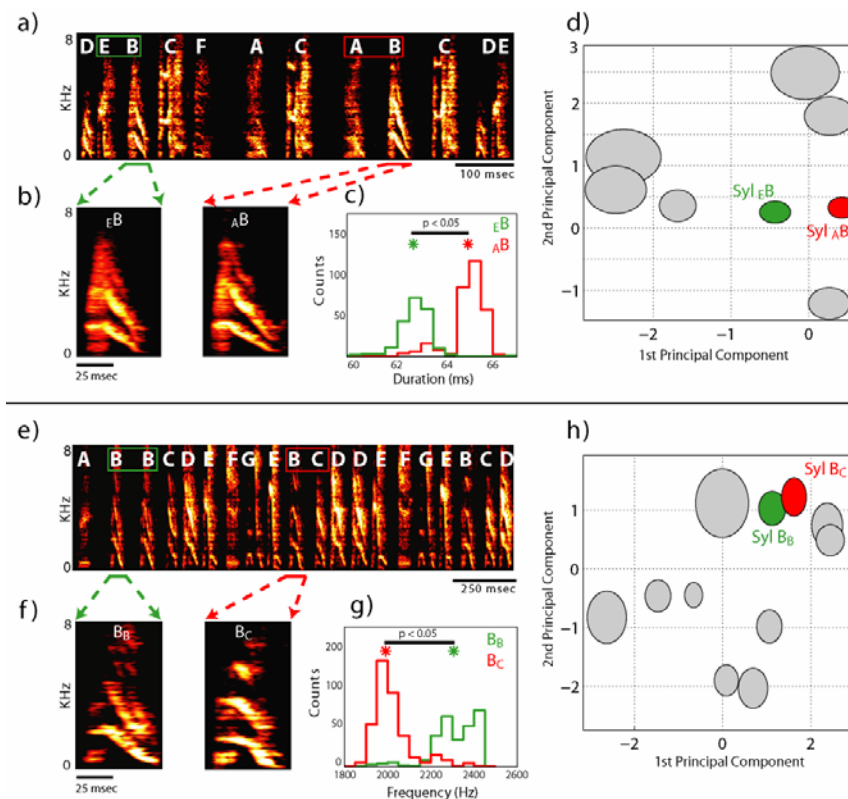
Figure 1.3: (a) Spectrogram of approximately 1.5 seconds of song, with syllable labels in white. The convergence on syllable B is highlighted: E-B in green, and A-B in red. (b) Average spectrograms of 10 examples of syllable B in each sequence, note the shorter duration of syllable B when it follows E. (c) Histogram of the durations of approximately 150 renditions of syllable B in each sequence (colors as in the rest of the figure). (d) Mean +/- one standard deviation ellipses for the first two principal components of syllable B found in each sequence, as well as all other syllables (in gray) for comparison from the bird's repertoire. (e) Spectrogram of approximately 1.5 seconds of song, with syllable labels in white. The divergence from syllable B is highlighted: B-B in green, and B-C in red. (f) Average spectrograms of 10 examples of syllable B in each sequence, note the higher frequency of syllable B when it precedes another B. (g) Histogram of the

Figure 1.3 continued pitch of syllable B in each sequence (colors as in the rest of the figure). (h) Mean +/- one standard deviation ellipses for the first two principal components of syllable B found in each sequence, as well as all other syllables (in gray) for comparison from the bird's repertoire.

The first instance is from the song of BF8 (Figure 1.3a), and it illustrates the two convergent sequences for syllable B: E-B and A-B. The duration of syllable B is different depending upon whether it follows an E or an A: cases of B following an E are significantly shorter than if they follow an A (average spectrograms in Figures 1.3b, quantification of duration in Figure 1.3c). Although the length of syllable B is different depending upon the sequence, the two instances of syllable B are still more similar to each other than to any other syllable in the bird's repertoire. Shown in Figure 1.3d are the mean +/- 1 S.D. ellipses ($_E$B in green, $_A$B in red) for the first two principal components of each rendition of syllable B, as well as those for the other syllables in the bird's repertoire (in grey).

Spectral modifications can also go in the reverse direction (right-to-left). The song of BF10 is shown in Figure 1.3e. This bird's song contains the divergent syllable B, which precedes either another B, or a C. For this divergence point, an alteration in pitch is observed between the two sequences: syllable B before a C is lower in pitch than it is before syllable B (average spectrograms in Figure 1.3f, quantification of pitch in Figure 1.3g). As in the above example, there are significant differences in one particular syllable feature (pitch), but if many spectral features are measured, the two examples of syllable B in each sequence are quantitatively more similar to each other than they are to another syllable in the bird's repertoire. Shown in Figure 1.3h are the mean +/- 1 S.D. ellipses for the PCA values for syllables B_B and B_C, as well as for the other syllables in

the bird's songs. The ellipses for syllable B in each sequence are overlapping and distinct from the PCA values for other syllables. The examples for convergent and divergent syllables demonstrate the efficacy of using PCA to measure syllable similarity, and the degree of similarity between these syllables as compared to differently labeled syllables.

When the same syllable is found in different sequences, we found subtle but significant differences in production. We wanted to determine how large these differences are with respect to the underlying variability of song production (Figure 1.4).

Figure 1.4: (a) Example spectrogram of a song from BF3, with the syllable labels along the x-axis. Highlighted in green, red, blue, and black, respectively, are the four different relationships among syllables: same syllable same sequence (SS), divergent syllable (DIV), convergent syllable (CONV), and different syllables (DIF). (b) Cumulative

In order to do so, it was necessary to analytically determine the upper and lower boundaries on COM distances for our similarity analysis. The lower boundary is designed to estimate the maximum *similarity* between two syllables. Data from the same syllable in the same sequence was randomly split into two groups, and the COM distances between these random groups were calculated. From this analysis, it is possible to put a lower limit on the COM distances between two syllables (Figure 1.4b, green line). The upper boundary was used to estimate the maximum *difference* between two syllables (or the upper limit on COM distance). The upper boundary was defined as the COM distances between syllables that were labeled with different letters (Figure 1.4b, black line). With these two analyses, we can determine how significant the production differences are for convergent and divergent syllables.

The COM distances for the same syllable in the same sequence, and for differently labeled syllables, were compared to those for syllables at convergent and divergent points in song (example song in Figure 1.4a). COM distance values for convergent and divergent syllables are significantly greater than those of same syllables found in the same sequences (Figure 4b, both are p < 0.0001, KS-test), and are significantly less than the distances between differently labeled syllables (Figure 1.4b, both are p < 0.0001, KS-test). Interestingly, the COM distances for convergent syllables are greater than those between divergent syllables (Figure 1.4c, p < 0.003, KS-test).

33

These results demonstrate that sequence can cause subtle but consistent effects upon phonation.

DETERMINING THE CAUSES OF CONVERGENT AND DIVERGENT SYLLABLE DIFFERENCES

In an effort to understand the causes of spectral differences between sequences, we asked whether the local motor context is predictive of convergent and divergent syllable similarity. In terms of left-to-right effects at convergence points, the prediction is that greater differences between the prior syllables will result in a greater difference between the convergent syllables. Similarly, at divergence points, sequencing effects should cause divergent syllables to be less similar when transitioning to more similar syllables. We analyzed both convergent and divergent syllables for any effects of adjacent syllable similarity (Figure 1.5). For convergence points, we compared the similarity of the two syllables preceding convergence to the similarity of the two convergent syllables; and for divergent sequences, we compared the similarity of the two syllables following divergence to the similarity of the two divergent syllables.
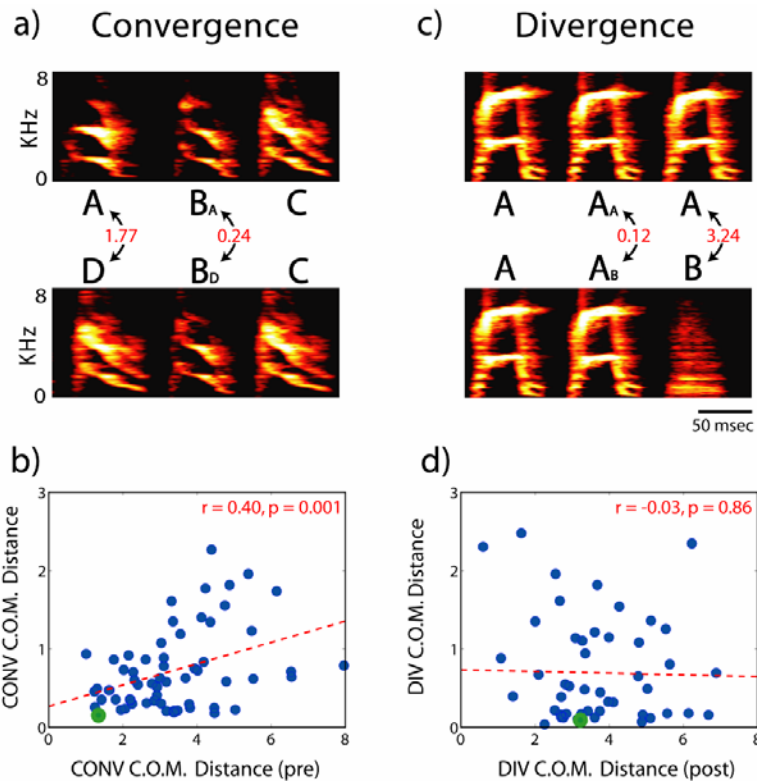
Figure 1.5: (a) Demonstration of how pre-note similarity is compared to convergence point similarity. Syllable B is found after either syllable A or D in the songs of BF6. The COM distance between syllables A and D is compared to the COM distances between B_A and B_D. This data point is in green in the 6b. (b) Pre-note similarity and its effects upon convergence point similarity. There is a significant, positive relationship between pre-note similarity and convergence point similarity ($r = 0.40$, p 0.001). (c) Demonstration of how post-note similarity is compared to divergence point similarity. Syllable A is found preceding either syllable D or E in the songs of BF13. The COM distance between syllables A_B and A_A is compared to the COM distances between A and B. This data point is in green in the 6d. (d) Post-note similarity and its effects upon divergence point similarity. There is a no measured relationship between post-note similarity and divergence point similarity ($r = -0.03$, p 0.86).

Our data show that the distance between convergent syllables is correlated with the similarity of the two preceding syllables (Figure 1.5b). We find a significant, positive relationship between the COM distance of the two syllables occurring before convergence, and the COM distance of the following convergent syllables (Figure 1.5b, r = 0.40, p = 0.001). Although there are differences in structure for divergence points, the

35

predictive relationship found for convergent syllables was not found for divergent

syllables (Figure 1.5d).

We have found that syllables immediately adjacent to transition points in song

(after convergence and before divergence) have differences in production. In an effort to

determine the temporal extent of sequence effects upon syllable structure, we analyzed

syllable similarity multiple positions away from transition points (i.e. the syllables

occurring after convergence, and those before divergence, see Figure 1.6a for schematic).

For convergent sequences, we found that increases in the distance from convergence

corresponded with increases in syllable similarity (Figure 1.6b, blue line). There is a

significant decrease in COM distances from the first syllable after convergence to the

second ($p = 0.003$, Wilcoxon signed-rank test), but no such decrease in similarity was

found between the second to the third syllable ($p = 0.56$, Wilcoxon signed-rank test).

For divergence points, distance from divergence did not correlate with any change in

syllable similarity; the COM distances were statistically indistinguishable for syllables

three positions leading up to divergence ($p > 0.25$ for all combinations, Figure 1.6b, red

line).
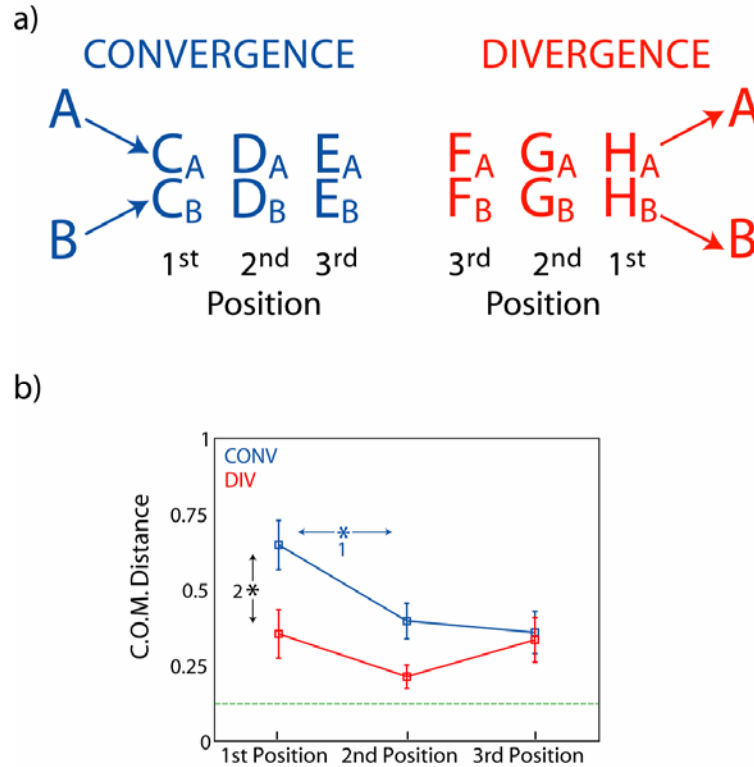
Figure 1.6: (a) Schematic of what is meant by 1st, 2nd and 3rd positions after convergence (blue) and before divergence (red). (b) Extent of effects upon syllable phonology at convergence and divergence points. Shown are mean +/- standard error values for syllables one, two, three positions removed from transition points (convergence, notes following transition; divergence, notes before transition). Significant differences in C.O.M. distances are numbered (1: $p = 0.003$, Wilcoxon signed-rank test; 2: $p = 0.008$, all KS-test). The green-dotted line denotes the mean COM distance value between SS syllables.

Although COM distances were similar for convergent and divergent syllables multiple positions away from the transition point ($p = 0.81$, KS-test), they were still significantly greater than distances between control syllables (same syllable, same sequence; Figure 1.6b, green line, $p < 0.0001$ for convergent and divergent syllables, KS-test). The residual differences in structure for syllables multiple positions away from the transition suggests that the motor program for song might also vary over a longer time-scale around convergence and divergence points. Adjacent syllable effects gradually diminish for convergence points, but the syllables never fully *converged*. From these data, we can say

that the local effects of adjacent syllable structure at convergence points diminish with distance from the transition point, but that syllables maintain some phonological discrepancies that seem to be driven by more temporally extended song features.

EXTENDED PATTERNS IN SYLLABLE SEQUENCING AND PHONOLOGY

The results above (Figure 1.6) suggest that phonology and sequence might co-vary over extended time-scales. In order to evaluate the relationship between syllable structure and sequence over longer time-scales, we looked at the effects of successive instances of convergence and divergence points upon syllable phonology. Our reasoning is that if syllable production and sequence are correlated with one another, extended patterns in the sequencing of syllables will affect their phonology. As an example, a segment of song is displayed in Figure 1.7a. The convergent syllable D is shown in two different sequences: C-D and G-D; and the divergent syllable A is shown in two sequences: A-B and A-F. For convergence, we asked if production of syllable D in the sequence C-D was different depending upon whether the next syllable D transition was C-D or G-D (Figure 1.7a, blue highlighted syllables). An example for divergence is also displayed in Figure 1.7a. Syllable A is found in the divergent sequences of A-B and A-F. We are interested in whether syllable A in the sequence A-B varied depending upon whether the next syllable A transition was A-B or A-F (Figure 1.7a, red highlighted syllables).
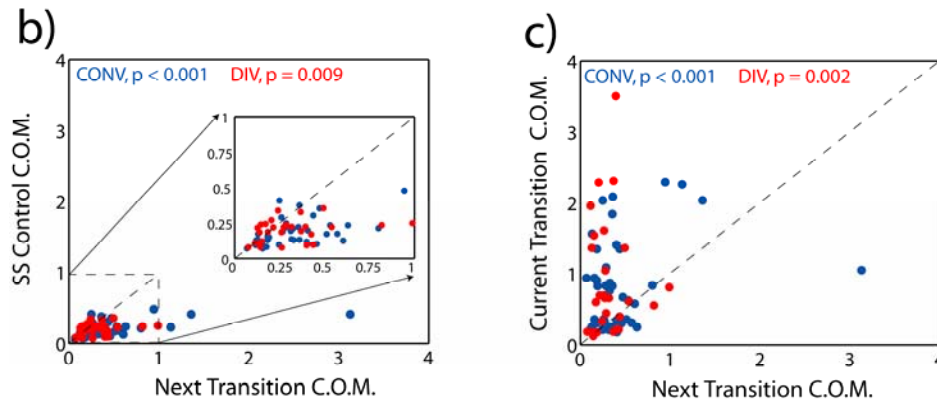
Figure 1.7: (a) Schematic of how successive sequence syllables are defined. In red are two different DIV syllable transitions (A-B and A-F) separated by several other syllables. This is an example of a successive divergence relationship between syllables. In blue is an example of two successive convergence syllables: two convergence points involving the same convergent syllable separated by other unrelated syllables. (b) Successive sequence similarity compared to SS control values in a pair-wise fashion. Convergent and divergent successive sequence syllable similarities were significantly greater than paired SS control values (convergence, $p < 0.001$; divergence $p = 0.009$, Wilcoxon signed-rank test). Inset is a zoom for COM values less than 1. (c) Successive sequence similarity compared to convergent and divergent syllable control values in a pair-wise fashion. Convergent and divergent successive sequence syllable similarities were significantly less than paired convergent/divergent syllable values ($p < 0.001$ for convergence, and $p = 0.002$ for divergence, Wilcoxon signed-rank test).

We call these syllables 'successive transition points' (successive convergent syllables or successive divergent syllables), and they were analyzed for the temporally extended correlations between sequence and phonology discussed above.

Two analyses were performed to estimate upper and lower bounds on COM distances for successive transition syllables. A random sampling method was used to determine a lower boundary (same syllable, same sequence control, or SS Control), and

COM distances from convergent and divergent syllables were used as an upper boundary. We find that for both convergence and divergence points, the distances between successive transition syllables are larger than the same syllable, same sequence control distances (Figure 1.7b, $p < 0.0001$ for both, Wilcoxon signed-rank test); and that they are smaller than distances between convergent and divergent syllables (Figure 1.7c, $p < 0.001$ for convergence, and $p = 0.004$ for divergence, Wilcoxon signed-rank test). These results are consistent with the hypothesis that long time-scale correlations exist between syllable phonology and sequence transition. If no extended correlation existed between phonology and sequence, syllables in transition points would not vary depending upon future transitions, and would be as similar as the same syllable found in the same sequence. Instead, we have found that phonology does indeed fluctuate depending long-range patterns in sequence.

The results above on successive transition points provide more insight into some of the previously discussed results. The successive transition data lends further support to the claim that the component of syllable similarity gained between the 1st and 2nd positions of convergent sequences (Figure 1.6b, blue line) is caused by a change in the current motor history, not the more extended relationship between syllables and sequence. With regards to the degree of dissimilarity in COM values for divergent syllables multiple positions away from divergence (Figure 1.6b, red line), the successive transition analysis suggests that this phenomenon is the result of long-range correlations between sequence and phonology, and not caused by differences in the immediate positioning of the vocal musculature. If there had been any significant change in

similarity as the bird approached the divergence point, it would have shown some local modifications are made. The implication of no significant increase or decrease in syllable similarity leading up to divergence is that the differences in syllable structure preceding divergence are instead a result of long-range correlations between sequence and phonology, and not immediate motor planning. And lastly, the findings presented above also explain why residual differences in syllable production remain up to three positions removed from a transition point (Figure 1.6b, 3rd position, red and blue lines). The successive transition results suggest that the bird is on slightly different motor trajectories when singing one sequence versus another. Because of these long-range correlations, syllables multiple positions away from convergence and divergence are altered as a result of song patterning.

The differences in convergent and divergent syllables similarity immediately allow us to disambiguate the short versus long-range effects of sequencing. The average COM distance for convergent syllables is 0.66, divergent syllables are 0.37, and for the same syllable in the same sequence, the average COM distance is 0.12. From these numbers, we have calculated that 54% of the difference in phonation at convergence points is due to adjacent syllable production, and the other 46% is a result of correlations between sequence and phonology.

MOTOR VERSUS AUDITORY CONTRIBUTIONS

Convergence points are different from divergence points in that differences in both prior

motor and auditory history can affect syllable production. We have shown that

convergence point COM values are higher than divergence point values (Figure 1.4c). It

is possible that a portion of convergent syllable differences can also be attributed to

differences in auditory history. Song nuclei in the bird brain have been shown to be

responsive to a bird's own song (Katz and Gurney, 1981; Williams and Nottebohm,

1985; Margoliash, 1986; Doupe and Konishi, 1991; Vicario and Yohay, 1993), and

differences in auditory feedback could affect the descending motor program. In order to

assess the effects of auditory input upon syllable phonation, we analyzed the songs of

Bengalese finches before and after deafening. We assessed the similarity of convergent

syllables in the songs before deafening, and then compared their similarity to the

similarity of the same convergent syllables in post-deafening songs (Figure 1.8).
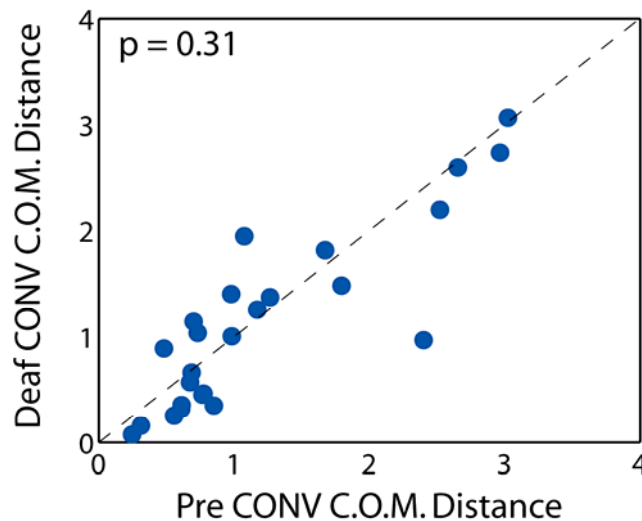
Figure 1.8 C.O.M. distances for convergence points before and after deafening. Data are from 4 birds, no significant difference is found between paired convergence points syllable similarities after before and deafening (p = 0.31, Wilcoxon signed-rank test).

We found no significant change in the similarity of individual convergence points as a result of deafening (p = 0.31, Wilcoxon signed-rank test). These data imply that differences in convergent syllable production are more influenced by differences in motor and/or proprioceptive history than past auditory discrepancies.

Discussion

Our results demonstrate that the production of syllables is modified by events occurring on two different time-scales: the left-to-right effects of immediately adjacent syllables (convergent syllable analysis, Figure 1.5b), and the long-range correlations between syllable sequence and phonology (successive sequence analysis, Figure 1.7). We have also shown how the influence of adjacent syllables gradually decreases as the distance from the transition point increases. Some spectral differences in syllable structure remained well after convergence, however, and these residual differences are most likely the result of the long-range correlation between phonology and sequence. Unlike convergent syllables, divergent syllables were not affected by the structure of adjacent syllables. This finding suggests that syllable production is more associated with the history of the system than with upcoming differences in motor control. Finally, as a result of the limited effects of deafening upon convergence point similarity, we believe motor and/or proprioceptive contributions are more dominant than auditory differences in driving the short-term effects on phonology observed at convergence points.

In regard to differences in divergent and convergent syllable phonation, we would like to be able to disambiguate the short-term effects of adjacent syllables from those that are a result of long time-scale correlations between phonology and sequence. Previous research studying syllable motor control suggests that when the same syllable is found in different sequences, it is treated as the same motor gesture. In brown-headed cowbirds, the control of bronchial airflow and air sac pressure during renditions of similar song

phrases is highly correlated both within and between birds (Allan and Suthers, 1994), and similar results have been found for syringeal control in brown thrashers (Goller and Suthers, 1996). Furthermore, research on mockingbirds has demonstrated that mimicry of another bird's vocalization involves of very similar set of syringeal muscle patterns (Zollinger and Suthers, 2004). These data suggest that the neural activity controlling the musculature for sequentially unique syllables may also be comparable across sequences. Muscle activation patterns used to control syllable structure arise from neural activity in the robust nucleus of the arcopallium (RA), and this activity has been shown to be very similar for the same syllable in different sequences (Leonardo and Fee, 2005). The implications of these findings are that any differences in phonation between sequences for the same syllable are *not* the result of central motor planning, and that there is no relationship between sequence and phonology at the periphery. These results, taken together with those of the current study, indicate the following: first, the short time-scale changes to convergent syllables may not be the result of differences in muscle activation patterns arising from the central nervous system; and secondly, the long-time scale correlations may relate to a shared control between sequence and phonology. The contribution of short and long time-scale affects can be disentangled by comparing the results of convergent and divergent syllables. Convergent syllables are not as similar as divergent syllables (Figure 1.4c), and this is most likely the result of the left-to-right effects of adjacent syllables we have discussed. Divergent syllables have similar motor histories, and any differences between them would not be from differences in prior state. The difference in similarity between convergent and divergent syllables would therefore be a result of the short-term effects of adjacent syllables, and not the long-time scale

correlations between sequence and phonology. Using the mean differences in similarity (Figure 1.6) allowed us to calculate the contribution of each of these effects upon phonology. From this analysis, we conclude that about half of the difference in phonation at convergence points is due to adjacent syllable production, and the other half is a result of correlations between sequence and phonology.

We have shown that syllable structure is modified on two different time-scales: there are long-range correlations between syllable phonology and song patterning, as well as the more local effects of adjacent syllables at convergence points. In order to gain an understanding of the extended relationship between syllable sequence and phonology, we determined the temporal extent of vocal changes in production associated with changes in motor behavior. We found that syllable phonology was correlated with temporally distant syllable transitions. Long-term correlations between phonology and sequence suggest that the 'state' of the bird may have an effect upon syllable production. The term 'state' can be defined as any number of possibilities: differences in neurotransmitter levels could bias one sequence over another and in turn modify syllable structure; the drive of the bird to sing can have far reaching effects; and/or variation in the circuit properties of the song system could cause prolonged differences in syllable production. Previous work on singing behaviors in male birds has shown that songs can be modified on an extended time basis. The differences in song structure when males sing to a female (directed song) rather than singing alone (undirected song) are one example of a temporally extended effect upon phonology. Male Bengalese finches sing a more stereotyped version of their song when they are singing to female, and these changes

occur over the length of the song (Sakata, Hampton and Brainard, 2008), and these alterations occur at both the phonological and sequential level of song organization. Differences in directed and undirected song are likely to be mediated by a circuit property-type phenomenon (Jarvis et al, 1998; Hessler and Doupe, 1999), and the same may be true of the temporally extended correlations between sequence and phonology.

The circuit properties responsible for sequence generation are thought to arise in the motor nucleus HVC, which are then imposed upon RA (Yu and Margoliash, 1996; Hahnloser et al, 2002). This connection was until recently thought to be entirely feed-forward, but recent work has found a projection from RA back to HVC (Roberts et al, 2008). Although the effects we have measured on syllable production could arise entirely in RA, this recently discovered connection may also mediate the circuit phenomenon that combines phonological and song sequence information. Conclusions based upon all of these results suggest that the control of phonology and sequence are more intimately linked during song production, and that a hierarchical model of song assembly is not an accurate description of Bengalese finch song generation.

With respect to the more local effect on sequencing syllables, the relationship between the similarities of adjacent syllables at convergence points suggests that the limitations of the song system have a component that is more peripheral than central. We found that when the bird is transitioning from two very different syllables to the same syllable, the two convergent syllable variants are less similar to one another. Consider this phenomenon in terms of the syringeal muscles: if two especially disparate muscle

states precede the same syllable, it may be more complicated for the syrinx to reach the same final muscle configuration during syllable production. If instead a sequence transition involves more similar syllables, the musculature may more easily achieve motor convergence. Parallel results have been found for left-to-right effects during human speech production. In spoken sequences of the form vowel-consonant-vowel (VCV), the amount of change of the final vowel as a result of carry-over effects from the preceding utterances is proportional to the differences in articulator positions throughout the sequence (Recansens, 1984; Öhman, 1966). Research on human language is pertinent to our results on the sequencing of Bengalese finch song because we also see carry-over effects of previous motor acts. Considering our results in the context of human coarticulation, we propose that song production is also modified by local limitations of the avian song system.

Bearing in mind the importance of history for vocal motor behaviors, we wanted to know if previous differences in audition could influence motor planning. It has been found that the motor nuclei of the song system are responsive to presentations of the bird's own song (Katz and Gurney, 1981; Williams and Nottebohm, 1985; Margoliash, 1986; Vicario and Yohay, 1993). Differences in auditory feedback percolating through this pathway could therefore alter the motor program during convergent syllable production. In order to assess the influence of auditory information on convergence point similarity, we analyzed songs from Bengalese finches before and after deafening. Removing auditory feedback had little effect upon convergence point similarity, implying that audition is not intimately involved in the online control of syllable phonology at

transition points. Research testing non-hearing humans has shown reduced effects of coarticulation in these individuals as compared to those who can hear (Rothman, 1976). These studies were confounded by the fact that the deaf individuals had not recently lost their hearing. There was sufficient time for central rearrangements to occur, so it is difficult to determine what short-term changes deafening has upon vocal coarticulation. From the results of our experiment, we conclude that the effects of syllable sequencing are relatively similar before and after deafening. These data imply that the control of syllable phonology for variably sequenced syllables is not significantly influenced by auditory input on a moment-by-moment basis.

The effects of sequence upon syllable structure are analogous to human coarticulation, but they are also more generally related to issues of motor sequencing. Many different motor systems exhibit variably sequenced motor acts. For instance, during finger spelling for sign language, the joint and hand positions for one sign can have carry-over effects that are similar to vocal coarticulation (Jerde et al, 2003). Similar results are seen for joint positions during piano playing (Engel, Flanders and Soechting, 1997). Each of these sequential motor acts, including birdsong, requires both the momentary control of muscles for the current behavior, as well as the more long-term control of transitioning between muscle states. The control for actions such as speaking, piano playing and finger spelling is refined with practice. We know that the control of phonology for sequentially variable syllables also goes through a similar refinement process. Understanding how birds maintain phonetic consistency for the same syllable

across sequences is generally informative for studies on motor sequencing, and is of

specific relevance to the understanding of sequencing in human language.

CHAPTER 2

THE RELATIONSHIP BETWEEN RA ACTIVITY AND SYLLABLE STRUCTURE

<u>Abstract</u>

Previous research has suggested that the control of syllable sequencing and syllable production are independent in the bird brain, with sequence being encoded by the motor nucleus HVC, and syllables encoded by the robust nucleus of the arcopallium (RA). The current model of song production proposes that timing signals from HVC are sent to RA, and then RA translates these signals into continuous motor commands for the vocal musculature. The model was derived from research on the zebra finch, a bird with highly stereotyped syllable sequencing. We have analyzed the songs of the Bengalese finch, a bird with variability in syllable sequencing, in order to more explicitly disambiguate activity in RA responsible for sequence from activity responsible for syllable phonology. We have found that mean changes in patterns of RA activity correlate with mean changes in syllable structure, suggesting that there is a phonological relationship between the firing of RA neurons and the structure of resultant syllables. These results are complementary to those found previously by correlating variability in RA with variability in syllable structure. In prior studies, syllable variability around the mean was analyzed with respect to RA activity; in the current study, we have found that mean changes in syllable structure syllables correlate with mean changes in RA activity. This is the first demonstration of a relationship between patterns of RA activity and then general phonology of syllables.

Introduction

      The vocalizations of birds are composed of individual vocal units called syllables, arranged into longer sequences called motifs and songs. For all songs, and especially those with variable sequencing, motor planning involves both the phonation of individual syllables, as well as the ordering of those syllables throughout songs. Previous birdsong research suggested that the control of syllable sequencing is independent of the control of syllable phonology (Vu et al. 1994; Yu and Margoliash, 1996). Song sequence is thought to be controlled by the motor nucleus HVC (Vu et al, 1994; Hahnloser et al, 2002); while phonological encoding occurs in the robust nucleus of the arcopallium (RA) (Vu et al, 1994; Leonardo and Fee, 2005). In current model of song production, timing signals are sent from HVC to RA, and RA translates the timing commands into a continuous stream of motor commands for the vocal and respiratory muscles (Hahnloser et al, 2002; Leonardo and Fee, 2005). Each syllable in song is thought to be controlled by a different ensemble of neurons in RA, with little to no relationship between patterns of RA activity and the spectral properties of syllables (Leonardo and Fee, 2005). These studies were conducted in zebra finches (*Taeniopygia guttata*), a bird with very little variability in syllable sequencing (Catchpole and Slater, 1995). In order to more clearly disambiguate signals controlling sequence from those involved in phonology, it is helpful to study a birdsong that has variable sequencing of song syllables. We have therefore used the Bengalese finch, a bird with syllables that are often found in different sequences (Okanoya, 2004; Sakata et al, 2008), in an effort to understand the influence of sequence and phonology on the activity of RA neurons.

We have shown in the previous chapter the relationship between sequence and phonology on a behavioral level (Chapter 2). We found that syllable phonology is subtly but consistently modified if the same syllable is found in different sequences. These are different changes than those observed between iterations of the same syllable in the same sequences of syllables (Sakata et al, 2008; Kao and Brainard, 2006; Sober et al, 2008). The effects of sequence upon syllable structure result in differences in the mean spectral parameters, whereas changes across iterations of a syllable in the same sequence result in variability around mean spectral parameters. We are interested in how RA activity is affected by these mean changes in syllable structure across sequences.

Previous work on the effects of variably sequenced vocalizations in humans has shown that there is both a central and peripheral component to changes in phonology (Daniloff and Hammarberg, 1973). At the peripheral level, the rapid changes in muscle position can result in adjacent vocalizations influencing one another. An example of this type of phenomenon is seen in the pronunciation of 's' in *cats* and *dogs*. The position of the tongue for the letters 't' and 'g' is different, and the 's' in *dogs* is pronounced more like a 'z' than an 's.' There are also instances where the language commands originating in the central nervous system (CNS) are different (Daniloff and Hammarberg, 1973), resulting in differences in phonation for the same vocalization found in different sequences. Song syllables of the Bengalese finch also show consistent modifications when found in different sequences. Are these differences a result of differences in neural activity in RA? Or, is RA activity the same during each rendition of a syllable, and the changes in syllable production are a result of peripheral effects?

We have three predictions for how RA may encode the same syllable found in different sequences:

1. RA activity is identical for the same syllable in different sequences.

2. RA activity is completely different for the same syllable found in different sequences.

3. Differences in RA activity scale with the differences in phonology for the same syllable found in different sequences.

If the first prediction is true, it suggests that all the changes in the phonation of syllables across sequences originate in the periphery. Prediction 2, on the other hand, would suggest that these changes are controlled by the CNS. If the third prediction is true, there are several conclusions that can be made. First, it would imply that the differences in syllable structure are at least partially correlated with changes in CNS activity; and secondly, Prediction 3 is consistent with a phonological relationship between RA activity and syllable structure.

Research on RA activity during zebra finch song suggested that there is no relationship between syllable phonology and activity patterns in RA (Leonardo and Fee, 2005). A previous study found a relationship between RA activity and spectral variability at specific time points during the production of a syllable (Sober et al, 2008; Chapter 3). The current report is different in that mean differences in syllable phonology were analyzed with respect to RA activity, not the spectral variability around the mean. The current study also examines the structure of the entire syllable in an effort to

understand how broader patterns in RA activity relate to syllable structure. Our

hypothesis is that Prediction 3 is the most accurate description of syllable encoding in

RA. If substantiated, the results of this study would be informative about how patterns of

RA activity related to syllable production.

Methods

Thirteen adult (>100 days), male Bengalese finches were used in the following study (labeled BF1 through BF13 in the rest of the text). The males were reared in the colony until the time of experimentation. They were then housed in individual cages in soundproof chambers. During recording sessions, the birds were isolated in a soundproof box where neural and acoustic recordings were made. Undirected songs (songs sung in isolation) were collected while we concurrently recorded single and multi-unit brain activity in the Robust Nucleus of the Arcopallium (RA, Figure 1a). Food and water were provided *ad libitum* throughout the course of the experiment. All procedures were performed in accordance with established animal care protocols approved by the University of California, San Francisco Institutional Animal Care and Use Committee.

Electrophysiology: After a period of food and water restriction (>1.5 hours), the birds were anaesthetized (20 mg/kg ketamine and 1.5 mg/kg midazolam; followed by 1.0 to 2.0% isoflurane), and an electrode microdrive was affixed to their skull. A small craniotomy was opened above RA in one hemisphere (10 over the right hemisphere, and 3 over the left hemisphere), and an array of 3 to 5 electrodes was implanted several hundred micrometers above RA. The birds were left to recover for several days before any further experimentation was conducted. The birds were then isolated in the experimental soundproof box, and neural recordings from RA were collected while simultaneously recording the bird's acoustic output. At the end of an experiment,

57

electrolytic lesions were made at several depths in order to reconstruct the electrode

trajectory.

An analytical technique described previously was used to classify single versus

multi-unit recordings (Chapter 1). Briefly, a principal components analysis was

performed (PCA) on the collected spike waveforms, and the extent of overlap across the

first two principal components was measured. Distributions with an overlap less than

0.01 were classified as single units, and those with larger overlaps were classified as

multi-unit recordings. In total, we recorded from 25 single neurons and from 120 multi-

unit sites. For the most part, the results discussed below were quite similar for single and

multi-unit recordings, but we did find a few important differences.

The majority of cells (both single and multi-unit) fired tonically before song,

transitioned to a more bursty mode during song, and then returned to tonic firing after a

brief period of inhibition after song (Figure 2.1b). Infrequently, another variety of cell

was recorded that was mostly inactive before and after song, but fired vigorously during

singing. It is generally accepted in the literature that these cells represent interneurons,

and the cells that fire tonically between song bouts are projection neurons (Spiro et al.

1999; Leonardo and Fee, 2005). Unfortunately, the distributions of spike characteristics

for each qualitative variety of neuron (rise time, slope, time to half-peak width, time of

firing) were not distinct, and we could not reliably categorize cells as either one class or

the other (data not shown). Because of the few numbers of recordings from putative

interneurons, and the overlap of distributions for spiking characteristics, we only included

those cells which we could reliably classify as putative projection neurons.

The bursts of activity during singing were generally associated with song features

(Figure 2.1c and 2.1d). Cells fired at multiple times points across the duration of a

syllable or its interval. Average activity patterns were analyzed for all cells recorded in

order to make a qualitative assessment of the neuron's functional relationship to the song.

Only those recording sites that were anatomically localized in RA (through histological

examination of electrode tracts), and modulated by song (through qualitative assessment

of spike rasters), were used for analysis.

Data Analysis for neural recordings: Once each recording site was categorized as either

a single-unit recording or a multi-unit recording (see Appendix 3 for a complete

description of classifying a recording site as single or mulit-unit), a series of analyses

were used to measure the similarity of two different spike trains. The first analysis was

performed to determine the relevant window of neural activity to be analyzed. The songs

of Bengalese finches are composed of short vocalizations (syllables) surrounded by brief

periods of silence (intervals). As a first measure of the premotor window in RA

responsible specifically for syllable production, we measured the covariation of neural

activity with the timing of syllable production. A cross-covariance analysis was

performed between the spike times of RA neural traces with the onsets of syllable

production. The second measure of the premotor window was calculated by using call

notes. Call notes are unlearned vocalizations that both male and females use that are

sung in relative isolation compared with the syllables of songs. Because of this acoustic isolation, we can measure both the start and size of a syllable's premotor window.

We also made a functional determination of the premotor window using the sequence variability inherent to Bengalese finch song syllables. Sequence variability comes in two varieties: convergence points and divergence points. Convergence points are syllables in song that can be preceded by at least two different syllables, and divergence points are syllables in song that can be followed by at least two different syllables. At convergence points, the motor plan for vocal production has to transition from two different plans to the same one in order to produce the common, convergent syllables. It is at that point of neural convergence that the RA activity is in fact controlling production of the convergent syllable. The point of RA neural convergence between two syllable iterations was measured by the use of a d-prime statistic. The d-prime statistic describes the discernability of two neural traces. The equation is as follows:

$$D' = (\text{mean}(a) - \text{mean}(b))/(\text{sqrt}(\text{var}(a)/\text{var}(b))/2)$$

where *a* and *b* represent matrices of neural traces (spike times smoothed with a 5 millisecond window). This analysis results in a point-by-point estimate of similarity. We used the absolute value of the d-prime numbers as a measure of similarity between two neural matrices. Larger d-prime numbers indicate an increase in the difference between two neural traces, and smaller d-prime values represent more similar neural traces. We

aligned the neural traces for all convergent sequences at the onset of the convergent

syllables (100 milliseconds before onset, and 100 milliseconds after). In order to

determine when the premotor activity was responsible for the convergent syllable

production, we compared the d-prime values after syllable onset to those before syllable

onset. We have made the assumption that the premotor activity in RA immediately after

syllable onset can only be responsible for the production of the convergent syllable. As a

result, this portion of the d-prime trace represents a control value for how similar two

neural activity traces can be for convergent syllable production. To perform this analysis,

the d-prime values in a 30 millisecond window after syllable onset were compared to the

d-prime values in successive 5 millisecond bins before syllable onset. A Kolomogrov-

Smirnov test was used to determine when the 5 milliseconds bins before onset were

significantly different than the d-prime values after syllable onset. Because there is some

fluctuation of the neural trace, we used the time when two consecutive 5 millisecond bins

were significantly above ($p < 0.05$, KS-test) the post-onset d-prime values as the

premotor start time. This was done for both single and multi-unit recordings.

Once a premotor window was determined, we compared the similarity of RA

neural activity to the spectral similarity of the syllables produced. For this analysis, we

took the average of the point-by-point d-prime numbers in a window of equal length to

the syllable in question (and starting a set amount of time before syllable onset). By

averaging across the entire pre-motor window, we generated a single numerical value to

describe the similarity of two neural matrices. The similarity of convergent and divergent

syllables was calculated and compared to the similarity of two other syllable categories:

syllables that are always found in the same sequence (same syllable, same sequence), and differently labeled syllables. These two syllable categories were used as lower and upper boundary (respectively) estimates for the similarity of convergent and divergent syllables. The lower boundary control was calculated by taking the data from the same syllable in the same sequence, randomly splitting it into two groups, and calculating the d-prime value between the two halves. For all of the analyses mentioned, only data sets with at least thirty iterations were used. The d-prime numbers for neural activity were then compared to the similarity values calculated for the resultant syllables.

Data analysis for acoustic recordings: The acoustic analysis of syllable similarity was described at length in a previous chapter (Chapter 2). Briefly, 40 songs were randomly selected from the data set of each bird, and segmented into syllables using both amplitude and temporal thresholds. Syllables were then visually labeled using a different letter for each unique syllable in the bird's songs, and eight different spectro-temporal measurements were made for each syllable. The eight measurements used were: duration, pitch, time to half-peak amplitude, frequency slope, amplitude slope, entropy of the spectral density, entropy of amplitude versus time, and entropy of the full spectrum. Once these measurements were made for all song syllables, a principal components analysis (PCA) was then performed on all values from all birds. The distance between syllables in PCA space was used as a measure of similarity. We used the Euclidean distance between the centers-of-mass of two syllables in six-dimensional space (the first six principal components accounted for 95% of the variability in the data set after the PCA was performed) as our similarity measurement (referred to as COM distance in the

rest of the paper). Syllables that are more spectrally different have larger COM values than those that are more similar to each other. COM values assigned a single numerical value to the similarity of two syllables, and could therefore be used in comparisons of neural and acoustic similarity.

Results

Thirteen (13) adult (> 100 days), male Bengalese finches were used for the study. Neural activity in the robust nucleus of the arcopallium (RA) was recorded simultaneously with the songs of the bird (Figure 2.1a). Songs and the corresponding RA activity were then analyzed for any correlations between syllable sequencing and changes in brain activity.
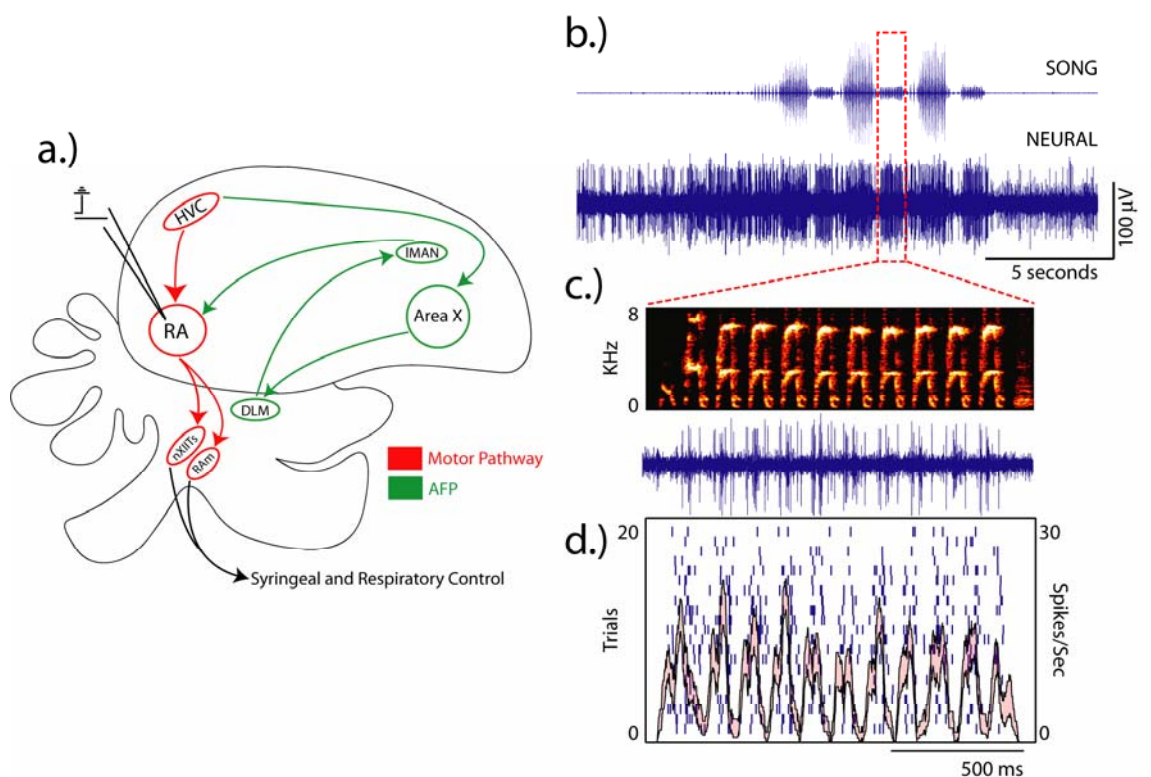
Figure 2.1: (a) Saggital view of the avian song system. In red is the motor pathway. Chronic neural recordings were made from the robust nucleus of the arcopallium (RA). (b) Song oscillogram (top) and corresponding neural activity in RA (bottom) during song from BF1. (c) Spectrogram (top) of a repeated syllable in BF1, and the neural activity in RA during that syllable (bottom). (d) Spike raster (blue tick marks) of 20 iterations of the repeated syllable shown in C, as well as the average spiking behavior (in pink).

DESCRIPTION OF BEHAVIORAL VARIABILITY

In Bengalese finch song, there are two varieties of sequence variability that affect syllable phonology. We will use the terminology *local effects* and *global effects* in reference to each type of change in sequence because they can be differentiated based upon temporal characteristics. These effects have been described in a previous chapter (Chapter 1). Briefly, local effects are those that describe the influence of adjacent syllables upon one another. These effects were most pronounced for convergence points, suggesting that motor history influences current vocalizations. An example of the differences in convergent syllable production, as well as the underlying RA neural activity is shown in Figure 2.2a and 2.2b, respectively.
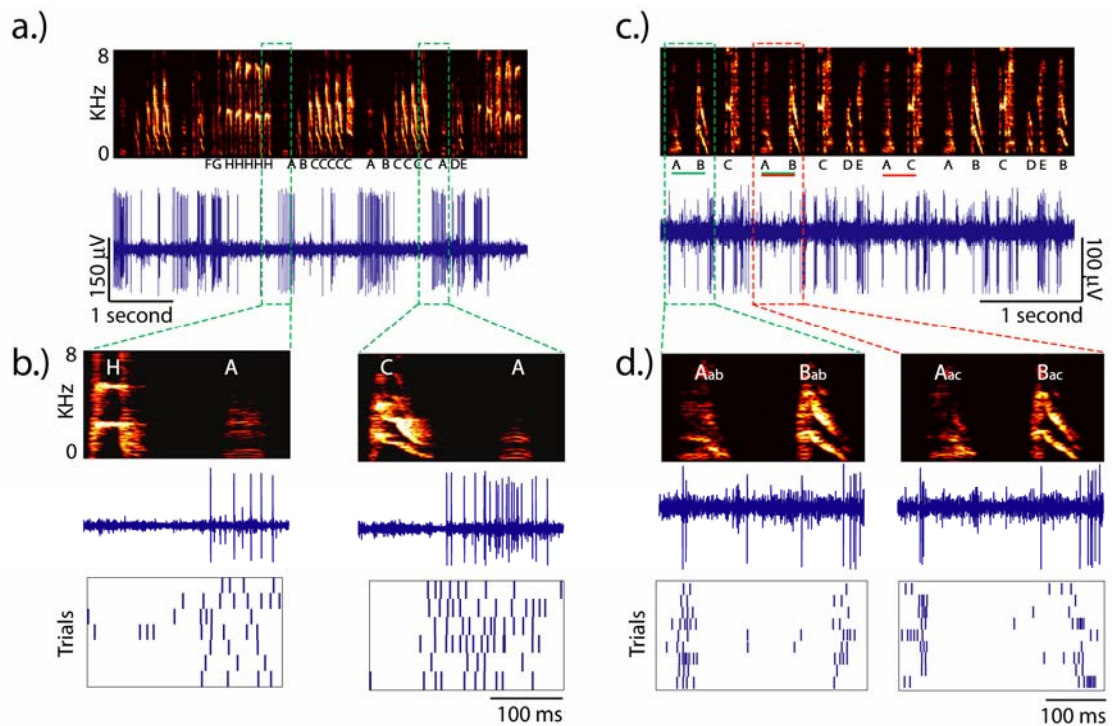
Figure 2.2: (a) Segment of song from BF2 (spectrogram, top; raw RA neural trace, bottom) showing the convergent sequences of syllable A. (b) Zoom of the convergent sequences H-A (left) and C-A (right). Shown are spectrograms (top), RA activity (middle), and spikes rasters (bottom) for each syllable sequence. (c) Song of BF3 showing the ordering of divergent sequences A-B and A-C throughout song (top, spectrogram; bottom, RA activity). (d) Zoom of the divergent sequences A-B when another A-B follows (left), and A-B when A-C follows (right). Shown are spectrograms (top), RA activity (middle), and spikes rasters (bottom) for each syllable sequence.

Shown are the two convergent sequences H-A and C-A (syllable A being the 'convergent' syllable). The differences in RA neural activity are immediately apparent for this syllable, with a robust increase in firing for syllable A in the C-A sequence (Figure 2.2b, middle). Each example of convergent and divergent syllables was analyzed with respect to the patterns of premotor activity in RA as a way of quantitatively demonstrating the relationship between RA activity and syllable phonology.

Whereas 'local effects' describes the influence of immediately adjacent syllables, 'global effects' are those associated with patterns across many more syllables (mean = 8 syllables, max = 25 syllables). Global effects relate to the ordering of specific convergent and divergent syllables throughout song. An example of global sequence effects is shown in Figure 2.2c. In this example song, there are two different divergent sequences: A-B and A-C. The analysis on global effects only evaluates syllable A in the sequence A-B, but focuses on whether the next syllable A divergent sequence is A-B or A-C. If the next divergent sequence is A-B, then successive instances of this divergence point are considered the same. If, on the other hand, the next divergent sequence is A-C, then successive instances are considered to be different. Instances of A-B where the next divergent sequence is the same are compared to instances of A-B where the next divergent sequence is different. For these syllables, the 'local effects' are the same (because the immediate sequences are identical), but extended patterning of sequences in song are different. In the example shown in Figure 2.2c and 2.2d, the burst of RA

activity preceding A-B is slightly broader when the next transition is A-B instead of A-C. The local and global effects of sequence upon RA activity will be discussed at length below.

DETERMINING THE RA PRE-MOTOR WINDOW

In order to compare the neural activity associated specifically with syllable production, it was necessary to determine the temporal boundaries of a syllable's premotor window (Figure 2.3).
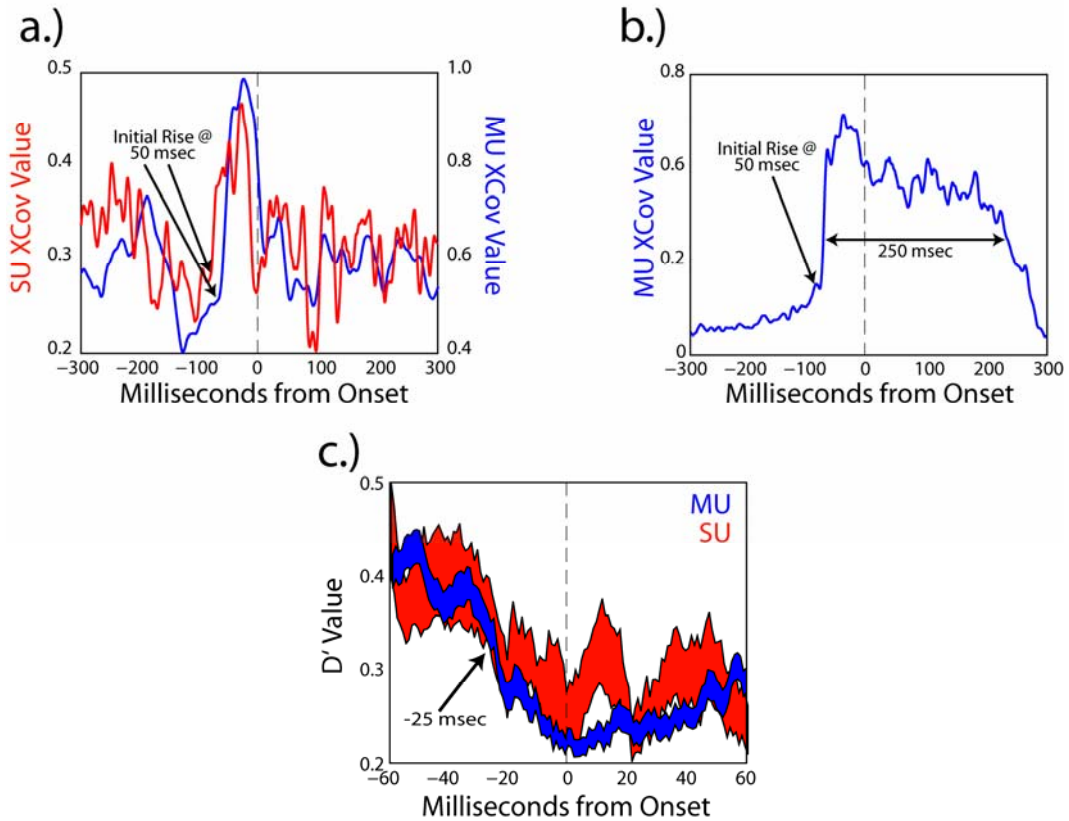
Figure 2.3: (a) Cross covariance of single (red) and multi-unit recordings (blue) in RA and the onset times of syllables. (b) Cross covariance of multi-unit recordings in RA and the calls. (c) Alignment of all point-by-point d-prime statistics for single (red) and multi-unit (blue) convergence point syllable onsets.

Bursts of activity in RA occur at multiple time points throughout a syllable's duration, as well as during inter-syllable intervals, and it is therefore not straight-forward to determine exactly what burst of RA activity is responsible for syllable production. As a result, we have employed multiple methods in order to accurately determine the start and size of the RA premotor window. First, we performed a cross covariance analysis between the traces of RA neural activity during songs, and the onset times of the syllables (Figure 2.3a). This analysis works under the assumption that syllable production requires activity in RA, and that the initiation of a syllable should correlate with an increase in activity in RA. By cross-covarying spike times with syllable onsets, we determined a start time for the premotor window of approximately fifty (50) milliseconds before the beginning of a syllable.

We also analyzed the covariance of premotor activity in RA with the timing of calls. By using calls, it is possible to provide an estimation of both the size and start of the premotor window because of the relative temporal isolation of these vocalizations. The average analyzed call was 200 milliseconds, and the width of the correlation peak was approximately 250 milliseconds (the start of the rise of the covariation peak was again at 50 milliseconds prior to syllable onset, Figure 2.3b). These data support a window length size that is at least as long as the length of the syllable being analyzed, and further substantiate the premotor window start time of 50 milliseconds prior to syllable onset.

As a final method of defining the premotor window, we analyzed neural convergence for convergent syllables as a way of functionally determining the premotor window. If neural convergence does indeed occur at behavioral convergence points, then the time at which that occurs can be used as a functional determination of the premotor start time. The rationale is that for two convergent syllables, the moment when neural activity is indistinguishable across each sequence is an indication of the time when RA activity has shifted to the control of the common syllable. We compared the similarity of the neural traces (using the d-prime statistic) after syllable onset to the similarity before onset. This analysis was designed to determine if and when the two convergent syllable neural traces become as similar as two neural traces responsible for the same acoustic output. We have found that RA activity for convergent syllables does become very similar *prior* to syllable onset, at a time approximately 25 milliseconds before syllable onset (Figure 2.3c). Combining the results of three analyses discussed above (with a bias toward being conservative and not choosing the extremes of the analysis), leads us to conclude that an appropriate premotor window for RA activity begins 30 milliseconds prior to syllable onset, and extends for the length of the syllable being analyzed.

RA ACTIVITY FOR THE SAME SYLLABLE IN MULTIPLE SEQUENCES

Now that we have a good estimate of the start and length of the RA premotor window for syllable production, we can begin to analyze patterns of neural activity associated with song behavior. Our initial question was how RA activity compared for the same syllable found in different sequences, and we proposed three possibilities for how this may occur (see Introduction). In an attempt to address the first possibility (RA

69

encodes the same syllable in different sequences identically), we compared the similarity

of neural activity for convergent and divergent syllable production to the similarity of RA

activity responsible for the same syllable in the same sequences (lower boundary

control). We found, for both single and multi-unit data, that RA activity for convergent

and divergent syllables was less similar than RA activity during production of the same

syllable in the same sequence (Figure 2.4b and 2.4c, green, red and blue lines, KS-Test, p

< 0.0001).
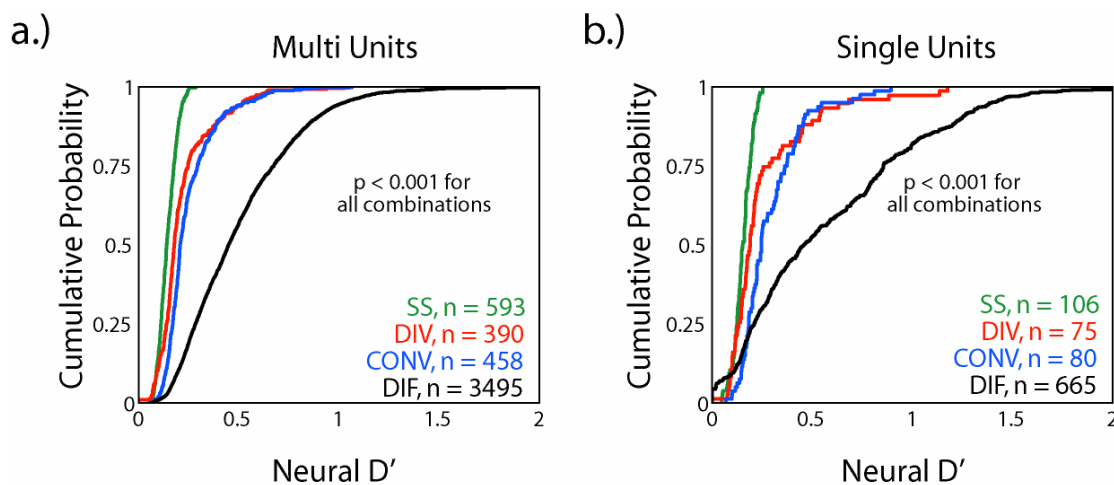
Figure 2.4: (a) Cumulative distribution function (CDF) of d-prime statistics of multi-unit activity for each syllable category: same syllable, same sequence (green); divergent syllables (red); convergent syllables (blue); and differently labeled syllables (black). All syllable categories are significantly different from one another (KS-Test, p <0.0001). (b) CDF of d-prime statistics of single unit activity for each syllable category: same syllable, same sequence (green); divergent syllables (red); convergent syllables (blue); and differently labeled syllables (black). All syllable categories are significantly different from one another (KS-Test, p <0.0001).

From this result, we conclude that the first possibility for how RA may encode the same

syllable in different sequences is incorrect: the activity is *not* identical across sequences.

The second possibility previously outlined for RA may encoding of the same

syllable in different sequences states that the activity will be completely different across

sequences. To examine this option, neural activity for convergent and divergent syllables

was compared to the RA activity for differently labeled syllables (upper boundary

control). The RA activity responsible for convergent and divergent syllables was

significantly more similar than that for differently labeled syllables (Figure 2.4b and 2.4c,

red, blue and black lines, KS-Test, $p < 0.0001$). This result rules out the second

prediction on RA encoding of syllables in different sequence: RA activity is not

completely different for convergent and divergent syllable production.

In broad strokes, the results of the analysis above illustrate that the similarity

across categories of syllables scales with the similarity of the underlying RA activity.

We found that neural activity in RA is most similar when producing the same syllable in

the same sequence, followed by divergent syllables, convergent syllables and differently

labeled syllables. The order of neural similarity across these four syllable categories is

identical to progression of syllable similarities at the behavioral level (see Appendix 2,

Figure 2.2 for syllable similarity analysis). This conclusion lends support to a

phonological component to the RA code (Prediction 3).

PHONOLOGICAL ENCODING IN RA

In the above analysis, we found that neural and syllable similarity scale with one

another across syllable categories. In order to look at the issue of phonological encoding

on a finer scale we compared the acoustic similarity of each convergent or divergent

syllable to the similarity of the neural activity responsible for its production. Syllable

similarity was measured as described previously by using PCA of eight different syllable

features (Chapter 2). These similarity values were plotted against the corresponding

neural d-prime numbers for each recording site from each bird (Figures 2.5a and 2.5b).
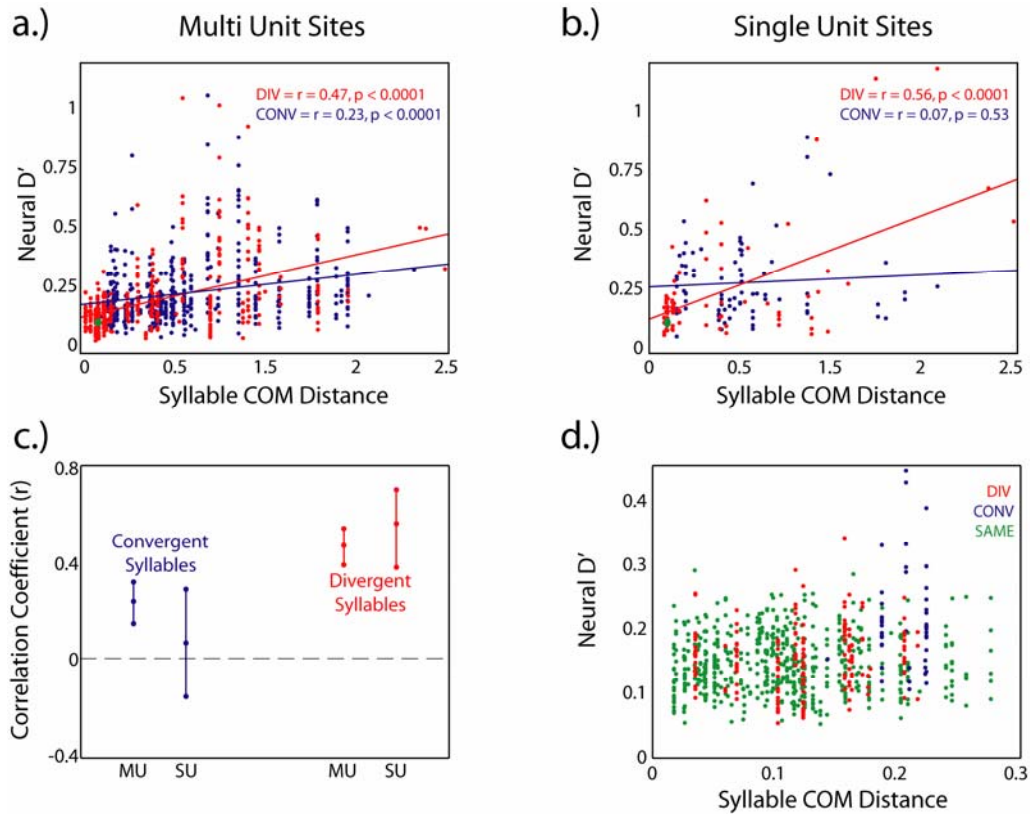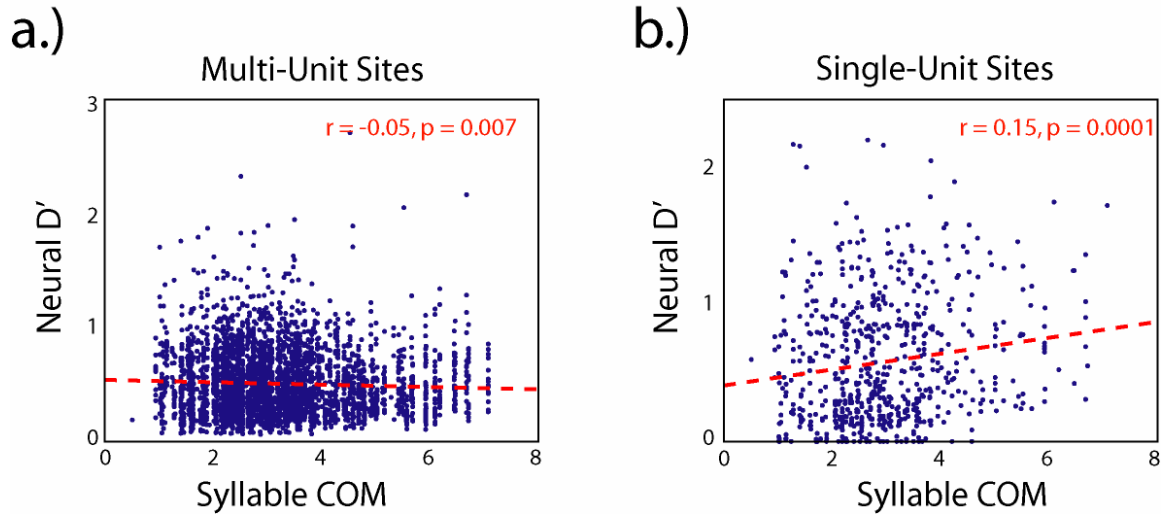

Figure 2.5: (a) Syllable similarity versus multi-unit neural similarity for convergent (blue) and divergent syllables (red). There is a significant, positive relationship between divergent syllable similarity and RA similarity ($r = 0.47$, $p < 0.0001$), as well as between convergent syllable similarity and RA similarity ($r = 0.23$, $p < 0.0001$). The green dot is the mean +/- 1 standard error ellipse for the same syllable in the same sequence as comparison. (b) Syllable similarity versus single unit neural similarity for convergent (blue) and divergent syllables (red). There is a significant, positive relationship between divergent syllable similarity and RA similarity ($r = 0.56$, $p < 0.0001$), but not between convergent syllable similarity and RA similarity ($r = 0.07$, $p = 0.53$). The green dot is the mean +/- 1 standard error ellipse for the same syllable in the same sequence as comparison. (c) Comparison of 95% confidence intervals for the single and multi-unit analysis of convergent (blue) and divergent (red) syllables. The 95% confidence intervals are overlapping for single and multi-unit data. (d) Data from convergent and divergent syllables less than the mean +2 S.D. of the same syllable, same sequence COM distance. Divergent syllables values (red) for neural activity are not significantly different than neural similarity values of the same syllable in the same sequence (green) ($p = 0.38$, t-test). Convergent syllable (blue) neural similarity is significantly greater than same syllable, same sequence similarity ($p < 0.0001$);

We compared multi-unit and single unit data separately in order to determine if there

were any differences in encoding when more or less neurons were recorded. For multi-

unit sites, we found a significant positive relationship between convergent and divergent syllable similarity and the similarity of neural activity in RA (Figure 2.5a, divergent in red, r = 0.47, p < 0.0001; convergent in blue, r = 0.23, p < 0.0001). The same was true for divergent syllable single unit recordings (Figure 2.5b, red line, r = 0.56, p <0.0001), but for convergent syllables, no relationship was found between single-unit neural activity and syllable similarity (Figure 2.5b, blue line, r = 0.07, p = 0.53). The 95% confidence intervals for the correlation coefficients of single and multi-unit recordings were overlapping (Figure 2.5c), however, suggesting that the relationship between RA activity and syllable structure is similar for single unit and multi-unit recordings. These results imply that phonology is indeed encoded in RA, with the underlying activity for more disparate syllables being less comparable than that for spectrally similar syllables. Changes in neural activity are therefore correlated with proportional changes in syllable phonology, lending further evidence in support of phonological code in RA (Prediction 3).

In order to test whether the relationship between syllable phonology and RA activity holds for more disparate syllable types, we analyzed the RA activity responsible for differently labeled syllables. In analyzing both multi and single unit sites for differently labeled syllables, we found a significant, but weak (low r^2) relationship between syllable similarity and neural similarity (Figure 2.6a, multi-units, r = -0.05, p = 0.007; Figure 2.6b, single units, r = 0.15, p = 0.0001).

Figure 2.6: (a) Syllable versus multi-unit neural similarity for differently labeled syllables. There is a small, but significant relationship (r = -0.05, p = 0.007). (b) Syllable versus single unit neural similarity for differently labeled syllables. There is a small, but significant relationship (r = 0.15, p = 0.0001).

These results suggest that although a significant correlation exists between convergent/divergent syllable phonology and RA activity, extrapolating to more disparate syllable types attenuates this relationship, especially for multi-unit sites.

SEQUENCE VS. PHONOLOGY IN RA ACTIVITY

The analysis on convergent and divergent syllable encoding in RA implies that syllable phonology differs, so to does the neural activity in RA responsible for production. We were interested in whether the patterns of RA activity are only influenced by phonology, or if other factors, such as sequence, may influence neurons in RA. In order to examine whether RA activity was solely controlling phonology, we compared the convergent and divergent syllables analyses to that of the same syllable in the same sequence. The mean plus one standard error ellipses for syllable and neural similarity of the same syllables in the same sequences are plotted in Figure 2.5a and 2.5b (green ellipses near the origin). By examining whether the linear regressions for

74

convergent and divergent syllables (single and multi-units) intersect with the values for

the same syllables in the same sequences, we can determine the contribution of

phonology to RA activity. This analysis allows us to extrapolate from the convergent and

divergent syllable data to down to similarity values that are equivalent to the same

syllable in the same sequence. If the linear regression intersects the values for the same

syllable in same sequence, it suggests that RA is mostly involved in phonology. On the

other hand, if the linear regression does not intersect those values, it suggests that some

other song feature beyond phonology is causing differences in RA activity. In other

words, if the extrapolated d-prime values for low COM convergent and divergent

syllables are still higher than d-prime values for the same syllable in the same sequence,

it suggests that there is some residual difference in RA activity even when all differences

in production have been accounted for. After taking into account the 95% confidence

intervals for the linear regression (Figure 2.5c), we find that the regression lines for

convergent and divergent syllables do intersect the values for the same syllable found in

the same sequence. These results imply that RA activity is primarily controlling syllable

phonology, providing additional support for Prediction 3.

As a further test of the influence of phonology and sequence upon RA activity, we

compared the data from the same syllable in the same sequence to restricted data sets of

convergent and divergent syllables. These data sets were restricted to those divergent or

convergent syllables that were as similar as the same syllable in the same sequence. This

was accomplished by only including divergent and convergent syllables with COM

distances less than the 2 standard deviations above the mean of the same syllable in the

same sequence. By comparing only very similar renditions of divergent syllables, we found that the underlying neural activity was as similar as RA activity responsible for the same syllable in the same sequence (Figure 2.5d, red versus green, $p > 0.05$, t-test). The differences in neural activity responsible for similar renditions of convergent syllables, on the other hand, was significantly greater than that for the same syllable in the same sequence (Figure 2.5d, blue versus green, $p < 0.0001$, t-test). The source of these differences between convergent and divergent syllable encoding will be discussed below. From this analysis, we can conclude when divergent syllable phonology is controlled for, differences in RA neural activity do not pertain to sequence.

GLOBAL EFFECTS OF SEQUENCE UPON RA ACTIVITY

The global effects upon syllable structure discussed above imply that sequence and phonology interact in multiple ways. In order to address whether any correlation can be found between sequence and the premotor activity of RA, the global effects of sequence upon phonology were analyzed with reference to changes in RA activity. We analyzed RA firing for any effects of global song sequencing, and found that premotor activity for convergent and divergent syllables was modified by these temporally distant patterns in sequencing. This result was true for both multi-unit and single unit sites (Figure 2.7b and 2.7c; in blue, convergences > same syllable same sequence, $p < 0.0001$; in red, divergences > same syllable same sequence, $p < 0.0001$).

Figure 2.7: (a) Schematic of how successive sequence syllables are defined. In red are two different DIV syllable transitions (A-B and A-F) separated by several other syllables. This is an example of a successive divergence relationship between syllables. In blue is an example of two successive convergence syllables: two convergence points involving the same convergent syllable separated by other unrelated syllables. (b) CDF of d-prime statistics of multi-unit activity for each successive syllable category: same syllable, same sequence (green); divergent syllables (red); convergent syllables (blue); and differently labeled syllables (black). All syllable categories are significantly different from one another (KS-Test, p <0.0001), except for successive divergence points and divergence point (p > 0.05). (c) CDF of d-prime statistics of single unit activity for each successive syllable category: same syllable, same sequence (green); divergent syllables (red); convergent syllables (blue); and differently labeled syllables (black). All syllable categories are significantly different from one another (KS-Test, p <0.0001), except for successive divergence points and divergence point (p > 0.05).

Remember that syllable phonology is also altered by distant patterns in sequencing, and as such, the analysis of global effects upon RA activity does not necessarily imply that sequence influences neurons in RA. Therefore, the global effects of sequence on RA activity lend further evidence to Prediction 3 being the most accurate description of how RA encodes the same syllable in different sequences.

__Discussion__

Previous work on the song motor pathway in birds has suggested that syllable sequence and phonology are controlled independently: HVC encodes sequence, and RA is involved in syllable production (Vu et al, 1994, Yu and Margoliash, 1996; Leonardo and Fee, 2005). These experiments were performed in zebra finches, a bird with highly stereotyped syllable sequencing. By studying the variable songs of Bengalese finch, we had more opportunities to disambiguate sequence from phonological encoding in the motor pathway. We focused on RA, and found further evidence that RA is primarily involved in syllable production. Unlike previous reports, however, we found a phonological relationship between different patterns of activity in RA: as neural activity in RA becomes more different, so to does the phonology of the resultant syllable. These results substantiated Prediction 3 (see Introduction), which states that the similarity of RA encoding for the same syllable in different sequences is proportional to the similarity of the syllables.

We were able to find a relationship between RA activity and phonology for several reasons. We found that in 'neighborhoods' of similar syllables (convergent and divergent syllables), that there is a relationship between syllable structure and RA activity. By making our measurements across similar syllables, we increased the likelihood of finding a phonological relationship across RA activity patterns. This is because similar pools of neurons are most likely active in RA for these syllables. If the same pool of neurons is active in RA, a similar pattern of activity is being sent to the motoneurons in the brainstem. Subtle changes in the activity of the same populations of

neurons may not drastically alter motoneuron firing as much as the interactions of different populations of neurons. On the behavior side, the changes between the same syllable in different sequences are much smaller than that between differently labeled syllables. The actual real difference in muscle activations at the motor periphery may just be the alteration of a few aspects of the vocal apparatus (expiratory pressure, tension on one muscle, etc). We are measuring features of the spectrogram as proxy for motor commands to the syrinx and respiratory centers, and this transformation may introduce some non-linearities to our analysis. When similar pools of neurons are active, on the other hand, we may be in a more linear range of the relationship between neuronal firing and the spectrogram. Because of the simple changes to both RA and the syllables at convergent and divergent syllables, the restricted range of our measurements increases our ability to find a relationship between neural activity and behavior.

The production of more disparate syllable types, on the other hand, may result in different pools of neurons being active, along with more complex changes at the periphery for syllable production. This results in several non-linearities being introduced into our neuronal and behavioral measurements. First, because of the great convergence of RA inputs onto the brainstem, the interactions between different pools of neurons may have more complicated effects on motoneuron firing than similar pools of neurons. Secondly, the differences in muscle and respiratory control for different syllables involve the interaction of many vocal effectors, and their interactions may further obscure our ability to correlate brain and behavior. The result of these added layers of effects may be that it is difficult to find a relationship between RA activity and syllable phonology for

disparate syllables. This effect is evident in our analysis of differently labeled syllables, as well as in previous studies correlating RA activity with spectrograms of songs (Leonardo and Fee, 2005).

The relationship between syllable phonology and RA activity was stronger for divergent syllables than it was for convergent syllables (Figure 2.5). Beyond these syllables having differences in similarity, another source of RA differences can be understood by looking at the motor history of convergent and divergent syllables. Through our functional determination of the RA premotor window using convergent syllables (Figure 2.4c), we found that complete neural convergence did not occur until 25 milliseconds before syllable onset in some cases. The premotor window used for analysis started 30 milliseconds before onset, which may include some time when the song system musculature is still transitioning from two different states to the same muscle state. The premotor window therefore included some time when different control signals were being employed. The motor history leading up to divergent syllables is very similar, however, meaning that the muscle configurations are already the same well before production initiates. At convergence points, our premotor window necessarily has some component that is involved in different muscle activations, whereas the pattern of muscle activations before divergent points is the same. The result of this phenomenon is that the relationship between RA activity and divergent syllable phonology should be stronger than the relationship between RA activity and convergent syllables. Throughout our analysis, divergent syllable/neural relationships were stronger than those for convergent syllables.

The discussion thus far has been concerned with identifying the phonological component of premotor activity in RA. We are also interested in determining if any signature of syllable sequence can be found in the RA code. We found in a prior chapter that phonology can be altered by the way specific sequences of syllables are ordered throughout song (Chapter 1). In this situation, the same syllable found in the same sequence (identical motor contexts) can vary significantly in structure due to the influence of syllables many syllable positions away (global effects). We believe that these effects are due to some underlying long term change in the circuit properties of song system, possibly comparable to changes such as those associated with female-directed versus undirected song (Sakata, Hampton, and Brainard, 2008). We found that RA activity was influenced by the global effects of sequencing, but because the syllable themselves are also different, it is not possible to determine if there are any changes specific to sequence. The phonological relationship between RA activity and the structure of convergent and divergent syllables leads us to believe that the differences in RA as a result of global sequence effects are primarily driven by changes in phonology (not sequence). This implies that the activity upstream of RA, possibly in HVC, is influenced by global sequence effects, and these differences in turn affect RA activity.

By viewing the current results in the broader framework of previous studies on RA activity in singing birds, we can begin to make a more complete model of RA encoding during song production. In a previous report, we found that trial-by-trial variability of acoustic structure for the same syllable found in the *same* sequence was

81

correlated with the firing of RA neurons (Sober, Wohlgemuth and Brainard, 2008, Chapter 3). The current results add a new layer onto this earlier finding. In this study, we found a correlation between mean changes in syllable structure and mean changes in RA activity, which is different than finding a correlation between variability in RA and variability in syllable parameters. The current result indicates that the motor code in RA is strictly phonological, and that by measuring changes in RA activity, we should be able to predict changes in syllable structure.

The relationship between RA activity and phonology also speaks to whether the differences in syllable production across sequences were the result of central or peripheral effects. Considering the phonological relationship between RA activity and syllable structure, we conclude that some portion of the changes in phonation due to sequence is driven by changes in RA activity. While we cannot rule out the possibility of sequence interactions at the peripheral level, a CNS source of syllable changes suggests that birds can actively shape syllable transitions. As with other song parameters, the fine motor control necessary for syllable transitioning must be learned and maintained. Birds therefore retain the ability to make adjustments to their syllable transitions throughout life.

The above results demonstrate two new findings about the control of syllable production in RA. We found that there is a direct relationship between syllable phonology and the underlying patterns of neural activity for vocal production. We also found evidence that similar pools of neurons are active when the same syllable is sung in

multiple sequences. Our data delve deeper into the activity patterns of RA in an attempt

to more fully describe the involvement of RA during song production. With these results,

we hope to provide new insights into how RA activity shapes syllable production on a

moment-by-moment basis.

CHAPTER THREE

CENTRAL CONTRIBUTIONS TO SYLLABLE VARIABILITY

Abstract:

Birdsong is a learned behavior remarkable for its high degree of stereotypy. Nevertheless, adult birds display substantial rendition-by-rendition variation in the structure of individual song elements or 'syllables.' Previous work suggests that some of this variation is actively generated by the avian basal ganglia circuitry for purposes of motor exploration. However, it is unknown whether and how natural variations in premotor activity drive variations in syllable structure. Here, we recorded from the premotor nucleus RA (robust nucleus of the arcopallium) in Bengalese finches and measured whether neural activity co-varied with syllable structure across multiple renditions of individual syllables. We found that variations in premotor activity were significantly correlated with variations in the acoustic features (pitch, amplitude, and spectral entropy) of syllables in roughly a quarter of all cases. In these cases, individual neural recordings predicted 8.5 +/- 0.3% (mean +/- S.E.) of the behavioral variation, and in some cases accounted for 25% or more of trial-by-trial variations in acoustic output. The prevalence and strength of neuron-behavior correlations indicate that each acoustic feature is controlled by a large ensemble of neurons that vary their activity in a coordinated fashion. Additionally, we found that correlations with pitch (but not other features) were predominantly positive in sign, supporting a model of pitch production based on the anatomy and physiology of the vocal motor apparatus. Collectively, our results indicate that trial-by-trial variations in spectral structure are indeed under central neural control at the level of RA, consistent with the idea that such variation reflects motor exploration.

Introduction:

The acquisition of any complex sensorimotor skill – whether learning to speak or to throw a curveball – is associated with a gradual decrease in motor variability. Our initial attempts are usually variable and inaccurate, apparent in the babbling speech or wild pitches familiar to anyone who has observed motor learning in action. With practice, however, we learn to control our motor system until we can reliably produce the desired output.

Despite these dramatic changes during learning, even well-practiced movements retain some variability. This trial-by-trial variation might originate in the motor periphery, reflecting unreliability either at the neuromuscular junction or in the muscles themselves. In this case, the central nervous system (CNS) might encode the same motor output during each rendition of a well-learned task, with behavioral variability resulting solely from downstream noise.

Alternately, variability might be generated centrally as variations in the motor command. Recent studies in the oculomotor system have demonstrated that trial-by-trial variations in the activity of single neurons correlate with measurable variations in eye movement (Medina and Lisberger, 2007). Furthermore, variation in the activity of cortical neurons can predict the kinematics of arm movements even before reaching movements begin (Churchland et al., 2006b). These results suggest that some of the "residual" variation in well-learned skills is driven by the CNS.

Birdsong is an excellent model system for studying the neural processes underlying motor control and motor variation. Song learning begins with exposure to the song of an adult male "tutor" and is characterized by a dramatic reduction in variation as

the developing song becomes more similar to the tutor's (Arnold, 1975; Tchernichovski et al., 2001; Kittelberger and Mooney, 2005). By adulthood, song in many species becomes highly stereotyped or "crystallized," and remains so throughout the bird's life. Three examples of crystallized song from a Bengalese finch are shown in Figure 3.1a to illustrate the consistency of vocal output across multiple renditions.
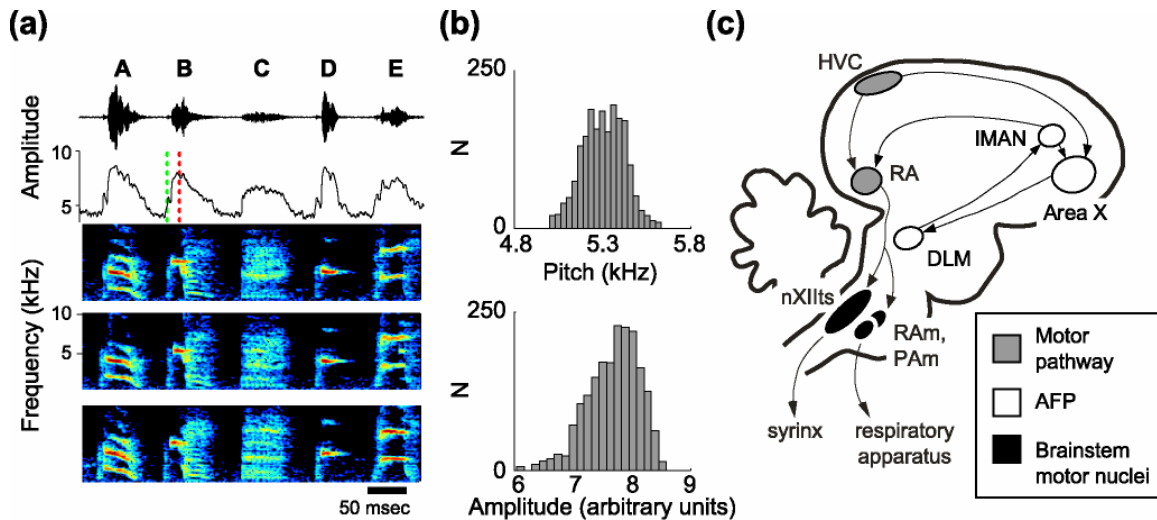
Figure 3.1. Acoustic variation in Bengalese finch song. (a) Top, raw sound amplitude waveform of five syllables (labeled ABCDE) from the song of an adult Bengalese finch. Middle, smoothed rectified sound amplitude waveform. Bottom, spectrograms of three different iterations of this motif. Spectrograms show the power at each frequency (color scale) as a function of time. The topmost spectrogram corresponds to the example sound waveform. Acoustic features such as pitch and amplitude were measured at a fixed time (red dashed line for syllable "B") relative to syllable onset (green dashed line). (b) Distributions of the pitch (top) and amplitude (bottom) of syllable "B" across 1919 renditions of the syllable. (c) The song system is composed of a direct motor pathway consisting of nuclei HVC and RA and an anterior forebrain pathway (AFP) containing Area X, the medial portion of the dorsolateral thalamus (DLM) and lMAN. RA sends projections to motor neurons in the tracheosyringeal portion of the twelfth motor nucleus (nXIIts), which innervates the muscles of the syrinx, and to motor nuclei retroambigualis (RAm) and paraambigualis (PAm), which innervate the respiratory musculature (Vicario and Nottebohm, 1988; Wild, 1993; Reinke and Wild, 1998)

However, despite the stereotypy of crystallized song, adult birds still demonstrate significant variation in acoustic output across multiple renditions of the same syllable (Tchernichovski et al., 2001; Kao et al., 2005; Olveczky et al., 2005; Kao and Brainard, 2006; Sakata et al., 2008). We refer to such cross-rendition variation in syllable structure

as "trial-by-trial" variation. Figure 3.1b shows the trial-by-trial variation of two acoustic

parameters (pitch and amplitude) for one syllable of crystallized song.

Recent work in the zebra finch has suggested that a significant component of trial-

by-trial variation in syllable structure is central in origin. The anterior forebrain pathway

(AFP), via its output nucleus lMAN (lateral magnocellular nucleus of the anterior

nidopallium), sends input to the motor pathway nucleus RA (Figure 3.1c). Several lines

of evidence implicate lMAN as a source of trial-by-trial variation: lesions or inactivation

of lMAN dramatically reduce variation (Kao et al., 2005; Olveczky et al., 2005; Kao and

Brainard, 2006), stimulation in lMAN affects the pitch and amplitude of individual

syllables, and the level of variation in lMAN activity correlates with the level of variation

in behavior (Kao et al., 2005). An intriguing implication of these findings is that lMAN

might contribute to behavioral variation by injecting neural variation into RA, and that

the role of the AFP in the adult bird might include adding variation to song for purposes

of motor exploration during learning (Doya and Sejnowski, 2000; Kao and Brainard,

2006; Fiete et al., 2007; Tumer and Brainard, 2007).

It is unclear, however, how variation generated by the AFP passes through RA

and into the motor periphery, or indeed whether trial-by-trial neural variation in RA has

any behavioral consequences at all. The activity of RA neurons is distinguished by an

extremely low level of variability and is far more precise than that observed in neurons of

the primate motor cortex (Chi and Margoliash, 2001; Leonardo and Fee, 2005).

Individual RA neurons fire stereotyped patterns of bursts during stereotyped sequences of

song syllables. Moreover, variations in the timing of these RA bursts covary with

differences in the timing of syllable features (Chi and Margoliash, 2001). The result of

this temporal covariation is that each syllable is consistently preceded by the same pattern of RA bursts, regardless of whether it occurs slightly earlier or later than average. This precise temporal alignment, however, begs the question of whether and how trial-by-trial variations in the number of spikes in each burst leads to trial-by-trial variations in the acoustic structure of syllables.

If RA and upstream areas are to drive trial-by-trial changes in behavior, then RA activity must vary from trial to trial and these variations must in turn drive variations in song. Such variation in neural activity might be distributed across RA in several different ways. In one model, variation might be restricted to a small subpopulation of RA neurons, which exert powerful control over trial-by-trial variations in behavior (Figure 3.2a).
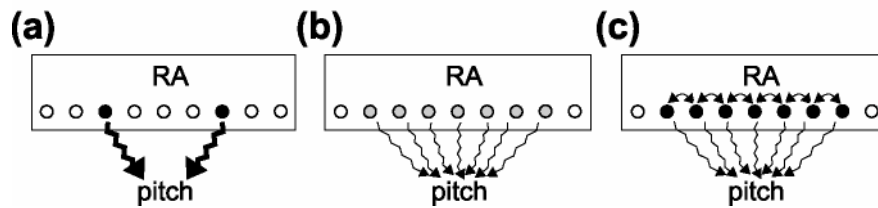
Figure 3.2: Three models of how RA might drive trial-by-trial variation in song. Circles represent populations of RA projection neurons and wavy arrows represent their contribution to variations in syllable structure (pitch in this case). In one model (a), a small number of independently varying RA neurons each drive a substantial amount of pitch variation (heavy arrows). In this case, a small proportion of RA neurons would exhibit correlations with pitch, and these correlations would be quite strong (indicated by filled black circles), reflecting the substantial influence of each neuron. In a second model (b), large numbers of independently varying RA neurons each make small contributions to pitch variation (thin arrows), with pitch variations resulting from the sum of independent modulations in firing. In this case, the activity of many neurons would be correlated with pitch, but these correlations would be weak (indicated by the lightly shaded circles), since each neuron is responsible for only a small fraction of behavioral variation. In a third model (c), large numbers of RA neurons generate correlated activity (curved arrows), driving pitch modulations with coordinated changes in activity. In this case, many neurons might exhibit correlations with each acoustic feature, and the strength of these correlations could be quite high (black circles), since variations in the firing of any one cell are correlated with variations in the entire population. In the Discussion, we consider the implications of our results in discriminating between these models.

Alternately, every neuron in RA might vary its activity independently, influencing motor output with the sum of many independent fluctuations (Figure 3.2b). In a third model, trial-by-trial variations across the RA population might be correlated such that variations in behavior result from coordinated changes in firing across many neurons (Figure 3.2c).

We hypothesized that a component of the acoustic variability observed across multiple iterations of the same syllable is indeed the consequence of trial-by-trial variations in RA activity. To test this, we recorded from RA neurons during singing in adult Bengalese finches and asked whether variations in spiking activity across multiple renditions of individual syllables could account for variations in the pitch, amplitude, and spectral entropy of those syllables. Furthermore, by examining the prevalence, signs, and strengths of correlations between premotor activity and acoustic output we investigated how variation is distributed across the population of RA neurons and compared how pitch, amplitude, and spectral entropy are encoded in RA.

Methods:

Adult (>100 days old) Bengalese finches (*Lonchura striata* var. *domestica*) were bred in our colony and housed with their parents until at least 60 days of age. Following electrode implantation, birds were isolated and housed individually in sound-attenuating chambers (Acoustic Systems, Austin, Texas) with food and water provided *ad libitum*. Unless otherwise specified, all recordings presented here are from undirected song (i.e. no female was present). All procedures were performed in accordance with established animal care protocols approved by the University of California, San Francisco Institutional Animal Care and Use Committee (IACUC).

Electrophysiological data collection:

Birds were anesthetized (induction with 20 mg/kg ketamine and 1.5 mg/kg midazolam, maintained with 0.5-2.0 % isoflurane) and a lightweight microdrive (Hessler and Doupe, 1999) was positioned stereotactically over RA in one hemisphere (10 implants over right RA, 3 over left RA) and secured to the skull with epoxy. Each microdrive carried a custom-made array of 3-5 high-impedance microelectrodes (Microprobe WE1.5QT35.0A3) with all electrode tips grouped within 300 μm of each other. After recovery from surgery, birds resumed singing within 1-3 days. After singing resumed, electrode arrays were lowered into RA. Extracellular spike waveforms as large as 4.5 mV were recorded during and between bouts of singing. RA recording sites were identified by the presence of characteristic changes in activity associated with the production of songs and calls and by *post hoc* histological confirmation of trajectory of

each electrode array. In a subset of birds (which contributed ~31% of all recorded units), we were able to estimate the dorsal-ventral position of the array on each recording. We observed no significant differences between dorsal and ventral RA with respect to the prevalence or sign of correlations between neural activity and acoustic output.

We used a quantitative technique (see "Quantifying unit isolation" in Appendix 3) to measure the isolation of spike waveforms. Briefly, we performed principal components analysis (PCA) on recorded voltage waveforms, examined their projections along the first two components, and quantified the extent of overlap between waveform clusters. Recordings yielding clusters with overlaps of less than 0.01 were classified as single units, and recordings with larger overlaps were classified as multiunit clusters, reflecting the potential contribution of several neurons to each recording. This technique yielded isolation estimates that agreed well with both qualitative assessments of isolation and estimates based on spike refractory periods. As described below, data from single- and multiunit recordings yielded nearly identical results. In total we collected 145 RA recordings (25 single-unit, 120 multiunit) from 13 birds. Unless the level of isolation is specified, the term "unit" in this paper refers to either a single unit or to one multiunit cluster.

Spiking statistics:

In quantifying the firing statistics of single units (n=25) in the Bengalese finch, we performed analyses similar to those in a prior study of zebra finch RA (Leonardo and Fee, 2005) to facilitate comparison between the two species. Instantaneous firing rate was defined as the inverse of the interspike interval (ISI) and computed for all single

units using only data recorded during singing. Based on the bimodal distribution of instantaneous firing rate shown in Figure 3.3c, we selected a 50 Hz threshold dividing bursting intervals (with higher firing rates) from nonbursting intervals (with lower rates). Using this criterion, we assigned every ISI either to an ongoing burst or to an inter-burst pause based on its instantaneous firing rate. Instantaneous firing rate distributions from individual single units resembled the pooled distribution shown in Figure 3.3c, with thresholds distributed around 54.0 +/- 19.9 Hz (mean +/- SD).

Only single units were used in the analyses shown in Figure 3.3c. However, the distributions of burst durations from single-unit and multiunit recordings were highly overlapping (29 +/- 25 msec for single-unit, 33 +/- 37 msec for multiunit, mean +/- S.D.). While the true number of neurons contributing to a multiunit signal is difficult to ascertain, these results suggest that many of our multiunit recordings reflect the activity of a relatively small number of neurons.

Acoustic features:

To characterize the relationship between neural and behavioral variation, we measured the acoustic properties of each syllable rendition as well as the premotor spiking activity before each syllable. Because of the complex acoustic structure of song, we must take care that the acoustic parameters being quantified reflect important dimensions of behavioral variation, since failing to do so could cause us to underestimate the contributions of RA activity to behavioral variation. Previous work on the song system has identified fundamental frequency (pitch), amplitude, and spectral entropy as potentially important axes of behavioral variation, since they are refined during song

learning, vary from trial to trial in the adult, and can be perturbed by electrical stimulation of the song system during singing (Tchernichovski et al., 2001; Fee et al., 2004; Kao et al., 2005). To address this issue quantitatively, we performed a separate analysis in which we used principal components analysis (PCA) to ask which dimensions of spectral variation capture the greatest fraction of the total spectral variability of each syllable. As described in detail in Appendix 3 (see "Quantitative analysis of spectral variability" and Appendix 3, Figures 3.2-3.4), the PCA-based analysis revealed that in most cases, variations in pitch, amplitude, or entropy indeed captured the greatest fraction of the total behavioral variation. These acoustic features therefore dominate behavioral variability and as such constitute a reasonable choice of behavioral parameters that might reveal the influence of trial-by-trial variations in RA firing.

For each syllable, we defined a measurement time relative to syllable onset that corresponded to a well-defined spectral feature (e.g. the band of spectral power at ~5.3 kHz in syllable "B" shown in Figure 3.1a). Syllable onsets were defined based on amplitude threshold crossings. Pitch was defined as the fundamental frequency at the measurement time and was quantified by finding peaks in spectral power. Amplitude was defined as the value of the smoothed rectified amplitude trace at the measurement time. Spectral entropy was defined as the entropy of spectral power at the measurement time within one octave centered on the peak power. Entropy was quantified according to the equation $E = -\Sigma(p \log_{10} p)$, where p is the probability distribution of spectral power.

Premotor neural activity:

Premotor activity was quantified by measuring the number of spikes occurring in a "premotor window" prior to the time at which the acoustic properties of each syllable were measured. The timing of this window was chosen to reflect the latency at which RA activity influences the acoustic structure of song. Stimulation studies have shown that disrupting ongoing activity in RA during song perturbs motor output, although different groups have produced varying assessments of the nature and latency of these effects (Vu et al., 1994; Ashmore et al., 2005). Stimulation with a single pulse, however, has been shown to disrupt the pitch of an ongoing syllable at a short (~15 msec) latency without altering the sequence of syllables being produced (Fee et al., 2004). To allow for some uncertainty about the premotor latency (and for the possibility that different acoustic parameters may have different latencies), we measured premotor neural activity in a 40 msec wide window that ended at the time when acoustic parameters were measured. Premotor neural activity was measured by counting the number of spikes in this window. We validated this approach to quantifying neural activity by comparing several models of premotor encoding in which spiking activity is quantified on different timescales (see "Testing the timescale of premotor encoding" in Appendix 3).

Correlation analyses:

To examine the relationship between premotor activity and acoustic output, we computed the linear correlation between each acoustic feature and the number of spikes in the premotor window. Prior to computing correlations, we discarded outliers with acoustic feature measurements lying more than 4 standard deviations from the mean.

Inspection of audio recordings revealed that these outliers usually resulted from noise artifacts unrelated to vocal production. Additionally, we performed a partial correlation analysis in which the relationship between each acoustic parameter and neural activity was considered while controlling for correlations between neural activity and the other two acoustic parameters.

Proportion of cases correlated:

We define the prevalence with which neural activity is correlated with a given acoustic feature in terms of the proportion of cases with significant correlations. One "case" is defined as one unit (that is, one single unit or one multiunit cluster) being active before one syllable. A unit is defined as active prior to a syllable if it fires on average at least one spike in the 40 msec premotor window, corresponding to a mean rate of ≥ 25 Hz. For a given acoustic parameter, therefore, one unit will contribute multiple cases if it is active before more than one syllable. For each acoustic parameter, we found the proportion correlated by dividing the number of cases in which the acoustic parameter was significantly correlated (at $p<0.05$) with neural activity by the total number of cases.

Multiple comparisons:

We expect that even if no relationship between premotor neural activity and song existed, some correlations would be significant by chance (~5% of all correlations, corresponding to our significance criterion of $p<0.05$). We used a permutation technique to quantify whether the observed proportion of correlations between neural activity and acoustic output was significantly greater than chance. Briefly, we created an artificial

dataset in which all correlations of interest (those between premotor activity and acoustic features) are broken, but all other correlations (such as those between different acoustic measures and between neural activity on consecutive syllables) are preserved. We then performed correlation tests on the artificial dataset and noted the proportion of cases with significant correlations. By performing this procedure 1000 times, we estimated the distribution of proportions of significant correlations under the null hypothesis, and then asked whether the proportion of significant correlations in the real dataset exceeded the 95th percentile of this distribution. See "Multiple comparisons" in Appendix 3 for a detailed explanation of this technique and a discussion of related issues.

Fraction of units active:

To quantify the fraction of the population active at a given time during song, we considered only birds from which we had recorded at least 10 units (n=7). The mean neural activity of each unit was quantified in 1 msec bins across the duration of a frequently occurring syllable sequence (motif). Repeating this analysis for each recorded unit allowed us to infer the mean activity across the population. The fraction active at a given time was defined as the percentage of recorded units with mean rate greater than 25 Hz. For the plot shown in Figure 3.4b, the fraction active is averaged over a sliding 5 msec time bin.

Timewarping:

Piecewise-linear timewarping (Leonardo, 2004) was used to create Figure 3.4a. Briefly, spike times were aligned to the mean durations of syllables and inter-syllable pauses by "stretching" them linearly. These small adjustments allow for easy comparison of neural activity across trials and units by eliminating small (typically 1-6% in our data) variations in song tempo. Note however that timewarping was used only for display purposes and was not applied as part of any of the analyses described above, in which the 40 msec premotor window was applied on a syllable-by-syllable basis.

Results:

Our recordings revealed that neurons in Bengalese finch RA fire distinct patterns of activity for distinct syllables (as previously observed in the zebra finch), consistent with the accepted role of RA in controlling syllable structure (Yu and Margoliash, 1996; Leonardo and Fee, 2005). Close inspection of firing patterns, however, revealed trial-by-trial variations in the number of spikes in each burst (Figure 3.3b). The analyses described below investigate whether this neural variation results in trial-by-trial variation in the acoustic structure of individual syllable.

Neural activity in Bengalese finch RA:

Neurons in RA exhibited characteristic patterns of activity before, during, and after song. The majority of RA units displayed regular tonic activity of 20-50 Hz during rest (Figure 3.3a) and fired syllable-locked bursts during singing (Figure 3.3b). Following song offset, the resting tonic activity of most units was transiently inhibited (see Appendix 3, Figure 3.7). Additionally, a small subset of recordings (1 single-unit, 3 multiunit) had very low or no spiking activity when the bird was at rest, displayed bursty spiking activity during song, and had narrower spike widths than the rest of the population. Based on these criteria (Spiro et al., 1999; Leonardo and Fee, 2005), these 4 units (representing 3% of the total dataset) were classified as putative interneurons and were excluded from further analysis (see "Putative interneurons vs. putative projection neurons" in Appendix 3). The results that follow describe the properties of putative projection neurons.
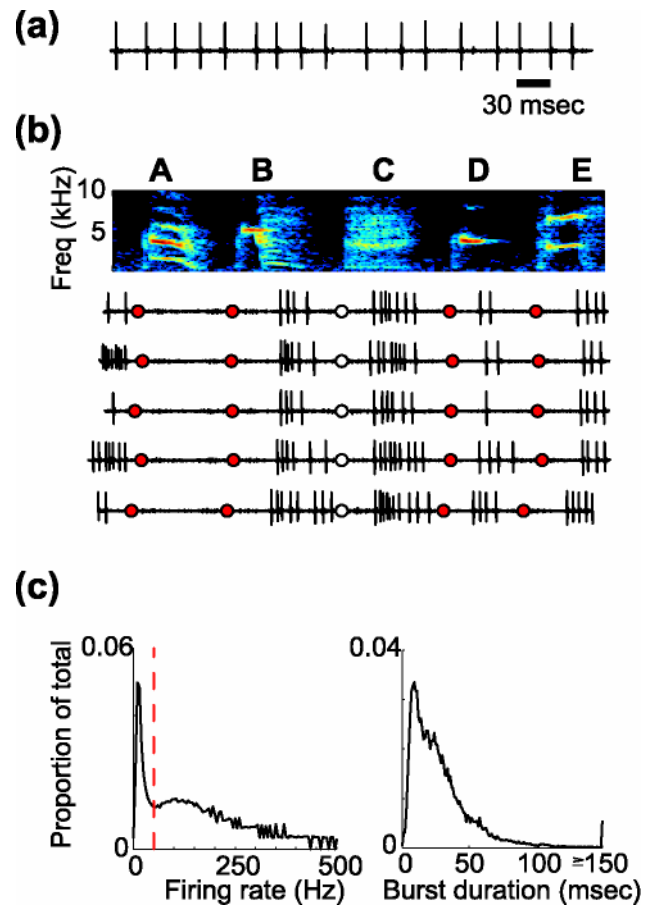
Figure 3.3. Neural activity in Bengalese finch RA. (a) Extracellular recording of an RA single unit in a nonsinging, awake bird. (b) Firing of the same unit during singing. Neural recordings are shown for five repetitions of the song motif "ABCDE," shown in the spectrogram at top. The five neural traces are aligned to the onset of syllable "C" (white dots). Onsets of the other syllables are shown as well (red dots). Data in (a) and (b) are plotted with the same time scale. (c) Left, instantaneous firing rates (1/ISI) during song for all single units. A 50 Hz rate threshold (dashed red line) was used to segregate bursting from nonbursting epochs in order to compute the distribution of burst durations during singing (right).

The qualitative impression that RA neurons fire in bursts during singing was confirmed by the bimodal distribution of instantaneous firing rates (Figure 3.3c, left), defined here as the inverse of interspike intervals (ISI). The peaks of the firing rate distribution (at approximately 12 and 110 Hz) correspond to periods between and within bursts, respectively. The trough between these peaks, which was centered at 50 Hz (red line in Figure 3.3c, left), suggested a criterion for assigning ISIs either to an ongoing

burst (if they were less than 1/50 Hz = 20 msec long) or to an inter-burst pause. Using

this criterion, we computed the distribution of burst durations shown at right in Figure

3.3c.

Serially recording from many units in the same bird allowed us to investigate how

premotor activity is distributed across the population of RA neurons. Figure 3.4a shows

examples of spiking activity recorded from 25 units in a single bird.
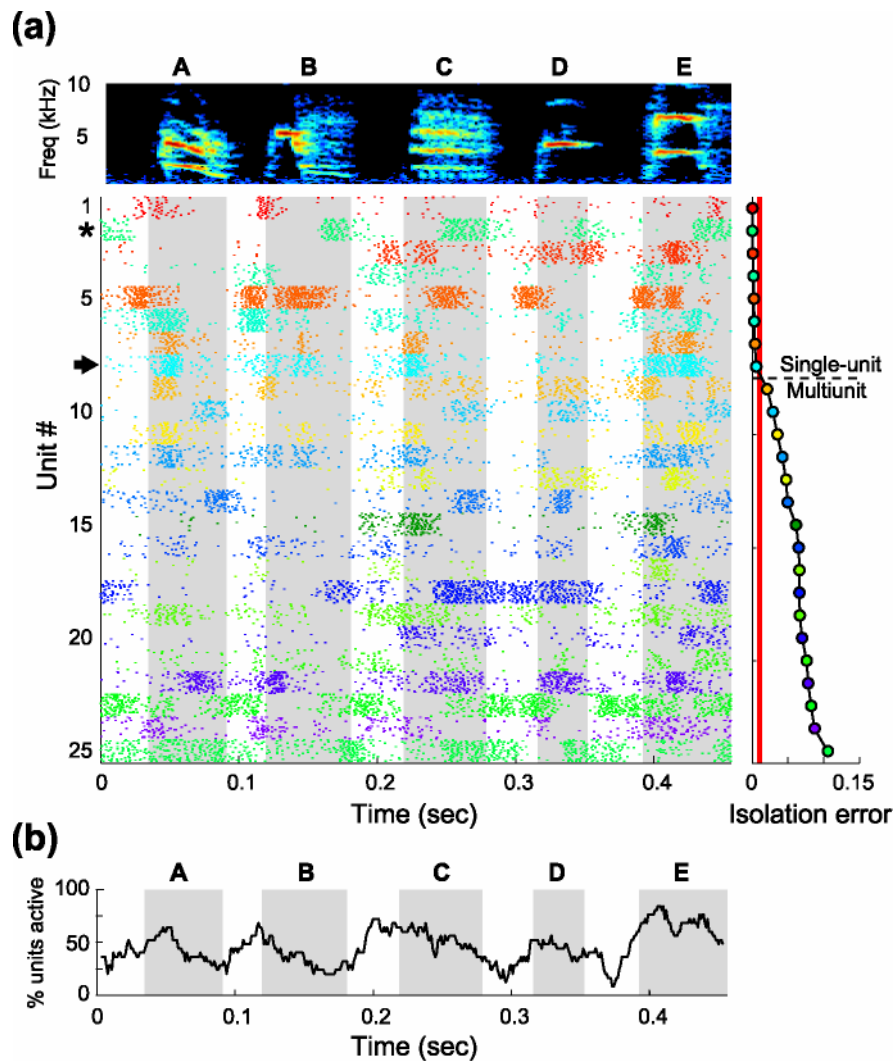
Figure 3.4. (a) Population activity and unit isolation estimates of 25 units recorded from
Bird 1. Each tick mark represents one spike, each row represents the activity during one
iteration of the song motif "ABCDE," and 20 iterations from each unit are shown. Colors
differentiate recordings from the 25 units. Mean syllable durations are shown as gray
boxes. The unit shown in Figure 3a,b is unit 2 (* at left), and the unit shown in Figure 5a
is unit 8 (arrow at left). Plot at far right shows the estimated isolation error

Figure 3.4 continued (see Methods). Units with isolation errors of less than 0.01 (red line) were classified as single units. (b) Fraction of units active (mean rate $\geq$ 25Hz) as a function of time.

Spike times (colored tick marks) are aligned relative to the mean duration of syllables (gray boxes) and inter-syllable pauses using piecewise-linear timewarping. Both single- and multiunit recordings displayed bursts of spikes throughout the song. As described in Methods, we computed the proportion of units active at each time during song. Figure 3.4b shows that this proportion varied considerably over time. Averaging across time and combining data across birds, we found that 58 +/- 19% (mean +/- SD) of the recorded population was active at any given time during singing.

Due to the temporally sparse nature of RA activity, not all units were active preceding all syllables. For example, unit 2 in Figure 3.4a (indicated by an asterisk at left and also shown in Figure 3.3b) was not active prior to syllable "B". When comparing neural variability to acoustic variability, we therefore restricted our analysis to cases in which the recorded unit was active prior to the syllable in question.

Prevalence and strength of neuron-behavior correlations:

We quantified premotor spiking activity and three acoustic measures (pitch, amplitude, and spectral entropy; see Methods) for each recorded unit and syllable. Previous studies have shown these three acoustic parameters to be under the control of the song system (Tchernichovski et al., 2001; Fee, 2002; Fee et al., 2004; Kao et al., 2005), and our PCA-based analysis shows that in most cases these features dominate the trial-by-trial spectral variation of each syllable (see "Quantitative analysis of spectral variability" in Appendix 3). Premotor neural activity was quantified by counting the

number of spikes in a 40 msec premotor window (Figure 3.5a). Figure 3.5b shows the

distribution of pitches of syllable "E" in Bird 1.
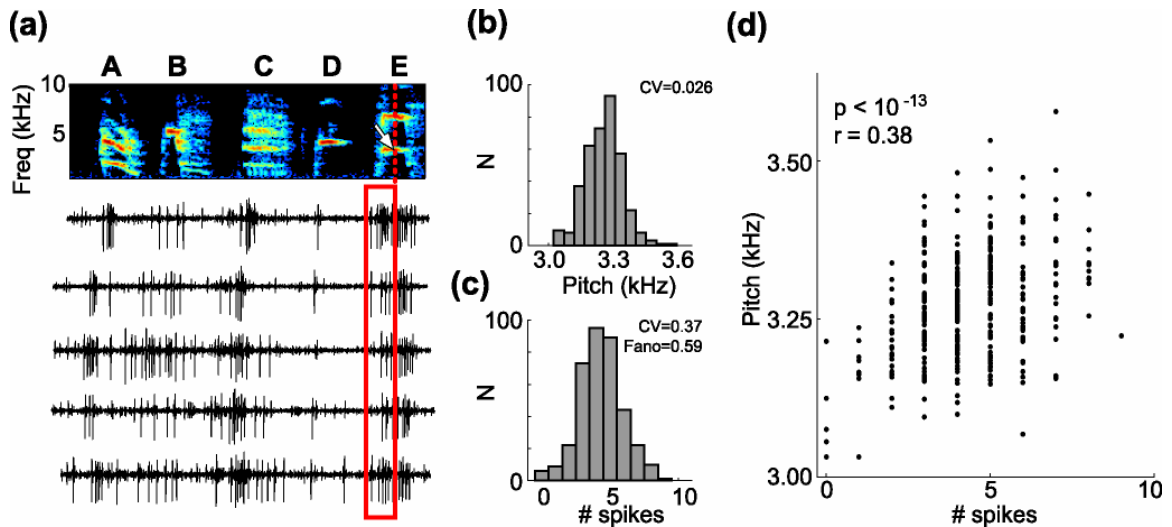
Figure 3.5.   Correlation between neural and behavioral (pitch) variation.   (a) Spectrogram of a song motif from Bird 1. Below the spectrogram are five neural traces showing the activity of unit 8 from this bird (see arrow in Figure 4a). Premotor neural activity for this unit was quantified in a 40 msec window (red box) preceding the time of pitch measurement (dashed red line).   White arrow indicates the measured pitch of syllable "E". Neural traces are aligned on the onset of syllable "E". (b) Distribution of pitches measured from syllable "E" while recording from the unit shown in (a). (c) Distribution of the number of spikes within the premotor window preceding syllable "E". (d) Plot showing the relationship between pitch and the number of spikes in the premotor window for each rendition of syllable "E" (n=371), along with the correlation coefficient (r) and p-value of the correlation.

The insets in Figure 3.5b and c quantify pitch and neural variation using the coefficient of

variation (CV, equal to S.D./mean) and Fano factor (variance/mean). The fundamental

question we address in this study is whether some of the observed behavioral variation

can be explained by the variation in premotor neural activity illustrated in Figure 3.5a and

quantified in Figure 3.5c. To do this, we measured the correlation between pitch and

premotor neural activity (Figure 3.5d). The example shown here yielded a highly

significant ($p<10^{-13}$) positive correlation with an r^2 value of 0.15 (indicating that premotor neural activity could account for approximately 15% of the behavioral variation).

We repeated this analysis for each acoustic parameter and for each case in which a recorded unit was active prior to a given syllable. The distributions of CVs and Fano factors across all syllables and neurons are summarized in Table 3.1.

Measure	Mean +/- SD
CV, spike #, single units: (across syllables)	0.46 +/- 0.23 (across recording sites and syllables)
CV, spike #, multiunits:	0.60 +/- 0.21
Fano factor, spike #, single units:	0.66 +/- 0.55
Fano factor, spike #, multiunits:	0.98 +/- 0.41
CV, pitch:	0.053 +/- 0.052 (across syllables)
CV, amplitude:	0.056 +/- 0.026
CV, entropy:	0.22 +/- 0.14

Table 3.1: Quantification of the degree of neural and behavioral variation. Measures of spiking variation were calculated from the distribution of spike counts in the 40 msec premotor window. A separate CV (SD/mean) and Fano factor (variance/mean) was computed for each case in which a unit was active (mean rate $\geq$ 25 Hz, see Methods) prior to a syllable (n = 705 cases; the mean +/- SD of each distribution is reported). Measures of behavioral variation were calculated by pooling measurements of pitch, amplitude, or entropy from all recorded renditions of each syllable (n = 83 syllables).

Figure 3.6 shows the distribution of significant correlations in Bird 1 for each syllable, recorded unit, and acoustic feature (similar representations of data from the other 12 birds in our study are provided in Appendix 3, Figure 3.9). In the example shown in Figure 3.6a, there were 91 cases (the total number of dots) in which units were active before a syllable. Of these 91 cases, significant correlations were found in 17.6% (16/91) of

cases. These are indicated here as green dots for cases with significant positive

correlations and red dots for cases with significant negative correlations.
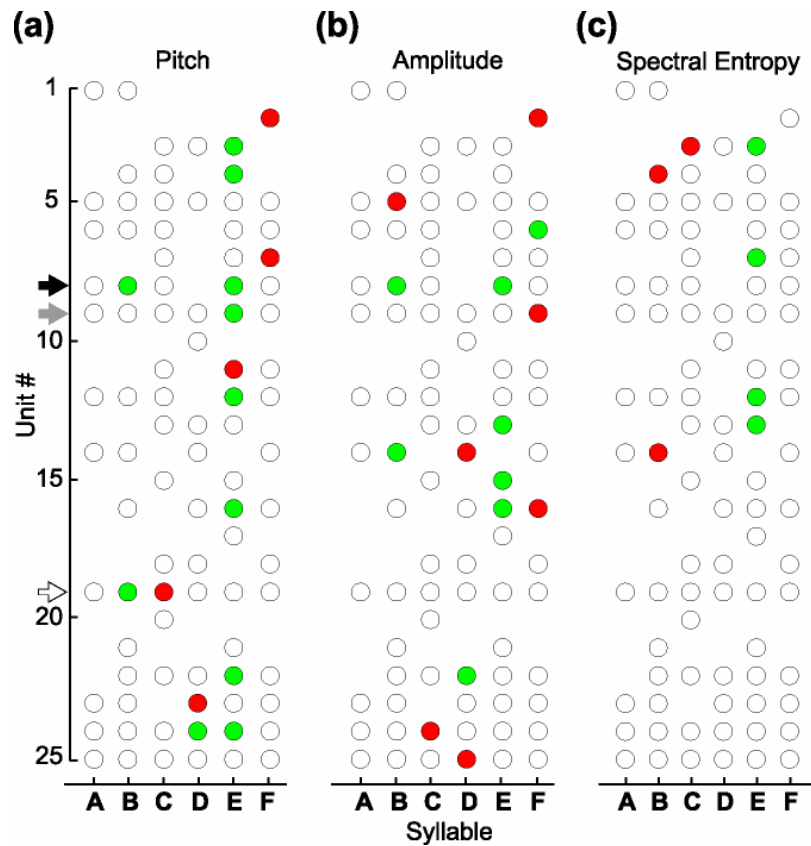

Figure 3.6: Significant correlations of premotor neural activity with acoustic structure across a population of RA units. (a) Correlations with pitch for all units recorded in Bird 1. Each row represents data from one single- or multiunit recording (see Figure 4), and each column represents firing in the premotor window before one of the six syllables produced by the bird (as indicated at bottom: ABCDEF). A dot indicates that the unit in question was active (mean rate $\geq$ 25 Hz in the premotor window preceding a syllable). Dot color indicates whether neural activity was positively correlated (green), negatively correlated (red), or uncorrelated (white) with pitch. Black, gray, and white arrows indicate units 8, 9, and 19, respectively, as referred to in text. (b,c) Same conventions as (a), but dot color signifies correlations between premotor neural activity and amplitude (b) or spectral entropy (c).

Since we expect that some comparisons would yield significant correlations by chance,

we used a resampling-based technique (see Methods and "Multiple comparisons" in

Appendix 3) to ask whether the proportion of significant correlations was itself

significantly greater than chance.   Applying this test to the example of pitch in Bird 1,

we found that the proportion of significant correlations was highly significant (p<0.001).

When data from all birds were combined, the proportion of correlated cases was

significantly different from chance for all measured acoustic parameters (p<0.001 in all

three cases). As shown in Figure 3.7a, premotor activity was correlated with pitch and

amplitude in 26.1 and 26.6% of cases, respectively. Both of these parameters were

correlated with premotor neural activity significantly more frequently than was entropy

(20.8% of cases).
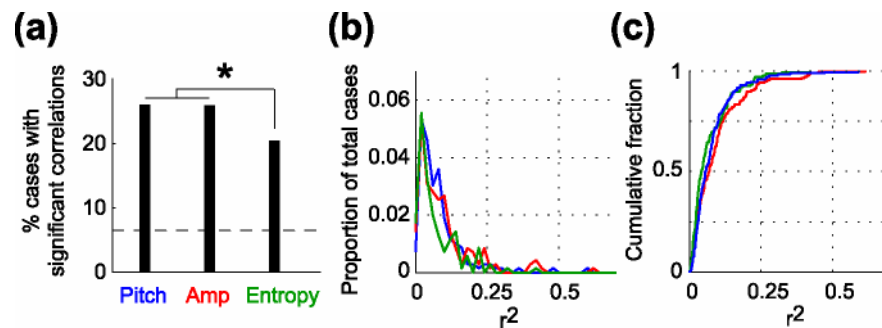
Figure 3.7. Prevalence and explanatory power of neuron-behavior correlations. (a) Bar plot of the proportion of cases in which neural activity was significantly (p<0.05) correlated with each of the three acoustic parameters measured. Note that one "case" corresponds to a unit's being active in the premotor window before a single syllable, and that one unit can therefore contribute several cases to each acoustic feature. Black lines above bars indicate that the proportions for pitch and amplitude were both significantly greater than the proportion for entropy (p<0.05, Z-test for proportions). The dashed line shows the significance threshold for the proportion of cases correlated exceeding chance at p<0.05, as determined by a permutation test (see Appendix 3). All three acoustic parameters were correlated with premotor activity significantly more often than chance. (b) Probability density functions and (c) cumulative distributions of r^2 values for the three acoustic parameters, corresponding to the fraction of behavioral variability in each case that can be accounted for by neural activity. Color conventions are the same as in (a).

In principle, a given unit might have a consistent relationship to an acoustic

feature (such as pitch) across all syllables for which it was active. This was not the case.

Firing rates were typically correlated with an acoustic parameter in some contexts but not

others. For example, unit 8 (black arrow in Figure 3.6a) was positively correlated with

pitch during syllable "B" and "E," but not during other syllables. Also, neural activity was sometimes correlated with different acoustic parameters in different contexts. This pattern can be seen in unit 9 (gray arrow in Figure 3.6a), which was positively correlated with pitch during syllable "E" and negatively correlated with amplitude during syllable "F." Finally, for a single acoustic parameter, some units displayed correlations of opposite signs during different syllables. For example, unit 19 (white arrow in Figure 3.6a) was positively correlated with pitch during syllable "B" but negatively correlated during syllable "C". Across the entire dataset, the sign of a significant correlation between premotor activity and an acoustic parameter during one syllable was not predictive of the sign of significant correlations during other syllables. See "Sparse distribution of significant correlations" in Appendix 3 for a fuller discussion of these issues.

The r^2 values for significant correlations showed that appreciable amounts of behavioral variability could be predicted from the activity of individual units. Figure 3.7b and c show the probability density functions and cumulative distributions of r^2 values from significantly correlated cases. For all acoustic parameters, r^2 values were distributed densely below 0.15 and had a long tail of larger values extending beyond 0.25. Mean r^2 values for pitch, amplitude, and entropy were 0.08, 0.09, and 0.07, respectively. The prevalence (proportion of cases correlated) and explanatory power (r^2 values) of correlations with pitch, amplitude, and entropy show that variations in the activity of individual units can predict a substantial amount of trial-by-trial behavioral variation across a range of acoustic features. Together, these data indicate that a component of motor variation is centrally generated at the level of RA.

Additionally, we collected a small amount of data during female-directed song in order to compare the overall level of neural variation across social contexts (see "RA activity in directed vs. undirected song" in Appendix 3). Previous studies have established that directed song is less variable than undirected song (Sossinka and Bohner, 1980; Kao et al., 2005; Kao and Brainard, 2006; Sakata et al., 2008), suggesting that trial-by-trial variability in RA activity might similarly be reduced in the directed condition. As shown in Appendix 3, Figure 3.10, neural variability was indeed significantly lower during directed song. While preliminary, these results suggest that not only do trial-by-trial variations in RA activity drive variations in song (the main finding in our study), but also that modulations in the overall level of RA variability can account for social context-dependent changes in song.

The analyses presented here examine correlations between acoustic features and the amount of neural activity in an immediately preceding 40 msec premotor window. This window was chosen based on previous studies to reflect the likely causal delay between RA activity and the acoustic structure of song (see Methods). However, in principle correlations between activity and behavior could extend across neighboring syllables. For example, if the pitches of successive syllables were serially correlated, then RA activity correlated with the pitch of one syllable would also be expected to correlate with the pitch of neighboring syllables. We found that serial correlations in behavior were indeed common in Bengalese finch song, and correspondingly we sometimes found significant correlations between premotor activity measured for one syllable and the acoustic features of neighboring syllables. However, the prevalence of significant neuron-behavior correlations was sharply peaked for the neural activity

occurring in the 40 msec window immediately preceding measured acoustic features (see "Correlations extend across time" in Appendix 3), indicating that neural activity in the premotor window used in this study had significantly more predictive power than neural activity preceding adjacent syllables.

Despite differences in unit isolation, single-unit and multiunit recordings were correlated with acoustic parameters in roughly equal proportions and yielded correlations with similar explanatory power. Across the entire dataset, the proportions of cases from single-unit and multiunit recordings with significant correlations (27.9 and 24.1%, respectively) were not significantly different (Z-test for proportions, p=0.24). Also, distributions of r^2 values for the two classes of recordings were not significantly different for any of the three acoustic parameters (2-tailed t-tests, smallest p-value = 0.44) or when all r^2 values were pooled (p=0.60).

Positive correlations with pitch:

Inspection of data from both single-unit and multiunit recordings revealed a strong asymmetry in correlations with pitch but not with other acoustic parameters. In the example shown in Figure 3.6a, positive correlations with pitch (green dots) were present in greater numbers than negative correlations (red dots). In 5 out of 5 birds (including this one) where there was a significant difference in the number of positive and negative correlations with pitch, positive correlations outnumbered negative ones. When data were combined across all birds, this asymmetry was significant for data from both single-unit and multiunit recordings (Figure 3.8a). This asymmetry in the sign of

correlations with pitch supports a model of pitch production based on recent anatomical and physiological studies (see Discussion).
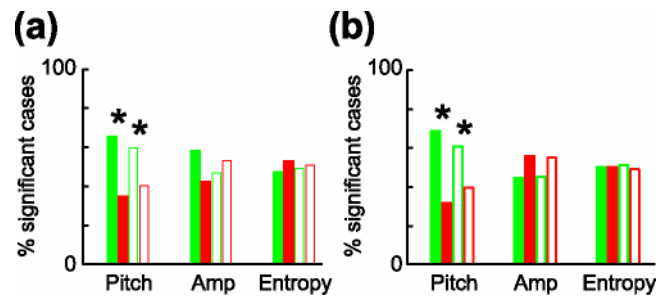
Figure 3.8. Correlation signs. Bars show the fraction of significant correlations with positive (green) and negative (red) slopes for single-unit (filled) and multiunit (empty) recordings. Asterisks indicate a ratio of positive to negative correlations significantly different from equality (p<0.05, binomial test). (a) shows results from the primary analysis, (b) shows results from the partial correlation analysis (see Methods).

In the analyses presented thus far, the relationship between neural activity and each of three acoustic parameters was assessed in separate correlation tests. However, examination of our behavioral data revealed that in many cases, the measured acoustic features were significantly correlated with each other (not shown). Some of the neuron-behavior correlations revealed in our initial analysis might therefore arise because of correlations between behavioral measures. For example, a correlation between neural activity and syllable amplitude might result from the combined effects of a correlation between neural activity and pitch and a correlation between pitch and amplitude. To disambiguate the correlations between premotor activity and each acoustic parameter, we performed a partial correlation analysis, in which the relationships between neural activity and each behavioral measure was assessed while controlling for the influence of correlations with the other two acoustic parameters (see Methods). This alternate analysis yielded nearly identical results as the primary analysis (Figure 3.8b).

Discussion:

Our results show that trial-by-trial variations in RA activity predict a significant component of acoustic variation in song syllables. We found correlations between premotor activity and all three acoustic parameters examined, although correlations with pitch and amplitude were found significantly more often than correlations with spectral entropy. Additionally, correlations with pitch had a positive sign in a significant majority of cases. Together, these results provide strong evidence that trial-by-trial variations in syllable structure result in part from variations in the motor command.

By exploiting trial-by-trial variability at particular times during song, our analysis provides the first description of covariation between RA activity and syllable structure. Using a contrasting approach, Leonardo and Fee (2005) found that mean population activity was on average uncorrelated with mean spectral output across different times in song, demonstrating that similar acoustic patterns can be produced by unrelated ensembles of RA cells. Our results are complementary to these prior findings. We show that within each ensemble of active neurons, trial-by-trial variations in activity can account for variations in behavior.

Comparison of our results with recent studies in primates suggests similarities between the neural control of birdsong and primate reaching movements. The activity of single neurons in motor, premotor, and parietal cortex is often correlated with multiple parameters describing reach kinematics (Fu et al., 1995; Buneo et al., 2002; Wang et al., 2007), just as RA neurons often appear to encode multiple acoustic parameters. Furthermore, although correlations between cortical activity and kinematic parameters

(such as hand position or velocity) vary widely in strength, the r^2 values we report in RA fall within the range reported for several areas of motor and premotor cortex (Carmena et al., 2005; Stark et al., 2007). These similarities suggest that the generation of variable motor commands using populations of neurons each moderately correlated with several task parameters might be a general principle of skilled motor control.

Our results provide an initial characterization of RA in the Bengalese finch. Consistent with recordings in zebra finch RA (Yu and Margoliash, 1996; Leonardo and Fee, 2005), the neurons described here are tonically active at rest, fire syllable-locked bursts during song, and are transiently inhibited after song offset. In contrast, RA neurons in the Bengalese finch have a lower peak firing rate during bursts and fire bursts of greater duration (Figure 3.3c) than their counterparts in the zebra finch (Leonardo and Fee, 2005).

Our analysis describes correlations between RA activity and acoustic output. However, the strength of such a correlation does not necessarily reflect a neuron's causal influence. Neurons in RA might covary such that an increase in one cell's firing is often accompanied by an increase in the firing of other cells, which make their own contributions to acoustic output. The measured correlation between neural activity and a behavioral parameter therefore depends both on the neuron's ability to drive changes in behavior and on its correlation with other RA neurons.

At one extreme, neural activity could vary independently across RA neurons. In this case, the correlation between each neuron's premotor activity and pitch (for example) would accurately reflect that cell's contribution to the total behavioral variability. If a small pool of independently varying neurons controlled pitch, these correlations would be

strong, reflecting the small number of neurons governing behavior (Figure 3.2a). Conversely, if pitch were controlled by a large number of independent neurons, correlations between activity and pitch would be weak (Figure 3.2b).

At the other extreme, neural activity could be strongly correlated across many RA neurons. In this case, the measured correlation between any one cell and pitch would include the contributions of the entire correlated ensemble. These correlations could be quite strong, since they reflect the contributions of many neurons (Figure 3.2c).

We can distinguish between these possibilities (Figure3. 2a-c) by estimating the number of neurons that control acoustic variation. In the zebra finch, right and left RA each contain ~8,000 neurons that project to brainstem motor nuclei (Gurney, 1981). Assuming a similar figure for Bengalese finch RA, which is of comparable volume to zebra finch RA (Tobari et al., 2005), we can estimate the number of neurons controlling pitch during each syllable. Of 16,000 total projection neurons, our data indicate that about 60% are active at any given time, and that of these approximately 25% make significant contributions to the control of pitch. Assuming that our recordings represent a uniform sampling, we can therefore estimate that about $16,000 \times 0.60 \times 0.25 = 2,400$ RA projection neurons control pitch at any given time. (A similar figure is obtained for the number of neurons controlling amplitude, and a smaller number for spectral entropy, reflecting the smaller proportion of cases with significant correlations.) If each of these neurons contributed equally to pitch, then each would contribute 1/2,400 of the total behavioral variation (in the absence of downstream motor noise). If the activity of all neurons were independent, the measured correlation between each unit's activity and pitch would have an r^2 value of $1/2,400 = 0.000417$. Alternately, if RA neurons were

strongly correlated, recording from any one neuron could have as much predictive power as recording from the entire population, and r^2 values would be far higher than those expected from an independent population.

The measured distribution of r^2 values (Figure 3.7b,c) suggests that covariation between RA neurons is common. We found r^2 values far larger than expected from a population of independent neurons, with a mean r^2 value (0.08 for pitch) nearly 200 times larger than the value predicted by the independent-activity model (0.000417) shown in Figure 3.2b. Put another way, only 13 independent RA neurons with r^2 values at the mean of our observed distribution could in principle account for 100% of the behavioral variation. Since the number of neurons correlated with each acoustic parameter is far larger than this (~2,400 neurons), some of the explanatory power of the measured correlations must arise from covariation between RA neurons (Figure 3.2c), ruling out a model in which a small number of independent neurons drive behavioral variation (Figure 3.2a). Covariation across RA might rely on networks of inhibitory interneurons that coordinate the activity of spatially separated projection neurons (Spiro et al., 1999). Although our calculations are based on rough estimates of neuron number and the prevalence of significant correlations, the difference between the empirical r^2 values and those expected from independent neurons is large enough to allow robust conclusions. The prevalence and strengths of neuron-behavior correlations therefore point to a model of motor variation in which "cooperating" (that is, covarying) assemblies of a few thousand neurons produce trial-by-trial modulations of song (Figure 3.2c). To the extent that acoustic variations are driven by the AFP (Kao et al., 2005; Olveczky et al., 2005),

our results also suggest that lMAN drives coherent modulation of a pool of RA neurons rather than injecting independent noise across RA.

Although a single unit's activity could account for as much as 40% of the variation in an acoustic parameter (Figure 3.7b,c), the fraction of behavioral variation controlled by the entire RA population is unknown. While RA is the sole output nucleus of the motor pathway, the brainstem motor nuclei controlling song additionally receive inputs from other parts of the brain (Wild, 2004), which may also contribute to premotor variation. Furthermore, peripheral motor noise presumably contributes to song variability as well. Note that if RA drives less than 100% of the behavioral variation, the r^2 value predicted by the independent-activity model would be even lower than 0.000417.

The observed predominance of positive correlations with pitch (Figure 3.8) is consistent with the functional anatomy of the descending motor system. As schematized in Figure 3.9, it is likely that increases in RA activity ultimately result in a net increase in the pitch of song.
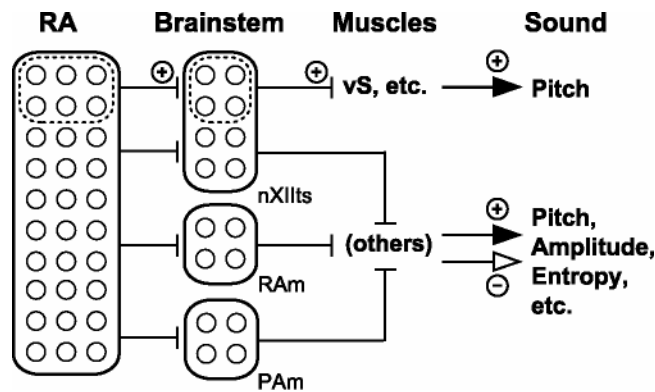
Figure 3.9. Acoustic control in the descending motor pathway. In our schematic, neurons are represented as white circles, synaptic connections are represented as lines connecting brain regions and muscles, and causal influences on acoustic structure are represented as arrows. RA projection neurons excite motor neurons in brainstem nuclei nXIIts, RAm, and PAm, which in turn activate the muscles controlling the syrinx and respiratory apparatus (Sturdy et al., 2003). Recordings from syringeal muscles reveal strong positive correlations between muscle activity and pitch, and one muscle in

The observed surplus of positive correlations may therefore reflect a subpopulation of RA cells responsible for activating (via the brainstem) muscles that drive increases in pitch. Our data suggest that birds modulate song by distributing variation across a few thousand neurons, thereby allowing them to explore the sensory consequences of varying the motor command. This motor exploration might be guided by differential reinforcement signals related to overall song quality (Tumer and Brainard, 2007). Alternately, by listening to these variations, adult birds could monitor the relationship between small changes in neural activity and small changes in acoustic structure. Knowing this relationship constitutes a local (that is, local to a single syllable) model of motor production. Maintenance of such a model might be necessary for the animal to adapt to changes in the strength of motor effectors as the bird ages or to changes in synaptic strength or connectivity over time. Song deteriorates dramatically when auditory feedback is removed in adulthood (Nordeen and Nordeen, 1992; Okanoya and Yamaguchi, 1997). Such deterioration might result from the inability of the bird to hear the consequences of motor exploration and thus maintain motor performance in adulthood.

General Discussion

The focus of the current study was to explore the generation of variability in adult birdsong. We used the songs of the Bengalese finch because they exhibit variability on two levels: individual syllables exhibit variability in production from iteration-to-iteration, and the syllables themselves are variably sequenced during song production (Okanoya, 2004). The occurence of both sources of variability in Bengalese finch song provides us with a particularly good model for studying the production of human language: they are constructed of elemental vocal units, called phonemes, which are used in a variety of sequences to produce words and sentences (Daniloff and Hammarberg, 1973). We have examined the song motor nucleus RA in an effort to understand the influences of both varieties of variability upon neuronal activity in the avian CNS.

Implications of Chapter 1:

We chose to focus our research on the Bengalese finch because of the variable nature of its song. As in human language, the Bengalese finch sequences elemental vocal units (syllables) into long, variable strings of vocalizations (songs). The effects of variable sequencing in human vocalizations have been studied extensively, but it was unknown if syllables in the songs of birds were similarly modified by sequence. In the variably sequence songs of the Bengalese finch, we found left-to-right 'carry-over' effects of sequence, but not any right-to-left 'anticipatory' effects.

The results of our behavioral analysis of Bengalese finch song suggest that syllables are modified by the history of the song system. The lack of any anticipatory effects at divergence points implies that the birds are not actively planning for upcoming motor transitions. It is still unclear whether our birds are actively controlling convergent syllable transitions, or if the effects upon syllable production are completely the result of a lack of control at these song junctions. For either scenario, we can assume that syllables are affected by those syllables that preceded them, and the rules governing these transitions must therefore be learned throughout song development.

It would be interesting to probe these local effects of syllable sequence further by manipulating song feedback during singing. We found that the removal of auditory input did not change the local effects of sequencing. This suggests that proprioception may be more involved in syllable transitions than audition. By modifying proprioceptive feedback (changing air sack pressure, beak gape, syringeal tensions, etc.), we may be able to determine if and how birds make compensations for syllable sequencing. If birds do compensate for exogenous manipulation of proprioceptive feedback, it suggests that the vocal musculature is actively shaping convergent syllable transitions. Bengalese finches may therefore be very useful in developing behavioral manipulations designed to modify vocal sequencing in other systems (i.e. human language and other variable sequenced behaviors).

We also found that syllable structure covaried with sequence over much longer time frames than a single syllable. These 'global effects' encompassed upwards of 25

syllables at times. We believe that the global effects upon syllable production are the result of different motor plans. Although the syllables analyzed for global effects were found in the same immediate sequences, the fact that their structure is still significantly different suggests that at some level in the CNS the encoding is different. Considering the hypothesized role HVC plays in sequence generation (Hahnloser et al, 2002), neural recordings from this motor nucleus would be informative about the source of global sequence effects.

Future experiments may be able to look more explicitly at the control of syllable transitions by analyzing HVC activity in light of the global sequence effects upon syllable production. We predict that HVC activity will be subtly different for these syllables that are essentially in the same motor context. If HVC activity is different, it suggests that birds have a way of keeping track of long-range patterning in syllable sequence. This information may be used by the bird to know what sequences have been sung, and what sequences should be sung.

Implications of Chapter 2:

The second chapter investigated the relationship between RA activity and syllable phonology. We examined mean changes in the patterns of RA activity and syllable structure, and asked if there was a general relationship between RA activity and syllable structure. We found that general patterns in RA activity did correlate with changes in spectral structure, suggesting phonological encoding at the level of RA.

Previous research on the role of RA during syllable production concluded that RA activity has no relationship to syllable structure (Leonardo and Fee, 2005). The current study suggests that there is a relationship, and these results in combination with those of the first chapter provide a more complete picture of syllable phonation during song production. In the first chapter, we found that RA could change syllable features by modifying the firing rate of neurons. This analysis was performed in restricted windows for both the neural activity and the syllable. In the third chapter, we instead measured patterns of neural activity and spectral structure over the duration of the syllable. From this analysis, we were able to provide a more general perspective on syllable production. Our interpretation of the results of Chapter 1 and 3 is that at any time point in song, RA activity can manipulate specific features of the acoustic output, and these changes are then tiled across time to produce the appropriate syllable.

There are different ensembles of neurons active in RA for each time point, and they work in concert to produce all of the command signals necessary for the syrinx and

respiratory centers. In Chapter 1 we found that the effects of any one neuron might change from syllable to syllable, so for each time point in song, each neuron in the active ensemble will make a different contribution to the spectral output. These ensembles of neurons are temporally arranged in order to provide a continuous command for the vocal musculature. When we record from one particular neuron, or collection of neurons, we are recording an ever-evolving program for muscle activations. In our analysis examining the entire premotor window and the entire syllable, we are measuring these evolving global patterns in RA activity. Our extended analysis of neural and behavior in Chapter 3 found a relationship between the changing actions of a single pool of RA neurons and the changes in vocal output.

The influence of any one pool of neurons is altered by the activity of other pools of neurons at different time points. RA neurons encode muscle tensions and respiratory pressure and not spectral features; but at one specific time point the activity of a given ensemble of neurons has a reliable effect upon spectral features. We were able to extrapolate to broader patterns of neural and syllable structure in Chapter 3 because we restricted our analysis to syllables with relatively small mean changes in production (convergent and divergent syllables). For these syllable varieties, we believe that the same ensemble of RA neurons is active. The mean changes in syllable structure for these syllables are produced by a correlated amount of change in the activity of that specific ensemble of neurons. For more disparate syllable varieties, the many different interactions downstream of RA complicate the neural-behavior relationship.

We were able to find a weak relationship between single neuron activity and more disparate syllable types, but not for the multi-unit recordings. We believe this is the case for several reasons. In the multi-unit recordings, each neuron that is contributing to the summed signal has a different influence upon the vocal musculature at any given time point. If one recording includes the activity from 5 neurons and another is a single unit recording, at each time point in song the multi-unit signal is going to have at least five-times as many different downstream interactions as the single-unit recording. As a result, we can still measure a weakened relationship for a single unit across different syllables, but not for multi-unit recordings.

The results of Chapter 2 could be expanded upon by making recordings from the vocal musculature and RA during song. By using a known model of the syrinx (Suthers and Sollinger, 2004), simultaneous recordings in RA and the periphery may reveal a relationship between ensembles of RA activity and syllable structure at every time point in song. These results would be useful in furthering our understanding on how the bird brain directs specific changes to song throughout life. A bird is constantly integrating sensory and motor information, and understanding how motor changes are made may bring to light how auditory information can drive those changes.

Implications of Chapter 3:

Understanding the sources and effects of variability in adult songs reveals two different facets of avian song control: the online control of song features, as well as

possible mechanisms for changing those song features. In order to look at the control of syllable structure, we examined the relationship between the activity of neurons in RA and specific changes to the spectral parameters of song (Chapter 3). In a significant proportion of the RA neurons assayed, we found a direct relationship between the magnitude of neural activity and specific changes in the pitch, amplitude and entropy of syllables. These results demonstrated a central source of the iteration-to-iteration variability of syllables. As mentioned previously, if the residual variability in adult song has some central component to it, it can then be used to modify syllable features. The discovery of RA's involvement in adjusting syllable features provided the first information on how a bird can make specific changes to a syllable by manipulating neural activity in the CNS.

In previous research, the AFP nucleus lMAN has been investigated for its role in generating song variability. It has been shown in both juvenile and adult birds that lMAN is important to syllable variability (Brainard, 2006; Bottjer, 1994; Scharff, 1991). lMAN activity in turn affects RA, and it is believed that most of its effects upon song are mediated through this interaction (Wild, 2004). The results of Chapter 3 provide the first demonstration of how differences in lMAN activity may act to change song features. Our results on the activity of RA neurons represent a method by which changes in lMAN firing can affect syllable production. Patterns of variable activity in lMAN can increase or decrease the amount of activity in RA. We now have an understanding of how RA neurons can directly change song features, and have the first components of a model for how lMAN and RA cooperate to modify syllables. Also informative to this model are the

effects of microstimulation of lMAN during song. Stimulation of lMAN causes both increases and decreases in the magnitude of syllable features (Kao et al, 2005). Our own research on RA implies that the excitation or inhibition caused by lMAN's input can affect syllable structure by simply adjusting the firing rate of RA neurons. The results of these studies propose the following model: changes in lMAN activity directly influence the excitability of RA neurons, and RA activity then drives specific changes in syllable spectral features.

The interactions between lMAN and RA that serve to maintain adult song fidelity may have a more instructive role during song development. Recent work has shown a greater influence of the AFP during early song production than the motor pathway (Aronov et al, 2008). These results imply that the AFP may be the primary instructor to the motor pathway during song learning. The influence of lMAN upon RA would therefore be extremely important during the process of honing juvenile song into adult song. The results of Chapter 3 suggest that lMAN makes adjustments to the firing of individual neurons in RA, thereby altering features such as pitch, entropy, and amplitude of the syllables. Over time, adjustments by lMAN become less and less necessary as the bird practices his song. The motor pathway eventually takes over as the primary circuit for song production. lMAN still retains the ability to make changes to song throughout life, but the influence of the AFP is reduced with age. In this way, the interactions of lMAN and RA serve two different, but related functions in song generation for juvenile and adult birds.

In order to further explore the interactions of lMAN and RA during song production, there are several possible experiments that could be performed. As a first experiment, it would be interesting to record from RA while temporarily inactivating lMAN. These technologies have already been utilized separately (current study and Olveczky et al, 2005), but the combination of both would elucidate the general influences of lMAN onto RA neurons. We predict that the immediate effect of lMAN inactivation upon RA would be to globally decrease the variability in firing of RA neurons. It has been shown that lMAN inactivation decreases syllable variability, and our results suggest that this is mediated by changing the firing of RA neurons. An experimental demonstration of this hypothesis would expand our understanding of the AFP's influence on the motor pathway.

A logical next step after inactivation studies would be to make simultaneous recordings in lMAN and RA during song production. For this experiment, it would be particularly beneficial to record from connected areas of lMAN and RA. Utilizing the topography of the connection between these two nuclei (Lou et al, 2001), along with antidromic stimulation, would aid in the localization of both recording electrodes. The results of this experiment would expand our understanding of how lMAN neurons shape the excitability of RA neurons on a moment-by-moment basis. Analysis of the contribution of lMAN to RA firing will also help disambiguate the input from HVC. With these experiments we could provide a more complete picture of song control by the motor and anterior forebrain pathways.

General Implications

The results presented above demonstrate that syllable production is affected by syllable sequencing, that these differences in production correlate with differences in activity in RA, and that specific changes to syllable structure correlate with specific changes in RA activity. The general interpretation of these data is that RA is implicitly involved in the generation of spectral features of song syllables. The results of Chapter 3 indicate a method by which the song system can make subtle adjustments to syllable structure, and Chapter 2 demonstrates the broader relationship between RA similarity and syllable structure. The way in which syllable are modified around the mean in Chapter 3 may therefore be indicative of how more drastic changes in syllable structure can be made like those observed in Chapter 1. The rules governing spectral modifications in adulthood are also a likely parallel to the sensory-motor refinement seen in juvenile birds.

Dissertation Bibliography:

Introduction

Bishop DV. (1999). How does the brain learn language? Insights from the study of children with and without language impairment. Developmental Med & Child Neurology 42(2): 133-142.

Bottjer SW. (1994). Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science 224: 901-903.

Bradbury JW, Vehrencamp SL. Principles of Animals Communication. Sinauer Associates, Sunderland, MD. (1998).

Brainard MS. (2004) Contributions of the anterior forebrain pathway to vocal plasticity. Ann N Y Acad Sci 1016:377-394.

Brainard MS, Doupe AJ. (2000). Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. Nature 404(6779): 762-766.

Catchpole CL, Slater PJB. 1995. Birdsong: Biological Themes and Variations. Cambridge University Press, 2nd Edition.

Daniloff, RG, Hammarberg, RE. 1973. On defining coarticulation. Journal of Phonetics, 1: 239-48.

Doupe DJ, Kuhl, PK. (1999). Birdsong and speech: common themes and mechanisms. Annual Reviews in Neuroscience 22:567-631.

Duquid I, Sjostrom PJ.(2006). Novel presynaptic mechanisms for coincidence detection in synaptic plasticity. Current Opin Neurobiology 16(3): 312-322.

Hahnloser, RH, Kozhevnikov, AA, Fee, MS. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature 419(6902): 65-70.

Hessler NA, Doupe AJ (1999) Singing-related neural activity in a dorsal forebrain-basal ganglia circuit of adult zebra finches. J Neurosci 19:10461-10481.

Kao MH, Brainard MS (2006) Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. J Neurophysiol 96:1441-1455.

Kao MH, Doupe AJ, Brainard MS. (2005). Contributions of an avian basal ganglia circuit to real-time modulation of song. Nature 433(7026): 638-643.

Konishi M (2004) The role of auditory feedback in birdsong. Ann N Y Acad Sci 1016:463-475.

Mitra, S, Amazeen, PG, Turvey, MT. 1998. Intermediate motor learning as decreasing active(dynamical) degrees of freedom. Human Movement Science, 17: 17-65.

Mooney R, Konishi M. (1991). Two distinct inputs to an avian song nucleus activity different glutamate receptor subtypes on individual neurons. PNAS 88(10): 44075-4079.

Nottebohm F. (1968). Auditory experience and song development in the chaffinch, Fringilla coelebs. Ibis 110:549-569.

Öhman, SE. 1966. Coarticulation in VCV utterances: spectrographic measurements. Journal of the Acoustic Society of America, 74:1115-1125.

Olvecsky BP, Andalman AS, Fee MS. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biology 3(5): e153.

Roberts, TF, Klein, ME, Kubke, MF, Wild, JM, Mooney, R. (2008). Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. Journal of Neuroscience 28(13):3479-89.

Sakata JT, Hampton CM, Brainard MS (2008) Social modulation of sequence and syllable variability in adult birdsong. J Neurophysiol 99:1700-1711.

Scharff C. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system. J. Neuroscience 11: 2896-2913.

Stark LL, Perkel DJ. (1999). Two-stage, input specific synaptic maturation in a nucleus essential for vocal production in the zebra finch. J. Neuroscience 19(20): 9107-9116.

Sohrabji F. (1990). Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. Behav. Neural Biol 53: 51-63.

Suthers RA, Goller F, Wild M. (2002). Somatosensory feedback modulates the respiratory program and crystallized birdsong. PNAS 99(8): 5680-5685.

Voronin LL. (1983). Long-term potentiation in the hippocampus. Neuroscience 10: 1051-1069.

Wild JM. (1993). Descending projections of the songbird nucleus robustus archistriatalis. Journal of Comparative Neurology 338(2): 225-241.

Wild JM (2004) Functional neuroanatomy of the sensorimotor control of singing. Ann N Y Acad Sci 1016:438-462.

Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. Science 273:1871-1875.

Yuste R, Majewska A, Cash SS, Denk W. (1999). Mechanisms of calcium influx into hippocampal spines, coincidence detection by NMDA receptors, and optical quantal analysis. J. Neuroscience 19(6): 1976-1987.

Chapter 1

Arnold AP (1975) The effects of castration on song development in zebra finches (Poephila guttata). J Exp Zool 191:261-278.

Ashmore RC, Wild JM, Schmidt MF (2005) Brainstem and forebrain contributions to the generation of learned motor behaviors for song. J Neurosci 25:8543-8554.

Buneo CA, Jarvis MR, Batista AP, Andersen RA (2002) Direct visuomotor transformations for reaching. Nature 416:632-636.

Carmena JM, Lebedev MA, Henriquez CS, Nicolelis MA (2005) Stable ensemble performance with single-neuron variability during reaching movements in primates. J Neurosci 25:10712-10716.

Chi Z, Margoliash D (2001) Temporal Precision and Temporal Drift in Brain and Behavior of Zebra Finch Song. Neuron 32:899-910.

Churchland MM, Santhanam G, Shenoy KV (2006b) Preparatory activity in premotor and motor cortex reflects the speed of the upcoming reach. J Neurophysiol 96:3130-3146.

Doya K, Sejnowski T (2000) A computational model of avian song learning. In: The New Cognitive Neurosciences (Gazzaniga M, ed), pp 469-482. Cambridge, MA: MIT Press.

Fee MS (2002) Measurement of the linear and nonlinear mechanical properties of the oscine syrinx: implications for function. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 188:829-839.

Fee MS, Kozhevnikov AA, Hahnloser RH (2004) Neural mechanisms of vocal sequence generation in the songbird. Ann N Y Acad Sci 1016:153-170.

Fiete IR, Fee MS, Seung HS (2007) Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. J Neurophysiol 98:2038-2057.

Fu QG, Flament D, Coltz JD, Ebner TJ (1995) Temporal encoding of movement kinematics in the discharge of primate primary motor and premotor neurons. J Neurophysiol 73:836-854.

Gardner T, Cecchi G, Magnasco M, Laje R, Mindlin GB (2001) Simple motor gestures for birdsongs. Phys Rev Lett 87:208101.

Goller F, Suthers RA (1996) Role of syringeal muscles in controlling the phonology of bird song. J Neurophysiol 76:287-300.

Gurney ME (1981) Hormonal control of cell form and number in the zebra finch song system. J Neurosci 1:658-673.

Hessler NA, Doupe AJ (1999) Singing-related neural activity in a dorsal forebrain-basal ganglia circuit of adult zebra finches. J Neurosci 19:10461-10481.

Kao MH, Brainard MS (2006) Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. J Neurophysiol 96:1441-1455.

Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. Nature 433:638-643.

Kittelberger JM, Mooney R (2005) Acute injections of brain-derived neurotrophic factor in a vocal premotor nucleus reversibly disrupt adult birdsong stability and trigger syllable deletion. J Neurobiol 62:406-424.

Laje R, Gardner TJ, Mindlin GB (2002) Neuromuscular control of vocalizations in birdsong: a model. Phys Rev E Stat Nonlin Soft Matter Phys 65:051921.

Leonardo A (2004) Experimental test of the birdsong error-correction model. Proc Natl Acad Sci U S A 101:16935-16940.

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

Medina JF, Lisberger SG (2007) Variation, signal, and noise in cerebellar sensory-motor processing for smooth-pursuit eye movements. J Neurosci 27:6832-6842.

Nordeen KW, Nordeen EJ (1992) Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. Behav Neural Biol 57:58-66.

Okanoya K, Yamaguchi A (1997) Adult Bengalese finches (Lonchura striata var. domestica) require real-time auditory feedback to produce normal song syntax. Journal of Neurobiology 33:343-356.

Olveczky BP, Andalman AS, Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol 3:e153.

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

Sakata JT, Hampton CM, Brainard MS (2008) Social modulation of sequence and syllable variability in adult birdsong. J Neurophysiol 99:1700-1711.

Sossinka R, Bohner J (1980) Song types in the zebra finch poephila guttata castanotis. Zeitschrift fur Tierpsychologie 53:123-132.

Spiro JE, Dalva MB, Mooney R (1999) Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons. Journal of Neurophysiology 81:3007-3020.

Stark E, Drori R, Asher I, Ben-Shaul Y, Abeles M (2007) Distinct movement parameters are represented by different neurons in the motor cortex. Eur J Neurosci 26:1055-1066.

Sturdy CB, Wild JM, Mooney R (2003) Respiratory and telencephalic modulation of vocal motor neurons in the zebra finch. J Neurosci 23:1072-1086.

Suthers RA, Zollinger SA (2004) Producing song: the vocal apparatus. Ann N Y Acad Sci 1016:109-129.

Tchernichovski O, Mitra PP, Lints T, Nottebohm F (2001) Dynamics of the vocal imitation process: how a zebra finch learns its song. Science 291:2564-2569.

Tobari Y, Nakamura KZ, Okanoya K (2005) Sex differences in the telencephalic song control circuitry in Bengalese finches (Lonchura striata var. domestica). Zoolog Sci 22:1089-1094.

Tumer EC, Brainard MS (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature 450:1240-1244.

Vicario DS, Nottebohm F (1988) Organization of the zebra finch song control system: I. Representation of syringeal muscles in the hypoglossal nucleus. J Comp Neurol 271:346-354.

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

Wang W, Chan SS, Heldman DA, Moran DW (2007) Motor cortical representation of position and velocity during reaching. J Neurophysiol 97:4258-4270.

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

Wild JM (2004) Functional neuroanatomy of the sensorimotor control of singing. Ann N Y Acad Sci 1016:438-462.

Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. Science 273:1871-1875.

Chapter 1 Supplemental

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

Goller F, Suthers RA (1996) Role of syringeal muscles in controlling the phonology of bird song. J Neurophysiol 76:287-300.

Kao MH, Brainard MS (2006) Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. J Neurophysiol 96:1441-1455.

Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. Nature 433:638-643.

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

Olveczky BP, Andalman AS, Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol 3:e153.

Sakata JT, Hampton CM, Brainard MS (2008) Social modulation of sequence and syllable variability in adult birdsong. J Neurophysiol 99:1700-1711.

Sossinka R, Bohner J (1980) Song types in the zebra finch poephila guttata castanotis. Zeitschrift fur Tierpsychologie 53:123-132.

Spiro JE, Dalva MB, Mooney R (1999) Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons. Journal of Neurophysiology 81:3007-3020.

Tchernichovski O, Mitra PP, Lints T, Nottebohm F (2001) Dynamics of the vocal imitation process: how a zebra finch learns its song. Science 291:2564-2569.

Theunissen F, Miller JP (1995) Temporal encoding in nervous systems: a rigorous definition. J Comput Neurosci 2:149-162.

Tiesinga P, Fellous JM, Sejnowski TJ (2008) Regulation of spike timing in visual cortical circuits. Nat Rev Neurosci 9:97-107.

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

Westfall PH, Young SS (1993) Resampling-Based Multiple Testing. New York: John Wiley & Sons, Inc.

Wild JM (1993) Descending projections of the songbird nucleus robustus archistriatalis. Journal of Comparative Neurology 338:225-241.

<u>Chapter 2</u>

Allan, SE, Suthers, RA. 1994. Lateralization and motor stereotypy of song production in the brown-headed cowbird. Journal of Neurobiology, 25(9): 1154-1166.

Bishop DV. (1999). How does the brain learn language? Insights from the study of children with and without language impairment. Developmental Med & Child Neurology 42(2): 133-142.

Catchpole CL, Slater PJB. 1995. Birdsong: Biological Themes and Variations. Cambridge University Press, 2nd Edition.

Daniloff, RG, Hammarberg, RE. 1973. On defining coarticulation. Journal of Phonetics, 1: 239-48.

Doupe, AJ, Konishi, M. 1991. Song-selective auditory circuits in the vocal control system of the zebra finch. PNAS, 88(24):11339-43.

Engel, KC, Flanders, M, Soechting, JF. 1997. Anticipatory and sequential motor control in piano playing. Experimental Brain Research, 113(2): 189-99.

Goller F, Suthers, RA. 1996. Role of syringeal muscles in gating airflow and sound production in singing brown thrashers. Journal of Neurophysiology, 75(2): 867-76.

Hahnloser, RH, Kozhevnikov, AA, Fee, MS. 2002. An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature, 419(6902): 65-70.

Hessler, NA, Doupe, AJ. 1999. Social context modulates singing-related neural activity in the songbird forebrain. Nature Neuroscience, 2(3):209-11.

Jarvis, ED, Scharff, C, Grossman, MR, Ramos, JA, Nottebohm, F. 1998. For whom the bird sings: context-dependent gene expression. Neuron, 21(4): 775-88.

Jerde, TE, Soechting, JF, Flanders, M. 2003. Coarticulation in fluent fingerspelling. Journal of Neuroscience, 23(6): 2383-2393.

Katz, LC, Gurney, ME. 1981. Auditory responses in the zebra finches' motor system for song. Brain Research, 221(1):192-7.

Leonardo A, Fee MS. 2005. Ensemble coding of vocal control in birdsong. Journal of Neuroscience, 25(3):652-61.

Margoliash, D. (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. Journal of Neuroscience, 6(6):1643-61.

Mitra, S, Amazeen, PG, Turvey, MT. 1998. Intermediate motor learning as decreasing active(dynamical) degrees of freedom. Human Movement Science, 17: 17-65.

Öhman, SE. 1966. Coarticulation in VCV utterances: spectrographic measurements. Journal of the Acoustic Society of America, 74:1115-1125.

Okanoya K. (2004). The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. Ann N Y Acad Sci 1016: 724-735.

Sakata, JT, Hampton, CM, Brainard, MS. 2008. Social modulation of sequence and syllable variability in adult birdsong. Journal of Neurophysiology. 99(4):1700-11.

Recansens, D. 1984. Vowel-to-vowel coarticulation in Catalan VCV sequences. Journal of the Acoustic Society of America, 76(6):1624-35.

Rothman, HH. 1977. An electromyographic investigation of articulation and phonetic patterns in the speech of deaf adults. Journal of Phonetics, 5: 369-76.

Tumer, EC, Brainard, MS. 2007. Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature, 450(7173): 1240-4.

Vicario, DS, Yohay, KH. 1993. Song-selective auditory input to a forebrain vocal control nucleus in the zebra finch. Journal of Neurobiology, 24(4):488-505.

Williams, H, Nottebohm, F. 1985. Auditory responses in avian vocal motor neurons: a motor theory for song perception in birds. Science, 229(4710):279-82.

Yu, AC, Margoliash, D. 1996. Temporal hierarchical control of singing in birds. Science, 273(5283): 1871-5.

Zollinger SA, Suthers RA. 2004. Motor mechanisms of a vocal mimic: implications for birdsong production. Proceedings of the Royal Society – Biology, 7(271):483-91.

Chapter 3

Catchpole, CL, Slater, PJB. 1995. Birdsong: Biological Themes and Variations. Cambridge University Press, 2nd Edition.

Daniloff, RG, Hammarberg, RE. 1973. On defining coarticulation. Journal of Phonetics, 1: 239-48.

Hahnloser, RH, Kozhevnikov, AA, Fee, MS. 2002. An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature, 419(6902): 65-70.

Kao MH, Brainard MS (2006) Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. J Neurophysiol 96:1441-1455.

Kimpo, RR, Theunissen, FE, Doupe, AJ. 2003. Propagation of correlated activity through multiple stages of a neural circuit. Journal of Neuroscience, 23(13):5750-61.

Leonardo A, Fee MS. 2005. Ensemble coding of vocal control in birdsong. Journal of Neuroscience, 25(3):652-61.

Okanoya, K. 2004. The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. Annals of the New York Academy of the Science. 1016:724-35.

Roberts, TF, Klein, ME, Kubke, MF, Wild, JM, Mooney, R. 2008. Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. Journal of Neuroscience, 28(13):3479-89.

Sakata, JT, Hampton, CM, Brainard, MS. 2008. Social modulation of sequence and syllable variability in adult birdsong. Journal of Neurophysiology. 99(4):1700-11.

Spiro, JE, Dalva, MB, Mooney, R. 1999. Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons. Journal of Neurophysiology, 81:3007-3020.

Vu ET, Mazurek ME, Kuo Y. 1994. Identification of a forebrain motor programming network for the learned song of zebra finches. Journal of Neuroscience, 14: 6924–6934.

Yu, AC, Margoliash, D. 1996. Temporal hierarchical control of singing in birds. Science, 273(5283): 1871-5.

General Discussion

Aronov D, Andalman AS, Fee MS. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science 320(5876): 630-634.

Bottjer SW. (1994). Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science 224: 901-903.

Brainard MS. (2004) Contributions of the anterior forebrain pathway to vocal plasticity. Ann N Y Acad Sci 1016:377-394.

Daniloff, RG, Hammarberg, RE. 1973. On defining coarticulation. Journal of Phonetics, 1: 239-48.

Hahnloser, RH, Kozhevnikov, AA, Fee, MS. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature, 419(6902): 65-70.

Kao MH, Doupe AJ, Brainard MS. (2005). Contributions of an avian basal ganglia circuit to real-time modulation of song. Nature 433(7026): 638-643.

Lou M, Ding L, Perkel DJ. (2001). An avian basal ganglia pathway essential for vocal learning forms a closed topographic loop. J. Neuroscience 21(17): 6836-6845.

Okanoya, K. (2004). The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. Annals of the New York Academy of the Science. 1016:724-35.

Olvecsky BP, Andalman AS, Fee MS. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biology 3(5): e153.

Scharff C. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system. J. Neuroscience 11: 2896-2913.

Suthers RA, Zollinger SA (2004) Producing song: the vocal apparatus. Ann N Y Acad Sci 1016:109-129.

Wild JM (2004) Functional neuroanatomy of the sensorimotor control of singing. Ann N Y Acad Sci 1016:438-462.

APPENDICES

Appendix 1
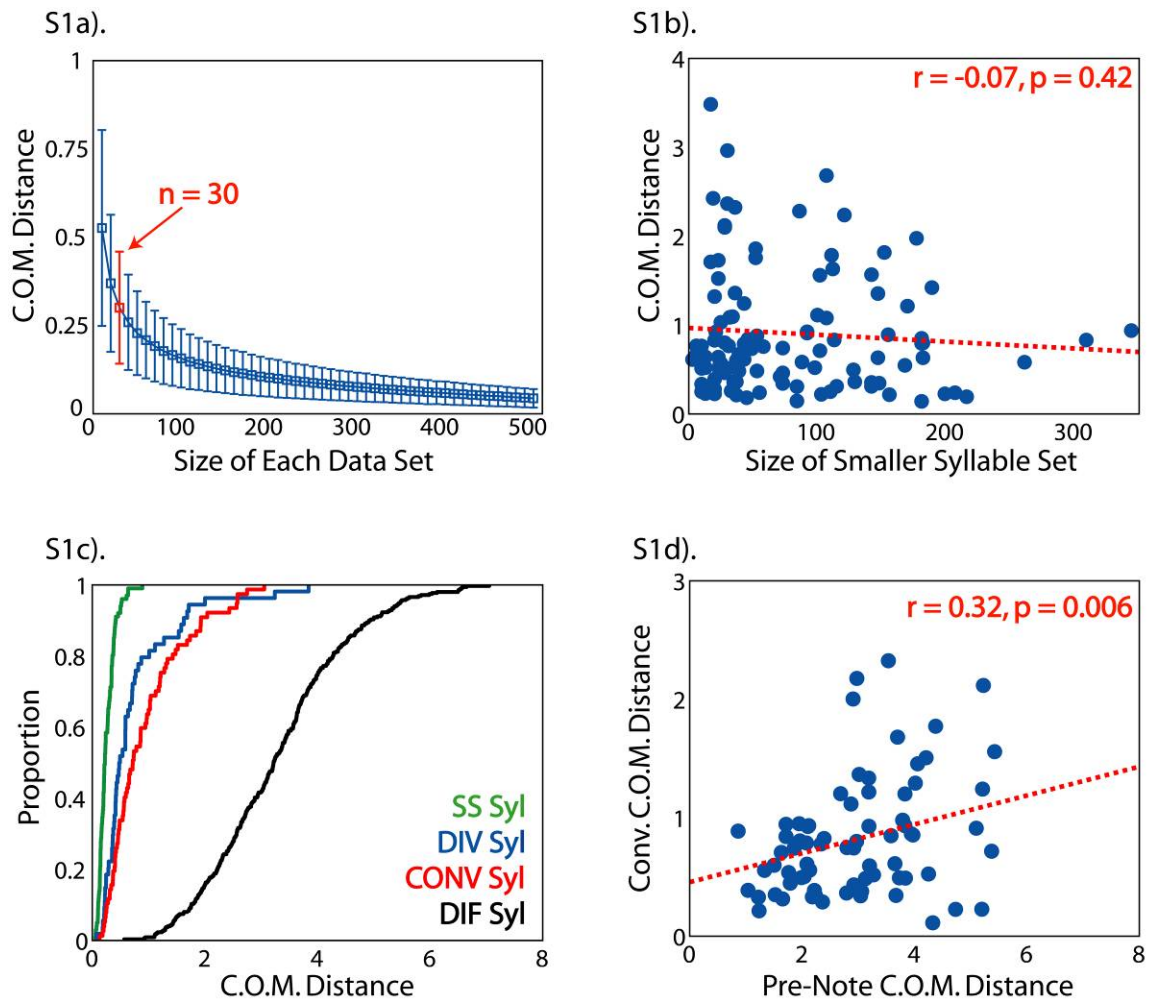
Appendix 1, Figure 1.1: Controlling for differences in data set size

(a) Average COM distance +/- standard deviation for different sized data sets of the SS syllables. Highlighted in red is the n=30 cutoff used for all analysis. Any syllables with fewer than 30 replicates were not used for analysis. (b) The relationship between data set size and COM distance for convergence points. Unlike as in S1A, there is no relationship between COM distance and data set size. (c) Cumulative distribution of COM distances for 30 randomly selected trials of SS, CONV, DIV, and DIF syllables. All syllable categories are significantly different from one another (p < 0.05, KS-test). (d) Effect of pre-note similarity upon convergence point similarity for a random selection of 30 replicates of each syllable sequence. A similar analysis to Figure 6b, but using only a random selection of 30 trials from each syllable. There remains a significant positive relationship for a random selection of 30 trials (r = 0.24, p = 0.02).
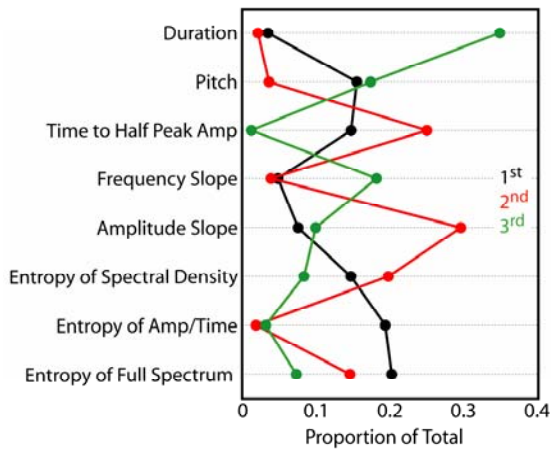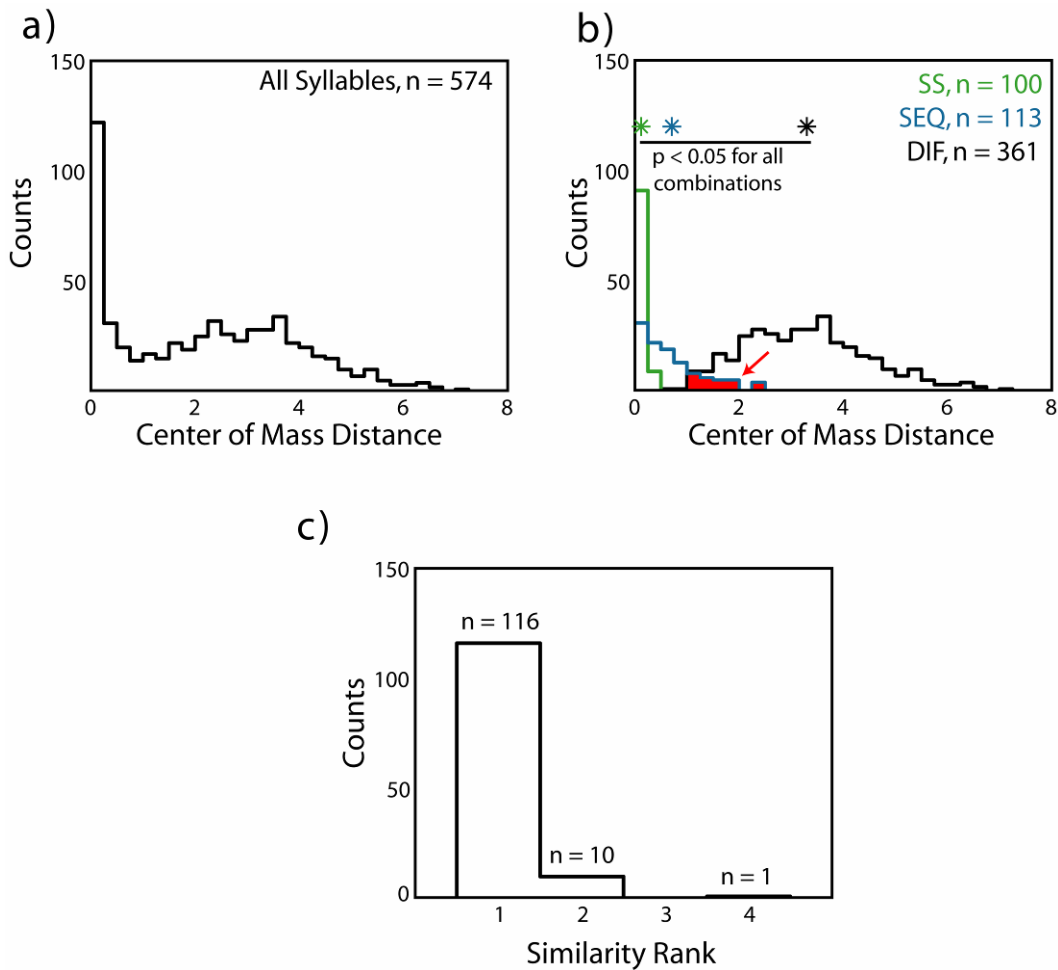
S2a). S2b).
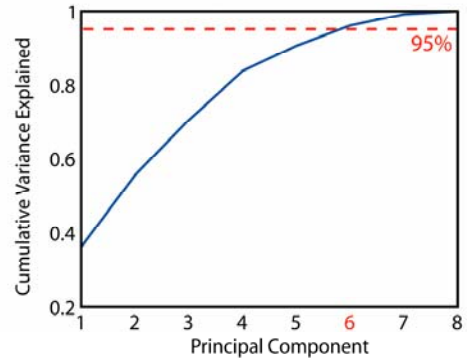
Appendix 1, Figure 1.2: Details of the PCA for syllable similarity

(a) The contributed proportion of each syllable feature to the first three principal components (black, red, green, respectively). (b) Percent variance explained by the addition of each principal component. 95% of the variance of the data set is explained with 6 principal components.

Appendix 1, Figure 1.3: Quantitative justification of visually labeled syllables

(a) Histogram of the COM distances from all birds, with the following types of relationships: the same syllable in the same sequence (SS), the same syllable in different sequences (SEQ), and differently labeled syllables (DIF). (b) The histogram in Figure 3A, color-coded for the three syllables combinations: green for SS, blue for SEQ, and black for DIF syllables. These three distributions are all significantly different from one another (p < 0.05, KS-test). The red highlights the proportion of SEQ distribution that overlaps with the DIF distribution. (c) Syllable similarity rankings from the label justification analysis. A rank of 1 means that two SEQ syllables were more similar to each other than they were to any other syllable in that bird's song. A rank of 2 means that there was one DIF syllable that was more similar to the SEQ syllable, and so on for ranks of 3 and 4. The graph shows that approximately 90% (116/128) of all SEQ syllable were more similar to each other than they were to any other syllable in that bird's repertoire.
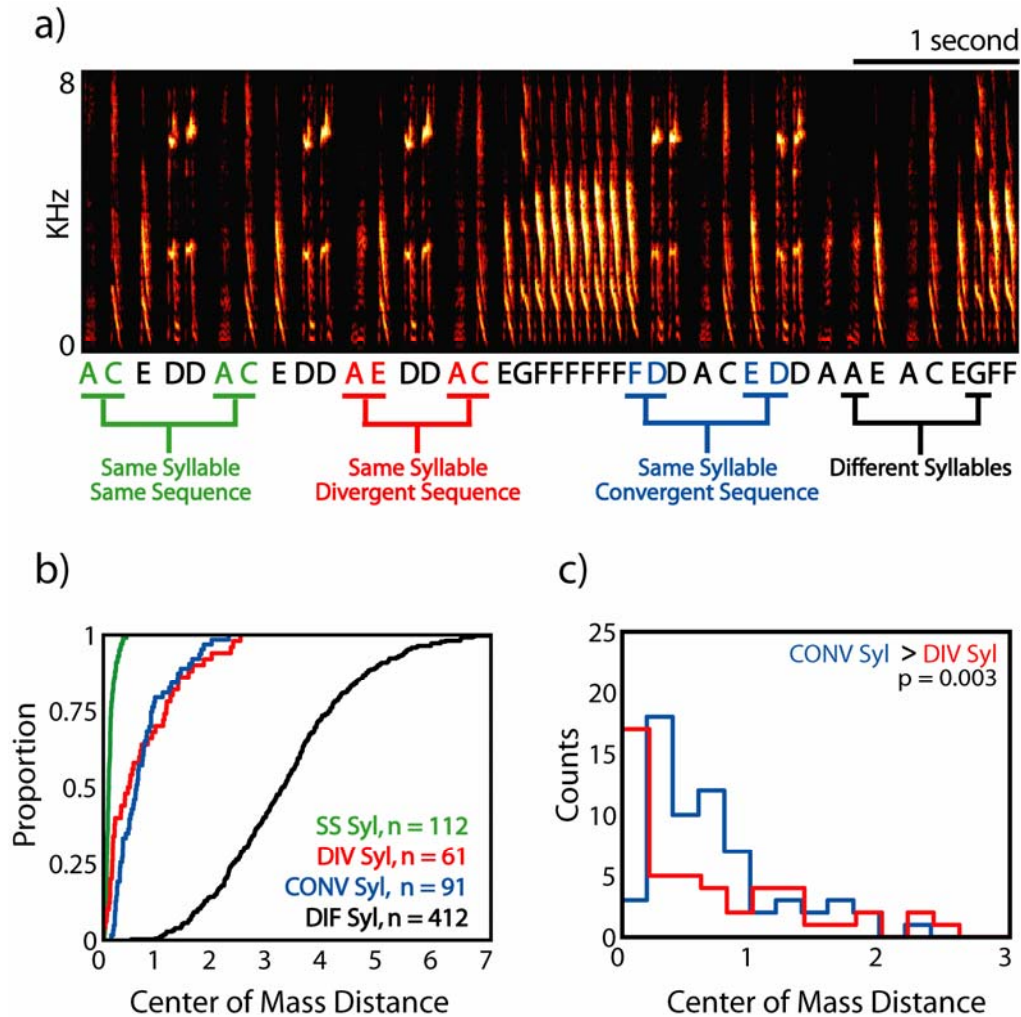
Appendix 2
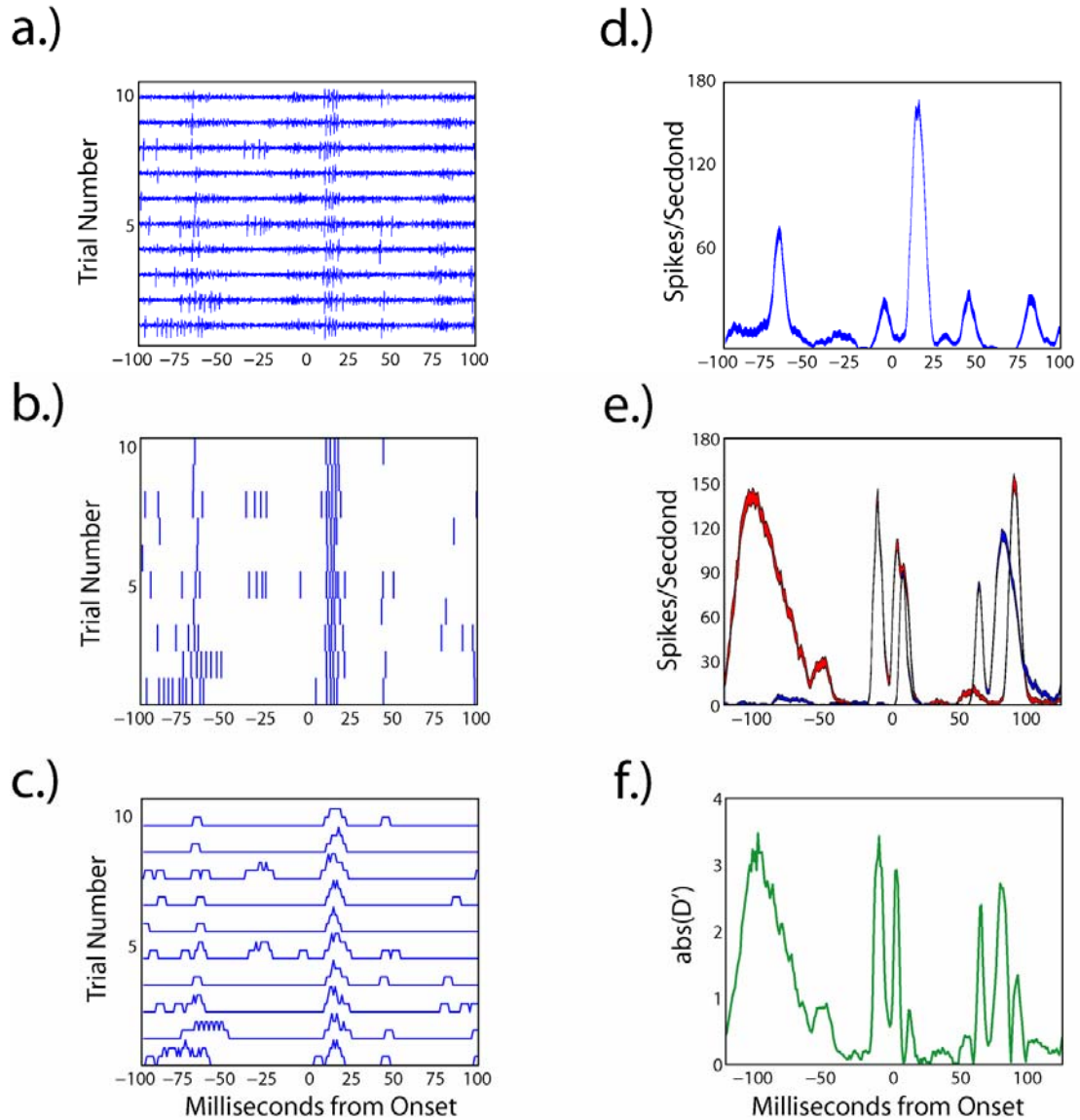
Appendix 2, Figure 2.1: Schematic of RA analysis

(a) Raw neural waveforms aligned at syllable onset (0 milliseconds). (b) Spike times aligned at syllable onset (0 milliseconds). (c) Spike times smoothed with 5 millisecond window aligned at syllable onset (0 milliseconds). (d) Mean +/- standard error for many trials aligned at syllable onset (0 milliseconds). (e) Comparison of two mean activity traces aligned at syllable onset (0 milliseconds). (f) Point-by-point d-prime between traces in E aligned at syllable onset (0 milliseconds).

Appendix 2, Figure 2.2: Syllable similarity across categories

(a) Example spectrogram of a song from BF4, with the syllable labels along the x-axis. Highlighted in green, red, blue, and black, respectively, are the four different relationships among syllables: same syllable same sequence (SS), divergent syllable (DIV), convergent syllable (CONV), and different syllables (DIF). (b) Cumulative distribution plots of the C.O.M. distances for each syllable relationship outlined in A (color convention the same). All distributions are significantly different from each other (p < 0.05 KS-test). (c) Distribution of COM values for CONV syllables and DIV syllables. COM distances for CONV syllables are significantly higher than between DIV syllables (p = 0.003, KS-test).
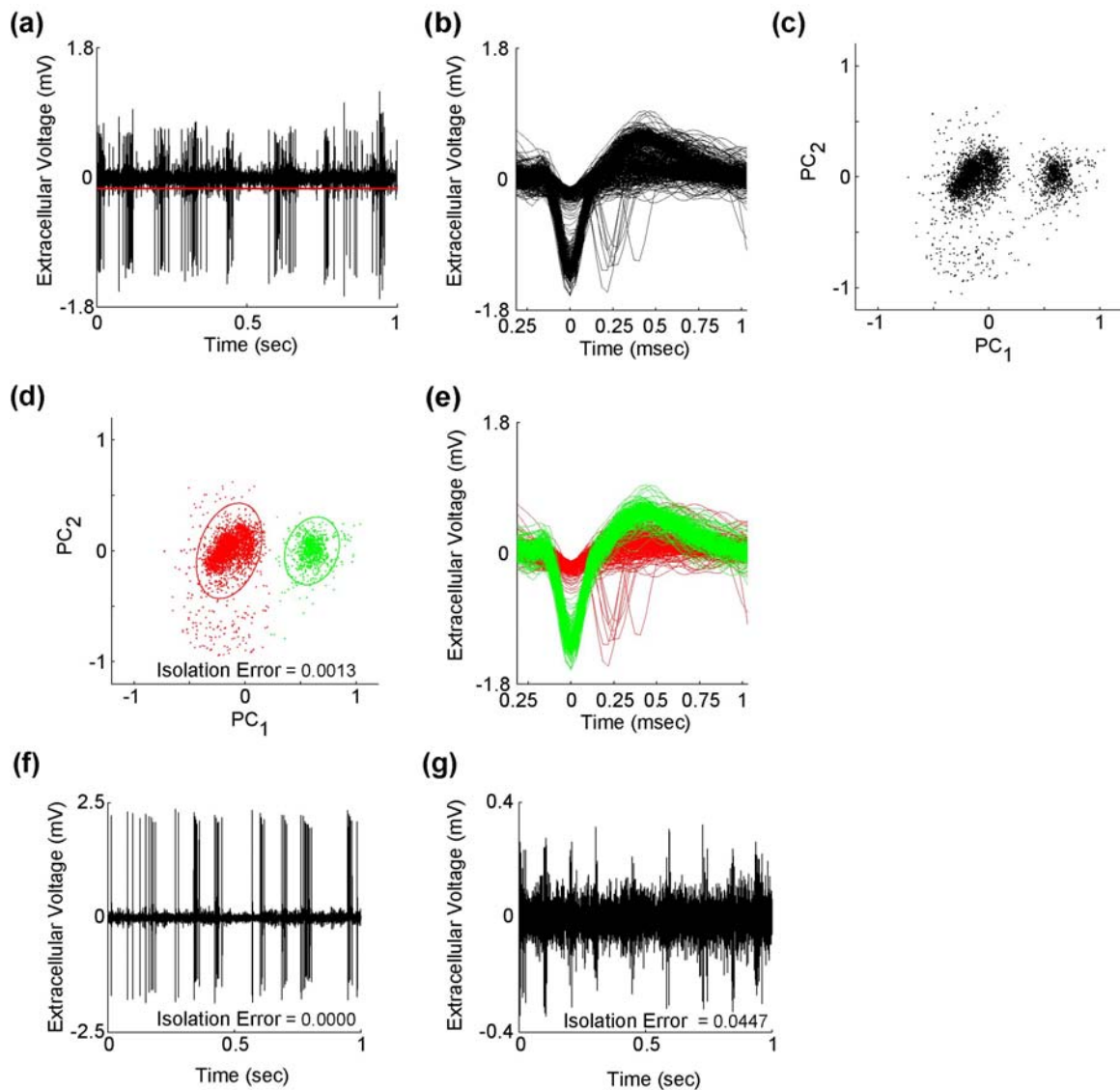
APPENDIX THREE

Quantifying unit isolation.

In order to make our criterion for classifying recordings explicit, we used a simple

method for quantifying the isolation of action potential waveforms. This technique is

based on principal components analysis (PCA), a mathematical technique commonly

used for waveform classification. As shown in Appendix 3, Figure 3.1a, a threshold (red

line) was used to select both spike and noise waveforms from singing-related activity.

These waveforms (Appendix 3, Figure 3.1b) were then analyzed using PCA, which

identifies the dimensions of variability (or principal components) that account for the

greatest amount of total waveform variability. By projecting each neural waveform onto

each of the two components that account for the most variability (PC_1 and PC_2), we

obtained a 2-dimensional representation of the distribution of waveforms (Appendix 3,

Figure 3.1c). Across the entire dataset, the first two components captured 64.0 +/- 7.2%

(mean +/- SD) of the total variance of waveforms in each recording (70.6 +/- 8.9% for

single units, 62.7 +/- 6.0% for multiunit sites, see below). An automated nearest-

neighbor clustering algorithm (kmeans.m in MATLAB, The MathWorks, Natick, MA)

was then applied to the 2-dimensional data to assign each waveform to a cluster. The

number of clusters used by the algorithm was set by hand. In the majority of cases, two

clusters (a "spike" cluster and a "noise" cluster) were selected. In a small number of

cases, three clusters were selected to capture waveforms belonging to two distinct spikes

in addition to the noise cluster. In these cases, only the waveforms belonging to the

larger spike were used. This cluster classification is represented by the colors of the

datapoints in Appendix 3, Figure 3.1d. Each cluster was then fit with a 2-D gaussian (ellipses in Appendix 3, Figure 3.1d) in order to estimate its mean and variance.

The goal of this analysis was to establish a scalar measurement of unit isolation. We found that the extent of overlap between the gaussian cluster fits served as a reliable indicator of unit isolation. Overlap was quantified by computing the probability with which a point from one cluster would be miscategorized by the nearest-neighbor algorithm as belonging to the other cluster. We generated 10,000 points from each cluster (according to that cluster's gaussian fit), reran the nearest-neighbor algorithm, and measured the frequency with which synthetic points were miscategorized. In the example shown in Appendix 3, Figure 3.1 a-e, the gaussian fits had an overlap (or "isolation error") of 0.0013. Examination of our data suggested that an isolation error of 0.01 was a reasonable threshold for classifying a recording as single-unit. Accordingly, clusters with isolation errors of less than 0.01 were classified as single-unit, and clusters with errors of greater than 0.01 were classified as multiunit. Two additional examples of isolation measurement are shown in Appendix 3, Figure 3.1 f and g.

Appendix 3, Figure 3.1. Isolation quantification technique. (a) One second of neural data from Bird 1, Unit 8. A threshold (red line) was used to collect voltage waveforms. (b) Collected waveforms. Times are relative to the peak of the negative excursion that crossed the voltage threshold. (c) Principal components representation of waveforms using the first two components. (d) A nearest-neighbor algorithm classified waveforms into two clusters, corresponding to "spike" (green) and "noise" waveforms (red). These clusters were fit with 2-D gaussians (ellipses show 2 S.D.) and cluster overlap (isolation error) was estimated. Since the isolation error is less than the threshold value of 0.01, this recording is classified as single-unit. (e) Waveforms colored by cluster assignment. Here the waveforms plotted in (b) are color-coded according to their cluster assignment. Note that while some "noise" traces (red) do contain spikes, these spikes were not responsible for the threshold crossing. and are separately represented as green traces with spikes centered at time zero. For clarity only 300 waveforms from each cluster are plotted in (b) and (e). Neural data and isolation errors for an additional single unit and one multiunit recording are shown in (f) and (g), respectively.

145

Quantitative analysis of spectral variability

In our analysis of RA's contribution to trial-by-trial variations in the acoustic structure of song syllables, it is important that we choose acoustic measures that capture a significant portion of behavioral variability. Failing to do so could cause us to underestimate RA's influence on behavior, since a correlation test requires the presence of both neural and behavioral variability in order to be meaningful. The spectral structure of song is extremely complex, offering many potential measures of vocal output (Tchernichovski et al., 2001). Although behavioral studies, stimulation experiments, and *in vitro* measurements of the syrinx have suggested that pitch, amplitude, and spectral entropy are tightly controlled song parameters (and therefore likely to be under RA's influence), we wanted to investigate explicitly the nature of spectral variation within each syllable.

With these issues in mind, we performed principal components analysis (PCA) on the power spectrum of each syllable in our dataset in order to characterize the important dimensions of behavioral variability. PCA identifies the axes of variation in a data set that account for the greatest amount variability. By performing PCA on the power spectra recorded from all renditions of a single syllable, we identify the axes of variation (that is, the deviations from the syllable's mean spectrum) that make the greatest contributions to the total spectral variation.

Appendix 3, Figure 3.2 (a) and (b) illustrate PCA as performed on two example syllables. For each syllable in our dataset, we measured the power spectrum at the same measurement time (dashed black lines at left) used to compute pitch, amplitude,

and entropy in the original analysis (see Methods). Power spectra were sampled at 71 evenly-spaced frequencies between 1-10 kHz. Mean spectra for the two example syllables are shown in the middle panels of (a) and (b). For each syllable, this procedure resulted in an n x 71 matrix, where n is the number of instances of the syllable in question. We then performed PCA on each matrix to derive a set of 71 orthogonal basis vectors (the principal components) in which PC1 is the direction along which spectral variance is maximized, PC2 is the second-most variable direction, and so on. The fraction of the total variance explained by each of the 71 principal components for the example syllables is shown at right in (a) and (b).

Identification of dominant axes of spectral variation:

These two examples are typical of our dataset in that the first 1-3 principal components accounted for much more of the total variance than did any of the remaining components (group data are shown in Appendix 3, Figure 3.2 c and d). These data indicate that while spectral variations were not easily reduced to a compact representation (as would have been the case if, for example, 95% or more of the cumulative variance had been accounted for by a small number of components), in each syllable a few dimensions of variability accounted for a relatively large amount of the total variability.

Quantifying similarity between PCs and changes in pitch, amplitude, and entropy:

We often found that the dominant principal components (those describing the greatest fraction of behavioral variation) resembled changes in pitch, amplitude, or

entropy. Below, we show components for which this was the case and illustrate how we quantified the similarity between principal components and these three acoustic features.

Appendix 3, Figure 3.3a shows and example in which the dominant principal component resembled a change in pitch. This can be seen by noting that positive weights along this component correspond to a downward shift in the spectral peaks of the syllable (red trace in the bottom panel, representing the mean spectrum + 2*PC1). Conversely, negative weights along this component reflect an upward shift in pitch (blue line in bottom panel).

To quantify the degree to which this principal component described a pitch shift, we measured its similarity to a "synthetic component" generated by explicitly varying the pitch of the mean power spectrum. Synthetic pitch components were generated by first linearly stretching (or contracting) the mean power spectrum across the frequency axis to simulate an increase (or decrease) in pitch. The mean spectrum was then subtracted from the pitch-shifted spectrum to generate a synthetic component representing the change in power at each frequency resulting from the pitch shift. The best-fit synthetic component (red dashed line in middle panel of Appendix 3, Figure 3.3b) was found by varying the stretching factor (pitch shift) until the closest possible match to the real component was achieved. The similarity between real and synthetic components was then computed using the cosine-similarity measure, which measures the cosine of the angle between the two vectors in 71-dimensional space (and ranges from -1.0 to 1.0). In the case illustrated in Appendix 3, Figure 3.3b, PC1 and the synthetic amplitude change have a cosine-similarity score of 0.99.

Appendix 3, Figure 3.3b illustrates the power spectrum (top) and first principal component (middle) for a case in which PC1 described a change in amplitude. Positive weights along this component reflect a subtraction of power at each frequency (red line in bottom panel). Conversely, negative weights along this component reflect an increase in amplitude (blue line in bottom panel). The blue dashed line in the middle panel of Appendix 3, Figure 3.3a shows the synthetic component explicitly simulating a change in amplitude that best fits PC1. Synthetic amplitude components were generated by simulating a constant amplitude offset at each frequency. The best-fit synthetic component was found by varying the offset until the closest possible match to the real component was achieved. In the case illustrated in Appendix 3, Figure 3.3a, PC1 and the synthetic amplitude change have a cosine-similarity score of 0.96.

Appendix 3, Figure 3.3c illustrates the power spectrum (top) and second principal component (middle) for a case in which PC2 described a change in spectral entropy. Positive weights along this component reduce the power at the spectral peaks (dotted vertical lines) and increase the power in the troughs between the peaks, as shown in the bottom panel. Although changes in spectral entropy (defined as $\Sigma \, p \log(p)$, see Methods) could in principle be achieved by many different changes to spectral structure, we consistently found principle components of the form shown in Appendix 3, Figure 3.3c. Synthetic entropy components were therefore generated by summing a scalar offset with gaussian distributions centered on each harmonic peak present in the spectrum:

$$PC_{synthetic\_entropy}(f) = \alpha + \beta \sum_{x} G(f, h_x, x, \sigma)$$

$$G(f, h_x, x, \sigma) = \frac{1}{\sigma\sqrt{2x\pi}} e^{\frac{-(f-h_x)^2}{(2x\sigma^2)}}$$

Where $PC_{synthetic\ entropy}(f)$ is the value of the synthetic component at a given frequency and h_x are the x harmonics of the syllable. The parameters of this equation are constrained such that (1) $|\alpha| < .66* |\beta|$ and (2) α and β are of opposite sign. These constraints are necessary to ensure that the gaussian components contribute to the overall shape of the synthetic component (i.e. they ensure that synthetic entropy components are not allowed to approximate synthetic amplitude components, which have the form $PC_{synthetic\ amplitude} = \alpha$). In the case illustrated in Appendix 3, Figure 3.3c, the real and synthetic components had a cosine similarity score of 0.93.

Classification of the dominant components:

We then asked how frequently the principal components explaining the greatest amount of behavioral variation were well-described by a synthetic pitch, amplitude, or entropy component. To do this, we restricted our analysis to PCs that explained at least 10% of the total variance of the syllable ($PC_{>10\%}$). For each $PC_{>10\%}$, we found the best-fit synthetic components describing changes in pitch, amplitude, and entropy and computed the cosine-similarity for each. If cosine-similarity exceeded a threshold of 0.8, the $PC_{>10\%}$ was classified as "congruent" to the relevant synthetic component.

To verify that our categorization of principal components as "congruent" with pitch, amplitude, or entropy is meaningful, we compared the amount of behavioral variation along the direction of each $PC_{>10\%}$ with the amount of behavioral variation along the direction of the best-fit congruent synthetic component. We found that the fraction of the total variability explained by the congruent synthetic components was not significantly different from the fraction explained by the corresponding $PC_{>10\%}$ (p=0.35, Wilcoxon signed-rank test).

If a $PC_{>10\%}$ was congruent with more than one best-fit synthetic component, then the component with the greater cosine-similarity score was selected. This occurred in 22% of cases and almost always resulted from a $PC_{>10\%}$ being found congruent with both amplitude and entropy. Using this criterion, we found that the majority (70.2%) of $PC_{>10\%}$ were congruent with either pitch, amplitude, or entropy (Appendix 3, Figure 4a), showing that the most dominant dimensions of behavioral variability usually reflect variations in one of these parameters. Moreover, in cases where $PC_{>10\%}$ were not congruent with pitch, amplitude, or entropy, the $PC_{>10\%}$ often resembled a linear combination of two of these features (not shown).

Conversely, Appendix 3, Figure 3.4b shows that when all PCs (not just those explaining more than 10% of the total variance) are classified by cosine-similarity, components congruent with pitch, amplitude, and entropy explain a much larger proportion of behavioral variation than do components that are not congruent with these parameters. This result demonstrates that among all dimensions of spectral variation yielded by PCA, the components describing changes in these three parameters had much more explanatory power than other components. Together, these results suggest that

pitch, amplitude, and entropy are indeed important components of behavioral variability, and therefore are reasonable features to use in characterizing the contributions of RA activity to behavioral variation.

Primary and PCA-based behavioral analysis yield similar results:

Finally, we wanted to ensure that our decision to quantify correlations between neural activity and the measured values of pitch, amplitude, and entropy (rather than correlations between neural activity and loadings along the $PC_{>10\%}$) did not result in our underestimating the strength of neural-behavioral correlations. Appendix 3, Figure 3.4c shows that the r^2 values of correlations between neural activity and the measured behavioral parameters are not significantly different from the strength of correlations between neural activity and loadings along the $PC_{>10\%}$.
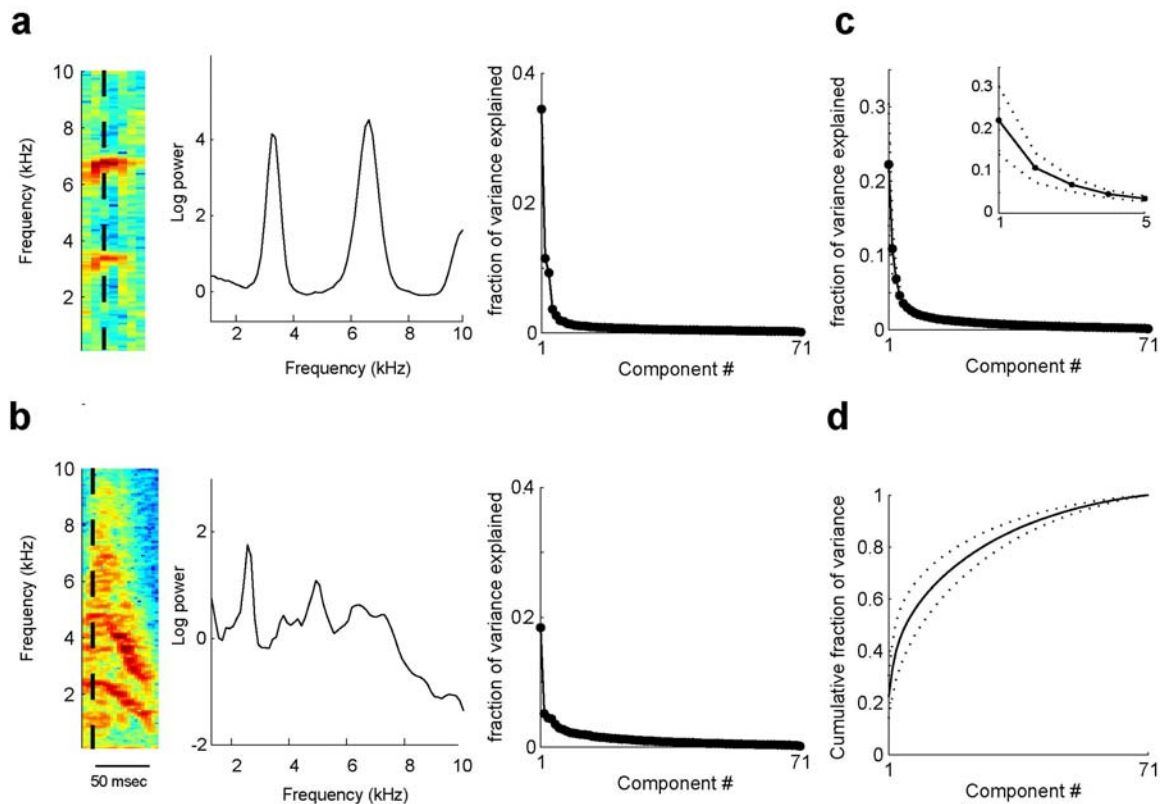
Appendix 3, Figure 3.2: PCA-based analysis. (a) and (b) show the spectra of two example syllables. Note that the syllable shown in (a) is syllable "E" from Bird 1, also illustrated in Figure 5 in the main text. For each syllable, we computed the power spectrum at the same measurement time used for the primary analysis (of pitch, amplitude, and entropy) in the main text (dashed lines in spectrograms at left). Each power spectrum consisted of measurements of the log spectral power at each of 71 equally-spaced frequencies between 1-10 kHz. The mean power spectra for the two example syllables are shown in the middle panels of (a) and (b). PCA was then used to find the dimensions of variation (or "components") that captured the greatest, second greatest, etc. fraction of the total variation of each syllable. The fractions of the total variance accounted for by each component of the two example syllables is shown in the right panels of (a) and (b). (c) shows the mean +/- SD of this measure for all 83 syllables in our dataset. Inset shows the first 5 components in greater detail. (d) shows the same data as (c) expressed as a cumulative fraction.
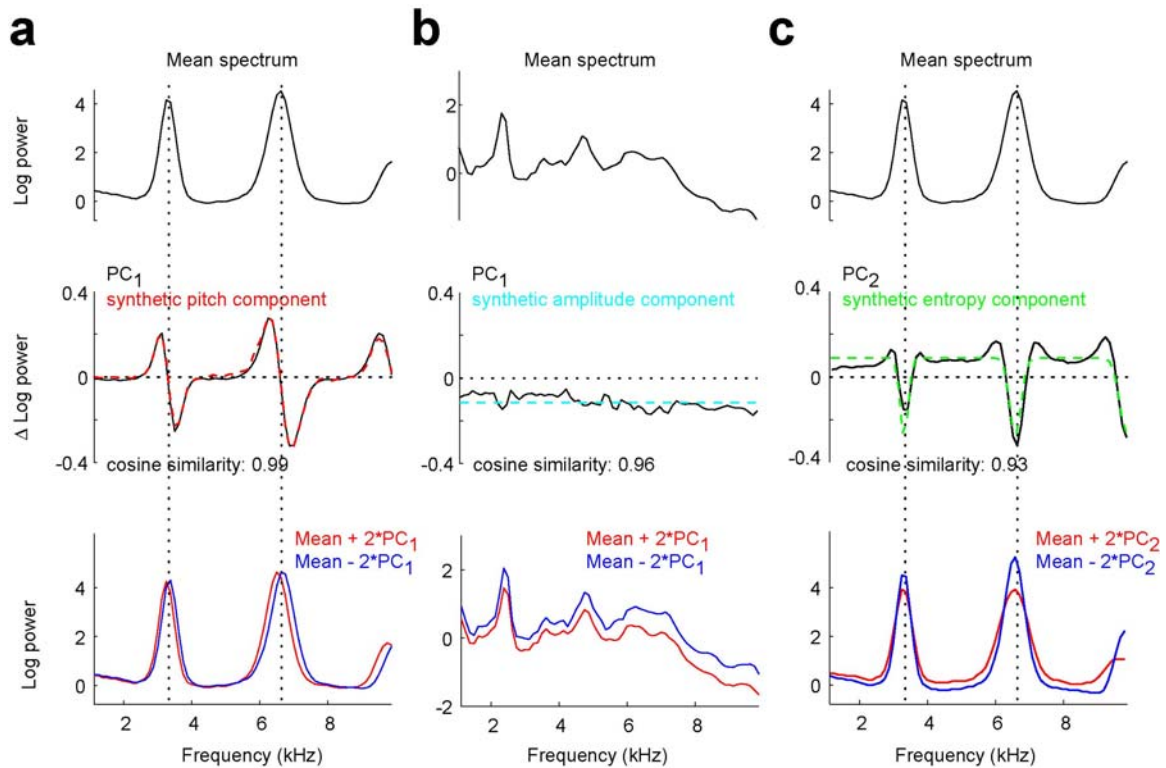
Appendix 3, Figure 3.3: PCs approximating changes in pitch, amplitude, and entropy. (a) shows the mean power spectrum (top), first principal component (middle, black trace), and deviations from the mean spectrum along the first principal component (bottom) for the syllable shown in Appendix 3, Figure 2a. The dashed red line in the center panel shows the synthetic pitch component that best fits the principal component (see text). Vertical dashed lines indicate spectral peaks at the first and second harmonics. (b) shows the same measures for the first principal component of the syllable shown in Appendix 3, Figure 2b. The dashed blue line shows the synthetic amplitude component that best fits the principal component. Other conventions as in (a). (c) shows the same measures from the second principal component of the syllable shown in Appendix 3, Figure 2a. The dashed green line shows the synthetic entropy component that best fits the principal component.
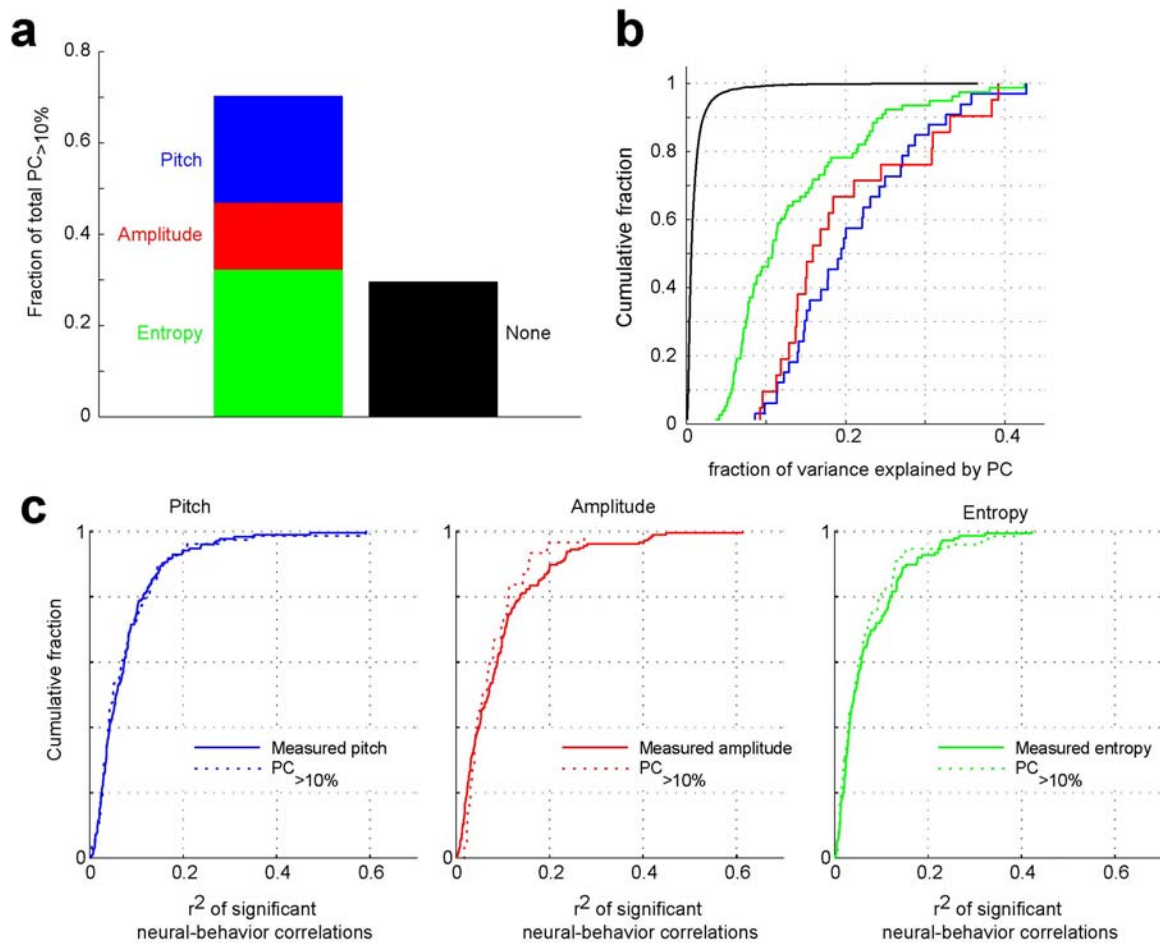
Appendix 3, Figure 3.4 (a) Distribution and classification of principal components explaining at least 10% of spectral variation (PC$_{>10\%}$). Blue shading indicates a PC$_{>10\%}$ classified as congruent with pitch, red indicates a PC$_{>10\%}$ congruent with amplitude, green indicates a PC$_{>10\%}$ congruent with entropy, and black indicates PC$_{>10\%}$ that were not congruent with any of the three parameters. (b) Cumulative distributions of the fraction of the total variance explained by all principal components. Data shown are for all principal components, regardless of how much spectral variance they explain. Components are divided into those congruent with pitch (blue), amplitude (red), entropy (green), and those not congruent with any of these three parameters (black). (c) Cumulative distributions of r^2 values for significant correlations between neural activity and acoustic output for each acoustic feature. In each plot, r^2 values are shown for two different methods of quantifying acoustic output on each trial. Solid lines show the correlation strengths resulting from measuring pitch, amplitude, or entropy as described

in the main text. Dotted lines show the r^2 values resulting from using the loadings along the $PC_{>10\%}$ congruent with the relevant acoustic measure.

Testing the timescale of premotor encoding

Work in diverse sensory and motor systems has shown that spike trains can carry information on multiple time scales (Theunissen and Miller, 1995; Tiesinga et al., 2008). At one extreme, sensory input or motor output might be encoded by the total number of spikes produced on a given trial (rate coding). At the other extreme, millisecond-scale differences in spike timing might encode important task parameters.

Here, we compare the predictive power of premotor encoding at a range of timescales. For each instance where a single unit was active prior to a syllable (one "case," see Methods), we fit the data with five different linear models that range from a simple rate coding scheme (counting the total number of spikes in the 40 msec premotor window) to a model that divides the premotor window into 1 msec-wide bins in order to ask whether fine temporal structure encodes motor output. The models take the form

$$b = \sum_{i=1}^{n} a_i s_i + c$$

where b is the behavioral parameter to be fit (pitch, amplitude, or entropy), s is the number of spikes falling in each of n consecutive bins spanning the 40 msec premotor window, and $[a_{1...n}\ c]$ are the fit parameters of the model. The five models tested have $n = [1,2,4,8,40]$, corresponding to bins with widths of 40, 20, 10, 5, and 1 msec respectively, as illustrated in Appendix 3, Figure 3.5a.

For each case and behavioral parameter, we used cross-validation to assess which model has the most predictive power while protecting against overfitting. We randomly

split the data into a training set consisting of 80% of the data used to fit the model parameters, which were then used to predict the data in a testing set consisting of the remaining 20% of the data. Repeating the procedure 1000 times for each model (splitting the data randomly on each iteration) yielded a distribution of error (mean squared error). The "best" model was defined as the model with the lowest mean error.

Appendix 3, Figure 3.5 shows that for all three behavioral parameters, counting the number of spikes in a single 40 msec-wide bin ($n = 1$ in the above equation) is the best predictor of the neural-behavioral relationship in a majority of cases, and represents the best model more than twice as frequently as any of the other tested encoding schemes. These results suggest that in most cases the fine temporal structure of spikes does not encode the acoustic structure of song (at least not in the linear fashion described by the above equation). Consequently, models that use fine temporal structure to predict behavior presumably produce higher cross-validation error because free parameters are fit to non-informative features of spike trains. Based on these results, we used spike count in a single 40 msec window – the tested model with the most predictive power in the majority of cases– to examine correlations between neural activity and song.
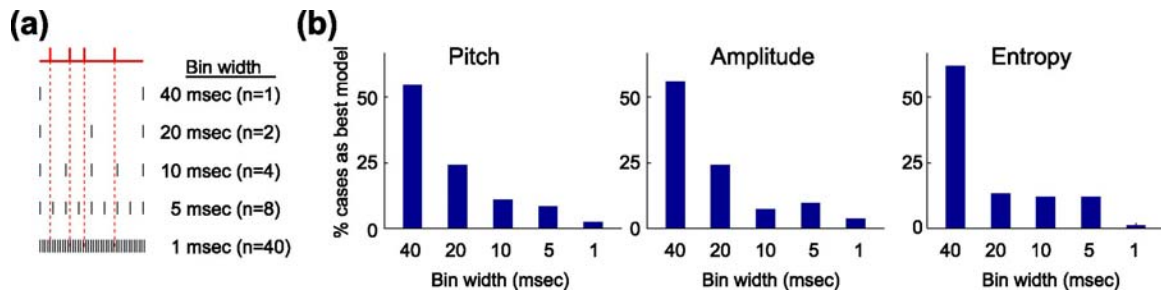
Appendix 3, Figure 3.5. Cross-validation analysis. (a) Five models with different bin sizes. Each model divides the 40 msec-long premotor window (horizontal red line) into a different number of bins. Song output is modeled as a linear function of the number of spikes falling within each bin (see text). For the example spike train (vertical red lines), the model with a single bin would count four spikes. For the model with two 20 msec-long bins, three spikes fall in the first bin, and one spike falls in the second bin, and so on. (b) Performance of the five models. As described in the text, we used a cross-validation approach to determine which model is best able to predict the neural-behavioral relationship in each case where a neuron is active prior to a syllable. Bar plots show the fraction of cases in which each model is the best predictor of pitch, amplitude, and entropy.

Multiple comparisons:

Our goal is to determine whether trial-by-trial variations in premotor neural activity are correlated with behavioral variations in a significant number of cases. This presents a multiple comparison problem: as the number of tests increases, so will the number of false positives (cases in which a significant correlation is found but none exists). For example, if we are performing k independent correlation tests (at $p<0.05$) on data in which no correlations exist, the odds of finding at least one spuriously significant correlation are $[1-0.95^k]$. Put another way (in terms of the expectation value for the number of false positives), for $k=100$ tests, we would expect to find 5 false positives.

There are several approaches to dealing with this problem. One approach (employed by the Bonferroni correction and related techniques) is to control the probability that <u>any</u> test is found to be significant. More formally, for k tests, these techniques control the probability of rejecting at least one of k null hypotheses when all k null hypotheses are true (Westfall and Young, 1993). With this type of correction, finding <u>any</u> test significant provides evidence (at a confidence level determined by the controlled false-positive rate) of a relationship between the variables in question.

An alternate technique is to test whether the <u>proportion</u> of significant correlations is greater than that expected by chance. As mentioned above, 100 independent tests of uncorrelated variables will on average yield 5 false positives. We could then use a binomial test to determine whether the proportion of significant correlations in the empirical data is significantly greater than the proportion expected by chance.

While both techniques are in principle correct (provided that all tests are independent, a topic discussed at length below), they answer slightly different questions

about our dataset. Applying the Bonferroni correction by reducing each p-value threshold within each neuron (e.g. by a factor of 18 if a neuron is subjected to 18 tests) still results in many significant correlations with pitch, amplitude, and entropy (76, 76, and 47, respectively). Furthermore, in the Bonferroni-corrected dataset significant correlations with pitch have a positive sign in 68% of cases (the asymmetry is significant at p=0.001), whereas no significant asymmetry is found in correlations with entropy and amplitude. Similar results are obtained from an even more conservative version of the Bonferroni correction, in which the threshold p-value for every test is reduced by a factor of 2115 (the total number of tests performed on all 145 neurons). These results indicate that at least some significant correlations are present in our data.

Because one goal of our paper is to estimate the proportion of RA units encoding the acoustic output at any given time, we also employed a proportion-based technique, since it tells us about the fraction of the neural population encoding behavioral variation. This approach is detailed below.

Non-independent tests:

While the proportion-based (binomial) test describe above is applicable when all tests being performed are independent, our analysis is potentially complicated by correlations among the three acoustic parameters measured from each syllable (behavioral-behavioral correlations) and by correlations between neural activity during consecutive syllables (neural-neural correlations) (see Results). Correlations of this sort raise the possibility that our tests for correlations between premotor activity and acoustic output (neural-behavioral correlations) might not be independent. In the extreme case, if

the pitch, amplitude, and entropy of each syllable (for example) were perfectly correlated, then the three correlations between neural activity and each parameter only truly constitute a single independent test. The apparent statistical power of a binomial test (to determine whether the proportion of correlations is significantly greater than 5%) might therefore be artificially inflated.

To illustrate this, consider a hypothetical case in which 20 correlation tests are performed, where all 20 tests are independent, and of which 2 are found to be significant. Although 2/20 > 5%, this proportion fails the binomial test (p=0.08). Now, imagine that for each of the 20 original tests, 9 other tests are performed with outcomes that are perfectly correlated with the corresponding test from the original 20. In this case, although there are only 20 truly independent cases, we would find that 20/200 tests were significant, a proportion that easily passes the binomial test (p=0.001).

Estimating the null distribution:

It is therefore necessary to correct for the consequences of neural-neural and behavioral-behavioral correlations. Our approach is to use a resampling technique to create an artificial dataset in which all neural-behavioral (NB) correlations are broken, but all neural-neural (NN) and behavioral-behavioral (BB) correlations are preserved. We then perform correlation tests on these resampled datasets and note the proportion of cases with significant correlations. By performing this procedure many times, we can estimate the distribution of proportions of significant correlations under the null hypothesis, and then ask whether the proportion of significant correlations in the real dataset is beyond the 95th percentile of the null distribution.

Details of the resampling-based approach:

First, consider a bird that sings only one syllable, from which we have recorded 10 neurons. For one of these neurons, let N and B be matrices of neural and behavioral data, respectively. The rows of N and B correspond to trials, and the columns correspond to the tests being performed. $N_{k,i}$ is therefore the premotor neural activity (# spikes) during trial k of test i, and $B_{k,i}$ is the value of the appropriate acoustic parameter during trial k. If there are n examples of the syllable, N and B will each have n rows and 3 columns, representing the three acoustic measurements (pitch, amplitude, and entropy). (Note that if the bird sings only one syllable, the columns of N are identical: $N_{k,i} = N_{k,j}$ for all k, since the three different behavioral measures are being compared to the same premotor neural activity.)

Our goal is to create a dataset in which any neural-behavioral correlation between columns N_i and B_i are broken, but correlations between columns N_i, N_j, $i \neq j$ (NN correlations) and B_i, B_j, $i \neq j$ (BB correlations) are preserved. To accomplish this, we permute (shuffle) the rows of N, resulting in N^p. (The same results could be achieved by permuting the rows of B.) We then compute correlations between the three paired columns N^p_i and B_i and count the number of tests that achieve significance (at $p < 0.05$).

Since our hypothetical dataset consists of ten such neurons, running the above-described procedure on each neuron would yield 30 tests, of which some subset will achieve significance. We then record this proportion and repeat the procedure many times to generate a distribution of proportions of significant correlations under the null

hypothesis. We can then reject the null hypothesis if the proportion of significant cases in the original dataset lies beyond the 95th percentile of the null distribution.

Multiple syllables:

Now we consider how to perform this procedure when a bird's repertoire consists of more than one syllable, as is the case in our data (Bengalese finch song typically contains 5-10 distinct syllables). Consider a case in which a bird's song contains 5 syllables, labeled "ABCDE". If the bird always sings these syllables in this order, an example song might be "ABCDE-ABCDE-ABCDE" (hyphens are inserted for visual clarity and do not represent syllables) and adapting the above procedure is straightforward. Both the N and B matrices, rather than having 3 columns (3 acoustic measurements of 1 syllable), will have 15 columns (3 acoustic measurements x 5 syllables). Each row of N and B will correspond to one rendition of the motif "ABCDE." Permuting the rows of N will remove NB correlations, while preserving BB and NN correlations. Note that when multiple syllables are analyzed, NN and BB correlations describe neural and acoustic correlations across, as well as within, syllables.

Our analysis is complicated, however, by the fact that syllable order in Bengalese finch song, while highly patterned, is seldom as stereotyped as in the above example. A more typical syllable order for a Bengalese finch song bout is "AB-AB-ABC-ABCDE-ABCDE-AB" (again, hyphens inserted for visual clarity only). Because of inconsistencies in the order and prevalence of syllables, it is not possible to gather all neural data into one neural matrix and one behavioral matrix. Instead, data from each syllable must be collected into a separate pair of matrices. Then, in each run of the

resampling algorithm, we permute the rows of all neural matrices. However, it would be incorrect to permute the neural matrices independently: to do so would destroy any NN and BB correlations between, for example, syllable "A" and syllable "B". Rather, we permute the neural matrices in a manner that, as closely as possible, preserves relationships between consecutive syllables in the original data.

Row permutation for complex syllable patterns:

For each recorded neuron, we divide the recorded syllables into "segments," where each segment is the longest possible sequence of syllables in which no syllable is repeated. Here, segment # is indicated with subscript for an example sequence:

A_1B_1 -A_2B_2 -$A_3B_3C_3$ -$A_4B_4C_4D_4E_4$ -$A_5B_5C_5D_5E_5$ -A_6B_6 −$A_7B_7C_7D_7E_7$

Then, on each resampling trial, we randomly permute the order of segments. For example, the segment reordering on one resampling trial might be: [6 2 5 7 4 3 1]. The reordered (permuted) song would look like this:

A_6B_6 - A_2B_2 - $A_5B_5C_5D_5E_5$ - $A_7B_7C_7D_7E_7$-$A_4B_4C_4D_4E_4$- $A_3B_3C_3$ -A_1B_1

Note that while the order of the segments has been permuted, local adjacencies (that is, the order of consecutive syllables within segments) are preserved.

We then use this reordering of segment sequence to shuffle the rows of the neural

matrices:

Permuted order for each syllable:

$A_6\ A_2\ A_5\ A_7\ A_4\ A_3\ A_1$ Reorder the 7 rows of N_A by [6257431]
$B_6\ B_2\ B_5\ B_7\ B_4\ B_3\ B_1$ Reorder the 7 rows of N_B by [6257431]
$C_5\ C_7\ C_4\ C_3$ Reorder the 4 rows of N_C by [3421]
$D_5\ D_7\ D_4$ Reorder the 3 rows of N_D by [231]
$E_5\ E_7\ E_4$ Reorder the 3 rows of N_E by [231]

The procedure allows us to permute the matrices of neural data in as consistent a fashion

as possible given that not every syllable is included in every song segment Note that in

cases where all segments are identical, as in

$A_1B_1C_1D_1E_1$- $A_2B_2C_2D_2E_2$- $A_3B_3C_3D_3E_3$ - ...

this method is equivalent to collecting all data into one neural matrix and one behavioral

matrix, and permuting the rows of the neural matrix.


Results of resampling analysis:

Using the above-described techniques, we permuted our entire dataset 1000 times,

creating a distribution of the number of significant correlations (out of the total of 2115

cases) expected under the null hypothesis. When tests for pitch, amplitude, and entropy

were combined, the 95th percentile of this distribution fell at 122 cases, (or 122/2115 =

5.8% of the total number of tests), as shown by the dashed black line in Appendix 3,

Figure 6a. When the combined resampled distribution was separated into tests for pitch,

166

amplitude, and entropy, the critical values were at 46, 44, and 45 significant tests out of 705 (6.5%, 6.2%, and 6.4%, respectively; dashed black lines in Appendix 3, Figure 6b,c, and d). Since these values are quite similar, we used the most conservative of these figures (6.5%) as the significance threshold in Figure 7a in the main text.

The number of significant correlations in our dataset exceeded the relevant significance threshold in all cases (169, 175, and 137 for pitch, amplitude, and entropy respectively; dashed red lines in Appendix 3, Figure 3.6b,c, and d). Furthermore, the number of significant correlations not only exceeded significance thresholds but fell beyond the range of all 1000 resampled datasets, whether tests for pitch, amplitude, and entropy were considered either together or separately. The observed proportions were thus significantly ($p<0.001$) greater than those expected by chance.
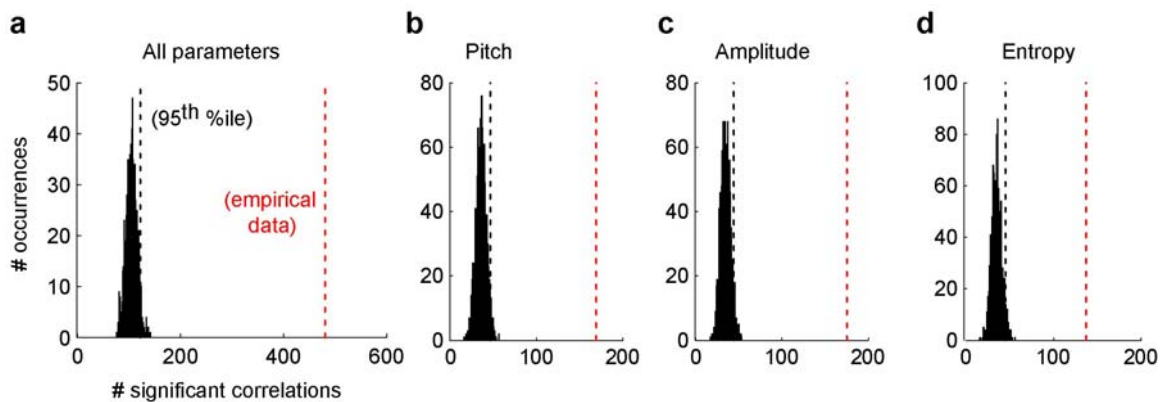

Appendix 3, Figure 3.6: Results of resampling analysis. Each plot shows the distribution of the number of significant correlations in each of the 1000 permutations of the empirical data. The 95th percentile of this distribution (dashed black lines) is the threshold for significance. In all cases, the number of significant correlations in the empirical data (dashed red lines) was beyond this threshold. Permutation tests were conducted using the full dataset, combining tests for pitch, amplitude, and entropy, as described in the text and shown in (a). (b), (c), and (d) show the results of separating the empirical and resampled data into separate distributions for each acoustic parameter.
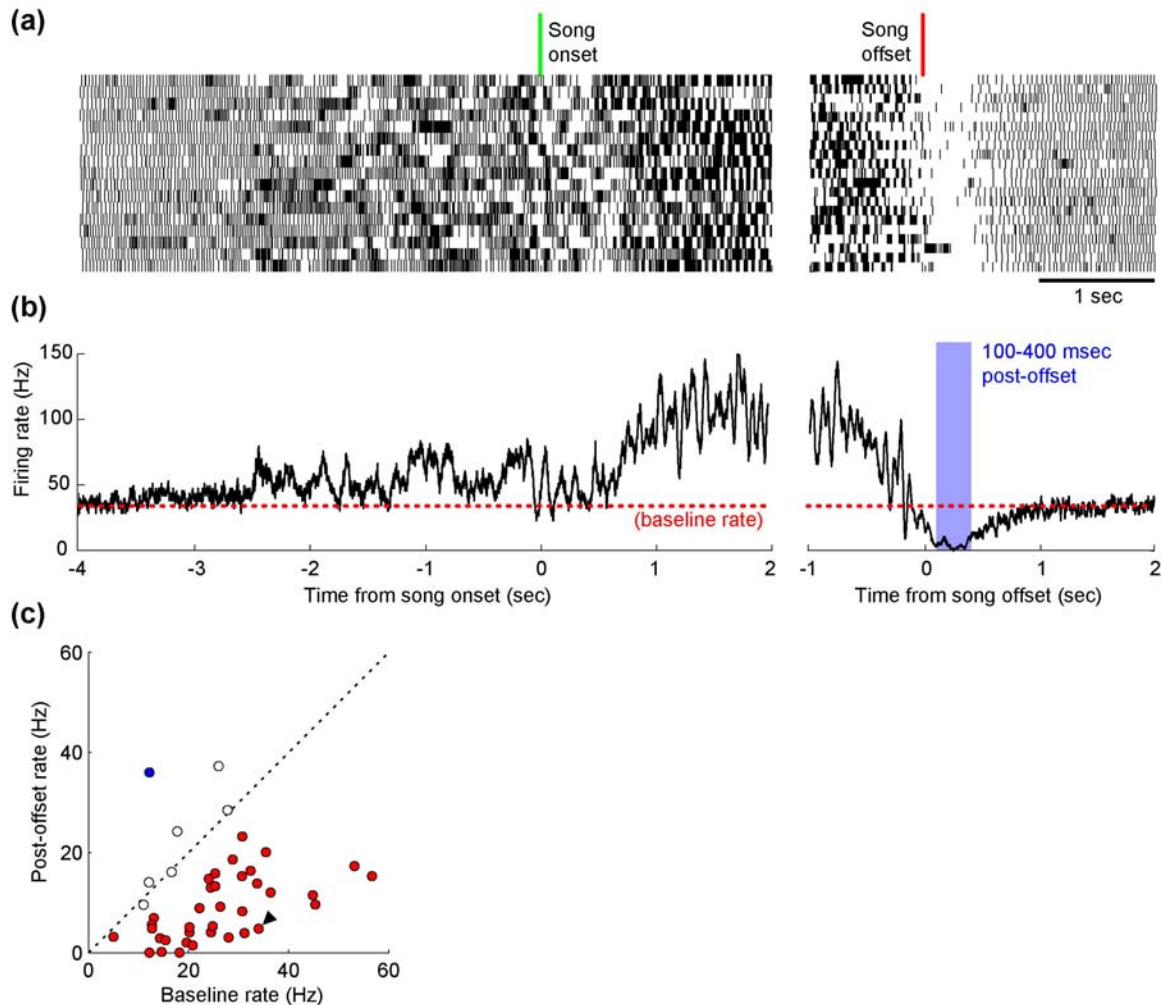
Appendix 3, Figure 3.7. Post-song inhibition. (a) Rasters show the activity of a single unit from Bird 4 aligned to the onset (left) and offset (right) of song. As was typical in our data, this unit switched from regular tonic activity (far left) to bursty firing several seconds prior to song onset. Following song offset, spiking was inhibited for approximately 500 msec, after which regular tonic firing resumed. (b) Mean firing rate for the unit shown in (a). Before song-related bursting begins, this unit had a baseline rate of 34 Hz (dashed red line). We quantified post-song inhibition by comparing the baseline rate to the mean firing rate in a window 100-400 msec after song offset (blue box). (c) Group data. We included only those units for which we collected at least 10 song offsets after which song did not resume for at least 3 seconds (n=43). Of these, 36 units (red dots) had significantly lower mean firing rates in the post-offset window, 1 unit (blue dot) had greater activity, and 6 units (white dots) were not significantly different (2-tailed t-test, p<0.05). An arrowhead indicates the unit shown in (a) and (b).

Putative interneurons vs. putative projection neurons

In contrast to the tonic activity characteristic of the majority of units (Figure 3.3a in the main text), a small subset of recordings (1 single-unit, 3 multiunit) had very low or no spiking activity when the bird was at rest and displayed bursty spiking activity during song. The spike width-at-half-height of the sole single unit of this type (103 μsec) was narrower than spike widths of all other single units (164 +/- 31 μsec, mean +/- S.D.). The similarity of the spike widths and activity patterns in this subset of units to a similar class in the zebra finch (Spiro et al., 1999; Leonardo and Fee, 2005) suggests that these four units are interneurons, and that the main body of our recordings are from projection neurons. The four putative interneurons were not included in further analysis. See Appendix 3, Figure 3.8b for an example of a recording from a putative interneuron alongside a recording from a putative projection neuron (Appendix 3, Figure 3.8a).
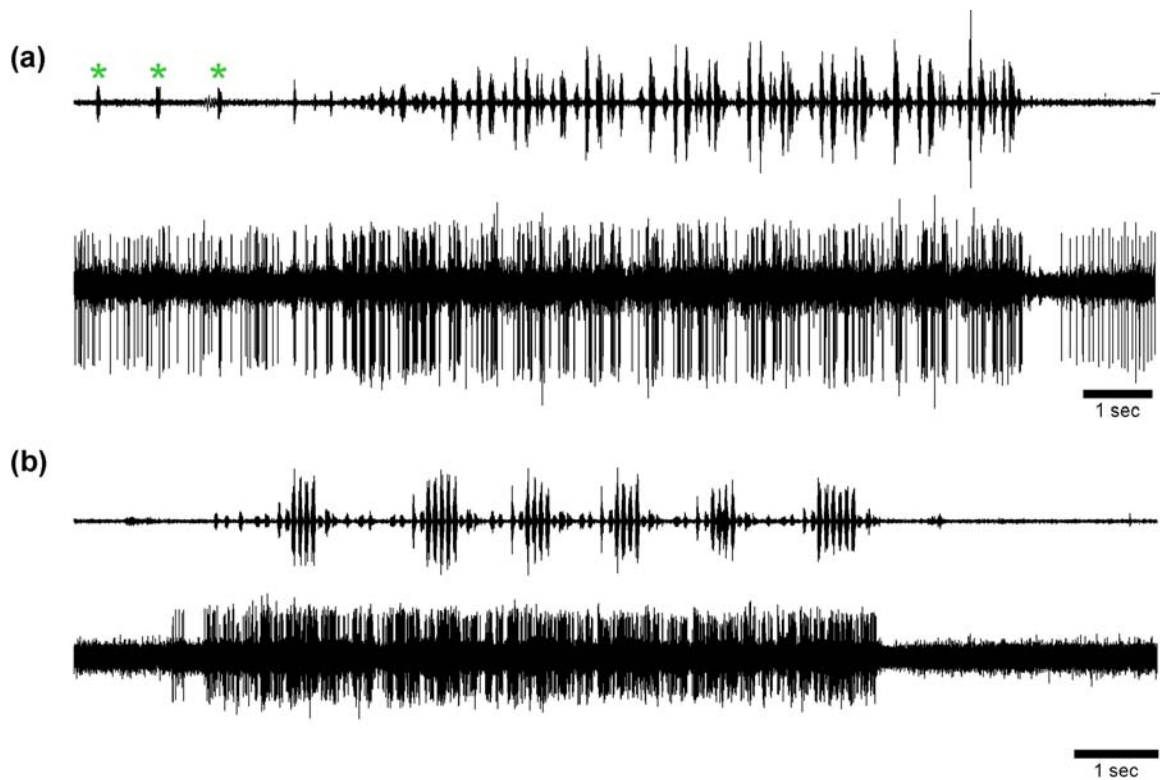
Appendix 3, Figure 3.8. Firing patterns of a putative projection neuron and a putative interneuron. (a) Raw sound amplitude trace (top) and a recording of a putative projection neuron (bottom) from Bird 2. Putative projection neurons are spontaneously active when the bird is not singing, displaying characteristic evenly spaced spikes, such as those seen in the last second of the neural trace. The bursty activity preceding the song is likely related to the production of three introductory notes, marked with green asterisks. (b) A putative interneuron from Bird 3. Neurons of this type are not active when the bird is at rest, but fire bursts of activity during song. Putative projection neurons and putative interneurons made up 97% and 3% of units sampled, respectively.

Sparse distribution of significant correlations

While most units are correlated with at least one acoustic parameter at some point during song, significant correlations are sparsely distributed (see Figure 3.6 in the main text and Appendix 3, Figure 3.9). That is, units are typically active before multiple syllables, but are significantly correlated with acoustic output in only a fraction of these cases. This sparse distribution might reflect dynamic changes in the strength of covariation between recorded neurons and the ensemble of RA neurons controlling each acoustic feature. Such changes would presumably lead to across-syllable differences in the correlation between that cell's activity and motor output.

The sparse distribution of significant correlations might also result from nonlinearities either in the brainstem targets of RA or in the syrinx itself. The activity of a group of brainstem motor neurons or a syringeal muscle might influence pitch (for example) during some syllables but not others. Variations in the activity of RA neurons driving these motor structures would therefore only produce pitch variations during a subset of syllables.

In addition to being sparsely distributed across syllables, significant correlations are also distributed across acoustic properties such that one unit can be correlated with more than one acoustic feature. This might reflect either the multiple actions of individual syringeal muscles or the connectivity of RA neurons to the motor neuron pool. EMG studies have shown that activation levels of the syringeal muscles controlling amplitude are also correlated with pitch (Goller and Suthers, 1996), suggesting that individual muscles can contribute to the control of multiple acoustic features. Furthermore, although RA has a roughly myotopic organization (Vicario, 1991), single

RA neurons might activate motor neurons controlling multiple muscles, thereby affecting

multiple aspects of song (Wild, 1993).

Appendix 3, Figure 3.9 (on following 2 pages) Significant correlations of premotor
activity with acoustic structure across all birds. Each box shows data from one bird in
our dataset, following the same conventions as Figure 6 in the main text (which showed
the data from Bird 1). Note that different birds can have different numbers of rows
(reflecting differences in the number of units recorded) and different numbers of columns
(reflecting differences in the number of syllables sung by each bird).

RA activity in directed vs. undirected song

Song is produced both in social isolation ("undirected song") and during courtship interactions ("directed song"). Previous studies have established that directed song is less variable than undirected song (Sossinka and Bohner, 1980; Kao and Brainard, 2006; Sakata et al., 2008). Furthermore, lesion, inactivation, and stimulation studies suggest that some of the increased behavioral variation observed during undirected song results from lMAN injecting neural variation into RA (Kao et al., 2005; Olveczky et al., 2005; Kao and Brainard, 2006). Together with the results of the current study, these prior findings suggest that the increased behavioral variability during undirected song might be driven by changes in the overall level of variability in RA activity across social contexts.

Although our dataset consists almost entirely of recordings during undirected song, in a small number of cases in one bird we were able to obtain both directed and undirected song while recording in RA. Two such recording sites yielded sufficient data in both conditions to allow for comparison across social contexts. As shown in Appendix 3, Figure 10, neural variability (CV) was indeed significantly lower in the directed condition. While preliminary, these results suggest that not only do trial-by-trial variations in RA activity drive variations in song (the main finding in our study), but also that modulations in the overall level of RA variability are responsible for social context-dependent changes in song.
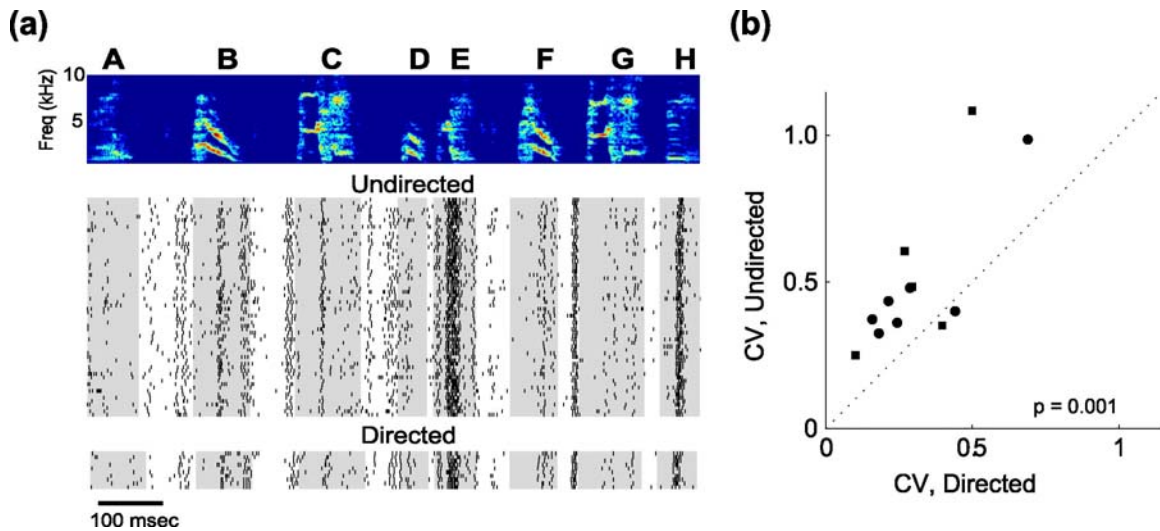
Appendix 3, Figure 3.10. Neural variability across social context. (a) Spiking activity during undirected (middle) and directed (bottom) song for Bird 3, unit 8. Bouts of directed and undirected song were interleaved during data collection and are plotted separately for visual clarity only. Other plotting conventions as in Figure 4 in the main text. (b) Variability of spiking activity in directed vs. undirected song. Each point plots the CV (SD/mean) of the number of spikes in the premotor window before one syllable. Only syllables with mean activity >25 Hz are analyzed (see Methods). Squares plot data from the unit shown in (a), circles plot data from Bird 3, unit 3. The p-value is from a Wilcoxon signed-rank test.

176

Correlations extend across time

When quantifying the relationship between neural activity and song output, we restricted our analysis of neural activity to a 40 msec premotor window preceding each syllable. This window likely encompasses the latencies with which spikes in RA directly (via synapses with motor neurons in the brainstem) influence behavior. Using this restricted time window allows us to examine the relationship between neural activity and acoustic output in the context of a single syllable.

However, variations in both premotor activity and acoustic output might have a timecourse longer than a single syllable (Glaze and Troyer, 2006), and an increase in firing rate or pitch (for example) in one syllable might correlate with a similar increase in the next syllable. In the extreme case, consecutive syllables could be perfectly correlated, both in terms of premotor activity and acoustic output. In this situation a 40 msec premotor window, although it includes the "true" causal latency of RA neurons, would have no more or less predictive power than any other window, and acoustic output would be predicted equally well by neural activity following, rather than preceding, the syllable. At the other extreme, variation in one syllable could be completely independent of variation in neighboring syllables. In this case, neural activity outside of the true premotor window would be a poor predictor of behavioral output.

To examine whether behavioral correlations extend across time, we asked whether the acoustic properties of a given syllable are correlated with the acoustic properties of the next syllable. We found this type of correlation to be widespread. The pitch of a given syllable was significantly ($p<0.05$) correlated with the pitch of the next syllable in 23.8% of cases. Similar measurements of amplitude and entropy yielded significant correlations in 48.5% and 21.4% of cases, respectively.

We then performed a similar analysis of premotor activity to ask whether neural activity is also correlated across time. As with acoustic output, we found a substantial number of correlations. The number of premotor spikes before a given syllable and the number of spikes before the next syllable were significantly correlated in 32.4% of cases.

The prevalence and strength of these correlations show that motor activity during consecutive syllables is neither perfectly correlated nor completely independent. Since consecutive syllables are not perfectly correlated, neural activity in the 40 msec premotor window we used should predict behavior better than activity taken from other times relative to the syllable. Since consecutive syllables are not independent, however, we expect that using other "premotor" windows to predict behavior would have some statistical power, though not as much as using the original premotor window.

We tested this prediction by repeating our analysis using two alternate premotor windows (Appendix 3, Figure 3.11). In one analysis, we examined correlations between acoustic data and premotor neural activity taken from the previous syllable (light gray bars). In a second alternate analysis, neural data were taken from the syllable after the one from which premotor neural data were taken (dark gray bars). Comparison of the results with those of the original analysis (black bars) confirms the prediction that these alternate premotor windows have some predictive power, but significantly less power than the window used in the original analysis.
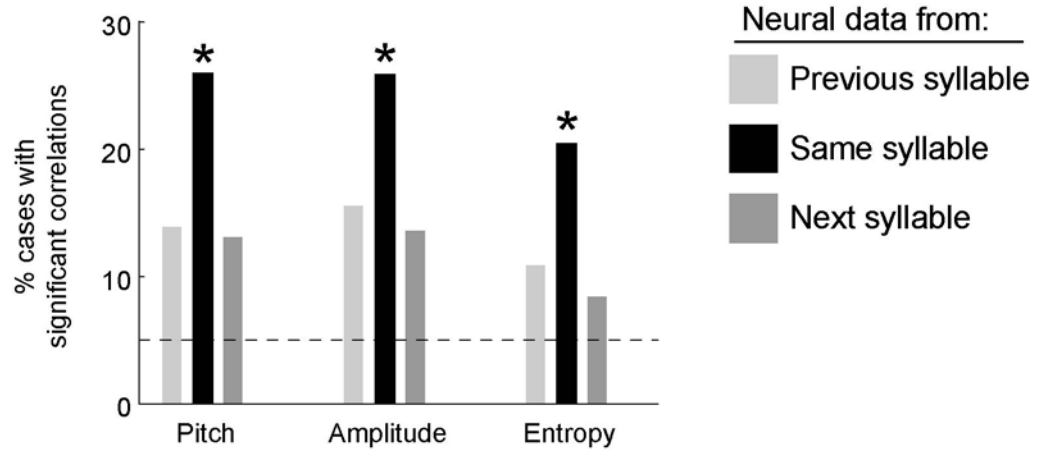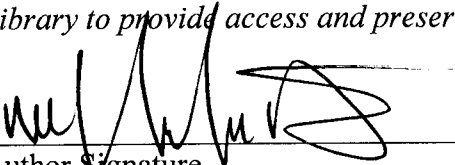
Appendix 3, Figure 3.11 Correlations with adjacent syllables. Each bar shows the percent of cases correlated with a particular acoustic feature. The black bars show the results of the main analysis (also shown in Figure 7a in the main text), in which the acoustic features of a given syllable are regressed against the premotor activity before that same syllable. Also shown are the results two alternate analyses, in which the neural data are taken from the premotor window preceding the previous syllable (light gray bars) or next syllable (dark gray bars). In all nine cases, the percent of cases correlated is significantly greater than chance (dashed line). Asterisks indicate that the proportion of cases correlated with neural data from the same syllable is significantly higher than those derived from using either of the two alternate premotor windows (Z-test for proportions, $p<0.0001$).

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

_____   12/12/08
Author Signature   Date

180