ABSTRACT

Title of dissertation: TEMPORAL STRUCTURE IN ZEBRA FINCH SONG: IMPLICATIONS FOR THE MOTOR CODE AND LEARNING PROCESS

Christopher Mulholland Glaze, Doctor of Philosophy, 2008

Dissertation directed by: Professor Catherine E. Carr, Department of Biology

One of the touchstone questions in neuroscience is how the nervous system encodes complex behavioral sequences such as speech. With experience-dependent learning, well-defined anatomy and complex temporal organization, zebra finch song has served as an excellent model system for these questions. Male songs are learned from older males during a sensitive period that includes song memorization and vocal learning guided by auditory feedback. Once learned, song acoustics are hierarchically organized into syllables, continuous stretches of vocalization separated by silent gaps, which are arranged into stereotyped sequences termed motifs; on a finer scale, syllables are composed of one or more notes, vocalizations with a homogenous spectral profile. Although much is known about the song system, progress has been limited by conflicting data on the neural basis of the acoustic hierarchy and the role this organization plays during learning: While behavioral and electrophysiological studies have suggested separate circuits and learning stages for individual syllables and syllable sequence, these models have been challenged by physiological evidence that songs are actually driven by a clock-like mechanism that does not segment songs into different units.

We have analyzed and modeled trial-to-trial timing variability in zebra finch song acoustics to investigate whether the hierarchy is in fact represented in the song system and learning process. Using automated template matching and dynamic time warping, we made millisecond-precise timing measurements in tens of thousands of recordings of both adult and juvenile song. In each adult song, we find rendition-to-rendition tempo variability that is spread across syllables and gaps; however syllable lengths stretch and compress with tempo changes proportionally less than gaps, *i.e.* they are less "elastic." Such non-uniformity is at odds with the simplest clock-based model in which songs are driven by a timing mechanism that paces song evenly across syllable-gap sequences. On the other hand, in a subsequent analysis we factored out tempo changes and used the remaining variability to investigate subsyllabic timescales that contradict the hierarchical model as well. Here, we find length variability that is specific to 10-msec song slices and independent of neighboring vocalization, yet correlated across motifs, providing the first behavioral evidence for a 5-10 msec timescale of song representation and an interaction with a neuromodulatory source operating on a much slower timescale. We have developed a model of song production constrained by the timing data; modeling suggests that adult song may be produced by an underlying chain of activity on a single 5-10 msec timescale, but with properties such as synaptic strength that do correspond to the acoustic hierarchy.

Finally, we analyzed juvenile song within the same framework and investigated how the timing properties we modeled may develop during sensorimotor learning.

The behavioral data indicate a period towards the end of learning in which syllable sequences become more stereotyped, tempo increases selectively among gaps, and independent timing variability falls two- to threefold across syllables and gaps. In remarkable contrast, over this same period we find no changes in patterns of global tempo variability or the fine timescale patterns indicative of chaining mechanisms. Overall, the developmental data suggest a final phase of song learning in which syllable-based representations are consolidated into the longer sequence-based chaining mechanisms proposed for the adult system. A similar process of linking simpler chains to form more functional activity patterns has been proposed for neocortex and other models of sequence learning in mammalian systems. In this respect, adult zebra finch song representations may be most analogous with procedural memory and overlearned sequences such as repetitive speech patterns.

TEMPORAL STRUCTURE IN ZEBRA FINCH SONG: IMPLICATIONS FOR THE ADULT MOTOR CODE AND LEARNING PROCESS

by

Christopher Mulholland Glaze

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Dr. Catherine E. Carr, Chair/Advisor
Dr. Todd W. Troyer
Dr. Robert J. Dooling
Dr. Jonathan Z. Simon
Dr. José L. Contreras-Vidal

Acknowledgments

First of all I would like to thank Dr. Todd Troyer, who has provided me with an endless list of supportive, brutally honest contributions to both my research and general intellectual development. I thank Dr. Catherine Carr, who has generously supported me during a crucial time as a graduate student. I would also like to express great gratitude to Dr. William Hall for guiding me towards the exciting field of birdsong. Thanks the rest of my committee for their time as well as useful feedback and encouragement; the Program in Neuroscience and Cognitive Science for financial support and an enriching environment; and the Center for the Comparative and Evolutionary Biology of Hearing for financial support. I also thank Sandy Davis, Chris Grogan and Pam Komarek for invaluable administrative support.

Thanks to Dr. Ed Smith for his indispensable contribution to the control analysis; Tina Johnson for helping me collect juvenile song; Ping Du, Guruprasad Kudva and David Sivakoff for support in the lab maintaining the colony and song recording system; the University of Maryland Animal Care and Use staff for keeping the birds alive and clean. A number of colleagues have provided useful feedback and proofing, including Dr. Michael Brainard, Dr. Alexay Kozhevnikov, Joanna Pressley, Dr. Marc Schmidt, Jon Raskin and Beth Vernaleo.

Finally, I want to give gratitude to my family. I cannot thank Violet enough for not only putting her career on hold so I could finish the thesis, but proofreading the final draft. Thanks to both Violet and Linus for simply being here when I needed their love and support. Thanks to my mom and dad for never giving up on me and encouraging me to think for myself growing up, I've found this to be one of the most important skills in science and life in general. Thanks to mom- and dad-in-law Pat and Drew for tolerating me while I went nuts in their basement processing data and working through equations.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

One of the touchstone questions in neuroscience is how the nervous system encodes complex behavioral sequences such as speech (Lashley, 1951; Keele et al., 2003; Rhodes et al., 2004). Birdsong has several important characteristics that make it an excellent model system for addressing these questions. In many species songs are learned by males via auditory exposure to their fathers or other males in the colony during a sensitive period in the first months of life (Konishi, 1965; Brainard & Doupe, 2000). Once learned, song temporal patterning is hierarchically organized into syllables, continuous stretches of vocalization separated by silent gaps, which are arranged into structured sequences termed motifs; on a finer scale, syllables are composed of one or more notes, vocalizations with a homogenous spectral profile. However, while birdsong and many other sequential behaviors contain obvious hierarchical organization in the behavior, these areas of research also share the fundamental problem that the neural bases for this organization remain elusive.

1.1 Adult song system

A range of lesion and electrophysiological studies have thus far established that forebrain nucleus HVC (used as proper name) and the afferent robust nucleus of the arcopallium (RA) compose the primary pattern generator for adult birdsong (Nottebohm et al., 1976; McCasland, 1987; Simpson & Vicario, 1990; Yu & Margoliash, 1996). Early multiunit recordings and microstimulation experiments suggested that the acoustic hierarchy is reflected over these circuits, with nucleus HVC being responsible for syllable sequence and nucleus RA representing individual syllables (Vu et al., 1994; Yu & Margoliash, 1996). The hierarchical representation of song is corroborated by behavioral studies in which flashes of light cause birds to interrupt their song at the boundaries of syllables and occasionally notes (Franz & Goller, 2002; Cynx, 1990). Furthermore, the respiratory pattern accompanying song production is as stereotyped as the acoustics, and consists of inspiration/expiration patterns that segment the song into syllables and acoustic gaps (Goller & Cooper, 2004; Suthers & Margoliash, 2002; Wild et al., 1998).

However, more recent physiological investigations into HVC and RA suggest that song may be driven by a precisely timed, clock-like mechanism that does not segment song into the acoustic units that are apparent in song spectrograms. During the production of a single motif, a given RA-projecting HVC neuron ($\mathrm{HVC_{RA}}$) bursts once and only once during motif production, and burst timing is locked to the same point in song with millisecond precision (Hahnloser et al., 2002); although the number of recorded neurons was limited, it was argued that the sparse bursting pattern across $\mathrm{HVC_{RA}}$ neurons serves as a clock-like mechanism that drives song without regard for syllable boundaries. Bursts in RA have also shown remarkable precision when aligned with song acoustics (Chi & Margoliash, 2001), and activity here appears to drive song with populations of neurons that are uncorrelated during any two 5-10 msec timepoints in a song motif (Leonardo & Fee, 2005). These findings in the forebrain have led to the proposal that song is driven by activity that operates on a single 5-10 msec timescale, without respect to different units in the behavior (Fee et al., 2004; Leonardo & Fee, 2005). Under this proposal, the song system involves no hierarchical division of motifs into syllables, silent gaps or notes.

1.2 Song learning

While the structure of adult song representation remains unclear, even less is known about the neural processes that subserve song development. In the zebra finch, song learning can be roughly divided into two overlapping processes (Immelmann, 1969; Marler, 1970; Konishi, 1985; Brainard & Doupe, 2000): "sensory acquisition," in which a young bird ∼20-65 days post-hatch (dph) is exposed to the song of one or more tutors and forms an auditory template; and "sensorimotor learning," in which the juvenile ∼35-90 dph learns to produce song based on that template. Learning involves the anterior forebrain pathway (AFP), which is homologous with basal ganglia circuits implicated in mammalian sequence learning (Doupe & Kuhl, 1999). Physiological evidence currently indicates that output nucleus LMAN actively induces variability in song production (Ölveczky et al., 2005; Kao et al., 2005). In contrast to HVC and RA, activity from LMAN shows no reliable correspondence with song output in either adults and juveniles, and LMAN lesions during sensorimotor learning prematurely crystallize song in a state that ceases to progress (Bottjer et al., 1984; Scharff & Nottebohn, 1991; Brainard & Doupe, 2001). It thus may be that the AFP is responsible for controlling an explore-exploit strategy in which the juvenile bird experiments with different premotor patterns and eventually arrives upon the template song; such a strategy has been suggested in other models of mammalian sequence learning and the role of striatal circuits in those systems (Graybiel, 2005; Barnes et al., 2005).

Far less is known about how juvenile songs are represented over the HVC-RA pathway so prominent in the adult song system. It has been proposed that the HVC timing mechanism proposed for adult song is present throughout development (Fiete et al., 2004); in this model, one of the chief tasks for the system is to map the clock to downstream neurons, *i.e.* modify HVC-RA synapses. This proposal is corroborated by behavioral evidence for "in-situ" syllable learning, in which syllables gradually emerge from a sequential pattern of vocalizations whose order never changes (Tchernichovski et al., 2001; Liu et al., 2003). On the other hand, it remains an open question as to whether there is in fact a distinct neural process associated with learning syllable sequences that is different from learning individual syllables (Troyer & Doupe, 2000b). In fact, several studies have suggested significant syllable-sequence variability in juvenile song that is sensitive to LMAN lesions (Scharff & Nottebohn, 1991; Bottjer et al., 1984; Ölveczky et al., 2005).

1.3 Motor variability

While stereotypy is a prominent characteristic of each adult zebra finch's song, the adult motor code is also capable of variability that may be driven by the same mechanisms responsible for song learning; the structure of this variability poses significant challenges to studies that rely on averaged burst sequences to make claims about the timescale of song representations (e.g. such as Hahnloser et al., 2002; Leonardo & Fee, 2005).

For example, experimental induction of reversible changes to song patterning can selectively divide out whole syllables from the motor code. These studies include temporary disruption to auditory feedback (Leonardo & Konishi, 1999) and obstruction of the syrinx (Hough & Volman, 2002), which change syllable sequence and acoustic structure, occasionally deleting entire syllables from the song; such changes are eventually reversed after restoration of normal peripheral functioning. Similar reversible song changes have been demonstrated with permanent unilateral Uva lesions (Coleman & Vu, 2005) and even microlesions to HVC (Thompson & Johnson, 2007); in both cases songs undergo temporary sequencing changes and degradation in acoustic structure but eventually recover, presumably due to the influence of a song template that is maintained elsewhere. In all of the above studies songs exhibited changes to temporal structure on both the sequence and syllable levels, suggesting that there may be neural representations specific to each timescale. Furthermore, it is worth noting that in some birds with unilateral Uva lesions or an obstructed syrinx, permanent sequence changes were also observed despite restoration of the acoustic structure of individual syllables (Hough & Volman, 2002; Coleman & Vu, 2005).

There is also naturally occurring variability in adult song that may provide insight into the motor code. Songs that are not directed towards a female tend to have slower tempo and greater rendition-to-rendition variability in acoustic features, song timing and syllable sequence (Sossinka & Bohner, 1980; Brainard & Doupe, 2001; Kao & Brainard, 2006; Kao et al., 2005). There is cursory evidence that this variability may be directly linked with HVC-RA burst sequences: RA burst-onset times during song have rendition-to-rendition variability that accumulates over a given motif rendition (Chi & Margoliash, 2001). This variability is strongly correlated with deviations in the timing of the acoustic song features that are driven by these bursts, suggesting that some component of timing variability in song acoustics may be directly tied to the timing variability in premotor burst sequences. Conceivably, this variability in burst onset times could have correlation structure that does correspond to the acoustic hierarchy even though averaged burst sequences give the appearance of clock-like activity.

It is worth noting that most of the variability discussed above appears to be gated by the AFP, the same circuit that is also crucial to song learning: LMAN lesions in adults prevent changes to song caused by deafening (Brainard & Doupe, 2001), denervation of the syrinx (Williams & Mehta, 1999) and microlesions to HVC (Thompson & Johnson, 2007). The increased variability observed during undirected song also appears to be gated by the AFP (Hessler & Doupe, 1999; Brainard & Doupe, 2001; Kao & Brainard, 2006; Kao et al., 2005). This suggests that the influences of social context and experimental perturbations may involve similar mechanisms, *i.e.* those involved in generating variability for song learning.

1.4 Outline of Thesis

Is the acoustic hierarchy represented in the adult song system and song learning? An overview of the aforementioned studies suggest that variability in the structure of individual songs may provide important clues to this question. In the following chapters we probe the adult motor code and the song learning process using naturally occurring variability in song timing. We use a combination of fine-grained measurements of song acoustics and computational modeling to investigate the song system with the following central argument: Timing deviations that originate downstream from the pattern generator must be offset by compensating deviations, provided that the pattern generator continues to pace activity independently of the sources of those deviations; thus, timing variability that accumulates over the length of a motif or song bout must be derived from the song pattern generator itself (Fig. 3.5). The strong link between accumulated timing variability in RA burst sequences and acoustic song features (Chi & Margoliash, 2001) suggests that the magnitude of such variance may be strong enough to uncover meaningful insights into song representations.

In Chapters 2 and 3 we probe the adult motor code to ask whether the motor code is in fact chunked into a hierarchy of different song elements. In Chapter 2 we used a semi-automated template matching algorithm to identify repeated sequences of syllables, and dynamic time warping (DTW) to make fine-grained measurements of the temporal structure of song. We find that changes in song length are expressed across the song as a whole rather than resulting from an accumulation of independent variance during singing. Song length changes systematically over the course of a day, and is related to the general level of bird activity as well as the presence of a female. The data also show patterns of variability that suggest distinct mechanisms underlying syllable and gap lengths: as tempo varies, syllables stretch and compress proportionally less than gaps, while syllable-syllable and gap-gap correlations are significantly stronger than syllable-gap correlations. There is also “identity-dependent” timing, especially strong positive correlations between the same syllables sung in different motifs that are an upwards of 2-3 seconds apart. Finally, there is increased temporal variability at motif boundaries, and evidence that syllable onsets may have a special role in aligning syllables with global song structure. Generally, the timing data in Chapter 2 support a hierarchical view in which song is composed of syllable-based units with distinct timing properties.

In Chapter 3 we investigate similar timing patterns on subsyllabic timescales and ask whether the identity-dependent syllable timing from Chapter 2 is itself due to a syllable-based code (as predicted by a hierarchical model) or representations on a finer timescale (predicted by the clock model). We find evidence for the latter. First, identity-dependent timing is dominated by independent variability in notes, finer song segments that compose a syllable; for example, the length of a note is no more correlated with other notes in the same syllable than it is with notes in other syllables. For a subset of notes, clear modulation in spectral structure allowed for accurate timing measurements on the 5-10 msec timescale. Temporal independence holds at this scale as well: the length of an individual 5-10 msec song slice is correlated with the same slice repeated 500-1000 msec later, yet is independent of neighboring slices. Overall, these data provide behavioral evidence for the ≤5-10 msec timescale proposed for the HVC-RA pathway and suggest fine-grained, persistent changes in song tempo that result from an interaction between slow modulatory factors and precisely timed, sparse bursting in HVC and RA.

Taken together, Chapters 2 and 3 provide a rich set of behavioral constraints on models for adult song production and suggest that neither the hierarchical nor clock models are adequate to capture the range of timing patterns we find. In Chapter 4, we attempt to synthesize our findings into a reduced synfire chain (Abeles, 1991) that had been previously proposed for adult song (e.g. Fee et al., 2004). Because song length variability reflects tempo changes shared across syllables and gaps, we focus on how modulatory changes in global network parameters affect the propagation of activity along the underlying chain. We find several influences on propagation speed that explain how timing patterns may be linked with the adult song system: First, tempo changes may be linked with fluctuations in neural excitation that are shared across the chain; this suggests neuromodulatory factors and is corroborated by the links between song tempo and ethological factors such as time of day and the presence of a female. Second, both average chain speed and elasticity (sensitivity to tempo changes) can be directly tied to a combination of the strength and density of synaptic connections between links in the chain, and the gain of inhibitory feedback from interneurons that have been found in premotor nuclei. Third, while the timing variability that remains after factoring out tempo is straightforward to model, the magnitude of this variability may require correlations within individual chain layers that are at odds with the simplest hypothesis that chains are anatomically distributed at random. Overall, the modeling suggests that while the timing variability we have found in adult song contradicts the clock model, it is consistent with the proposal that song is driven by a single chain of activity on a fine timescale. However, the data at minimum require physiological parameters (such as synaptic strength) that do correspond to the acoustic hierarchy by distinguishing syllable-based from gap-based chain segments.

How are the timing properties of adult songs learned? In Chapter 5 we probe the development of the song system from late plastic song through ∼1 year of age and ask how several of the key patterns of timing variability we found in Chapters 2 and 3 develop. We focus on average song speed and two basic kinds of timing variability: "global variability" that drives tempo changes spread across the song, and "local variability" that reflects the timescale of song representations across the chain. We also examine sequence variability, specifically, changes in syllable-transition probabilities, which could also conceivably change during late plastic song. As with adults, the developmental data distinguish different elements of the acoustic hierarchy: Average song tempo increases almost entirely during the gaps of silence between syllables. Most of this tempo increase occurs 65-90 dph, and on a gap-by-gap basis increased speed is linked with increases in the reliability of corresponding syllable transitions, as well as decreases in local timing variability. We also find interesting changes in timing variability itself: while the magnitude of local variability decreases two- to threefold through the first year of life, we find no significant changes in global timing variability that is spread across the song bout. Furthermore, we do not find any changes in elasticity patterns as far back as 85-90 dph, nor changes in the timescale of local variability. Overall, the developmental data suggest a late learning phase in which syllable-based activity becomes more organized into the tightly-timed, automated process indicated by adult song. This period of motor consolidation has been previously suggested for zebra finch song (e.g. Brainard & Doupe, 2001) and has been proposed for the formation of procedural memories such as sequential behaviors in mammalian systems (Rhodes et al., 2004).

As with the adult analysis, the developmental patterns of timing provide a rich set of behavioral constraints for models of the learning process. More generally, the timing variability we have analyzed provides a common language for synthesizing results from behavioral and electrophysiological studies into models that attempt to connect the two levels of analysis.

Chapter 2

Temporal Structure in Zebra Finch Song: Implications for Motor Coding

(Glaze CM, Troyer TW (2006), *J Neurosci* 26(3): 991-1005)

Most natural behaviors are arranged hierarchically, with complex actions composed of a serial combination of more basic motor gestures (e.g. Lashley, 1951; Miller et al., 1960). The learning of courtship song in birds presents an ideal model system for understanding the neural mechanisms underlying complex behavior. Birdsongs have a hierarchical structure spanning time scales from several milliseconds to several seconds, and are executed by well-delineated circuitry known as the song system (figure 2.1).

Zebra finch songs are highly stereotyped, making them especially well suited for in depth analysis. The acoustic structure of song is arranged in a hierarchy, with vocal units known as syllables strung together in sequences called motifs (figure 2.2). Several lines of evidence suggest this acoustic hierarchy is embedded within the underlying representation for song. Flashes of light cause birds to interrupt their song at syllable boundaries (Franz & Goller, 2002; Cynx, 1990), and the patterns of inspiration/expiration segment the song into syllables and acoustic gaps (Goller & Cooper, 2004; Suthers & Margoliash, 2002; Wild et al., 1998). Early electrophysiological experiments suggest this structure is reflected in the anatomical hierarchy in the forebrain, with nucleus HVC being responsible for syllable sequence and nucleus RA representing individual syllables (Vu et al., 1994; Yu & Margoliash, 1996).

This hierarchical view has been challenged by recent recordings of the HVC neurons projecting to RA ($\mathrm{HVC}_{(\mathrm{RA})}$ neurons, Hahnloser et al., 2002). During each motif, individual $\mathrm{HVC}_{(\mathrm{RA})}$ neurons produce a burst of spikes aligned to the acoustic output on the millisecond time scale. Although the number of recorded neurons was limited, the bursts did not appear to respect syllable vs. gap distinctions. These results have led Fee and colleagues to propose what we term the "music box" model for song production (Fee et al., 2004): HVC activity serves as the clock-like drum of the music box, and the HVC-RA synaptic connections trigger bursts in RA, which get read out by brainstem motor nuclei. Under this proposal, there is no hierarchical division of motifs into syllables and gaps. With its uniform representation, the music box model predicts that changes in song tempo should be accompanied by proportional scaling of all parts of the song, with no correlation structure that would delineate articulatory units (see Rhodes et al., 2004; de Jong, 2001; Heuer, 1988; Gentner, 1987, for discussions of proportional scaling in humans).

We addressed these issues by making fine-grained measurements of syllable timing within zebra finch songs and analyzing subtle patterns of variation within and across motifs. We find that song length changes systematically over the course of a day and these changes are expressed across the song as a whole rather than resulting from an accumulation of independent variance during singing. Our data also show patterns of variability that distinguish syllables and inter-syllable gaps

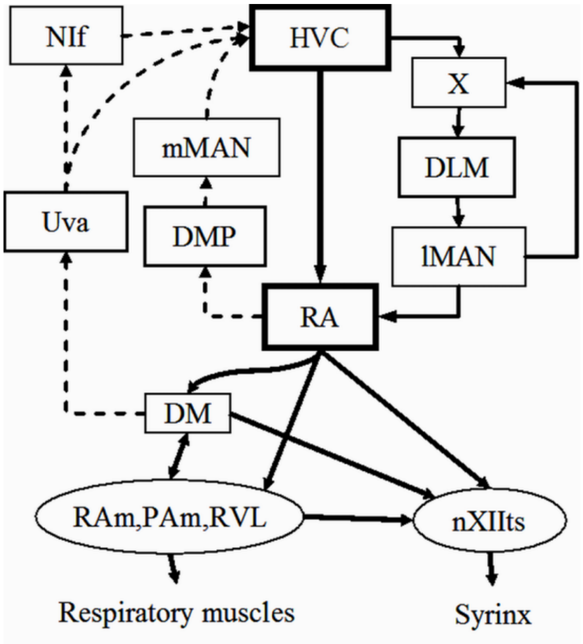

Figure 2.1: The song system. The premotor pathway consists of HVC (used as proper name) to RA (robust nucleus of the arcopallium) to brainstem nuclei RAm (retroambigualis), PAm (parambigualis) and RVL (ventrolateral nucleus of the rostral medulla), which project to respiratory motorneurons, and nXIIts (nervi hypoglossi, pars tracheosyringealis) which projects to the syrinx. RA can influence respiratory brainstem nuclei via alernative circuitry passing through the midbrain nucleus DM (dorsomedial intercollicular). DM is also involved in an ascending pathway that extends to Uva (nucleus uvaeformis), NIf (interfacial nucleus of the nidopallium), and back to HVC. HVC-RA activity is modulated by two pathways: (1) the descending anterior forebrain pathway (AFP), which consists of Area X, DLM (dorsolateral nucleus of the medial thalamus) and LMAN (lateral magnocellular nucleus of the anterior nidopallium), and (2) an ascending pathway from DMP (dorsomedialis posterior thalami) to MMAN (medial magnocellular nucleus of the anterior nidopallium).

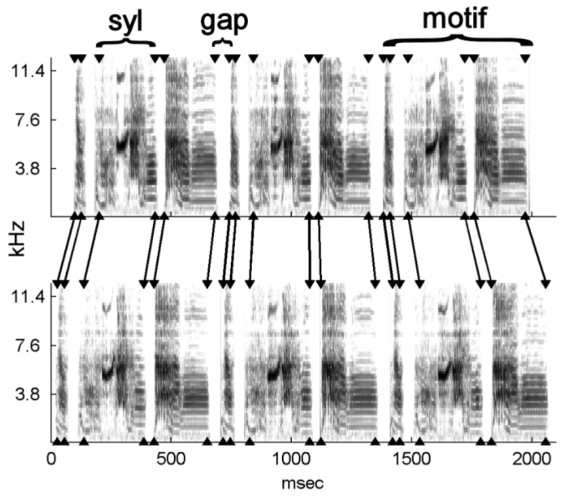

Figure 2.2: Spectrograms from the shortest (top) and longest (bottom) songs in the sequences recorded from bird 10. Arrow heads indicate each syllable's onset and offset measured by an automated algorithm. The algorithm marks syllable boundaries according to reliable peaks in the amplitude derivative, so that less reliable, small amplitude parts of some syllables fall outside these boundaries (see Methods).

and thus provide strong evidence for hierarchical structure in the song output. As tempo changes, syllables stretch and compress proportionally less than gaps, a violation of the proportional scaling implied by the simplest form the music box model. (Note that our data do not bear on the sparseness of the underlying representation; Hahnloser et al. (2002).) We also find increased temporal variability at motif boundaries and especially strong positive correlations between the same syllables sung in different motifs.

2.1 Materials and Methods

Terms Zebra finch song consists of several introductory notes followed by series of discrete vocalizations that is repeated several times (figure 2.2). We will refer to syllables as any vocalization delineated by silence on either side, motifs as stereotyped series of syllables, and songs as an uninterrupted series of motifs produced back-to-back and separated by silence on either side. Since we truncate songs to analyze identical sequences of syllables with a fixed number of motifs, we will generally use the term “sequence” rather than song.

2.1.1 Birdhousing and recording

Recordings

All care and housing of birds conformed to the procedures approved by the institutional animal care and use committee at the University of Maryland, College Park. Birds were maintained on a 14/10 hour light/dark cycle and given food ad libidum.

This paper focuses on the analysis of temporal variability in songs produced by zebra finches in the presence of other male birds (“undirected song”; Sossinka & Bohner, 1980). The majority of these recordings (86%) were obtained from adult birds that were acting as tutors for other developmental studies. Other recordings were made in the presence of another adult male or when the bird was alone. To minimize the effect of subtle changes in song that can occur in young adulthood, the data analyzed came from birds that were at least 400 days post hatch (dph). We also examined temporal structure within a smaller data set of songs recorded in the presence of a female (“directed song”). Unless explicitly specified, all statements pertain to the larger sample of undirected song.

For all recordings, birds were housed in (approx. 18x36x31 cm) cages within small sound isolation chambers (Industrial Acoustics, Bronx, NY). Cages were separated by ∼18 cm and two directional microphones (Pro 45; Audio-Technica, Stow, Ohio) were placed in this space. Between recording sessions birds were returned to the colony room where they were housed in larger cages with several (≤ 6) adult males or were paired with a female for breeding.

Real time signal processors (Tucker Davis, Alachua, FL) digitized the signal at 24414.1 Hz. Data were selected using a circular buffer and a sliding window amplitude algorithm (∼10 msec of below threshold sound needed to stop data recording). "Sound clips" selected by this algorithm that were separated by less than 200 msec were considered part of the same "recording" (candidate song) and were "glued" back together with the correct temporal alignment by filling periods between clips with zeros. Extensive examination of song recordings indicated that relative power in the two microphones averaged over the entire song is sufficient to unambiguously determine which bird is singing in the vast majority of cases. All signal analysis was performed in Matlab (Mathworks, Natick, MA).

Sample

To minimize the effects of extraneous behavioral variables, recordings of undirected song spanned at least 100 days and included data from at least 2 recording sessions with different juveniles. This left us with a universe of 20 birds. Of these, 1 bird was omitted because the song was deemed excessively variable and noisy and 5 were omitted because as juveniles they had learned most of their song from another adult in the sample. For each bird we gathered an initial pseudo-random sample of

1000 candidate songs that were at least 1200 msec long and had maximum power from the side on which the target adult was stationed. We omitted two additional birds because the template-matching algorithm (described below) failed to detect more than 100 sequences from those samples. This left us with a sample of 12 birds.

2.1.2 Sequence identification

All analysis described here concerns the main sample of undirected songs. Directed analysis is detailed below. In the first stage of analysis, repeated sequences of syllables were identified within the recording samples. A syllable was defined as any regular and continuous vocalization delimited by at least 5.243 msec of silence on either side. In a few cases, syllables occasionally split into parts separated by periods of silence longer than this criterion. In these cases, the entire period was analyzed as a single syllable. Mean syllable length was 110 ± 56 msec with a range of 37-294 msec. Recordings were analyzed using the complex amplitudes obtained from a fast-Fourier transform (FFT) using a 256-sample (10.486 msec) window advanced in 128-point steps.

Sequences were identified using a semi-automated procedure consisting of several steps: (1) a hand screening of song was done to determine the most common sequence of syllables produced by each bird; (2) a spectro-temporal template was developed for each syllable using an average of 2-3 syllable exemplars; (3) a modified sliding cross-correlation method was used to determine candidate matches of individual syllable templates to recorded data; (4) these candidate matches were used to select entire sequences using timing and syllable order information; (5) selected sequences were hand screened to ensure that they matched the template sequence. The goal of the template matching process was to obtain clean recordings from which accurate temporal measurements could be made using more sophisticated techniques. Candidate songs that did not have a good match to the template were not considered. This may have introduced an unknown amount of selection bias, although visual inspection revealed nothing obvious.

(1) Template sequences For 10 birds the most common sequence had 2-4 motifs produced back-to-back, or approximately 1300-2000 msec of continuous song; a greater number of motifs per sequence were chosen if that bird's motif was relatively brief. In two birds, sequences were more variable and the most common sequence was determined using syllable transition probabilities that had been calculated for other research. Sequencing in these birds was more variable overall so a higher percentage of those songs were excluded from the analysis. The average motif had 5 unique syllables (this factors out repeats; range 3-8) for a total of 60 unique syllables in the sample. Across motifs, each bird's sequence was comprised of a mean of 12 syllables (range 9-18; repeats counted multiple times) for a total of 146 syllables across birds.

(2) Syllable templates Syllable template spectrograms were based on 2-3 exemplars taken from a single song. The first exemplar syllable was arbitrarily used as a "proto-template" and other exemplar spectrograms were aligned to this proto-template using the syllable template matching algorithm described below. The template

spectrogram was obtained by averaging across exemplars the amplitudes for each time-frequency bin.

(3) Matching syllable templates and song data   For each syllable template, candidate matches were determined using a sliding window the same length as the template syllable. The spectrograms of the syllable template and each window of the song were normalized separately by first subtracting the mean from each time-frequency bin and then dividing out the root mean squared deviation. The distance between the normalized values was defined as the absolute value of the difference in each time-frequency bin, summed over all bins. The final match value was calculated as the reciprocal of this distance. This match value was observed to be significantly more accurate than a similar one using Euclidean (squared) distance, which gives more weight to larger differences. For each syllable, match values were computed across the bird's entire sample. The threshold for candidate matches was set at the 90th percentile of the distribution of match values. This criterion resulted in a higher threshold for syllables that had stronger matches to extraneous sound or other syllables. An initial set of candidate matches was then determined as above-threshold peaks in the match. Given similarities in acoustic structure between syllables, this process produces a number of false positives.

(4) Sequence selection   The next step in the analysis was to prune candidate matches using motif and timing information. For the purposes of the timing analysis detailed below, the only sequences we considered were stereotyped renditions. The selection

algorithm starts with the candidate syllable match that is most likely to belong to a target sequence and searches forward and backward to find syllables matches that are in the expected order and at roughly the expected time.

The main difficulty in sequence identification was false positives from syllable-by-syllable matches, so the algorithm started with the syllable that had the least number of matches. The location of these candidate syllables were defined as "anchors" for determining candidate motifs. From each anchor, a candidate motif was determined by serially moving forward/backward and looking for the appropriate next/previous syllable with the appropriate timing. The forward-searching routine was defined as follows: given that syllable n has just been found, look for a match to syllable $n + 1$ starting in a time window from t1 to t2, where t1 = onset(n) + .6*length(n), and t2 = onset(n+1) + 31 msec. In this way, 2 syllable matches could overlap by no more than 40% of the length of the first template, and an inter-onset interval could deviate positively from the motif template by a maximum of 31 msec. This upper bound was determined in a previous analysis of syllable onset deviations using a similar template matching algorithm. The backward-searching routine was a mirror-image of the forward-search. Both routines were looped until the boundaries of the sequence template were reached. If at any stage the algorithm failed to find a candidate match for the appropriate syllable that motif was deemed not to match the template and was not included in further analysis. In the two birds with non-identical motif structure, the entire song was treated as a single motif.

If any candidate motifs overlapped, the one with the highest average template match was chosen; this step was based on the assumption that an overlap would indicate that at least one of the matches was a false positive. If the gap between 2 identified motifs deviated positively by more than 25 msec they were partitioned into separate sequences; otherwise they were counted as part of the same sequence. For example, if in the same recording a bird produced a series of motifs, paused for a period of time and produced a second series, each series was counted as a different sequence. On average, about 1.1-1.2 separate target sequences per recording were identified; however, in one bird this ratio was closer to 2 because most sequences were separated by a variable number of call-like notes not defined in the sequence template.

Across birds the algorithm obtained a match to at least one motif in 56 ± 25% (std. dev.) of sampled recordings (range 17-95%). Most of the recordings without a match consisted of back-and-forth calling and wing flaps. The remainder of omissions was due to sequences of syllables that did not match the most common motif structure, or failures of the template-matching process – mostly due to acoustic interference from other birds in the recording chamber. Within a random sample of 140 sequences, visual inspection revealed that an estimated 16% of the recordings in which the algorithm failed to find a match actually contained a target sequence that was missed. Of the 4 birds with the fewest matches (fewer than 30% of recordings matched), 2 had songs with variable sequencing and the proportion of rejections due to non-standard sequences was much higher than the rest of the sample (estimated 40% and 80% compared with 2% for the rest of the sample). In the other 2 birds, the template matching process was unreliable, leading to above-average proportions of target sequences that were incorrectly rejected (estimated 50% and 70%). These

2 birds were excluded from our sample because the template matching yielded fewer than 100 sequences each. Finally, of the recordings that contained a match to at least one motif, a mean of 40% were omitted because they contained fewer motifs than in the target sequence for that bird.

(5) Hand screening All target sequences satisfying the initial template matching procedure were screened by visual inspection of log-transformed spectrograms to determine if they were suitable for high precision temporal analysis. Across birds an average of 40% were omitted, the majority due to mild acoustic interference from the other bird in the recording chamber.

2.1.3 Temporal analysis

In the second stage of analysis, the timing of each syllable onset and offset was determined by a modified dynamic time warping (DTW) algorithm applied to templates consisting of the derivative of the smoothed amplitude envelope of each syllable. Steps of this analysis were: (1) segment and select a portion of the original signal surrounding each syllable match obtained from the sequence matching procedure above; (2) calculate the smoothed amplitude envelope and take its derivative; (3) for each syllable, make a template using the averaged waveform from all examples of that syllable in the sample; (4) set onset and offset times by choosing peaks and troughs in the template waveform; (5) use DTW to identify onsets and offsets in the recorded data.

(1) Segmentation of the original signal The sequence matching procedure provided the best alignment of syllable template spectrograms with the song, defined using 128-point (∼5 msec) time bins. Songs with matched sequences were spectrally reanalyzed using an FFT with a 128-sample window slid forward in 4-point steps (yielding ∼.16 msec time bins). For each template match, we partitioned the song by selecting the portion of the original signal corresponding to the time period of the template, plus a buffer on either end. The preceding buffer extended from the onset of the current template to a point 30 sample points (∼5 msec) beyond the end of the previous template match. Similarly, the subsequent buffer extended from the offset of the current template to a point 30 sample points (∼5 msec) beyond the beginning of the subsequent template match. Information within these buffers was gradually discounted at a later stage of the analysis, but was included to allow the identification of syllable boundaries that fell outside the coarser template match.

(2) Amplitude waveform definition For each syllable partition, corresponding spectrograms were log-transformed and summed across the 1.5-7.1 kHz range to yield an amplitude waveform for each selected portion. The frequency range was chosen to encompass the regions of highest power and exclude higher frequencies where spectral features are less reliable (Chi & Margoliash, 2001). Amplitude waveforms were smoothed with a 64-point Gaussian window with a 25.6-point standard deviation (equivalent to ∼4 msec); this reduced the length of each buffer by 64/2=32 points. The amplitude derivative was calculated as the difference in adjacent amplitude values divided by the 4-sample time bin. Peaks and troughs of this derivative

were used to define onsets and offsets. These correspond to inflection points in the amplitude and proved to be more reliable than either zero-crossings in the derivative or heuristically defined threshold-crossings in the original amplitude waveforms. Information in the buffers was discounted in a gradual manner by multiplying the amplitude derivative by value that was equal to one at the edge of the original syllable template match, and ramped linearly down to zero at the edge of each buffer.

(3) Waveform templates Amplitude derivative templates were constructed from windowed waveforms as follows: (1) an initial mean waveform across songs was computed based on the initial spectrogram template alignment, (2) waveforms were aligned to the mean using the raw cross-correlation and a max lag of 1/2 the mean length, (3) the mean was recomputed across aligned waveforms, and (4) steps 2 and 3 were repeated for the aligned waveforms. This process amounted to a rudimentary bootstrapping method for computing the mean waveform without having to manually select an exemplar as a template. Step 4 proved to make a significant difference in the resolution of syllables with fast amplitude modulations. Syllables repeated in the same sequence but in different motifs were treated independently in this part of the analysis.

(4) Defining onset and offset times Template onsets were manually selected among positive peaks towards the beginning of the waveform template while offsets were selected among negative troughs towards the end of the syllable. Selections were based on a combination of how close peaks were to the beginnings/ends of syllables

plus height and regularity across songs defined by visual inspection. Frequently the most reliable onsets/offsets occurred after/before brief periods of low amplitude noise in the syllable so templates were slightly shorter than the actual periods of vocalization.

(5) Onset and offset identification Syllable onsets and offsets were identified by mapping individual syllable waveforms to amplitude derivative templates using a modified dynamic time warping algorithm (see Rabiner & Juang, 1993; Anderson et al., 1996). Our implementation was developed to match waveform peaks to corresponding templates by finding a warping of time that maximizes the average product of the template and candidate waveforms. Since the multiplication of large values dominates the matching, the algorithm is directed toward the alignment of peaks and results in a significant improvement in temporal alignment over the traditional DTW based on minimizing Euclidean distance. Individual waveforms were noisier than templates and often had multiple candidate peaks for matching. Visual inspection of several examples confirmed that the times of “double peaks” in syllable waveforms were “averaged” in the alignment with the corresponding peak in the template. Details of the DTW algorithm are presented in Appendix A.

Outliers and final sample

The template matching and DTW yielded a total of 3175 sequences. Those with at least one interval outside 5 standard deviations from its mean length were omitted. Across birds, 137 intervals were outliers and 103 sequences were omitted for this reason, ranging from 1-5% of each bird's sample. Conversely, roughly 5% of all intervals in omitted sequences were outliers, suggesting a weak tendency for outlying intervals to come from the same songs. Thus, the final sample for analysis consisted of 3072 sequences (106-515 per bird) and 72,192 intervals.

2.1.4 Directed song analysis

Female-directed songs were gathered from a subsample of 4 of the 12 birds. Recordings spanned two days in which males spent the night alone. At 11:30 each day a female was introduced into the same small cage within each recording chamber and remained there until 16:30. The male was observed periodically over the afternoon to ensure that songs were in fact directed towards the female, *i.e.* the male was in close proximity and facing the female while singing. Every detected song that we analyzed was at least 1200 msec long.

We analyzed these songs using a slightly abbreviated template matching procedure. Syllable templates were defined as described above, and an initial syllable identification was assigned based on the syllable template with the highest matching score. To maximize the number of sequences that could be analyzed, a spectrogram for each recording was visually inspected, and target syllables that were incorrectly matched were manually corrected using custom software. Matched sequences were excluded if it was thought that background noise or recording quality would yield erroneous temporal analysis.

Between 79 and 161 (mean of 108) recordings per bird were gathered; of these, roughly 50-70% (mean of 64%) had a target sequence that was matched. Very few songs in this sample had fewer motifs than were defined in the target sequence, and an estimated 5% of target sequences were excluded because of acoustic interference such as wing flaps. Note that a greater proportion of directed songs were retained for temporal analysis than in the sample of undirected song, most likely due to greater acoustic interference from juveniles than from females. A total of 266 sequences from directed songs were gathered and analyzed. Temporal analysis followed the procedure described above. As in the undirected sample, songs outside 5 standard deviations were omitted, yielding 259 sequences (38-87 per bird) and 5915 intervals.

2.1.5 Timing data analyses

Tempo change vs. accumulation of variance

One question we addressed was whether variations in sequence length represent tempo changes versus simply an accumulation of variance. To quantify these possibilities, we expressed sequence length as

$$z = \sum_{i=1}^{m} x_i$$

where x_i is the length of interval i and m is the number of intervals in the sequence. We expressed variance in sequence length as

$$var(z) = \sum_{i=1}^{m} var(x_i) + \sum_{i=1}^{m} \sum_{i \neq j}^{m} cov(x_i, x_j) \tag{2.1}$$

where *var* and *cov* denote variance and covariance. Dividing both sides by the sum of the individual variances, we derive a normalized quantity equal to the ratio of the variance in sequence length to the sum of the individual variances. We call this ratio the "gross covariance" and denote it as g. The gross covariance is equal to one plus a value that depends on the summed covariance of all intervals. Values greater than one indicate a net positive covariance among intervals, which is what one would expect from a change in overall tempo. Significant differences between sequence length variance and the sum of interval variances were tested using the 95% confidence interval for sequence length variance, which can be obtained from a chi-squared distribution.

Elasticity calculations

To quantify how tempo changes are related to changes in the length of individual intervals, we performed linear regressions of individual interval lengths x_i with overall sequence length z, i.e. we write

$$x_i = a_i + b_i z + \epsilon_i \tag{2.2}$$

where b_i is the regression coefficient and ϵ_i the residual. One of our primary interests was whether zebra finch song displays proportional scaling in which changes in song length are accomplished by a proportional scaling in the lengths of each of its intervals. If proportional scaling holds, the regression coefficient b_i should be equal to the ratio of the mean interval length to mean sequence length. We tested whether intervals violated proportional scaling by first calculating the standard error of a regression coefficient, which is given by

$$se_b = \sqrt{\frac{mse}{(n-2)var(z)}} \tag{2.3}$$

where n is the number of sequences and mse is the mean squared error from the regression ($\frac{1}{n}\sum \epsilon^2$). We then used a two-tailed t-test with a null hypothesis of proportional scaling in which the regression coefficient should be equal to the ratio of the mean interval length to mean sequence length, $b_i = \bar{x}_i/\bar{z}$.

Note that in the DTW algorithm used to measure syllable onsets and offsets local path constraints were slightly biased away from a slope of 1 (see appendix B). Thus, if there were any possible bias introduced into coefficients by our DTW, it would be in favor of syllables having larger regression coefficients.

Conceptually, b_i represents an interval's ability to stretch and compress with sequence length. Since we were most interested in whether an interval stretched proportionally more or less than the entire song, we normalized b_i by dividing out the value expected for b_i under the condition of proportional scaling such that

$$\beta_i = b_i \frac{\bar{z}}{\bar{x}_i} \tag{2.4}$$

We term this normalized version of the regression coefficient β_i an interval's "elasticity". Proportional scaling implies uniform elasticity across intervals such that the normalized coefficient equals one for all intervals. For ease of comparison across intervals, throughout the paper we report normalized elasticity coefficients only.

For some analyses we were interested in whether intervals had correlations that were independent of song tempo. For these analyses, we calculated the correlation coefficients among the residual values ϵ_i.

Elasticity for pairs of intervals

For the purpose of determining whether syllable onsets have a special role in timing, we examined how the elasticity of each interval in a pairing of one syllable and one gap contributes to the elasticity in the sum of the two interval lengths. For two intervals x_i and x_j, we write the sum of the regression equation as

$$x_i + x_j = [\beta_i \frac{\bar{x}_i}{\bar{z}} + \beta_j \frac{\bar{x}_j}{\bar{z}}]z + a_i + a_j + \epsilon_i + \epsilon_j \tag{2.5}$$

Dividing the coefficent of z by $(\bar{x}_i + \bar{x}_j)/\bar{z}$, we find that the elasticity coefficient for the pair of intervals, $\beta_{i+j} = c_i\beta_i + c_j\beta_j$ where $c_i = \bar{x}_i/(\bar{x}_i + \bar{x}_j)$ and $c_j = \bar{x}_j/(\bar{x}_i + \bar{x}_j)$. That is, the elasticity of the joint interval is the average of the component intervals, weighted by their contribution to overall length. If onsets are especially tied to global tempo one would expect the elasticity coefficient for a pair of intervals consisting of a syllable and the following gap to be tightly clustered around 1.

Statistical analysis

Interval and song length data contained more than 100 sequences per bird. Since visual inspection indicated normality, the Pearson correlation coefficient was used to test the relationships among lengths. However, most of our inferential statistics compared distributions of variables that violate the independence assumption in t-tests, so we used more conservative nonparametric tests. Unless indicated otherwise, differences between two distributions were assessed using the two-tailed Wilcoxon rank sum test, and determinations of whether a given distribution differed from a specific value relied on the Wilcoxon ranked sign test. The Wilcoxon sign test was also used when comparing pairwise differences between two dependent samples. All mean quantities are reported with standard errors.

2.2 Results

The temporal structure of zebra finch song was examined by analyzing the temporal variability in the songs of 12 adult zebra finches. In the main sample we report throughout this section, the majority of recordings (86%) were made as part of other developmental studies in which an adult male tutor and a juvenile finch were housed in two small cages within a single recording chamber. Females were not present during any of these recordings.

Zebra finch song generally consists of several short introductory notes followed by a series of 2-7 motifs, each consisting of a stereotyped sequence of syllables (periods of vocal output separated by silence; figure 2.2). The present study focused on the temporal variability of syllable lengths and the lengths of the gaps between syllables; spectral structure is not analyzed in this study. Recordings containing the most common sequence produced by each bird were identified using an automated template-matching algorithm and manually screened (see Methods). Introductory notes were not considered and longer songs were truncated so that the data analyzed contained repeated renditions of an identical series of syllables. Since we did not analyze entire songs we refer to each series of syllables as a sequence.

Syllable onsets and offsets were recalculated at a finer temporal resolution with a modified dynamic time warping algorithm (see Methods). Generally, we will use the term “interval” to denote the time between adjacent syllable boundaries. We divided this into two interval types: syllables and the gaps between syllables. Hence, a sequence with n syllables had $n-1$ gaps and $2n-1$ intervals. Our sample had 280 distinct intervals (146 syllables and 134 gaps); in the main undirected sample we analyzed a total of 3072 sequences containing 72,192 intervals across birds. We also considered the intervals between the onsets of consecutive syllables; we denote these special intervals “inter-onset intervals” (IOIs). Syllables having the same spectro-temporal structure (e.g. syllable “A” in the first and second motifs) are said to have the “same identity” and constitute one “unique syllable” (total of 60 unique syllables across birds). In 2 of the 12 birds analyzed the sequence of syllables in the first and second motifs were not identical.

2.2.1 Descriptive Statistics

Table 1 summarizes the descriptive statistics across syllables, gaps, inter-onset intervals, motifs and sequences. Overall timing was very tight with the coefficient of variation (CV = standard deviation divided by the mean) for sequence length ranging from 1.1-1.7%. Half of all syllables and gaps deviated from their respective means by under 1.5 msec, or roughly 1.5-2.5% of mean length. Syllable lengths were slightly more variable than gaps, but this is due to the fact that syllables were significantly longer than gaps and that variability was positively correlated with interval length ($r = 0.376$, $p < 0.0001$). However, gap CV was generally about 1.5 times greater than syllable CV.

		mean	**st dev**	**n**
Mean (msec)	syllables	94.51	51.18	146
	gaps	49.87	16.10	134
	inter-onset	142.39	55.93	134
	motifs	627.68	195.21	10
	sequences	1706.80	263.15	12
Std Dev (msec)	syllables	2.54	0.92	146
	gaps	2.39	0.98	134
	inter-onset	3.18	1.12	134
	motifs	10.30	2.78	10
	sequences	24.28	4.78	12
CV (%)	syllables	3.28	2.05	146
	gaps	4.91	1.79	134
	inter-onset	2.34	0.87	134
	motifs	1.67	0.16	10
	sequences	1.42	0.17	12

Table 2.1: Summary statistics regarding the length of song components.

2.2.2 Changes in Song Length Reflect Tempo Changes

Changes in song length can be attributed to two basic sources. First, these deviations could result from small independent perturbations that accumulate during the production of an individual song. Alternatively, changes in song length may reflect song-to-song variations in a global tempo mechanism that exerts its effects throughout the song. In the music box analogy changes in tempo would correspond to variations in the overall speed of rotation in the music box drum.

To gain an understanding of how these scenarios would affect the accumulation of timing deviance, we constructed a simple model that can generate random interval lengths. By changing a single parameter, the model generates sequences based on these two scenarios, leaving other factors constant (see Appendix B). Sequences of interval lengths were generated such that the number of sequences, the variance of overall sequence length and the mean length of each interval were matched to experimental data from given bird. In the "independent intervals" version of the model, the length of each interval was chosen independently, whereas in the "tempo change" version, interval lengths were dominated by a global tempo factor shared by all intervals. Figure 2.3 shows a graphical comparison between the models and the data from bird 10. Each line depicts the accumulation of deviation from this average sequence for a given sequence. These plots indicate that global tempo has a strong influence on interval length.

To quantify the strength of global tempo we used a normalized measure termed "gross covariance", denoted g (see Methods). The measure g is equal to the sequence length variance divided by the sum of individual interval variances. If variations in song length are determined by an accumulation of independent interval deviations then sequence length variance is equal to the sum of interval variances and g is equal to one. If, however, song length is determined by some global tempo mechanism, then the positive covariance among intervals will increase the variance of the entire song causing g to be greater than one. In the 12 birds in our sample, we found gross covariance ranging from 1.16-6.55. These values were significantly greater than one

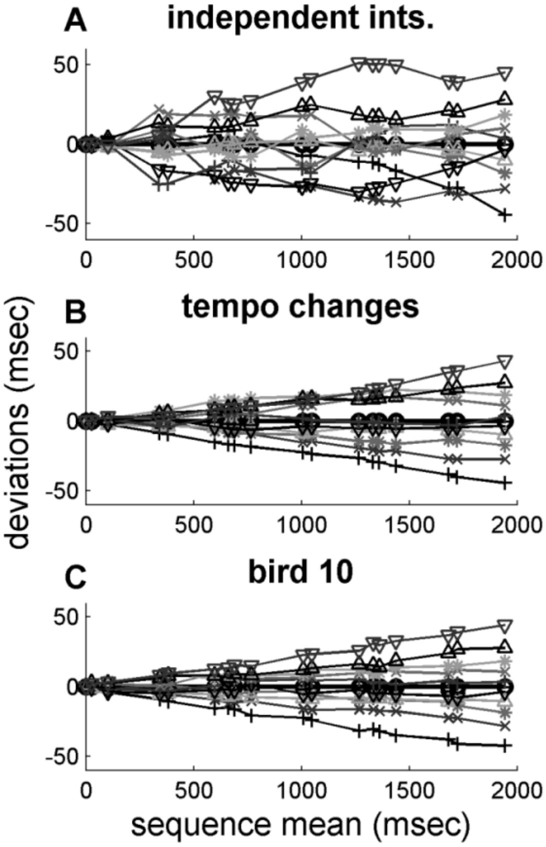

Figure 2.3: Qualitative analysis of interval length deviations. Each line indicates cumulative deviations from the mean for a single sequence. Markers indicate syllable onsets and offsets. The x-axis indicates the mean time from the beginning of the sequence for each onset and offset. The y-axis represents cumulative deviation from mean timing up to that point in the sequence. Sequences shown are at the 5th, 15th, ... , 85th, and 95th percentiles of the distributions of sequence length (thus, individual interval deviations do not necessarily reflect percentiles because they do not perfectly correlate with sequence length). (A) Simulated sequences in which interval deviations are independent. (B) Simulation sequences in which deviations are positively correlated. (C) Experimentally measured sequences from bird 10.

in 11 of these birds ($p < 0.05$, χ^2), and approached significance in the 12th. Note that measurement error in determining syllable onsets and offsets will suppress gross covariance by artificially increasing interval variances while having a minor effect on sequence variance. Overall, our data indicate that changes in sequence length are dominated by global tempo rather than an accumulation of local effects (*cf.* Chi & Margoliash, 2001).

Repeated motifs slow down

Our data are consistent with previous work demonstrating that motifs tend to slow down over the course of a song (Chi & Margoliash, 2001). In 9 of the 10 birds having consistent motifs, the second motif was longer by 5.9 ± 4.4 msec (range 0.04-12.7 msec) or $0.82 \pm 0.49\%$ (range 0.00-1.49%). In the one bird (bird 8) with a shorter second motif, the difference of -13 msec was entirely attributable to variation in the amplitude structure of the first syllable leading to a different choice for the onset-peak of this syllable in the first vs. subsequent motifs. In the 5 birds with at least three motifs (excluding bird 8), the third motif was longer than the second (range 1.2-3.7 msec). In 4 of these 5 birds, the difference between motifs two and three was smaller than between motifs one and two (range 35-67% of the difference between first two motifs); in the fifth bird, the difference between motifs two and three was actually 2.3 times greater. In all 5 of these birds, the gap between motifs two and three was longer than the gap between motifs one and two.

Tempo change correlates with behavioral factors

A number of studies suggest that singing behavior and song tempo may be affected by factors such as the presence of a female, hormone levels and circadian rhythm (Jansen et al., 2005; Deregnaucourt et al., 2005; Zann, 1996; Sossinka & Bohner, 1980; Ollason & Slater, 1973). To explore these factors we examined the relationship between sequence length, acoustic activity and time of day. Acoustic activity was assessed as the average number of recordings per minute on that bird's side of the cage in a 30 minute window centered on each sampled song. This measure is dominated by song production, but an undetermined number (roughly 30%) of these recordings consisted of non-song sounds such as wing flaps and repeated calls. We report time of day as hours from light onset (birds were housed on a 14/10 light dark cycle).

Average song length was shortest at hour 4 (late morning) and gradually increased until hour 11 (late afternoon); decreases in length until hour 4 were significant for 7 birds while increases between hours 4 and 11 were significant for 10 birds (Pearson correlation, $p < 0.05$; figure 2.4). Acoustic activity tended to decline steadily over the day beginning at hour 5 (roughly noon; Ollason & Slater, 1973). The decline was significant for 8 of the 12 birds (Pearson, $p < 0.05$; figure 2.4). As suggested by these results, songs tended to be faster when activity rates were higher, a relationship that was statistically significant for 8 out of the 12 birds ($p < 0.05$). Thus, it is possible that at least some of the tempo changes were due to general arousal state that in turn varied over the course of the day.

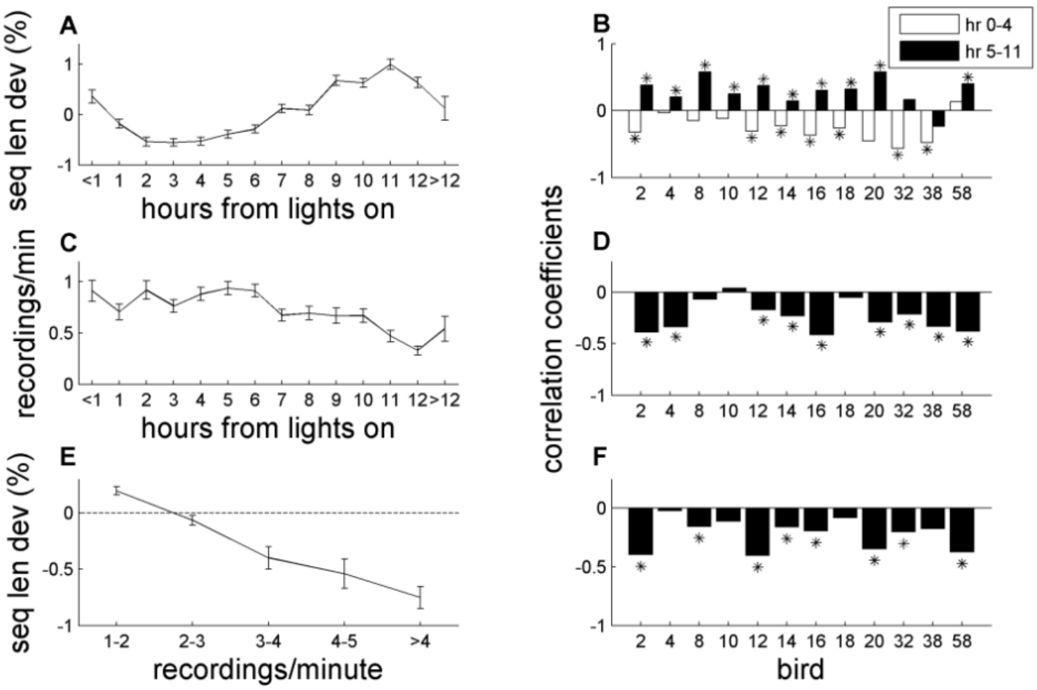

Figure 2.4: Behavioral factors and song tempo. (A,C,E) Means and standard errors across birds. (B,D,F) Pearson coefficients by bird for each relationship indicated directly to the left. Pearson's marked * are significant with $p < 0.05$. (A) Song tempo by time of day in hourly bins. (B) Strength of tendency for songs to speed up in the first 4 hours after lights on (white bars) and slow down between hours 5 and 11 (black bars). (C) Acoustic activity by hour of day. (D) Tendency for activity to decrease over the afternoon beginning at hour 6. Acoustic activity was defined as the number of recordings/minute in a 30 minute window centered on each song. (E,F) The relationship between sequence length deviation and acoustic activity binned in integers. Song tempo changes systematically during the course of the day and may be influenced by factors correlated with overall arousal.

2.2.3 Scaling of Syllable and Gap Length with Tempo

In the music box model, tempo changes are most easily accomplished by simply changing the speed of the underlying clock-like mechanism. Such a model predicts that song intervals should exhibit proportional scaling, *i.e.* they should stretch and compress the same amount per unit time as the entire sequence (Rhodes et al., 2004; Gentner, 1987). We examined whether zebra finch songs showed proportional scaling by performing a linear regression of each interval with overall sequence length (the slope of the regression line is known as the β coefficient). We normalized the coefficients so they reflect how much interval length stretched relative to the length of the entire sequence (see Methods). From this perspective, we will refer to the normalized coefficient as an interval's "elasticity". If $\beta > 1$, then the interval stretches and compresses relatively more than the overall sequence, if $\beta < 1$ it is relatively inelastic, and if $\beta = 0$ then the interval length is unrelated to the tempo. If song length is dominated by variability in the length of a few very elastic intervals, these intervals will have β values greater than 1 while all the other song segments will tend to have coefficients less than 1. In the two birds where the first and second motifs had different syllable sequences (4 and 32), the intermotif gap was very elastic. This suggests that other measurements based on a presumptive global tempo may be misleading in these birds so we restricted the rest of the elasticity analysis to the 10 remaining birds. Examination of the excluded data did not indicate significant differences from most of the other patterns presented.

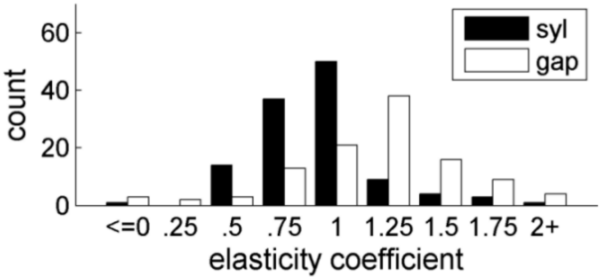

Figure 2.5: Elasticity by interval type. The elasticity coefficient measures the fractional change in interval length relative to the changes in sequence length. Syllables (black) are significantly less elastic than gaps (white). This violates the proportional scaling predicted by a simple "music box" model of song production.

Proportional scaling does not hold

Proportional scaling would imply that all intervals have the same elasticity and all β coefficients are equal to 1. Of all intervals analyzed, 60% of the corresponding coefficients differed significantly from the hypothesis of proportional scaling (2-tailed t-test, $p < 0.05$; see Methods). Therefore, our data strongly indicate that variability in song length is expressed unevenly across the course of the song, with some intervals being more elastic than others. This contradicts the simplest versions of the music box model in which song length is governed by the speed of a single underlying clock.

Breaking down the data by interval type, we found that elasticity coefficients for syllables was significantly smaller than for gaps (figure 2.5), $p < 0.0001$). (Significance tests regarding distributions of elasticity coefficients used nonparametric Wilcoxon tests; see Methods.) Seventy percent of syllables had coefficients less than 1 (mean $0.921 \pm .031$) while 75% of gaps had coefficients greater than 1 (mean $1.169 \pm .041$; figure 2.5). β coefficients for syllables were smaller than for gaps in 7 of the 10 birds analyzed, and this reached significance in 6 of these birds ($p < 0.025$). One bird (58) actually showed syllables to be significantly more elastic than gaps ($p < 0.025$).

Figure 2.6 shows the sequences of β coefficients for all birds in the sample, shown in order of decreasing gross covariance. Birds with greater gross covariance tended to show a stronger alternating pattern of coefficients, with greater elasticity among gaps than syllables. Moreover, the data in many birds appeared to show an identifiable pattern that was preserved across motifs.

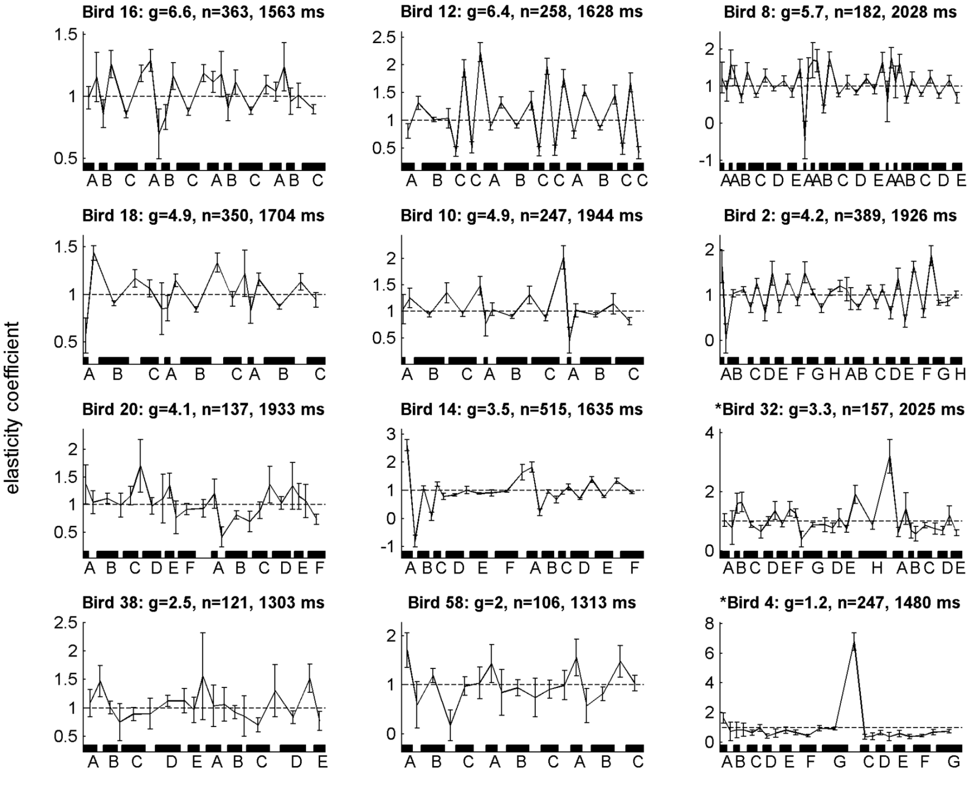

Figure 2.6: Elasticity by bird. Titles indicate bird, gross covariance (g), sample size and mean sequence length. Plots are sorted by g. Error bars show 95% confidence intervals for the sequence length regression. Coefficients are spaced along the x-axis according to mean interval lengths and onset times (syllables are demarcated with black bars).

Motif boundaries are more elastic and more variable

We also noticed that inter-motif gaps and the beginning syllables of motifs (we will call them 'syllable A') tend to have especially high elasticity (figure 2.7). The coefficients for syllable A were significantly greater than those for other syllables (means are $1.122 \pm .098$, and $0.854 \pm .020$, $p < 0.01$), while the difference between inter-motif gaps and other gaps was nearly significant (means are $1.381 \pm .094$ and $1.129 \pm .045$, $p = 0.064$). The elasticity of syllable A was not significantly different from within-motif gaps ($p = 0.511$).

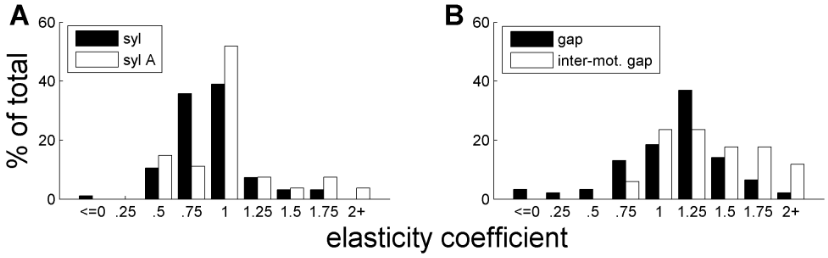

Figure 2.7: Elasticity for intervals at motif boundaries. (A) Distributions of elasticity coefficients for syllables that start a song motif ("syllable A"; white) and other syllables (black). (B) Distributions for gaps falling between motifs (white) and other gaps (black). Intervals at motif boundaries are more elastic.

In addition to being more elastic, inter-motif gaps and syllable A tended to be more variable than other respective intervals of the same type. To reduce confounds based on overall interval length, we only considered intervals between 50 and 80 msec long. This range avoided consideration of very small intervals, which are likely to be dominated by measurement noise, and longer intervals that contained very few syllable As. The resulting sample contained slightly more than one third of all syllable As ($n = 11$) and three quarters of inter-motif gaps ($n = 13$). Within this range, the distribution of CVs for syllable A was greater than for other syllables (means 4.993 ± .561%, $n = 11$, vs. 2.950 ± .002%; $p < 0.005$) and similarly for inter-motif gaps vs. other gaps (means 4.947 ± .468%, $n = 13$, vs. 3.588 ± .190%; $p < 0.025$).

2.2.4 Covariance Structure Follows Syllable/Gap Distinction

The fact that elasticity coefficients separate out by interval type suggests that intervals of the same type have some shared representation. However, the data can't distinguish between two scenarios that we call the "independent" and "grouped" scenarios. In the independent scenario song tempo is the only factor driving changes in interval lengths, while in the grouped scenario there are rich representations shared by intervals of the same type. Under the grouped scenario, one expects syllable lengths to covary independently of global tempo. To focus on this independent component of variance we examined the residual values obtained from subtracting off the linear regression with sequence length. We then computed all pairwise correlations among these residual values and sorted these pairs into three categories: gap-gap, syllable-syllable (syl-syl) and non-adjacent syllables and gaps (syl-gap). Adjacent syl-gap pairs were omitted because of possible confounds due to measurement error: any error in determining the boundary between adjacent intervals makes the measured length of one interval shorter and the other longer and contributes a negative correlation to adjacent intervals.

The three distributions of residual correlation coefficients are shown in figure 2.8. We found positive correlations among most intervals of the same type: 75% of

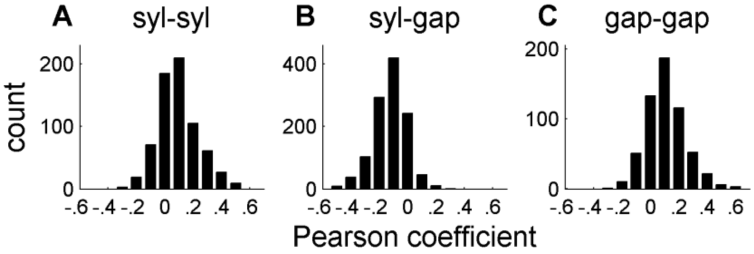

Figure 2.8: Pairwise correlations among and between syllables and gaps. Distributions of the correlation coefficient for all pairs containing (A) two syllables, (B) two gaps, or (C) a syllable and a gap. To remove the effects of song tempo, correlations were calculated between the residual values of regression of interval length vs. total sequence length. Directly adjacent syl-gap pairs were omitted since measurement jitter in the boundary between the pair will induce negative correlations. Stronger within-type vs. between-type correlations suggest shared neural mechanisms in addition to the differential dependence on song tempo shown in figure 2.5.

all syl-syl and 81% of gap-gap correlations were positive (means are $0.093 \pm .005$ and $0.111 \pm .006$). By contrast, 89% of non-adjacent syl-gap correlations were negative (mean $-.123 \pm .003$). Within-type correlations were significantly larger than syl-gap correlations ($p < 0.0001$). This relationship was quite reliable, holding for all 10 birds on an individual level ($p < 0.005$; significance tests regarding distributions of correlation coefficients used nonparametric Wilcoxon tests - see Methods). Thus, our data strongly indicate that syllables and gaps are grouped by properties affecting interval length in addition to the factors that determine elasticity.

2.2.5 Local Temporal Structure

Our analysis thus far has focused on syllables and gaps as groups. While there may be mechanistic representations specific to interval types as a whole, there might also be more specific temporal structure within song. We looked for 3 basic kinds of local structure: relationships between the syllables and gaps of the same identity across motifs, what we call "identity-dependence"; between intervals as a function of separation in sequence position, or "distance-dependence"; and temporal structure delineated across syllable onsets, *i.e.* inter-onset intervals.

Elasticity and correlation structure is identity-dependent

If unique intervals have specific representations then one might expect a similar degree of elasticity in all intervals of that identity produced across motifs (see Yu & Margoliash, 1996; Hahnloser et al., 2002; Leonardo & Fee, 2005 for physiological evidence that the same activity patterns underly the production of repeated motifs). Indeed, we found that β coefficients were significantly closer between intervals of the same identity than between any two coefficients of the same type but different identity (figure 2.9; $p < 0.0005$ for both syllables and gaps). Mean absolute differences were $0.238 \pm .037$ among same-identity syllable pairs and $0.330 \pm .014$ for syllable pairs with different identities. The corresponding values among gaps were $0.262 \pm .025$ and $0.415 \pm .017$. These differences did not reach statistical significance in any individual bird, however, probably due to the small number of same-identity pairs.

We also examined whether syllable and gap residuals of the same identity were more correlated than they were with other intervals of the same type (figure 2.9). In fact, lengths in same-identity syllable pairs were more strongly correlated than in different-identity syllable pairs, with 95% of same-identity syllable pairs having positive correlation (mean for same-identity was $0.239 \pm .013$, for different,

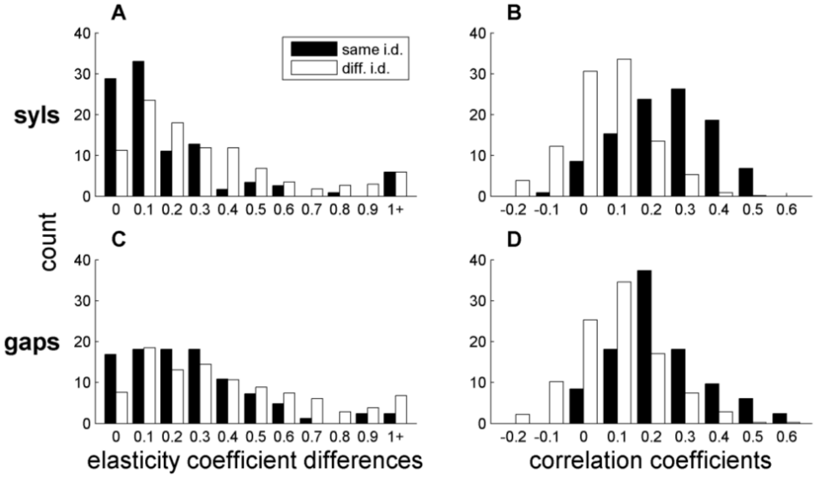

Figure 2.9: Intervals of the same identity are linked. Distributions of pairwise differences between elasticity coefficients (A and C), and pairwise Pearson coefficients (B and D), for syllables (A and B) and gaps (C and D). Black: pairs containing intervals of the same identity (e.g. syllable D in motfs 1 and 3, gap between B and C in motifs 2 and 3). White: pairs containing intervals of the same type (syl-syl or gap-gap) but different identity.

$0.056 \pm .005$, $p < 0.0001$). The same-identity gap correlations were 99% positive and significantly stronger than different-identity gap pairs (same-identity is $0.224 \pm .016$, different, $0.089 \pm .006$, $p < 0.0001$). Since we found no difference in the strength of this effect for syllable A and for intermotif gaps, these numbers include all intervals. The same-identity correlations were stronger than other within-type correlations in all 10 birds, but this difference did not reach statistical significance ($p > .05$) for syllables in 1 bird (38) and for gaps in 2 birds (10 and 20).

Two birds (8 and 12) sang repeated syllables. Overall, correlations between syllables of the same identity but in different motif position were significantly stronger than correlations among syllables of different identities ($p < 0.0001$). Although this suggests that the increased correlation between same identity syllables was not due simply to being in the same location within each motif, the sample is too small to make strong conclusions (*cf.* Leonardo & Fee, 2005).

Identity-dependence among gaps is independent of syllables

Even though identity-dependence was shown for both syllables and gaps, it is possible that the only direct linkage is between same-identity syllables; gap correlations could follow as a consequence. For example, a correlation between gap BC in motifs 1 and 2 could be due to correlated length changes in syllable B and syllable C across the same motifs. To control for such possibilities, we calculated pairwise correlations between gaps after subtracting off the influence of syllable length as well as sequence length. Specifically, we performed a multiple regression of each gap with sequence length and the length of each individual syllable in the sequence, and we calculated the correlation coefficient between the residual values for all pairs of gaps. Although one would expect the adjacent syllables to create the largest confound, all syllables were included in the regression to eliminate any contribution from syllable length. Among gaps of the same identity, 70% of these correlations were positive (mean $0.074 \pm .014$). By contrast, 80% of correlations among gaps of different identity were negative (mean $-.105 \pm .006$), and the 2 distributions were significantly different ($p < 0.0001$). Thus, the data do indicate that gaps of the same identity are more correlated than gaps of different identity, and this relationship is not simply a byproduct of the relationships among syllables.

Distance-dependent correlations

We also considered positional distance in the sequence as a possible factor affecting the correlation between intervals. For birds with short motifs, distance in the sequence can become confounded with syllable identity and motif position at relatively short distances; for example, in the sequence ABCABC syllable A in motif 1 could be more correlated with syllable B in motif 2 than C in motif 1 due to possible transitive correlations with syllable A in motif 2. Therefore, we confined our analysis to the 6 birds that sang at least 4 syllables per motif. For each syllable that did not begin or end the sequence, we computed pairwise correlations with all other syllables in the motif that were not repeated and compared the strength of correlations one position away with two positions away. We considered shared dependence on tempo to be a viable factor behind any such effect so we looked at both raw and residual Pearson coefficients.

Across birds, adjacent syllables had significantly stronger raw correlations than syllables 2 positions away (mean difference 0.041 ± .020; $p < 0.05$). However, there was significant inter-bird variability, with stronger correlations for adjacent syllables found in 4 of 6 birds, only two of which show significance (2, 8, 12 and 14; 8 and 12 significant, $p < 0.05$). In 2 birds (20 and 58), adjacent syllable correlations are actually weaker on average than the more distant correlations. There was also a trend in the correlations among residuals but this did not reach significance (mean difference 0.022 ± .019, $p = 0.24$). Thus, if there is distance-dependence in the correlation structure it may be related to systematic patterns of elasticity that span several intervals.

Syllable onsets are especially aligned to global tempo

We also looked for structure that might indicate something about how syllable onsets and offsets are coded. The simplest possibility is that a central tempo mechanism "triggers" syllable readout at times corresponding to syllable onsets. The syllable is then produced at a rate that is influenced by independent variability within the syllable mechanism as well as song tempo. If the production of a syllable does not influence the onset of the next syllable, any fluctuations in syllable length will come at the expense of the gap following that syllable, making the correlation between those two intervals more negative.

To look for these relationships we started by making a direct comparison of the strength of the negative correlation expected between a syllable and gap making up an inter-onset interval and a syllable and gap making up an inter-offset interval. We made the comparison for pairs that shared the same gap, since all gaps in a sequence have both a preceding and a following interval (which is not true for the first and last syllable in a sequence). The mean difference between the correlations for the inter-onset pairing and the inter-offset pairing was significantly different from 0 ($p < 0.0005$), with 73% of gaps showing a stronger anti-correlation in the inter-onset pairing (mean difference in Pearson $r = 0.145 \pm .025$). However, recall that measurement error is expected to make a significant negative contribution to the correlation between adjacent intervals. Thus, the above correlations may simply reflect a tendency for greater measurement error at syllable offsets than at onsets.

To circumvent this problem we returned to our focus to the elasticity calculations. Since overall sequence length is nearly unaffected by measurement error (except for the first syllable onset and last syllable offset), β coefficients should also be independent of such measurement jitter. We again looked at the trade-off between the syllable length and gap length, but for length differences due to differential elasticity. One can show that the elasticity coefficient for any interval formed by combining a syllable and gap is given by $\beta_{s+g} = c_s\beta_s + c_g\beta_g$, where the subscripts s and g denote syllable and gap and c is the ratio of each interval's mean length to the mean sum of both interval lengths (see Methods). We looked at the trade-off between syllable and gap length by plotting $c_s\beta_s$ versus $c_g\beta_g$ for three different pairings of syllables and gaps: syl-gap pairs making up an inter-onset interval, syl-gap pairs making up an inter-offset interval, and gaps paired with a random syllable in the sequence (figure 2.10). If syllable and gap elasticity exactly trade off in length then the elasticity coefficient for the pair $\beta_{s+g} = c_s\beta_s + c_g\beta_g$ should be exactly equal to 1 (figure 2.10, dashed line).

To quantify how closely the data cluster around this prediction, we computed the mean absolute deviation from $\beta_{s+g} = 1$. The mean deviation for inter-onset pairs was $0.108 \pm .012$, which was significantly smaller than the mean deviations for the latter two groups ($p < 0.005$). The mean deviations for inter-offset pairings were not significantly different than for random pairings ($p = 0.262$; means for inter-offset pairs was $0.142 \pm .013$ and was $0.167 \pm .015$ for random pairings). Inter-onset deviations from 1 were smaller than those for inter-offset intervals in 9 of 10 birds

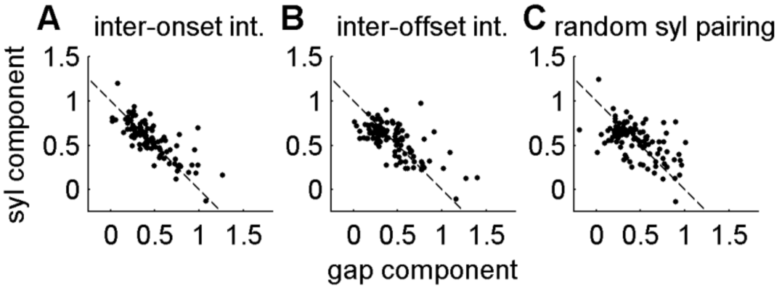

Figure 2.10: Tradeoffs in elasticity coefficients for different syllable-gap pairings. The summed length of a syllable (s) and a gap (g) has elasticity given by $\beta_{s+g} = c_s\beta_s + c_g\beta_g$ where c_s and c_g are relative lengths of each interval in the pair (see Methods). Plots show gap ($c_g\beta_g$) and syllable ($c_s\beta_s$) components along x and y-axes respectively. (A) Inter-onset pairs (gap and preceding syllable). (B) Inter-offset pair (gap and following syllable). (C) Gaps paired with random syllable. Tighter clustering around the $\beta_{s+g} = 1$ line (dashed) for the inter-onset pairings suggests that elasticity in a syllable comes at the expense of the following gap, *i.e.* syllable onsets are more closely tied to tempo than syllable offsets. The negative slope for random pairings is due to the fact that $c_s + c_g = 1$.

(exception is bird 8) and smaller than those for randomly paired gaps and syllables for 8 of 10 birds (exceptions are birds 8 and 10).

2.2.6 Effect Sizes

We have presented several factors that influence the lengths of intervals in zebra finches singing undirected song. To get a sense for the size of these influences, we performed for each interval a stepwise multiple regression with (a) sequence length alone, (b) sequence length and the sum of all intervals of the same type but different identity (“type sums”), (c) sequence length, type sums and the sum of all intervals of the same identity and (d) all previous factors and the sum of adjacent intervals. Inter-motif gaps were excluded because there were no other intervals of the same identity in the sequences with only two motifs. Across intervals, sequence length explained 24.4 ± 1.2% of variance. Factoring in the sum of all intervals of the same type but different identity explains an additional 5.8 ± .4%, the sum of all intervals of the same identity explains an additional 3.5 ± .2%, and adjacent intervals an additional 17.8 ± .8%. This makes for a total of 51.5 ± 1.1% of interval variances explained by the factors analyzed here. (The relatively large amount of variance explained by adjacent intervals is likely due to the fact that the errors in measuring syllable boundaries result in correlated length changes on either side of the boundary.)

2.2.7 Temporal Structure of Female-Directed Songs

It has been reported elsewhere that directed song, in which a male sings towards a female as a form of courtship, tends to be faster and shows different physiology (Sossinka & Bohner, 1980; Hessler & Doupe, 1999). To determine whether singing to a female altered the temporal structure of song, we examined directed songs from a subsample of 4 males (see Methods). Consistent with previous reports, we found that directed songs tended to be faster than undirected songs. In the four birds analyzed (birds 8, 10, 12, and 14), mean sequence length was shorter than undirected song by 16, 71, 7 and 48 msec respectively, or approximately 0.8, 3.8, 0.4 and 3.0%. Songs were produced between 11:30 and 16:30 so it is unlikely that time of day introduced these differences (if anything it may have muted them since undirected songs slowed down over the afternoon).

To determine whether singing to a female had a substantial effect on the fine-grained temporal structure of song, we compared grouped statistics regarding the elasticity and correlation structure of intervals broken down into syllable and gap categories. Overall, elasticity patterns were similar to what we found with undirected songs. Of the 98 intervals analyzed, 65% had β coefficients significantly different from 1 and syllables were significantly less elastic than gaps (means are $0.603 \pm .080$ and $1.600 \pm .090$, $p < 0.0001$). Given our small sample of birds, we did not test the motif boundary effect. We also found the same basic covariance structure among the directed songs. The correlation coefficient between residuals remaining after subtracting out the influence of tempo was positive within interval types and negative between types (syl-syl mean $0.073 \pm .012$, gap-gap $0.093 \pm .012$, non-adjacent syl-gap $-0.102 \pm .009$, $p < 0.0001$). Finally, all local structure reported for undirected songs was also found among directed songs.

We attempted an interval-by-interval comparison to determine if directed songs are simply sped up versions of undirected songs with the same individual temporal structure. Even though song spectrograms allowed for an easy identification of the same syllable in directed and undirected songs, roughly half of the scatter plots of sequence length vs. interval length for directed songs fell substantially outside what would be extrapolated from the undirected data (determined by visual inspection). Closer examination revealed that many syllables showed differences in the shape of the amplitude envelope that precluded an unambiguous matching of syllable onsets and offsets in directed and undirected versions of the song. Because of these complications, determining possible pairwise differences in the fine-grained temporal structure of directed and undirected songs of individual birds will require a more extensive analysis that is beyond the scope of this study.

2.3 Discussion

We have exploited the remarkable stereotypy of zebra finch song to analyze the temporal structure of repeated syllable sequences. For practical reasons, we did not address subsyllabic temporal structure and focused exclusively on the first several motifs sung by birds housed alone or in the presence of other males. The ability to collect hundreds of songs and analyze them with high precision allows us to distinguish temporal variations on the millisecond time scale. Our results indicate that song length is highly stable under these conditions, with a majority of deviations under 1.5%. Measurement of the “gross covariance” of song intervals indicates that changes in song length are dominated by global influences that differ from song to song rather than an accumulation of local jitter during song production. This is consistent with our demonstration that song length is correlated with the time of day, as well as with previous data showing a continuous “drift” in song production across repeated recordings (Chi & Margoliash, 2001).

We also investigated how the lengths of individual intervals correlate with these global tempo changes. We find that syllables tend to stretch and compress with tempo changes less than gaps. Additional song-to-song variability is shared by intervals of a given type, so that syllable-syllable and gap-gap correlations are stronger than syllable-gap correlations. Overall, these data suggest that (1) interval length correlations are induced by mechanisms that span the entire song and (2) syllable and gap lengths are driven by distinct components within the song circuit.

More detailed analysis reveals further structure at the local level. It appears that syllable length has a stronger trade-off with the length of the following rather than preceding gap, suggesting that onsets represent the preferred alignment of syllables with overall song tempo. Also, syllables of the same identity sung in different motifs show stronger correlations than other syllable pairs, consistent with physiological data showing that they are supported by similar patterns of neural activity (Leonardo & Fee, 2005; Yu & Margoliash, 1996).

Models of song production

We discuss our results in the context of three basic models for song production (figure 2.11). Although elements of all three models are likely to play a role, we discuss pure forms of each model for conceptual clarity. The first is the "music box" model in which song is driven by an underlying clock-like mechanism (Fee et al., 2004). The second is a "chain model" in which syllables and gaps are subserved by separate neural mechanisms and serially linked together in a chain. Song tempo is the byproduct of temporal fluctuations along the chain. Third is the "tempo and syllable model" in which syllables are integrated within a global temporal structure for song. In this scenario, gaps are simply the time left over between syllables.

The main prediction of the music box model is that changes in song tempo lead to a proportional temporal scaling of song elements. Our analysis clearly shows that proportional scaling does not hold for zebra finch song, contradicting the simplest

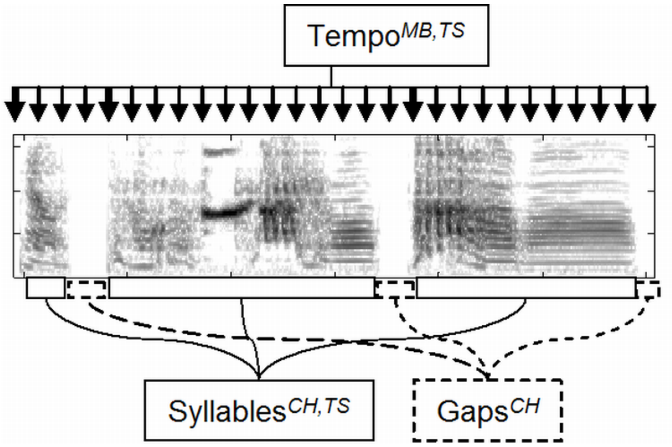

Figure 2.11: Functional elements possibly contributing to temporal structure. The "Tempo" represents a pattern generator that drives song, either continuously across the entire song, or at particular points such as syllable onsets (thicker arrows). The "Syllable" and "Gap" boxes represent mechanisms that control the temporal structure of the corresponding units of song. Superscripts indicate participation of these mechanisms in three basic models of song production. MB: music box model. A clock-like drum triggers the production of acoustic output on a fine time scale. The song is not decomposed into syllable or gap-based units. CH: chaining model. Temporal structure results from a chaining of syllables and gaps. Song length is a consequence of the combined action of the syllable and gap mechanisms. TS: tempo and syllable model. A tempo mechanism determines the overall rate of song production and triggers the action of a mechanism that produces syllables as units of song. Gaps are simply the intervals left over between syllables.

forms of this model. Furthermore, the distinction between syllables and gaps in the correlation structure suggests the presence of neural mechanisms specifically dedicated to the production of song syllables.

The other two models both presume the existence of syllable-based units but differ in whether gaps are units of song or simply left over from the interplay between syllable length and song tempo. One piece of evidence suggesting that gaps may be left over from tempo and syllable interactions is the anti-correlation in elasticity between syllable-gap pairs making up inter-onset intervals (figure 2.10). On the other hand, the relative inelasticity of syllables could be explained by the chaining model if this elasticity induces an active compensation in the subsequent gap length so that inter-onset intervals scale nearly proportionally with song length. There is actually a fundamental limitation in distinguishing these models in our analysis: song length is determined by the sum of syllable lengths and gap lengths. Given this mathematical relationship, any temporal structure can be explained by any two of syllable length, gap length and song tempo. Whatever the structure of the pattern generator for song, gaps must have some form of representation in the system since they correspond to activation of motor neurons driving inspiration (Goller & Cooper, 2004; Suthers & Margoliash, 2002; Wild et al., 1998).

Locus of hierarchical representations

Fee and colleagues have shown that song activity is driven by regular, clock-like bursting from $\text{HVC}_{(\text{RA})}$ neurons (Hahnloser et al., 2002). However, our behavioral data demonstrate that syllables and gaps scale differently with changes in song tempo, and that variations in syllable and gap lengths are correlated with other intervals of the same type. How can these two sets of data be reconciled?

One possibility is that the bursting of $\mathrm{HVC}_{(\mathrm{RA})}$ neurons does not act like the ticking of a single clock but rather as a series of bursts grouped into functional units. There are a number of candidate mechanisms within HVC that may subserve this grouping. For example, previous recordings within HVC showed modulations in firing rate that were tied to individual syllables and repeated by motif (Yu & Margoliash, 1996). Subsequent recordings suggest that this activity was most likely due to the spiking of inhibitory interneurons within HVC (Fee et al., 2004; Hahnloser et al., 2002). It is possible that these interneurons, via their projections onto $\mathrm{HVC}_{(\mathrm{RA})}$ neurons (Mooney & Prather, 2005), serve to organize the $\mathrm{HVC}_{(\mathrm{RA})}$ activity into functional groups. This proposal is supported by evidence from brain slice recordings showing that transient pulses delivered to HVC can induce inhibition-dependent rhythmic bursting whose timing roughly matches the rate of syllable production (Solis & Perkel, 2005).

Temporal grouping of activity might also be driven by afferent input to HVC. Bilateral HVC recordings show brief periods during each motif in which multi-unit activity becomes synchronized across hemispheres (Schmidt, 2003). Because there are no inter-hemispheric connections in the avian forebrain this synchronization must be induced by HVC afferents. Synchronization preferentially occurs at syllable onsets, consistent with our data that intervals may be grouped into inter-onset pairs. This view of HVC is consistent with data from the afferent nucleus uvaeformis of the thalamus (Coleman & Vu, 2005; Williams & Vicario, 1993).

Alternative hypotheses also exist. Bursting in $\text{HVC}_{(\text{RA})}$ neurons could scale proportionally while downstream mechanisms lead to the production of syllable-based units. One possibility is that output from an HVC clock encounters syllable-based representations within the premotor nucleus RA (or subsyllabic representations; see Yu & Margoliash, 1996). Like HVC, RA has a rich network of inhibitory interneurons, and these may give RA dynamic properties distinct from its HVC input (Abarbanel et al., 2004; Spiro et al., 1999).

Syllable-based representations could also be induced by afferent input to RA from LMAN, the output nucleus of an indirect pathway connecting HVC to RA that passes through the avian basal ganglia (figure 2.1). Spike timing is highly correlated across HVC, RA and LMAN (Kimpo et al., 2003), and a requirement for synchronous arrival of LMAN and direct $\text{HVC}_{(\text{RA})}$ input could lead to inelasticity of activity within RA. Moreover, this pathway has been implicated in syllable sequencing in Bengalese finches (Kobayashi et al., 2001), consistent with the proposed role of the basal ganglia in sequential behaviors (e.g. Fujii & Graybiel, 2005; Aldridge et al., 2004; Hikosaka et al., 2002).

The departure from proportional scaling could also occur when clock-like activity in RA projection neurons interacts with the brainstem pre-motor nuclei for song production. These nuclei have a complex network of interconnections (Sturdy et al., 2003; Wild et al., 1997) and are responsible for the production of the discrete unlearned vocalizations known as calls (Simpson & Vicario, 1990). It may be here that song is organized into discrete motor gestures at the syllable or subsyllable level.

Finally, it is possible that proportional scaling holds as far as the motor neurons driving the syrinx and respiratory muscles but physical constraints at the periphery lead to differential scaling of syllable lengths with song tempo. There are several reasons that this is unlikely to explain our data. First, while it is easy to imagine how peripheral dynamics might constrain syllables to show smaller changes than the overall song (β less than one), a number of syllables show changes in length that are proportionally greater than changes in tempo. Second, for several syllables, β is greater than one in the first motif and less than one in later motifs (figure 2.6).

Evolution of vocal behavior

It has been proposed that learned song evolved as birds became able to aggregate series of unlearned calls into organized sequences (Zann, 1993). Both the ability to learn from a song model and the ability to coordinate and elaborate a series of calls is likely to have required the involvement of complex and flexible sensorimotor circuits in the forebrain (Simpson & Vicario, 1990). This hypothesis is similar to suggestions that human speech evolved as complex cortical circuits built upon brainstem circuits gave rise to coordinated movements of the tongue, jaw and diaphragm (e.g. MacNeilage, 1998). Under these hypotheses, motor representations at the level of the forebrain would then evolve under the competing constraints of constructing global representations for vocal sequences and coordinating the production of the elements that constitute these sequences. Given these constraints, it may not be surprising to find a mixture of both global and local representational schemes within the forebrain circuits for complex vocal behavior.

Chapter 3

Behavioral Measurements of a Temporally Precise Motor Code in Birdsong

(Glaze CM, Troyer TW (2007), *J Neurosci* 27(9): 7631-7639)

How brains learn and produce complex sequences is one of the touchstone questions in neuroscience (e.g. Lashley, 1951; Hikosaka et al., 2002; Keele et al., 2003; Rhodes et al., 2004). Although many natural skills contain a hierarchy of subtasks (Miller et al., 1960), the units of behavior are not always clear. Zebra finch courtship song has several characteristics that make it an ideal model system for understanding sequence learning and production. Songs are learned, highly stereotyped, and have a hierarchical temporal structure spanning multiple time scales: songs consist of several repeats of 500-1000 msec long "motifs;" motifs consists of a stereotyped sequence of 3-7 "syllables," 50-250 msec long vocalizations separated by silence; many syllables can be further divided into 30-70 msec long "notes."

Many studies have proposed that the syllable is a basic unit of song production (e.g. Solis & Perkel, 2005; Williams, 2004; Zann, 1996; Yu & Margoliash, 1996). This view is supported by evidence that respiratory expirations accompany syllables and inhalations accompany silent gaps, while song interruption caused by strobe flashes or electrical stimulation tends to occur during gaps (Cynx, 1990; Vu et al., 1994; Wild et al., 1998; Franz & Goller, 2002). However, the syllable-based view is challenged by temporally sparse bursting in the premotor nucleus HVC. During each motif, HVC projection neurons produce a single burst of spikes time-locked to the song with millisecond precision (Hahnloser et al., 2002). Fee et al. (2004) have proposed that HVC acts like a clock, continuously pacing song behavior. Under this proposal, the 5-10 msec long burst is the fundamental unit of the song motor code, and slower acoustic changes result from convergent connections downstream of HVC (Fee et al., 2004; Leonardo & Fee, 2005).

In a previous study, we explored song temporal structure by closely examining natural variability in the lengths of syllables and the gaps of silence between them (Glaze & Troyer, 2006). At a slow timescale, length changes are dominated by modulatory factors that influence syllables and gaps throughout the song. At the syllable timescale, syllables are less "elastic" than gaps, *i.e.* they stretch and compress proportionally less with tempo changes, and syllable-syllable and gap-gap length correlations are stronger than syllable-gap correlations. Such syllable/gap differences contradict the hypothesis that song timing is driven by a uniform clock that continuously paces motor output. Importantly, syllable pairs consisting of the same syllable repeated across motifs were especially related, having strong length correlations and similar elasticity. This "identity-dependence" of temporal variability suggested that syllables may form a basic unit in the motor code for song.

Here, we extend our methods to examine the structure of song timing at timescales finer than the syllable. We find that the identity-dependent temporal structure of syllables is dominated by independent variability among constituent notes. Furthermore, for a subset of notes, we were able to reliably measure temporal variability within short, 10-msec slices of the song. Here, we find that identity-dependent temporal structure is dominated by independent variability among constituent 10-msec slices. Overall, we find timing variability on two widely divergent timescales: (i) slow modulations that result in song-to-song changes in tempo, and (ii) deviations at short timescales (as fast as 5-10 msec) that are reliably repeated across motifs (every 500-1000 msec). These patterns provide the first behavioral evidence for a fine-grained motor code on a timescale comparable to that found in forebrain premotor nuclei.

3.1 Materials and Methods

Analysis was based on the songs from 9 adult males >400 days-post hatch (dph). Birds were recorded while serving as tutors for juvenile birds as part of other developmental studies. All care and housing was approved by the institutional animal care and use committee at the University of Maryland, College Park. All analysis was performed in Matlab (Mathworks, Natick, MA), and all template matching and dynamic time warping algorithms were written as C-MEX routines.

3.1.1 Song collection

During recordings, birds were housed individually in (approx. 18x36x31 cm) cages. Recordings were made from sound isolation chambers (Industrial Acoustics, Bronx, NY), which contained two cages separated by 18 cm and two directional microphones (Pro 45; Audio-Technica, Stow, Ohio). Signals were digitized at 24,414.1

Hz, and ongoing data were selected using a circular buffer and a sliding window amplitude algorithm. "Sound clips" separated by <200 msec were included in the same "recording" and clip onset times were indicated by filling the gaps between clips with zeros.

For each bird, we gathered an initial random sample of 1000 recordings that were >2 sec long and had maximum power from the side on which the target bird was stationed. Recordings were analyzed using the log-amplitude of the fast-Fourier transform (FFT) with a 256-point (10.49 msec) window moved forward in 128-point steps. Frequency bins outside the 1.7-7.3 kHz range were excluded from all subsequent analysis because song structure is less reliable at the highest and lowest frequencies. We then used an automated template matching algorithm (detailed below) to select out recordings which contained repeated sequences of the most commonly produced motif and were relatively free of extraneous sound such as interfering vocalization from the other bird in the sound chamber.

A median 633 (range 411-862) recordings per bird had a template sequence. If a sequence contained an interval between adjacent syllable onsets that deviated from the mean by more than $\sim$30 msec the entire song was discarded under the assumption that the match was erroneous (median 22, range 6-97 songs per bird omitted for this reason). The vast majority of these deviations occurred when an introductory note was incorrectly identified as the first syllable in the song.

Template matching

Each recording was composed of a series of "clips," periods of sound separated by at least 10 msec of silence. The next stage of the analysis was to determine if the sound in these clips matched syllables in the bird's song (clips could also result from cage noise, production of non-song vocalizations or calls, and sounds produced by the juvenile bird in the same recording chamber). To do this, syllable templates were formed by aligning and averaging 4-5 manually chosen clips corresponding to each syllable; exemplars were aligned using the lag times corresponding to peaks in a standard cross-correlation.

These templates were then matched against each clip with a novel sliding algorithm: For each template and each time point (t) in the clip, a match score (c) was computed as the reciprocal of the mean squared difference between template and song log-amplitudes at each time-frequency point:

$$c(t) = n \times m / \sum_{i=1}^{n} \sum_{j=1}^{m} (s(i+t,j) - s'(i,j))^2$$

where s is the song spectrogram, s' is the template spectrogram, n is the number of time bins in the template, m is the number of frequency bins, i indexes time and j indexes frequency.

Candidate syllable matches were computed as peaks in the score vector c over a fixed threshold of .3 (manually chosen based on visual inspection). Based on the alignment giving the peak match, a clip was determined to potentially constitute a syllable if the onset and offset for the clip and template were matched to within

20 msec. If a clip had multiple syllable matches, the match with the highest peak value was chosen.

For each bird, the template song was based on the most common syllable sequence falling within the first ∼2 sec of that bird's song (this reflects a drop-off in available song recordings that are longer than 2 seconds). If the syllable-matching algorithm found a sequence of syllables matched to this template song, each clip corresponding to a syllable match was selected out for further analysis.

3.1.2 Song timing calculations

Timing variability was then analyzed with a more fine-grained algorithm; each syllable in the song sequence was independently analyzed in this part of the analysis. Analysis can be divided into the following steps: First, all identified clips from song sequences were reprocessed using the log-amplitude FFT with a 128-point window slid forward in 4-point steps, yielding 0.16 msec time bins. Second, the resulting spectrograms were smoothed in time with a 64-point Gaussian window that had a 25.6 (∼5 msec) SD. Third, time derivative spectrograms (TDSs, calculated as differences in log-amplitude in time-adjacent bins) were computed and used in the rest of the analysis; the TDS has proved to yield more reliable data on timing than the amplitude spectrogram. Fourth, syllable templates were recomputed by averaging syllable TDSs across songs, aligning each TDS to this mean, re-averaging aligned TDSs, and repeating this process twice. Fifth, each TDS was then mapped to its template using a dynamic time-warping (DTW) algorithm (Anderson et al., 1996; Glaze & Troyer, 2006, see Appendix C). If the algorithm failed to map a syllable onset or offset, the entire song was omitted (median 2, range 0-176 songs per bird). One bird had a large number (176) of songs omitted because the first syllable of each motif had variable and noisy onsets. A final sample of 411-877 sequences per bird resulted from the process described above.

Note boundaries were manually determined based on large and sudden changes in spectral properties within syllable templates (Fig. 3.1A,B). Note lengths were then determined as the interval between points mapped via DTW to corresponding boundary times in the template.

3.1.3 Timing analysis

The first part of our analysis concerned the measurement of two latent factors we hypothesized to explain song-to-song variability in note lengths: a "note-specific" factor that makes the same note especially related to itself across motifs, and a "syllable-specific" factor that makes different notes in the same syllable especially related. A linear regression of each note length with total sequence length was used to extract two significant components of song-to-song variability: (1) the normalized regression coefficient, "elasticity," which represents the ability to stretch and compress with global tempo changes, and (2) the residuals from the regression, which represent length components independent of global tempo (Glaze & Troyer, 2006).

We looked for note-specific and syllable-specific factors by examining pairwise differences among elasticity coefficients and pairwise Pearson's correlation coefficients among residuals. Here we describe calculations for correlation coefficients; calculations for elasticity coefficients were analogous. We compared distributions of correlation coefficients between renditions of the same note produced in different motifs ("same-id"), between different notes in the same syllable across motifs ("same-syl"), and between different notes in different syllables across motifs ("diff-syl"). Because the residual lengths of a given syllable are related across all motifs in a song (Glaze & Troyer, 2006), pairwise measurements from different motif pairings were not independent (e.g. same-id measurements between the same note in motifs 1 and 2 and in motifs 3 and 4). This means that distributions containing all pairwise correlations have repeated measures that invalidate statistical tests. Therefore, we calculated three measures for each unique note from the first motif: mean correlation with the same note in all subsequent motifs, mean correlation with different notes in the same syllable in subsequent motifs, and mean correlation with notes in different syllables across subsequent motifs. We then tested whether the distribution of pairwise differences per note were different from zero using the Wilcoxon signed-rank test. We excluded syllables with only one note in order to have a "same-syl" measurement for each note. Across birds, the sample included 122 notes (11-19 per bird), and of these, 109 (10-17 per bird) were in syllables that included more than one note.

The second part of our analysis involved an analogous set of questions on a finer timescale. Specifically, we examined whether two ∼10-msec "slices" of a note were any more related to each other than they were to other 10-msec slices in other notes and syllables. Here, we focused the analysis on “amplitude-modulated” (AM) notes which have an even distribution of power across frequencies at any given time point, but fast changes in power across time (Fig. 3.3C). This type of note lends itself well to an analysis of temporal stretching and compressing on the 5-10 msec timescale. Three birds had at least one unique AM note, and one bird had two in the same motif, yielding a total of 4 unique AM notes across birds.

We divided each AM note into slices defined as peak-to-peak intervals in the spectral time-derivative (Fig. 3.3C). This yielded a total of 15 unique AM slices, 3-4 slices per note. We then performed the same linear regression and pairwise statistics for each AM slice as we did for notes. Most of this analysis compared “same-slice” and “same-note” relationships because only one bird had multiple AM notes in the same motif.

3.2 Results

We analyzed subsyllabic temporal structure in zebra finch song from 9 adult males that were tutoring juveniles in a larger development study. Adult song acoustics are organized hierarchically (Fig. 3.1A): A bout of song generally consists of several motifs, defined as stereotyped sequences of syllables. Syllables, distinct vocalizations separated by gaps of silence, can in turn be divided into notes, segments with distinct acoustic structure. Adults tend to produce a variable number of motifs within a single song bout. For the purposes of this study we gathered from each song recording a manually defined syllable “sequence,” a fixed number of back-to-back motifs, range 2-4 per bird. Across birds the sample included 41 distinct syllables and 122 distinct notes within motifs. The final sample included 411-877 sequences per bird, for a total of 5745 sequences, 69,434 syllables, and 205,146 notes across motifs and recordings.

We had previously found tempo changes that are shared by all song segments across the sequence, and two measures of syllable length from a linear regression with sequence length that distinguished syllables from each other (Glaze & Troyer, 2006): residual length correlations that remain after factoring out tempo, and elasticity coefficients, the normalized regression coefficient which measures the ability to stretch and compress proportionally with sequence length. That analysis indicated that syllables could be distinguished from each other in each of these measures: the residual length of a given syllable is more correlated with the same syllable in other motifs, and elasticity coefficients among the same syllables produced across a sequence are significantly more similar to each other than they are to the coefficients of other syllables. For simplicity we refer to these two patterns together as "identity-dependence."

Here, we extend our methods to probe the timescale of identity-dependence by analyzing temporal variations within syllables. There are two basic types of subsyllabic organization that one may expect to explain identity-dependent patterns. First, if the syllable is a cohesive unit of behavior, subsyllabic segments within that syllable should share the properties of the syllable. In this "grouped" scenario, two subsyllabic intervals in the same syllable will have similar elasticity coefficients and a stronger correlation than two intervals coming from different syllables (Fig. 3.1C).

Alternatively, the syllable might simply be the concatenation of independent components of the motor code whose efferents converge on a continuous representation of the syllable at the periphery. In this "independent" scenario, the elasticity coefficients and the length deviations for two intervals in the same syllable will be no more related than for two intervals coming from different syllables (Fig. 3.1D).

All statistics were based on notes from the first motif, while correlations and elasticity comparisons were made with notes in subsequent motifs (see Materials and Methods). Unless indicated otherwise, all values reported below include standard errors (mean±SE). However, since many of the distributions showed significant skew, statistical significance was assessed using the Wilcoxon signed-rank (WSR) test for pairwise comparisons.

3.2.1 Note-based analysis

To distinguish the grouped and independent scenarios, we first segmented syllables into notes based on sudden changes in spectral profile (Fig. 3.1A,B). Across birds, 70% of the syllables in our sample had more than one note (range 1-9). Across all notes, mean note length was 36.40 ± 18.62 msec, while mean standard deviation was 1.70 ± 0.70 msec, coefficient of variation (standard deviation / mean), $5.58 \pm 3.88\%$; after factoring out sequence length in the regression, residual standard deviation was 1.61 ± 0.88 msec (all ranges ± standard deviation).

For each note of the first motif, we made three measurements with notes in subsequent motifs: average correlation with (1) notes of the same identity, "same-

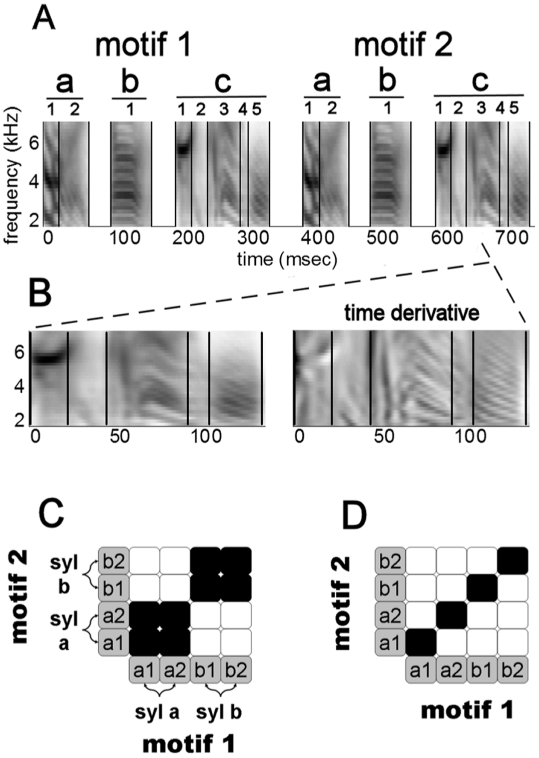

Figure 3.1: Song hierarchy and timing models. A: Template spectrogram from bird 16 (first two motifs only), restricted to 1.7-7.4 kHz where spectral cues are most reliable (see Materials and Methods). Song are segmented at three levels of organization: songs are divided into motifs; motifs are divided into syllables (lowercase labels); and syllables are divided into notes (numeric labels, vertical lines indicate segmentation). B: Template spectrogram (left) and time-derivative spectrogram (right) of the last syllable in the motif. C-D: Schematic covariance matrices. Grey squares represent notes. Covariances are represented as either strong (black) or no/weak covariance (white). Pairwise covariance between two syllables is equal to the sum of all pairwise covariances among constituent notes. In C, note timing variability is grouped by syllables, so that notes from the same syllable are strongly correlated. In D, there is no grouping by syllable, and syllable length deviations stem from the accumulated deviations of individual notes that are correlated across motifs.

id" notes; (2) different notes in the same syllable, "same-syl" notes; and (3) notes in different syllables, or "diff-syl" notes. To test the grouped and independent scenarios, we then asked whether a given note was more correlated with same-id notes than it was with same-syl and diff-syl notes, and whether average correlation with same-syl notes was stronger than average correlation with diff-syl notes. We based all statistics on pairwise comparisons of measures for each note (109 notes in multi-note syllables).

A representative correlation matrix from a single bird is shown in figure 2G; the prominent off-diagonal structure shows that correlations across motifs are dominated by notes of the same identity. Across all notes, mean same-id correlation was 0.18 ± 0.01 (range [0.11,0.31] averaged by bird); diff-id, same-syl, -0.02 ± 0.01 ([-0.05,0.01] by bird); diff-syl, 0.01 ± 0.003, ([-0.01,0.03] by bird). Same-id correlations were significantly stronger than same-syl and diff-syl correlations ($p < 0.0001$ WSR; Fig. 3.2A,C,E). Unexpectedly, same-syl correlations were actually slightly more negative than diff-syl ($p < 0.0001$ WSR).

We then asked why notes are significantly more anti-correlated with those in the same syllable than with those in other syllables. Although our analysis excluded directly adjacent note pairs, the same-syl group did include notes with adjacent motif positions (for example, the correlation between note a1 in the first motif and note a2 in the second motif). Qualitative analysis suggested that these "motif-adjacent" pairs accounted for the difference, which the statistical tests supported: Focusing on syllables with more than 2 notes (n=80 unique notes, 5-13 per bird), non-adjacent notes in the same syllable were just as correlated as notes from different syllables ($p = 0.96$ WSR; non-adjacent correlation was 0.01 ± 0.01), while notes were significantly more anti-correlated with motif-adjacent notes than they were with non-adjacent in the same syllable ($p < 0.0001$ WSR; motif-adjacent correlation was -0.06 ± 0.01). In these data it is impossible to discern whether the motif-adjacent anti-correlations reflect a real tradeoff in variability or correlated measurement error (see Discussion).

We also measured absolute differences among elasticity coefficients and found patterns that were qualitatively similar to the correlation structure (Fig. 3.2B,D,F). Overall, mean same-id difference was 0.20 ± 0.02, same-syl 0.60 ± 0.06, diff-syl, 0.64 ± 0.05. Same-id coefficients were significantly closer than same-syl and diff-syl coefficients ($p < 0.0001$). Elasticity coefficients were also slightly closer among different notes in the same syllable than they were among notes from different syllables (WSR, $p = 0.012$). However, the effect was inconsistent with mean difference showing the opposite trend in 4 out of 9 birds (*i.e.* mean same-syl differences greater than diff-syl differences).

In the aggregate, the timing data indicate that the temporal relationship among syllables of the same identity is dominated by note lengths. While syllables of the same identity undergo unique, correlated changes in length across motifs, this reflects an accumulation of independent deviations in note lengths. Furthermore, while syllables of the same identity have similar elasticity coefficients, this similarity is also dominated by similarities on the note level.

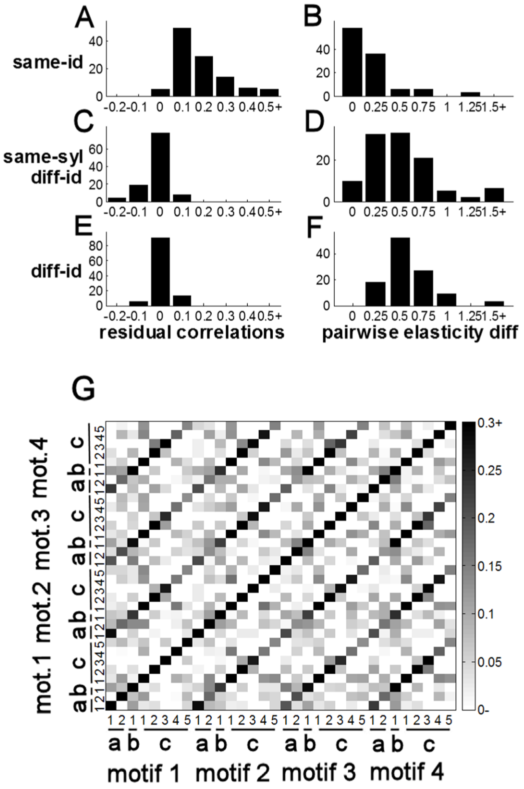

Figure 3.2: Timing is note-based. A,C,E: Correlation coefficients among notes after factoring out sequence length, either between notes of the same identity across motifs (A), notes of different identity in the same syllable (C), and notes from different syllables (E). B,D,F: Pairwise absolute elasticity differences (see Materials and Methods for definition) organized as in A,C,E. G: Correlation matrix from bird 16.

Spectral type does not explain note-based data

Previous studies have classified notes on the basis of more abstract acoustic properties, or spectral "type." It is possible that note type may explain some (or all) of identity-dependence as well if these correlated timing deviations are tied to mechanisms that directly represent song acoustics. We focused our analysis on the subset of notes that allow for clear classification, using previously defined categories (Williams et al., 1989; Williams & Staples, 1992; Sturdy et al., 1999): "harmonic stacks," which have a clear fundamental frequency in the 500-1000 Hz range that remains fairly constant throughout the note; "high notes," which exhibit peak power in the 3-7 kHz range; "sweeps," which show a continuously decreasing fundamental frequency; "short noisy sweeps," which typically constitute introductory-like notes; "noisy, low-amplitude" notes; and "amplitude-modulated" (AM) notes, which show fast, regular changes in amplitude across frequency bins. We found that the length distribution for harmonic stacks is bimodal with a relatively clean break at 60 msec. Using this threshold, we divided this category into "short stacks" and "long stacks." In total, 64% of notes were classified as one of these types, with 4-31 unique notes in each category. Of these, 47 notes from 7 birds (3-14 per bird) had at least one other note of the same type in that song.

When we examined relationships between note pairs of the same spectral type, the same-type relationships closely resembled the same-syl and diff-syl data and not the same-id data: Correlations were not significantly different between same-type and diff-type pairs ($p = 0.16$ WSR; means $0.01 \pm .01$ vs. -0.01 ± 0.01), nor were elasticity similarities ($p = 0.63$ WSR; means 0.83 ± 0.15 vs. 0.68 ± 0.11). These data suggest that the identity-dependent temporal structure is unrelated to acoustic type.

3.2.2 Fine timescale for identity-dependence

We have used a note-based analysis to argue that identity-dependence among syllables is dominated by independent patterns at sub-syllabic timescales. In fact, identity-dependence among notes could stem from patterns on an even finer scale, such as the accumulation of variability in the 5-10 msec bursting patterns found in premotor nucleus HVC (Fee et al., 2004; Hahnloser et al., 2002). In this case, we should find independent timing variability in any given 5-10 msec "slice" of song. If notes are not distinguished as cohesive units in the motor code, then two different slices from the same note should in fact be as unrelated to each other as they are to slices from different notes (Fig. 3.3B). On the other hand, if the motor code does distinguish notes, length variations in different slices from the same note should be especially related to each other as well (Fig. 3.3A).

However, there is a limit to testing these patterns using our methods. To measure song-to-song timing variability, the DTW algorithm depends on regular changes in the spectral profile of a syllable (see Materials and Methods and Appendix C). The precise tracking of timing in 5-10 msec slices is thus impossible within notes that are temporally homogeneous (*e.g.* harmonic stacks), and unreliable within notes that have a strong spectral component of song-to-song variability. Instead, the current analysis depends on spectrotemporal features that can be reliably identified

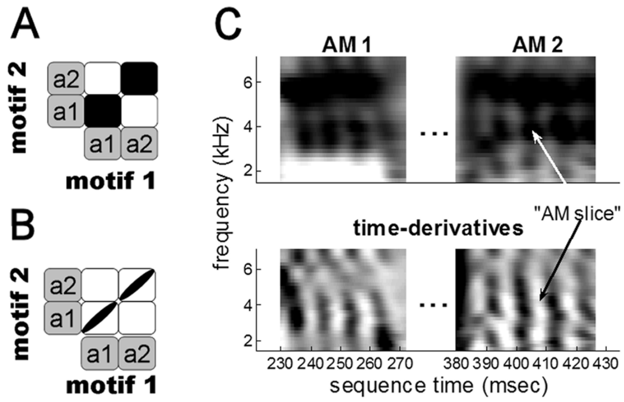

Figure 3.3: Timing on a fine scale. A-B: Schematic covariance matrices for two notes in the same syllable produced in two different motifs. As in figure 1, covariances are represented as either strong (black) or no/weak covariance (white). In A, notes are encoded as cohesive units, so the covariance between the same note across motifs is shared by all portions within the note. In B, notes are not encoded cohesively and the covariance between same-id notes reflects the accumulation of independent covariances on a finer timescale. C: Top, the mean spectrograms for two AM notes produced by bird 10. Bottom, corresponding spectral time-derivatives used to divide notes into roughly 10-msec slices of song.

song by song on a fine timescale (*cf.* Chi & Margoliash, 2001).

Thus, we focused our analysis on amplitude-modulated notes (defined above), which allowed for the accurate measurement of ∼10 msec temporal slices corresponding to pulses in song amplitude (Fig. 3.3C). We divided these notes into intervals defined peak-to-peak in the spectral time-derivative (mean length 9.37 ± 0.38 msec). Across the sample, the analysis included 4 unique notes from 3 birds, with 3-4 slices per note and a total of 15 slices.

Variance on the 10-msec scale

We found that a 10-msec slice was indeed more correlated with itself across motifs than it was with other slices of the same note (Fig. 3.4A,C, $p < 0.001$ WSR;

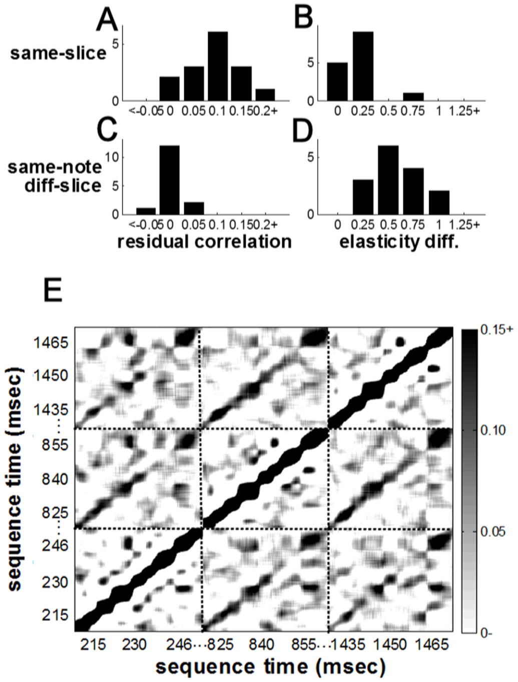

Figure 3.4: Temporally precise deviations in AM notes. A,C: Correlation coefficients among 10-msec AM slices after factoring out sequence length, either between slices at the same position across motifs (A), or slices from different motif positions in the same note (C). B,D: distributions of pairwise absolute elasticity differences, organized as in A and C. E: Correlation matrix for the same AM note produced by bird 10, repeated across 3 motifs. Each value in the matrix represents the correlation between two 5-msec stretches of song centered at the times corresponding to that locations's vertical and horizontal coordinate.

mean same-id correlation 0.10 ± 0.01, same-note correlation 0.004 ± 0.01). The mean difference held for all 4 notes individually. We did not find significant anti-correlation among motif-adjacent slices as we had among notes ($p = 0.626$ WSR; mean motif-adjacent and non-adjacent correlations were -0.001 ± 0.01 and 0.01 ± 0.01 respectively).

The elasticity coefficients showed the same pattern as correlations: slice elasticity was significantly closer to the same slice across motifs than it was to the elasticity of other slices in the same note ($p < 0.005$ WSR; mean same-id elasticity difference was 0.20 ± 0.04, same-note, 0.56 ± 0.06).

Are different 10-msec slices in the same AM note any more related than those from different AM notes? Two observations suggest not. First, bird 10 produced two different AM notes (shown in Fig. 3.3C), and here we found no effect of note identity

($p = 0.47$ WSR among both correlation and elasticity distributions). Second, the average relationship between two different slices in the same note is similar to the relationship between two different notes in different syllables (among correlation coefficients, 0.004 ± 0.01 vs. 0.01 ± 0.003, in elasticity similarity, 0.56 ± 0.06 vs. 0.64 ± 0.05).

While peaks in the time-derivative provided a convenient, systematic way to segment AM notes, there is no reason that this segmentation should necessarily correspond to a temporal segmentation in the underlying motor code. To look for structure in the underlying representation, we performed a continuous correlation analysis, calculating length correlations between 5-msec intervals centered on any two points within each AM note. Across notes we found that the correlation depended most strongly on the distance between the two intervals rather than any particular alignment with the discrete pulses of acoustic output (Fig. 3.4E). These qualitative results confirm a 5-10 msec timescale for the representation of song, but do not indicate that the amplitude pulses directly correspond to elements of the underlying motor representation.

3.2.3 Effect sizes

Correlation coefficients yield a normalized measure of the strength of identity-dependence across motifs. We also estimated this strength in units of real time. Because this analysis did not require pairwise statistical tests, we include all 122 notes in these estimates.

To isolate identity-dependence, we first regressed each note with (i) sequence length, (ii) the sum of all notes (except same-id notes), and, to factor out jitter that is correlated across motifs, (iii) adjacent intervals on either side of that note (previous and subsequent gaps included for the first and last notes of syllables). Among the residuals from this multiple regression, we estimate mean pairwise covariance among notes of the same identity at 0.18 ± 0.07 msec^2 (range [0.006,0.59] msec^2 by bird). Taking the square-root of positive covariances (111 out of 122), this comes out to an estimated 0.36 ± 0.03 msec of note length deviation that is correlated across motifs and independent of global factors and jitter.

We performed the same analysis with ∼10-msec AM slices, except we did not factor out adjacent intervals because we had found no significant anti-correlation as we had among notes. Here we find a mean 0.04 ± 0.02 msec^2 of variance that is shared between the same slice of song across motifs. Again taking the square-root of covariances (all $15 > 0$), we derive an average 0.18 ± 0.03 msec of deviation in the length of a 10-msec AM slice that is repeated across motifs.

Under the hypothesis that identity-dependence is simply the accumulation of similarity between fine timescale components, the covariance between intervals should be proportional to interval length. For all notes excluding AM notes, the average covariance per msec was 0.005 ± 0.001 msec^2/msec (for positive covariances, std. dev./msec $= 1.24 \pm 0.07\%$). Among AM notes and slices, where we know that covariance accumulated independently on the 10-msec timescale, the average covariance per msec was 0.004 ± 0.001 msec^2/msec (std. dev./msec $= 0.92 \pm 0.11\%$). The average covariance per msec among AM notes is similar to what we find in other notes (0.004 ± 0.001 vs. 0.005 ± 0.001); thus, the data as a whole are consistent with the proposal that identity-dependence is dominated by an independent accumulation of covariance on a fine timescale.

3.3 Discussion

We have probed subsyllabic timing in zebra finch song to test the hypothesis that song syllables constitute cohesive units within the underlying motor representation for song. We examined two measures of song-to-song variability which we had previously found to be similar for syllables of the same identity repeated across motifs (Glaze & Troyer, 2006): the ability to proportionally stretch and compress with tempo change (elasticity), and length correlations that remain after factoring out global tempo. In each of these measures, we find that identity-dependent similarity is dominated by smaller segments: length deviations among notes in the same syllable are no more correlated with each other than they are with notes in other syllables, and note elasticity is poorly predicted by the particular syllable in which it is produced. We then applied the same analysis to a subset of notes that allow accurate timing measurements of 5-10 msec sub-note slices. We found analogous patterns on this finer scale: timing deviations in a given slice are correlated with the same slice repeated in other motifs, yet are independent of other slices in the same note.

These results suggest that the song motor code has remarkably high fidelity; specifically, song segments as short as 5-10 msec are represented independently. The data also expose remarkably divergent timescales–temporal deviations in 5-10 msec segments are correlated over seconds. We hypothesize that the slow timescales in our data stem from modulatory factors that are spread widely through the song system and vary from song to song, whereas the fast temporal deviations are a direct behavioral expression of sparse and precise bursting activity that has been recorded in the premotor nuclei HVC and RA (Hahnloser et al., 2002; Fee et al., 2004; Leonardo & Fee, 2005; Yu & Margoliash, 1996).

Fast and slow timescales

One of the most remarkable aspects of our data is the contrast of fast and slow timescales. This mixture follows naturally from the hypothesis that sparse, fine-grained bursting in HVC reflects the activity of a synfire chain, in which neurons that are active during consecutive 5-10 msec slices of the song motif are linked by strong synaptic connections (Fee et al., 2004; Abeles, 1991). Modulatory changes that increase the excitability of neurons along this chain will lead to faster propagation of activity and hence a faster song tempo (Arnoldi & Brauer, 1996). If different neural groups undergo somewhat different changes in excitability, tempo changes will be spread unequally over the different links in the chain. Since repeated motifs are generated by repeated propagation along the same chain, these link-specific temporal deviations will persist over the time course of the slow modulation.

This explanation applies most cleanly in HVC, where individual projection neurons burst exactly once during each song motif, ensuring the independence of different links in the chain. However, RA bursting patterns are also repeated across motifs (Yu & Margoliash, 1996; Chi & Margoliash, 2001), with largely uncorrelated populations of neurons active during any two time points within the same motif (Leonardo & Fee, 2005). Therefore, slow modulation in RA could also contribute to temporally specific tempo changes that are repeated every 500-1000 msec.

Syllable-based representations

While our data provide evidence for a fine-grained motor code, other experiments suggest that song has a syllable-based representation. The respiratory pattern is highly stereotyped, and involves expirations that accompany syllables and short inhalations that accompany gaps (Wild et al., 1998; Franz & Goller, 2002). Respiratory nuclei in the brainstem are part of coordinated recurrent circuits that run through HVC and RA, and mediate inter-hemispheric coordination of HVC activity that is particularly pronounced at syllable onsets (Ashmore et al., 2005; Schmidt, 2003). Furthermore, birds interrupted by bright flashes of light or brief pulses of current tend to stop their songs at syllable (occasionally note) boundaries (Cynx, 1990; Vu et al., 1994; Franz & Goller, 2002), and *in vitro* data indicate that brief pulses delivered to HVC slices yield rhythmic bursting whose timing roughly matches the rate of syllable production (Solis & Perkel, 2005). Finally, syllables and gaps are distinguished from each other along two independent measures of timing: syllables are less elastic than gaps, and after factoring out tempo, syllable-syllable and gap-gap length correlations are significantly stronger than syllable-gap correlations (Glaze &

Troyer, 2006).

In reconciling the fine-grained and syllable-based views of the motor code for song, it is important to separate the notion that syllables form cohesive units from a simple distinction between portions of the song with and without vocal output. This distinction may reflect systematic differences between the bursting neurons that subserve vocalization and those that are active during silent gaps. Such differences in network structure could include a number of different factors, such as patterns of connectivity, the strength of synaptic connections, or the number of neurons active at a given point on the song. However, such differences need not entail syllable-based (or note-based) units, since neurons bursting at different times in the same syllable may be no more related than neurons bursting within different syllables. To establish a true syllable-based hierarchy, any experimental manipulation must lead to measurable changes that are shared across neurons coding for the same syllable, but are distinct from the changes in neurons coding for different syllables.

Peripheral vs. central representations

Our leading hypothesis is that fine-grained deviations are driven by sparse, precise bursting in the forebrain nuclei HVC and RA. However, it is possible that these deviations in fact stem from physiological mechanisms peripheral to the central pattern generator, or from inaccuracies in measurements of song timing. Two results make this unlikely. First, if similar peripheral mechanisms generate similar acoustic features, one would expect song acoustics and temporal variability to be

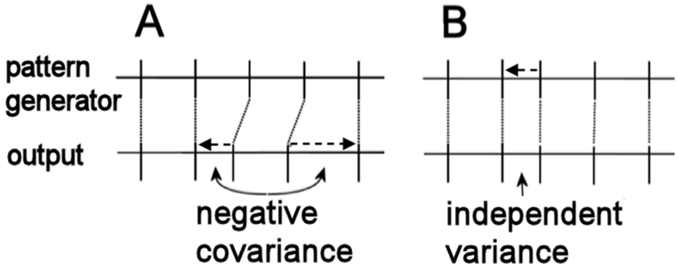

Figure 3.5: Schematic representing two hypothetical sources of length deviations. The factors determining measured timing values can be conceptually separated into those that depend on the central pattern generator for song and those determined by influences downstream of the pattern generator. The timing of song is indicated by vertical dashes, according to the coding within the pattern generator (top) or as measured in the song output (bottom). Dashed arrows indicate temporal deviations. A: Deviations originating downstream of the pattern generator. Because centrally coded timing continues unaffected, the deviations are eventually offset by equal and opposite deviations. This leads to patterns of negative covariance in measured timing. B: Timing deviations caused by the pattern generator. Here, the relative timing of subsequent output activity remains unaffected, leading to variance that can be independent of other timing deviations. Overall, independent length deviations in the behavior must reflect timing variability in the song pattern generator.

related. We detected no such relationship between timing variance and note type. Furthermore, different 10-msec pulses within AM notes have very similar acoustics, yet have independent length variation. Second, any timing deviation that originates downstream from the song pattern generator must be offset by compensating deviations, provided that the pattern generator continues to pace activity independently of the source of the deviation (Fig. 3.5). Thus, deviations that stem from either peripheral mechanisms or the measurement algorithm will induce patterns of negative correlation among neighboring song segments. Again, we detected no such correlation between slices within AM notes.

We did find negative correlations between a given note and adjacent segments in subsequent motifs. This could result from spectral deviations that are correlated across motifs. Because the DTW algorithm warps time to achieve maximal spectral matching, these spectral deviations would be converted to correlated temporal distortions. However, these negative correlations were nearly six times smaller than the positive correlations among same-id notes, again suggesting that the central pattern generator dominates timing variance on the note level. Indeed, the data suggest that linear warping techniques used to align spike-timing with song acoustics (e.g. Leonardo, 2004) may be improved if performed note-by-note rather than syllable-by-syllable.

Implications for perception and learning

Our results demonstrate that the vocal production system of birds operates with sufficient fidelity to repeat sub-millisecond temporal deviations in specific portions of the song. Tests of auditory discrimination have shown that zebra finches can perceive temporal changes on the millisecond scale (Lohr et al., 2006; Dooling et al., 2002), while HVC neurons in anesthetized birds shows similar auditory sensitivity (Theunissen & Doupe, 1998). From our data, we estimate that 0.5 msec^2 of identity-specific variance would accumulate during a 100 msec interval of song. This suggests that the temporal deviations specific to individual song syllables are near or above the detection threshold for zebra finches.

The timescales in these data may also have implications for song learning. Others have suggested that sparse representations may facilitate reinforcement-based learning strategies (Fiete et al., 2004). However, feedback delays are expected to be roughly 40-100 msec, complicating the use of matching signals to adjust temporally precise motor programs (Troyer & Doupe, 2000a). Because temporal deviations are repeated over multiple song motifs, evaluative signals from earlier motifs may be used to modulate neural plasticity triggered by premotor spike patterns that are repeated in later motifs. Such a strategy may be useful in wide array of sensorimotor learning tasks in which similar delay problems exist.

Conclusion

We have analyzed temporal variability in birdsong acoustics to reveal structure in the motor code on multiple timescales. The fine-scale data suggest a direct link between acoustics and premotor bursting patterns. The long timescale of correlations is suggestive of similar patterns found in behavioral studies on humans (Gilden, 2001). In general, timing variability provides a common language for synthesizing results from behavioral and electrophysiological studies, and also provides strong constraints for computational models that attempt to connect the two levels of analysis. Furthermore, the large samples of song acoustics that can be readily collected yield statistical power that is difficult to achieve in physiological investigations. This approach to song analysis may thus reveal subtle but important changes in song representation during different behavioral contexts and over the course of song development (Crandall et al., 2007; Cooper & Goller, 2006; Deregnaucourt et al., 2005; Hessler & Doupe, 1999; Ölveczky et al., 2005; Kao & Brainard, 2006; Tchernichovski et al., 2001; Brainard & Doupe, 2001).

Chapter 4

Timing Modulation in a Synfire Chain Model of Zebra Finch Song

Most proposals for temporal coding in the brain fall into two broad categories: strategies based on interacting patterns of neural oscillation, and strategies involving precise patterns of spiking on a millisecond timescale. Synfire chains comprise the primary mechanistic hypothesis for how neural circuits could generate and propagate precise spike patterns. Here, near-synchronous spiking in one group of neurons triggers spikes in the next group and so on in a chain-like manner (Abeles, 1991; Bienenstock, 1995; Diesmann et al., 1999). While multiple studies support the existence of synfire-like spiking in cortical circuits (Vaadia et al., 1995; Riehle et al., 1997; Beggs & Plenz, 2003; Ikegaya et al., 2004), the complexity of ongoing cortical activity has made statistical verification of these patterns quite difficult (Gerstein, 2004; Baker & Lemon, 2000; Oram et al., 1999).

The adult zebra finch song system is a good candidate for the identification of synfire chains. Song acoustics are composed of highly stereotyped, tightly timed sequences of discrete vocal gestures known as "syllables." Recent recordings from nucleus HVC suggest that the motor representation of adult song might be based on a precise timing code (Hahnloser et al., 2002). In contrast to the cortex, the bursting pattern of these neurons is "ultra-sparse": each neuron bursts exactly once during the singing of a stereotyped sequence of vocalizations known as a motif, with different neurons bursting at different times during the motif. Furthermore, spiking in any given RA neuron is tied to song behavior with millisecond precision (*cf.* Chi & Margoliash, 2001). These data led Fee and colleagues to propose that the temporal code in HVC acts much like a clock, pacing the output for each motif (Fee et al., 2004). At the circuit level, precise bursting in HVC was hypothesized to result from a synfire-like mechanism with neurons active in one 5-10 msec interval triggering a burst in the next set of neurons along the chain.

On the other hand, there is also evidence that song representations are chunked into discrete syllables. This segmented model is supported by the fact that syllables and the gaps of silence between them correspond to periods of expiration and inspiration, respectively (Franz & Goller, 2002), and birds startled by bright flashes of light tend to interrupt their songs at syllable boundaries (Cynx, 1990; Franz & Goller, 2002).

We have developed statistical models of song timing that provide support for both syllable-based representations and a temporally precise motor code. Our first set of results (Glaze & Troyer, 2006) showed tempo changes that were non-uniform across syllables and gaps. Specifically, syllable lengths scale with tempo changes proportionally less; *i.e.* they are less "elastic." Such non-uniformity is at odds with the simplest synfire model in which songs are driven by a clock-like mechanism that paces song evenly across syllable-gap sequences. On other hand, in a subsequent analysis we factored out tempo changes and used the remaining variability to investigate subsyllabic timescales (Glaze & Troyer, 2007). Here, we found length variability that was specific to 10-msec song slices and independent of neighboring vocalization, yet correlated across motifs. Such patterns of variability suggest an interaction between fine-grained representations and a neuromodulatory source operating on a much slower timescale.

Here we use computational models to link the hypothesis that song timing is driven by synfire-like bursting in nucleus HVC (Fee et al., 2004) with the constraints provided by our behavioral data (Glaze & Troyer, 2006, 2007). Song length variability reflects tempo changes shared across syllables and gaps; thus, we focus on how modulatory changes in global network parameters affect the propagation of activity along the underlying chain. We find two main influences on propagation speed: First, changes in speed can be directly tied to the strength of synaptic connections between links in the chain, so that the modulation of excitatory parameters result in distinct patterns of tempo change. Second, while it is straightforward to model the fine grained noise that remains after factoring out tempo, the magnitude of measured noise may require correlations within individual chain layers that are at odds with the simplest hypothesis that chains are anatomically distributed at random. In addition to leading to specific hypotheses regarding song production, our results may have important implications for the measurement and detection of synfire chains in cortical circuits.

4.1 Methods

We used a leaky integrate-and-burst model based on the standard leaky integrate and fire (LIF) model neuron, $\tau_m \; dV/dt = -V + V_{rest} + I(t)$. Synaptic input was modeled as a current rather than a conductance. Our main intent in studying simplified chaining models was to clarify how chain timing depends on network function, and this level of modeling has the basic advantage of providing a linear framework in which different effects can be modeled and explored independently. For example, increasing the strength of synaptic input simply scales the resulting post-synaptic potential (PSP), whereas with conductance-based synapses increasing input strength changes both the size and shape of the PSP. Initial simulations using conductance-based synapses support the qualitative conclusions from the current-based models.

We used the following parameters: resting potential, $V_{rest} = -65$ mV; membrane time constant, $\tau_m = 20$ msec; and spike threshold, $V_{thresh} = -55$ mV. When a neuron crossed threshold, it burst with 5 spikes over 6 msec in 1.5 msec intervals with 0.2 msec jitter; these parameters were based on previous research (Hahnloser et al., 2002; Fee et al., 2004). All synaptic currents ($I(t)$) had an instantaneous rise and exponential decay ($\tau_{syn} = 5$ msec). Resulting EPSPs from single-spike inputs had a 9.27 msec time to peak, similar to what has been previously reported for the postsynaptic potentials evoked from a spike from another HVC_{RA} neuron (Mooney & Prather, 2005).

Each neuron was also subject to synaptic background input. This input was modeled as independent Gaussian noise, which approximates weak Poisson input from a large neural population. Typical noise levels yielded fluctuations in resting potential with a standard deviation of $1 - 3$ mV.

4.1.1 Synfire chain model

A synfire chain involves synchronous spike volleys over a sequence of discrete neural subpopulations, connected serially with many-to-many connectivity (Fig. 4.1; Abeles, 1991; Diesmann et al., 1999). We explored chains with a minimum of 100 neurons per layer; however, most simulations used 200 neurons per layer, the estimated number of neurons that burst during a given point in the song sequence (Fee et al., 2004). A given neuron projected to 20-100 neurons in the next layer (20 projections for most simulations, see Results), and each projection had a fixed axonal transmission delay picked from a random distribution of $0.5 - 1.5$ msec. In conjunction with burst variability (see above), this yielded a floor of ~ 0.5 msec burst onset jitter per layer without any background noise or external noise source.

Throughout, we refer to a chain "link" as the set of all connections between two adjacent synfire layers. We calculate the overall timing of a given chain layer in a given trial as the average burst onset time across all neurons in that layer, and calculate the latency T between two adjacent chain layers simply as the difference between both averaged burst onset times. Note that since propagation speed is proportional to the inverse of the latency, both terms will be used to describe synfire timing.

The primary goal of this investigation was to link synfire chain parameters with our previous measurements of the variability in song timing. Therefore, we assume a direct link between the length of a song interval and the sum of all latencies across the chain subserving that interval's production. The central parameter we explore here is the distance between membrane potential and spike threshold at the time a neuron is signaled to burst during song; we denote this distance θ. Lowering θ for all neurons will increase chain speed (*i.e.* tempo), while increasing θ slows it down (Wennekers & Palm, 1996).

4.1.2 Calculating Elasticity

In the behavioral data, we measured the elasticity for each song interval with respect to song length, defined as the change in fractional length of that interval relative to fractional changes in song length. In an analogous way, we now define the elasticity of a synfire chain with respect to changes in distance to threshold, E_T^θ, as the change in the latency to threshold T for each unit change in distance to threshold θ, expressed as a fraction of the average latency to threshold, $\overline{T}$, *i.e.* $E_T^\theta = (dT/d\theta)/\overline{T}$. In our simulations, we measured $dT/d\theta$ by systematically varying θ from $5-15$ mV in 1-mV steps over 100 chain links, calculating the slope $dT/d\theta$ at each link, and averaging across the chain. For clarity, we will refer to behavioral measurements using song length as "song elasticity" to distinguish them from timing changes with respect to physiological parameters in the chain model.

4.2 Results

The primary objective of this project was to use a synfire chain to simulate several key patterns of timing variability we have found in adult zebra finch song acoustics reported in Chapters 2 and 3 (Glaze & Troyer, 2006, 2007). Songs consist of sequences of vocalizations known as "syllables," interspersed with brief silences we have termed "gaps." We will also use the generic terms "interval" and "segment" to refer to either syllables or gaps. On a longer scale, syllables are arranged into stereotyped sequences called "motifs"; we will also use the term "sequence" to refer to a set number of motifs produced back-to-back. One previous analysis (Glaze & Troyer, 2006) indicated very tight timing in adult songs, with the majority of sequence length deviations under 1.5% and interval length deviations under 2-3 msec. However, behind this tight timing we have found statistical structure in rendition-to-rendition timing variability that suggests how song representations are organized.

The goal of the current inquiry is develop a synfire chain model that yields this statistical structure in song timing. The synfire chain is a computational mechanism that has been hypothesized to subserve highly regulated behaviors such as zebra finch song (Diesmann et al., 1999). Chains were based on known connectivity and physiological parameters in premotor nucleus HVC (see Methods).

Throughout, we equate song timing with propagation speed across the chain, and focus on latency T from the time of burst onsets in a previous layer until neurons in the current layer reach spike threshold and begin to initiate their bursts. Specifically, we examined (1) what may determine the sensitivity of T to rendition-to-rendition perturbations in network properties and (2) how different patterns of network perturbations across the chain may drive the patterns of covariance in T we find across song.

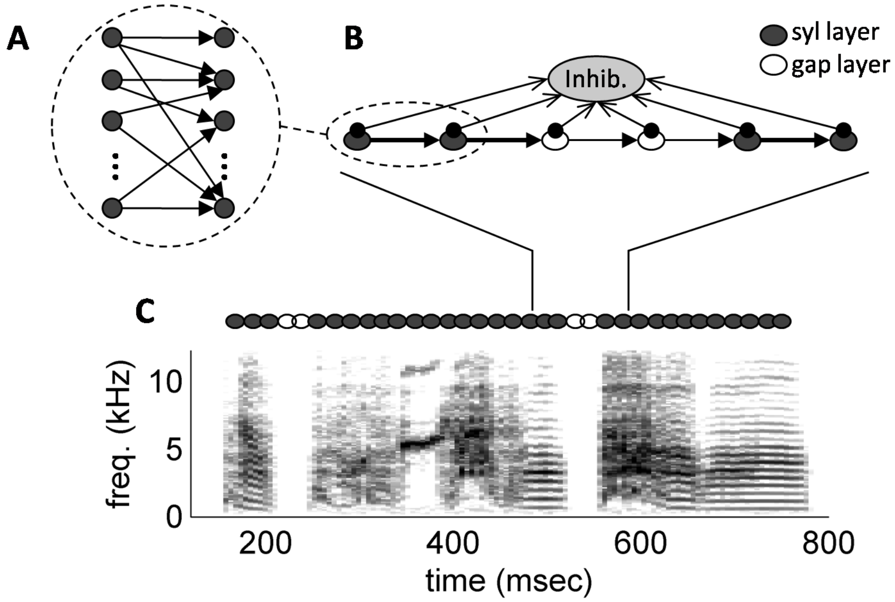

Figure 4.1: Schematic of the synfire chain model we explore. A, two adjacent layers of the synfire chain. Layers are linked with many-to-many projections from the first to the second. B, A sequence of six layers in which syllable-based subpopulations are represented by filled circles, gap-based by empty circles. The chain receives inhibitory feedback from a population of inhibitory interneurons (see Results for details). C, Spectrogram of a song motif with corresponding synfire layers above.

4.2.1 Tempo changes and elasticity

In one previous behavioral analysis (Glaze & Troyer, 2006) we found tempo changes that were strong enough to induce a net rendition-by-rendition positive covariance across the lengths song intervals, suggesting global neuromodulatory factors shared by all song representations. Perhaps the most basic parameter affecting propagation speed is the distance between membrane potential and spike threshold; we denote this distance θ. We modeled tempo by allowing all neurons to share a common source of variability in θ: lowering θ increases tempo, while increasing θ slows it down (Wennekers & Palm, 1996). Biological factors that can affect θ include changes in background levels of synaptic input, modulation of conductances contributing to the resting potential, and modulation of one or more spiking conductances.

Our previous analysis also indicated that tempo changes were non-uniform across the song. One of the most striking patterns here was a robust difference between syllables and gaps. Specifically, syllables tend to strength and compress with total song length proportionally less than gaps, *i.e.* syllables have lower song elasticity coefficients (see Methods). How can measured song elasticity be linked with the network parameters determining chain elasticity?

Our central hypothesis is that the distinction between syllables and gaps does not come from different modulations in neural excitability (represented by θ), but rather differences in network parameters that determine the sensitivity of neuronal timing to modulations spread evenly across the chain. In particular, we hypothesize

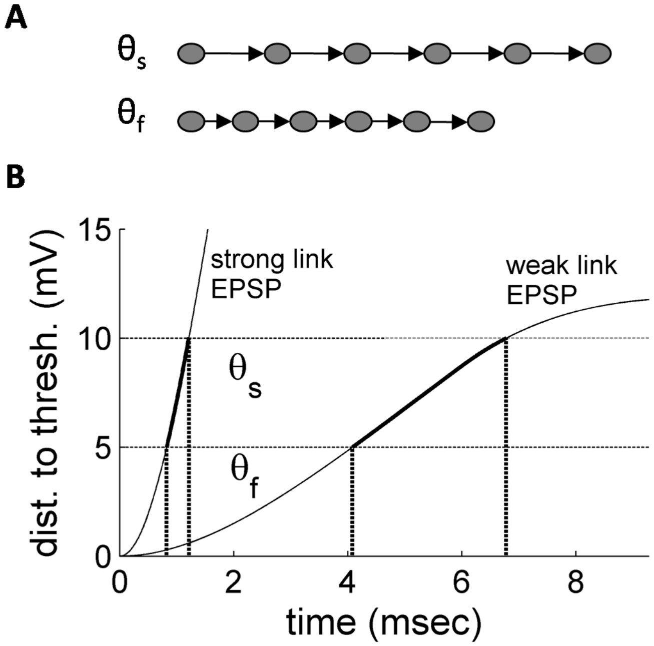

Figure 4.2: Conceptual model of tempo changes and elasticity. A, Circles represent chain layers and arrow length represents latencies to spike threshold. Top, distance to threshold θ_s is relatively high so chain activity is slow. Bottom, θ_f is low, yielding faster chain speed. B, EPSP plots for strong and weak chain links. Two distances to threshold (θ_s and θ_f) are indicated by horizontal dotted lines, while spike latency is the intersection with the EPSPs, indicated by vertical dotted lines. Strong-link EPSP slope is steeper, yielding a smaller difference in latency with respect to θ_s and θ_f, *i.e.* lower elasticity.

that the elasticity of a chain segment is in fact determined by the aggregate slopes of the EPSPs elicited across all neurons in that segment. If perturbations in θ are small enough, the sensitivity of the latency to changes in distance to threshold ($dT/d\theta$) is inversely related to the slope of the EPSP near threshold (dV/dt). The rise of stronger EPSPs will have a steeper slope and thus yield latencies with less temporal sensitivity to changes in θ (Fig. 4.2).

However, elasticity depends on the change in latency relative to the mean latency. Since stronger EPSPs will reduce the mean latency $\overline{T}$ as well as reducing the sensitivity $dT/d\theta$, the relationship between EPSP size and elasticity ($= (dT/d\theta)/\overline{T}$)

is not direct. For example, if the rising portion of the EPSP can be approximated by a power function, $V(t) = mt^{\alpha}$, elasticity is entirely independent of EPSP magnitude m:

$$\frac{dT/d\theta}{\overline{T}} = \frac{1/(dV/dt)}{\overline{T}} = \frac{1/\alpha\overline{T}^{\alpha-1}}{\overline{T}} = \frac{1}{\alpha m\overline{T}^{\alpha}} = \frac{1}{\alpha\overline{\theta}} \tag{4.1}$$

where the last equality comes from the fact that $m\overline{T}^{\alpha} = V(\overline{T}) = \overline{\theta}$. Hence $E_T^{\theta} = 1/\left(\alpha\overline{\theta}\right)$ is independent of the EPSP size m. Thus, any relationship between elasticity and EPSP magnitude thus requires that the rising portion significantly deviate from the closest power function approximation.

Elasticity is associated with link density and synaptic strength

We first explored our hypothesis by varying the magnitude of the compound EPSP in two different ways. First, we changed the synaptic strength of each neuron. Stronger synapses yield larger excitatory postsynaptic current (EPSC) for each presynaptic burst, leading to a larger compound EPSP. Second, we changed the density of synaptic connections between to chain layers: the more projections a neuron receives from the previous layer, the stronger the compound EPSP. To calculate elasticity at each set of parameters, we simulated uniform chains of 100 links with realistic levels of background noise, burst variability, transmission delays and jitter, and calculated changes in latency when θ was varied from $5 - 15\ mV$ in $1\ mV$ steps (see Methods).

Indeed, elasticity decreased with both stronger EPSCs and denser links (Fig. 4.3C,D): for a 1 mV increase in membrane potential, latencies decreased by 6% among the

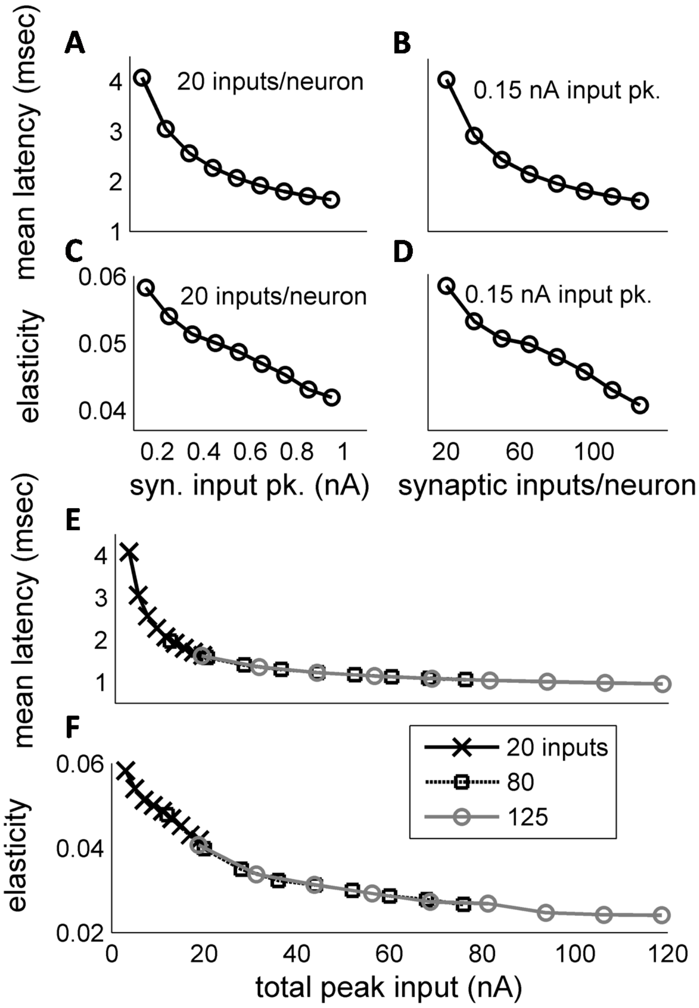

Figure 4.3: Mean latencies and elasticity with respect to synaptic strength and connection density. A & C, relationship between peak current per synaptic input and latency and elasticity respectively. Simulations used the weakest connection density considered, 20 inputs/neuron. B & D, relationship between connection density and the same timing properties, with the weakest synaptic input peak considered. E & F, latencies and elasticity as a function of total synaptic input peak, calculated as inputs/neuron X input peak. Results indicate that these two parameters have independent and similar influences on timing properties; thus, the key functional parameter is total synaptic input peak.

weakest chains and 4% among the strongest. Latencies also decreased with increased link strength (Fig. 4.3A,B), ranging from ~ 4 msec among the weakest links and ~ 1.5 msec among the strongest. As expected, the simulations indicated that the relative influences of EPSCs peak and link density were independent and very similar to each other; this suggests that the main functional excitatory parameter is overall link strength, *i.e.* compound EPSC peak (Fig. 4.3E,F).

Inhibitory feedback gain influences elasticity

We have thus far shown that stronger synapses and a greater number of synaptic connections per layer can decrease both elasticity and average latencies. However, there is also indirect evidence that $\mathrm{HVC_{RA}}$ bursting patterns may be influenced by inhibitory feedback from interneurons (denoted as $\mathrm{HVC_{int}}$; Mooney & Prather, 2005; Hahnloser et al., 2002), suggesting that the properties of synaptic connections between interneurons and RA-projecting neurons may influence chain timing.

To explore the possible influences of inhibitory input on chain timing, we explored the gain of inhibitory feedback from $\mathrm{HVC_{int}}$ to $\mathrm{HVC_{RA}}$ subpopulations. There is evidence that singling-related activity among inhibitory interneurons may help shape $\mathrm{HVC_{RA}}$ patterns with tonic, high-frequency spiking and no adaptation (Kozhevnikov & Fee, 2007; Mooney & Prather, 2005; Hahnloser et al., 2002); furthermore, $\mathrm{HVC_{int}}$ neurons receive large-amplitude, short-latency input from $\mathrm{HVC_{RA}}$ and show a steep, linear relationship between firing rate and injected current (Mooney, 2000; Mooney & Prather, 2005). Thus, we used a simple firing rate model for inhibition, modeling inhibitory activity as $I_{inh}(t) = \gamma s(t)$, where $s(t)$ represents the input from $\mathrm{HVC_{RA}}$ neurons and γ is the gain of inhibitory feedback. Since the rate of spiking activity in $\mathrm{HVC_{RA}}$ neurons will be inversely proportional to latency ($s(t) \propto 1/T$), we modeled tonic inhibitory activity simply as γ/T.

A common simplifying assumption is that inhibition is “global” and all inhibitory interneurons can be lumped into a single population providing inhibitory feedback to all excitatory neurons. However, there is little evidence for such robust synchrony, and one of the most thorough studies to date has shown only weak correlations among firing patterns of $\mathrm{HVC_{int}}$ neurons during song (Kozhevnikov & Fee, 2007, average pairwise correlation = 0.05). We thus make an alternative simplifying assumption, and model inhibitory feedback $I_{inh}(n)$ to a given chain n layer as driven solely by the previous latency $T(n-1)$ (which depends on activity over layers $n-2$ and $n-1$):

$$I_{inh}(n) = \gamma/T(n-1) \tag{4.2}$$

We explored the influence of this inhibition on mean latency and elasticity in the same chains used above (Fig.4.4). As expected, stronger inhibition increases average latencies by pushing average membrane potential further away from threshold, and thus decreases chain speed. Furthermore, gain suppresses elasticity across different link strength parameters. This is because any change in distance to threshold θ is partially offset by an opposing change in inhibitory feedback. For example, although decreasing θ shortens T, a shortened T in turn represents an increase in $\mathrm{HVC_{RA}}$ firing rates which will increase inhibitory feedback and thus will lengthen T

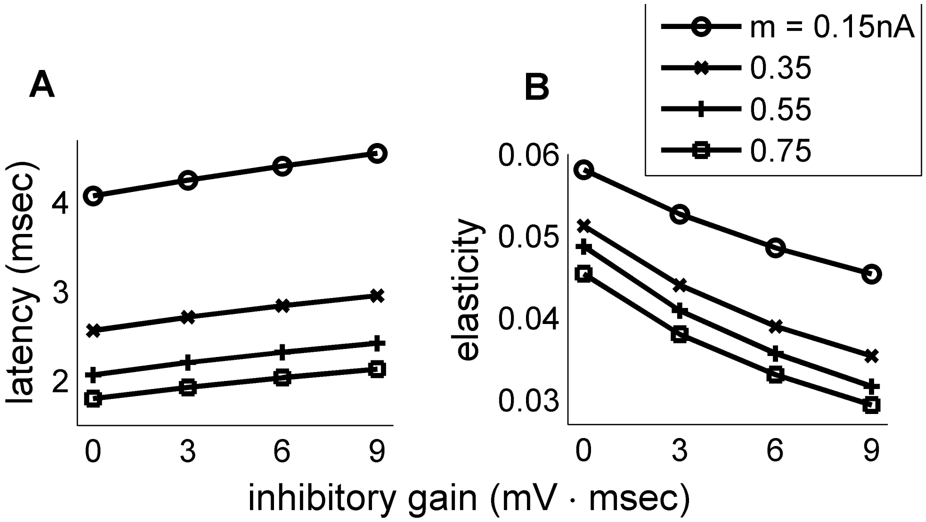

Figure 4.4: Elasticity with respect to the gain of inhibitory feedback across four simulations with varying synaptic strength. A, latencies increase with stronger gain. B, in contrast, elasticity decreases with increasing gain. Overall, the influence of feedback gain is similar regardless of synaptic strength.

down the chain. The steady state spike sequence will represent a tradeoff between these two opposing factors.

It is important to note that the influence of inhibitory feedback gain γ stands in stark contrast to link strength parameters: while increases in EPSP size decreases both latency and elasticity, increasing inhibitory feedback increases latencies while decreasing elasticity.

Quantitative matches to song elasticity

Thus far we have investigated the relationship between network parameters and the elasticity E^{θ} in latency T with respect to changes in distance the threshold θ. (We have dropped the subscript T in E^{θ} for convenience.) We use this understanding to provide quantitative estimates of timing variability within a chain modeling the entire song, and how this can be matched to the patterns of timing variability seen in our behavioral data. To be concrete, consider a song with two motifs and $Q_s = 4$ syllables and $Q_g = 4$ gaps per motif, *i.e.* ABCDABCD, where the last unique gap is the "inter-motif" gap between D and A. Across the entire song there are $2Q_s = 8$ syllables and $2Q_g - 1 = 7$ gaps, since the inter-motif gap does not get repeated, and we give P_s links in the chain to each syllable and P_g layers to each gap. The chain thus has a total of $Q_sP_s + Q_gP_g$ links and $Q_sP_s + Q_gP_g + 1$ layers, where activity in the first layer initiates the propagation down the chain. Activity in the last layer of the chain triggers activity to begin the second motif, that the song ends after the last syllable in the second motif. We assume that there are two sets of network parameters, one for syllable links and one for gap links. These lead to mean latencies of $\overline{T}_s$ for syllable and $\overline{T}_g$ for gap links, and an elasticity E_s^θ for syllables and E_g^θ for gaps. Thus, the average length of a song interval is $\overline{S} = P_s\overline{T}_s$ for syllables and $\overline{G} = P_g\overline{T}_g$ for gaps.

Our goal is to determine the relationship between the elasticities E_s^L and E_g^L of syllables and gaps with respect to changes in song length, and elasticities E_s^θ and E_g^θ of syllable and gap links with respect to changes in θ. First, we show that if one knows the ratio of elasticities with respect to song length, $r_E = E_s^L/E_g^L$, then one can uniquely determine each elasticity separately. To see this, consider a change in song length ΔL. By the definition of elasticity

$$\frac{\Delta S}{\overline{S}} = E_s^L \frac{\Delta L}{\overline{L}} \tag{4.3}$$

where ΔS is the change in length of that syllable. Using the similar expression for gaps and summing over all syllables and gaps, we can write

$$\Delta L = Q_s \; \Delta S + Q_g \; \Delta G = E_s^L \frac{\overline{S} \; \Delta L}{\overline{L}} + E_g^L \frac{\overline{G} \; \Delta L}{\overline{L}} \tag{4.4}$$

Now we divide out ΔL and use $E_s^L = r_E E_g^L$ to obtain

$$1 = r_E E_g^L \frac{\overline{S}}{\overline{L}} + E_g^L \frac{\overline{G}}{\overline{L}} = E_g^L \left(r_E \frac{\overline{S}}{\overline{L}} + \frac{\overline{G}}{\overline{L}} \right) \tag{4.5}$$

From that we solve for E_g^L, yielding $E_s^L = r_E E_g^L$. Thus, the ratio of syllable and gap song length elasticities uniquely determines both measurements, so we are now only concerned with the ratio.

If we divide both sides of equation 4.3 by the corresponding expressions for gaps we have

$$\frac{\Delta S / \overline{S}}{\Delta G / \overline{G}} = \frac{E_s^L}{E_g^L} \tag{4.6}$$

But by definition, $\Delta S = Q_s E_s^\theta \overline{T}_s$ and $\Delta \theta = E_s^\theta \overline{S} \; \Delta \theta$. Substituting the corresponding expressions syllables and for gaps into equation 4.6, we have

$$\frac{E_s^\theta}{E_g^\theta} = \frac{E_s^L}{E_g^L} \tag{4.7}$$

Thus, the basic syllable-gap pattern of song elasticity measured in the behavior can be directly modeled with network parameters with the sole requirement that chain-based elasticity and song-based elasticity ratios are equal.

4.2.2 Decomposing timing variability

So far, we have discussed global timing variations that are shared across the entire song. We now turn to two additional sources of timing variability revealed by our previous behavioral analysis (Fig. 4.5): First, after factoring out tempo song timing we have found "slow noise," variability that is independent for different subsyllabic notes, but correlated over 500-2000 msec for repeated renditions of the same note within a song (Glaze & Troyer, 2007). In fact, further investigation on a finer timescale revealed similar correlations among the same ∼10 msec song slices produced across motifs, but again, no special correlation among different song slices within the same note. Second, timing analysis has also uncovered faster noise that is specific to subsyllabic timescales and, in contrast to slow noise, does not repeat across motifs (Glaze & Troyer, 2006, see also Chapter 5 and Appendices D-E for a more direct analysis of fast noise). Since slow and fast noise sources are independent across subsyllabic intervals within a motif, we will refer to these collectively as independent noise sources.

To understand the differences among tempo variability and the independent noise sources, we distinguish between the timescales of chaining and neuromodulation. For our purposes, neuromodulation acts to change network parameters on a timescale that is slow relative to the production of song syllables. These changes are implemented as parameter changes that are constant across one song, but change from rendition to rendition. The chaining timescale is derived from the encoded burst sequence which drives song. This timescale corresponds to the time it takes to activate one to several links in the chain.

To model these three sources of timing variability we decompose variations in distance to threshold into three different components (Fig. 4.5B):

$$\theta_{\alpha ic} = \theta_{\alpha}^{tempo} + \theta_{\alpha i}^{slow} + \theta_{\alpha ic}^{fast} \tag{4.8}$$

The subscript α indexes the song, i indexes the neuron, and c indexes the motif (1 or 2) of the burst affected by the noise. Since $\Delta\theta_{\alpha}^{tempo}$ is shared across all neurons and lasts for all motifs, it only depends on the song α. $\Delta\theta_{\alpha i}^{slow}$ is slow noise that is sustained across motifs and hence does not depend on c. Finally fast noise is specific to the neuron, song and motif. To start, we assume that fast and slow noise are completely uncorrelated across all neurons in the chain; for example, this might be expected if neurons are anatomically distributed at random.

We model slow nosie as the interaction between the two contrasting timescales of chaining and neuromodulation (Fig. 4.5B): Because it is randomly distributed across chain layers, it is uncorrelated on subsyllabic timescales; however, because the noise source varies on a much slower timescale, each layer will sustain the influence of the noise over the production of an entire song bout. We accordingly modeled this variability by injecting noise that was specific to each chain neuron but lasted over an entire rendition. There are a number of ways in which this may occur in the system, but perhaps the most straightforward is if the neuromodulatory influence on tempo varies randomly across the network. Here, while two chain links may share a common modulatory source, they receive slightly different levels of excitation; the modulatory inputs change on a relatively slow timescale that spans multiple motifs, so the same link will nonetheless be significantly correlated with itself across seconds of song. The slow noise stands in contrast to faster noise that also accumulates over song but is completely uncorrelated across motifs (Fig. 4.5C). Fast noise also has a number of potential mechanisms, including background noise, or actively induced perturbations from LMAN (Kao et al., 2005; Ölveczky et al., 2005; Kao & Brainard, 2006).

For comparison with our behavioral analysis, we now compute the variance in interval length that can be attributed to each component of the noise. Our behavioral data indicate that independent noise alone yields an approximately $1-2\%$ CV in song interval lengths, with approximately twice as much fast noise as slow (Glaze & Troyer, 2006, Appendix E). For clarity, we will use language and notation specific to syllables, but the derivations are exactly the same for gaps. We consider a syllable with N neurons per layer that is driven by M links in the chain. We will assume that each neuron within that syllable has the same elasticity E^{θ}.

For tempo noise, all neurons receive the same change $\Delta\theta^{tempo}$, so each link will be scaled by a factor $E^{\theta}\overline{T}\ \Delta\theta^{tempo}$, and the tempo related variance of that interval will be equal to

$$\mathrm{Var}(\mathrm{S}^{\mathrm{tempo}}) = \mathrm{M}^2(\overline{\mathrm{T}}\mathrm{E}^{\theta})^2\mathrm{Var}(\theta^{\mathrm{tempo}}) \tag{4.9}$$

where $\mathrm{Var}(\theta^{\mathrm{tempo}})$ is the variance in distance to threshold induced by the tempo modulations across songs.

Both the slow and fast noise are assumed to be spread independently across

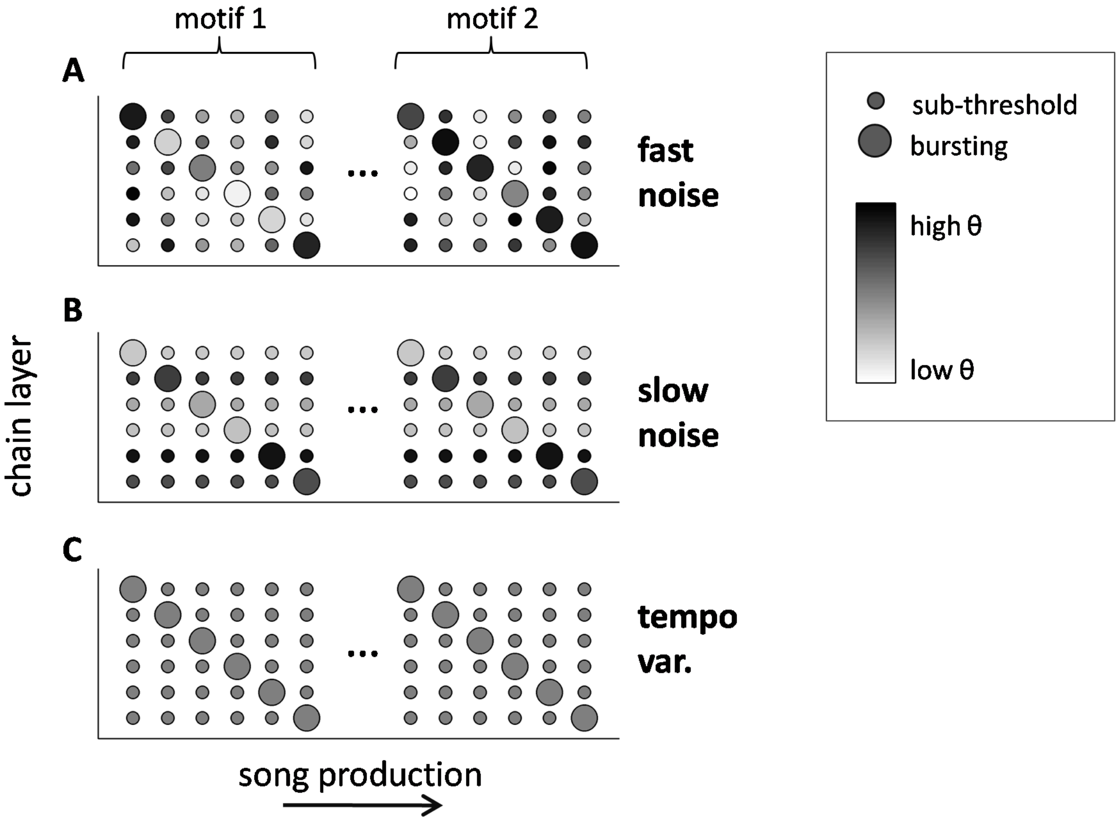

Figure 4.5: Schematic of three different types of timing variability and how they may be linked with chain propagation speed. Each circle represents a neural subpopulation dedicated to a chain layer, with larger circles representing a bursting population and smaller circles subthreshold neurons. Shading represents how far the subpopulation is from bursting threshold (θ), which in turn influences bursting latencies and thus chain speed (see Results). A, fast noise is uncorrelated across chain layers and varies on a relatively short timescale, yielding patterns of timing variability that are uncorrelated across motifs. B, slow noise is also uncorrelated across chain layers, but varies on a relatively long timescale that spans multiple motifs within the same song bout. Slow noise thus yields variability that repeats across motifs but is uncorrelated within each motif itself. C, tempo changes are correlated across all layers and also vary on a timescale that spans motifs. As a consequence, tempo induces timing variability shared across the chain.

all neurons in the chain. Since the latency for any layer in the chain is determined by the average of burst onset timing for neurons in that layer, the variance of link timing will be the variance of each neuron divided by the number neurons N. Since the length of an interval is the sum of the M link latencies composing that interval and each of these are independent,

$$\begin{aligned}\mathrm{Var}(\mathrm{S}^{\mathrm{fast}}) &= \frac{M}{N}(\overline{T}E^{\theta})^2\mathrm{Var}(\theta^{\mathrm{fast}}) \qquad (4.10)\\ \mathrm{Var}(\mathrm{S}^{\mathrm{slow}}) &= \frac{M}{N}(\overline{T}E^{\theta})^2\mathrm{Var}(\theta^{\mathrm{slow}}) \qquad (4.11)\end{aligned}$$

Note that expressions for the fast and slow variance are the same. The difference is that the slow variance is correlated for the same syllable sung in different motifs, whereas the fast variance is uncorrelated across motifs.

Comparing equation 4.9 with equations 4.10 and 4.11, we see that the global and independent variability factors have vastly different dependencies on the number of neurons in a layer and the number of links comprising an interval: global noise variance scales as M^2 whereas independent noise scales as M/N. For an interval consisting of $N = 100 - 200$ neurons per layer and $M = 15 - 50$ layers per interval, this requires anywhere between 40 and 100 times as much independent timing variance per burst onset as global tempo variance.

Can the magnitude of independent variability we find in the data be captured by the model? We investigated this question by analyzing the same chains used to derive the elasticity measurements (see Methods), only with 6 mV fluctuations in membrane potential; we have found that with a typical distance to threshold of 10 mV, noise much larger than this begins to lower spike probability below a level that is realistic for sparse, reliable, precisely timed bursts (Hahnloser et al., 2002). To provide a range of the upper limit on independent noise, we analyzed chains with both the weakest and strongest links from the parameters we have explored. In network simulations, these fluctuations yielded interval length variability of approximately $0.4-0.8\%$ CV depending on elasticity parameters. This is below the $1-2\%$ required by the behavioral data, and the noise leaves no room for variability linked with global tempo.

Thus, with the assumption that noise is uncorrelated across neurons in a layer, the variability required by the behavioral data is untenable. We thus injected noise that is correlated within layers (Fig. 4.6C); there are a number of potential mechanisms for such noise (see Discussion). In this case, all deviations in timing are also shared across a layer, removing the factor of $1/N$ in the variance expressions 4.10 and 4.11 that comes from averaging across independent within-layer latencies. Under this assumption, the correlated component of independent noise variance scales as M.

4.2.3 Full network simulation

We used the analysis detailed above as a guide in selecting parameter values to simulate typical zebra finch song timing patterns. We found several difficulties in precisely matching model timing to measured values obtained by averaging across birds. First, it was difficult to achieve the precise syllable-gap elasticity ratio while

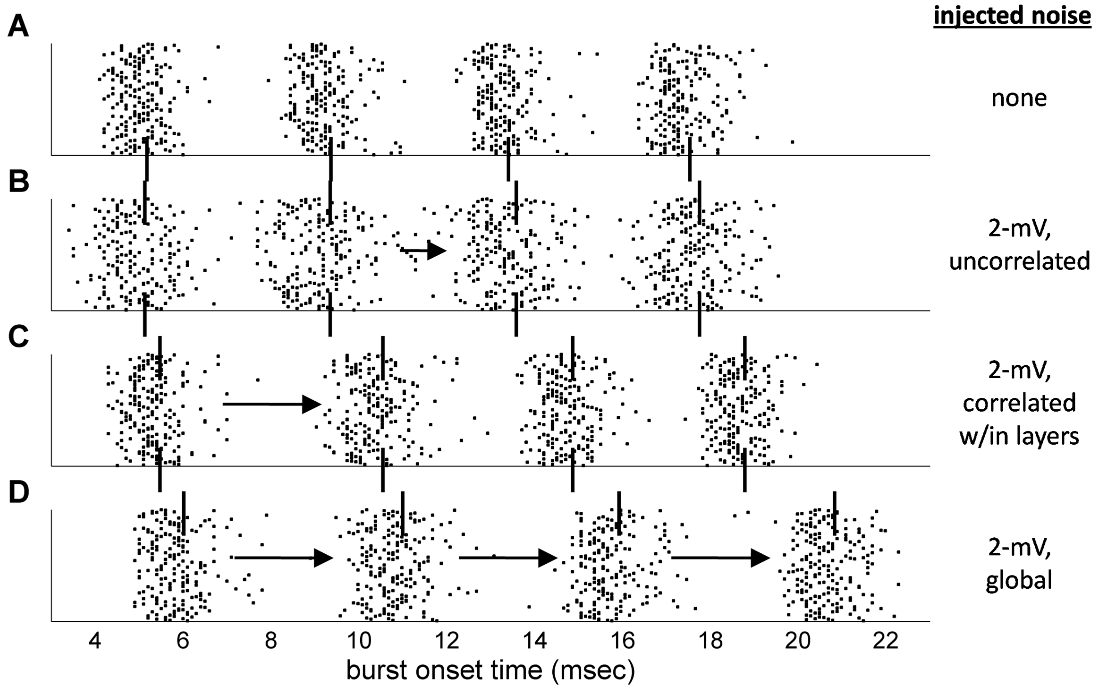

Figure 4.6: Burst onset times for different kinds of injected fluctuations in membrane potential. Times are aligned to the first layer of the sequence (not shown). Within each layer, neurons are spread out across the vertical axis for visual clarity. Tick marks indicate average onset time within each layer. A, Burst onset sequences in a simulation with no injected noise. Timing variability is due to variable connection density and latency that are fixed across all simulations. B-D, burst sequences from the slowest examples out of 50 trials per simulation type. Arrows represent significant deviations in averaged onset times that cause that burst sequence to be especially slow. B, 2-mV fluctuations in membrane potential are uncorrelated across all neurons. C, fluctuations of the same magnitude, but correlated within each layer, causing relatively larger shifts in timing deviations. D, the same fluctuations correlated across all neurons and layers, causing a consistent accumulation of variability across the sequence that is therefore the largest of all three scenarios.

preserving the relative amounts of timing variability across the three different noise sources. Although such a balance may be possible, this precision is outside the scope of this investigation given possible nonlinearities we have not accounted for in our simplified models. Second, while basic timing patterns tend to hold across birds, we have found considerable bird-to-bird variability in precise parameter averages; for example, syllable elasticity averages range from approximately $\sim 0.70 - 0.95$ by bird.

The aim of this simulation was thus to examine whether model parameters could be combined to yield timing variability within the range of behavioral data, rather than precisely matching behavioral averages. Specifically, we used equation 4.7 as a guide in choosing link strength parameters for syllables and gaps. Given the relationship between elasticity and both link strength and inhibitory gain (Figs. 4.3,4.4), there are several different ways in which network parameters may match behavioral data; here we present one possible set of network parameters (Fig. 4.2.3A,B): 20 synaptic inputs/neuron, 0.55 and 0.15 nA peak current per input for syllables and gaps respectively, 43 and 12 layers per syllable and gap, and inhibitory gain factor of 3. To match the range of rendition-to-rendition timing variability we find across birds, we injected 0.20 mV stdev variability global excitation, 1 mV slow and fast membrane potential fluctuations uncorrelated across chain layers, 1 mV fluctuations in slow modulation correlated within layers, and 2mV fluctuations in fast noise correlated within layers.

The simulation yielded a total song length CV of 1.01%. Average syllable length was 96.27 msec range $90.59 - 97.15$, while average syllable song elasticity

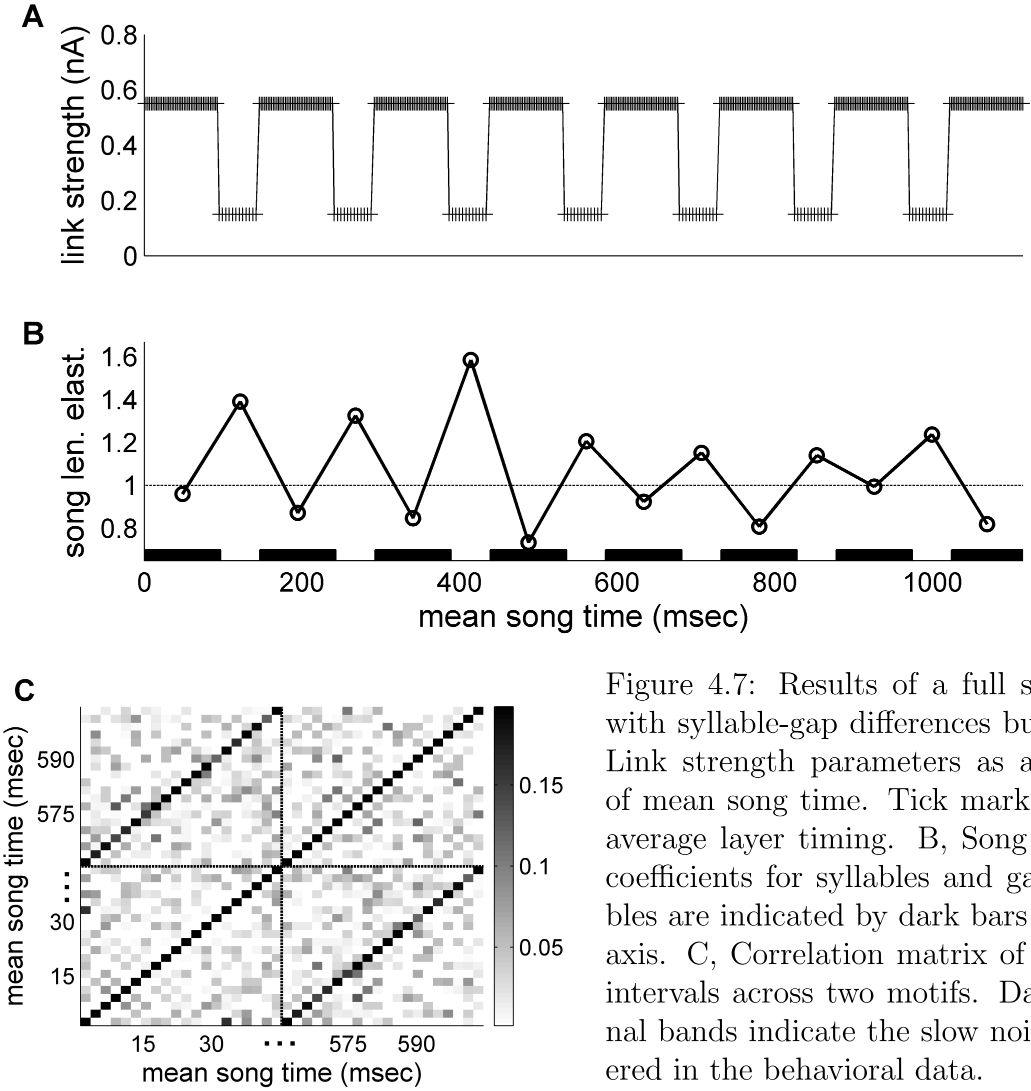

Figure 4.7: Results of a full simulation with syllable-gap differences built in. A, Link strength parameters as a function of mean song time. Tick marks indicate average layer timing. B, Song elasticity coefficients for syllables and gaps; syllables are indicated by dark bars on the x-axis. C, Correlation matrix of inter-link intervals across two motifs. Dark diagonal bands indicate the slow noise discovered in the behavioral data.

was 0.86, range $0.73 - 0.99$. Among gaps, mean length was 49.21 msec, range $49.11 - 49.31$, while mean song elasticity was 1.29, range $1.14 - 1.58$. Behavioral measurements had yielded a song length CV of 1.4%, respective mean lengths of 94.51 and 49.86 msec, and average song elasticities of 0.92 ± 0.03 and 1.17 ± 0.04 (Glaze & Troyer, 2006).

The simulations also generally matched slow noise measurements (Fig. 4.2.3C): regressing latencies with song length yields a residual covariance of 0.005 msec^2/msec (range $0.002 - 0.013$) shared between the same syllable-based links across motifs, vs. an average of 0.005msec^2/msec among 10-msec song slices in the behavior. As required by the behavioral data, fast noise yielded total variability in timing that was about twice the magnitude of what was associated with slow noise: syllable-based inter-link intervals had a standard deviation of 0.22 msec (range $0.19 - 0.24$), while taking the square root of cross-motif covariances yields approximately 0.10 msec.

Thus, simulations support the hypothesis that the timing variability we have measured in the behavioral data are consistent with a synfire chain model and may be linked with a range of basic network parameters such as synaptic strength and distance to threshold.

4.3 Discussion

Neurons in nuclei HVC and RA of songbirds produce bursts of spikes aligned to behavioral output with millisecond precision (Hahnloser et al., 2002; Yu & Margoliash, 1996; Chi & Margoliash, 2001), suggesting that synfire-like mechanisms may play an important role in the motor code for song (Fee et al., 2004). Fine-grained measurements of temporal variability in song acoustics have provided two basic sets of constraints on potential models of song production. First, the motor code cannot rely on a single uniform chain to pace song behavior (Glaze & Troyer, 2006, Chapter 2). Second, rendition-to-rendition variability in song timing reveals a striking contrast of timescales that suggests both temporally precise chain activity and broad neuromodulatory influences (Glaze & Troyer, 2007, Chapter 3). Here, we investigated how chain-like networks can produce patterns of elasticity and residual noise consistent with the behavioral data.

In an initial set of simulations, we investigated how the propagation speed along a synfire chain is influenced by perturbations in several system parameters. Over a range of plausible EPSP shapes, fluctuations in parameters determining neural excitability, *i.e.* resting potential, will induce proportionally greater changes in the timing of chains that have weaker links. Our simulations also indicate that greater inhibitory feedback gain can also suppress elasticity, suggesting that neural subpopulations with especially high timing variability may also receive fewer inputs from inhibitory interneurons. We next modeled the influence of slow and fast noise on chain timing by injecting varying amounts of variability in membrane potential on two different timescales. This analysis indicated that the degree of independent variability found in the behavioral data may require noise that is shared across neurons within each chain layer. However, with a modest degree of within-layer correlation, the model was able to capture the basic timing patterns we find in adult zebra finch song.

Syllable-based representations

In the simulations, neurons coding for the syllable portions of the song are presumed to receive stronger synaptic input and hence have shorter latencies to threshold. If song tempo is modulated by changes in overall levels of excitability, our simulations indicate that syllables will react to such changes relatively less than gaps, which is consistent with our behavioral measurements of temporal elasticity. Weaker connections between neurons coding for gaps is consistent with the fact that electrical stimulation and bright flashes of light tend to interrupt singing at syllable boundaries (Cynx, 1990; Franz & Goller, 2002). For example, if neurons involved in weaker links are less likely to reach threshold, then those chain segments will not only be more elastic, they will also be more likely to halt the entire song if bombarded with enough noise to disrupt the synchrony that is crucial to propagation (Diesmann et al., 1999).

Hence, we predict that propagation should be most subject to interruption in portions of the chain that show the greatest temporal elasticity with respect to the modulation of excitatory parameters.

Noise correlations

The simulations were unable to capture the magnitude of fast and slow noise measured in the behavior given the simplest assumption that chain layers are randomly distributed, and thus receive uncorrelated noise inputs. How could correlated noise inputs arise in the system? The simplest mechanistic explanation might involve "chronotopic" chain organization in which neurons encoding for the same time point are nearby and thus share a common source of variability in the spatial distribution of neuromodulator.

Chain-layer correlations may be also explained without such organization. For example, fast noise could be due to global fluctuations in excitation; here, links may share variability in the noise source, but the influence is expressed in a temporally precise pattern to the extent that at any given time song is driven by a single link. LMAN has been demonstrated to induce variability in song acoustics via projections to RA (Kao et al., 2005; Ölveczky et al., 2005; Kao & Brainard, 2006) and may thus be well positioned for such a role. However, the mechanistic basis for slow noise is more difficult to explain unless this same global pattern repeats across motifs. This could be possible if, for example, (1) inputs from LMAN drive this global noise pattern and (2) LMAN activity is itself correlated across motifs. Similar scenarios might be at play across inhibitory inputs from HVC interneurons as well. Overall, such mechanisms are plausible in that they require correlated connections from another nucleus that may arise via spike-time dependent plasticity during development. Although such explanations shift the onus of slow noise to another neural population, larger fluctuations in membrane potential may be more tenable in networks that do not show the same precise bursting patterns found over the HVC-RA pathway.

Beyond synfire chains

It is important to note that the model investigated here consists of a single feedforward pathway and positions the HVC network as the central timing mechanism that drives the accumulation of variability. However, while synfire activity may play a prominent role in timing, variability may be potentially derived from a number of different sources, and there is evidence for multiple functional loops in the song circuit which also have a strong influence (Vates et al., 1997; Ashmore et al., 2005). LMAN itself is the output nucleus of basal ganglia circuit that receives inputs from HVC, and as already discussed, may be crucial in certain types of song variability. Indeed, HVC activity itself is coordinated across hemispheres, particularly around syllable onsets (Schmidt, 2003), and forebrain-brainstem loops are well positioned to carry the signals for such coordination. Without a central timing mechanism, the accumulation of timing variability requires that noise from either the basal ganglia or brainstem-forebrain circuits at least propagate throughout the song system. Otherwise, activity across different nuclei would progressively become less synchronized as a song bout proceeded.

The ability to manipulate the behavioral state of the bird and the discrete nature of song system anatomy may thus provide important tools for addressing this circuit complexity. For example, female-directed songs are faster than undirected singing (Sossinka & Bohner, 1980). Cooper & Goller (2006) have shown that these speed increases are disproportionately reflected in the lengths of expirations accompanying syllables. The greater change in expirations between directed and undirected song contrasts with the smaller changes in syllable lengths during small tempo fluctuations within each condition (Glaze & Troyer, 2006). This suggests that the increase in song tempo for directed song is controlled by different circuit mechanisms than those leading to tempo variability within each behavioral condition. Kao & Brainard (2006) have shown that the directed/undirected tempo difference is eliminated by LMAN lesions.

Cortical Chains

Synfire chains have been proposed as a general model of computation in the cortex (Bienenstock, 1995). An important advantage to computations with temporal patterns is that functional interactions depend on pattern synchronization as well as synaptic weights. For example, compositionality can be achieved by synchronizing the activity within more primitive chains via weak cross-connections (Arnoldi & Brauer, 1996; Abeles et al., 1994). Our elasticity results demonstrate that global changes in a single network parameter can have differing effects on propagation speed depending on the properties intrinsic to different chains (such as EPSP size). It follows that the ability of two chains to synchronize will depend on levels of tonic input (Sommer & Wennekers, 2005) and/or the modulatory state of the network. Since multiple network parameters influence propagation, synchronization may require specific combinations of excitatory, inhibitory, and modulatory inputs to the network. This suggests that propagation speed may be an important mechanism for the modulatory control of functional connectivity within neural circuits.

Given the highly variable activity within cortical circuits, the study of synfire chains has largely centered on the ability to detect of precise sequences spikes separated by fixed intervals (reviewed in Gerstein, 2004). Timing variation is assumed to take the form of random “jitter” in the interval between any two spikes. Our behavioral data from the songbird suggests that interval variations may in fact be highly correlated. If so, these correlations could provide an alternative statistical argument supporting the existence of cortical synfire chains and the models presented here would make specific predictions tying these correlations to underlying system parameters.

Repeated patterns of precise spiking in cortical circuits have also been demonstrated in vitro (Beggs & Plenz, 2003, 2004; Ikegaya et al., 2004), raising the possibility that mechanistic models of spike propagation may be tested via pharmacological manipulation. One challenge for this approach is that large perturbations in the system may destabilize individual sequences, limiting these studies to making statistical statements about the presence or absence of particular sequences or subsequences. In contrast, temporal measurements are highly sensitive, and can yield a rich set of structural data concerning the circuits that drive sequential spiking without pushing the system outside of physiological ranges.

More generally we have shown that elasticity and noise structure may be powerful conceptual tools linking temporal measurements to biophysical mechanisms that generate precise patterns of spiking activity.

Chapter 5

Development of Sequence Structure and Timing Variability in Zebra Finch Song

Sequence learning is one of the touchstone questions in neuroscience (*e.g.* Lashley, 1951; Hikosaka et al., 2002; Keele et al., 2003; Rhodes et al., 2004), and the zebra finch song has served as an excellent model system for understanding sequence learning and production. Songs are learned, highly stereotyped, and have a well-defined temporal structure spanning multiple time scales. In the zebra finch, song learning can be roughly divided into two overlapping processes (Immelmann, 1969; Marler, 1970; Konishi, 1985; Brainard & Doupe, 2000): "sensory acquisition," in which a young bird ∼20-65 days post-hatch (dph) is exposed to the song of one or more tutors and forms an auditory template; and "sensorimotor learning," in which the juvenile ∼35-90 dph learns to produce song based on that template. Once learned, songs consist of several repeats of 500-1000 msec long "motifs;" motifs consists of a stereotyped sequence of 3-7 "syllables," 50-250 msec long vocalizations separated by silence; many syllables can be further divided into 30-70 msec long "notes."

Recent physiological studies have suggested that song is encoded over premotor nuclei HVC and RA on a single, 5-10 msec timescale; in this proposal, HVC drives RA patterns with sparse bursts that encode corresponding points in song independently of all other points (Hahnloser et al., 2002; Fee et al., 2004; Leonardo

& Fee, 2005). It has also been proposed that the HVC timing mechanism is present throughout development (Fiete et al., 2004); in this model, one of the chief tasks for the system is to map the clock to downstream neurons by modifying HVC-RA synapses. This proposal is corroborated by behavioral evidence for “in-situ” syllable learning, in which syllables gradually emerge from a sequential pattern of vocalizations whose order never changes (Tchernichovski et al., 2001; Liu et al., 2003).

On the other hand, there is also evidence for syllable-based representations in the adult code (Cynx, 1990; Franz & Goller, 2002; Schmidt, 2003; Solis & Perkel, 2005; Glaze & Troyer, 2006; Cooper & Goller, 2006), and it remains an open question as to whether there are in fact distinct neural processes associated with learning syllable sequences that is different from learning individual syllables (Troyer & Doupe, 2000b). In fact, at least several studies have suggested significant syllable-sequence variability in juvenile song that is sensitive to lesions of LMAN, the output nucleus of a basal ganglia circuit involved in song learning (Scharff & Nottebohn, 1991; Bottjer et al., 1984; Ölveczky et al., 2005).

To investigate these models of song learning, we analyzed the development of temporal structure in zebra finch song from late plasticity through ∼1 year of age. We had previously investigated rendition-to-rendition variability in adult song timing, and found several patterns that suggested how song representations are organized (Glaze & Troyer, 2006, 2007): First, song intervals share a common source of length variance that we term “global variability,” while syllables are proportionally less sensitive to this variability than gaps, *i.e.* they are less “elastic.” Second, after factoring out tempo, we found variability on subsyllabic timescales that was uncorrelated across different notes in the same syllable, but repeated seconds later with multiple renditions of the same motif in the same song bout; we term this “repeated variability.” Finally, timing analysis (Appendices D and E) has suggested that this repeated variability may be distinguished from what we term “independent variability” in this manuscript (not to be confused with independent variability in Chapter 4), timing variations that are unique to individual song notes and not repeated across multiple motif renditions in the same bout. We have synthesized these findings into a chain-based model of song production in which song is driven by a single chain of activity on a fine timescale, but with physiological parameters (such as synaptic strength) that do correspond to the acoustic hierarchy by distinguishing syllable-based from gap-based chain segments.

Here, we examine the development of these timing patterns in order to probe how the motor code may develop. We find increases in song tempo that occur almost entirely during the gaps of silence between syllables. Most of the tempo increase occurs 65-90 dph, and on a gap-by-gap basis increased speed is linked with increases in the reliability of corresponding syllable transitions, as well as decreases in local timing variability. We also find interesting changes in timing variability itself: while the magnitude of local variability decreases through 1 year of age, we find no significant changes in rendition-to-rendition global timing variability. Furthermore, we did not find any changes in elasticity patterns as far back as 85-90 dph, nor changes in the timescale of local variability, a parameter which had previously been linked with proposed chaining mechanisms over the premotor pathway in the adult song system.

Overall, the data suggest a phase of late development in which song production becomes faster and more automated, while the magnitude of fine-grained timing variability decreases two- to threefold. Such as process of motor consolidation has been previously suggested for zebra finch song (*e.g.* Brainard & Doupe, 2001); the current study provides a rich set of behavioral constraints for models of such consolidation.

5.1 Materials and Methods

Analysis was based on the songs from 7 male birds, recorded between 65 and 375 days-post hatch (dph). All care and housing was approved by the institutional animal care and use committee at the University of Maryland, College Park. All analysis was performed in Matlab (Mathworks, Natick, MA), and all template matching and dynamic time warping algorithms were written as C-MEX routines.

5.1.1 Song collection

Each bird was raised in a breeding cage with his parents and clutch mates until 25-30 dph. Juveniles were then housed with a tutor (either the father or another adult male) in a sound isolation chamber (Industrial Acoustics, Bronx, NY). Birds were housed individually in approximately 18x36x31 cm cages separated by 18 cm. Birds were paired with tutors in the same chamber until 150 dph, after which they were released back into an aviary with other birds. Subsequent adult (365+ dph) recordings were gathered by returning birds to a studio, paired with one other adult.

Studios were equipped with two directional microphones (Pro 45; Audio-Technica, Stow, Ohio). Signals were digitized at 24,414.1 Hz, and ongoing data were selected using a circular buffer and a sliding window amplitude algorithm. "Sound clips" separated by <200 msec were included in the same "recording" and clip onset times were indicated by filling the gaps between clips with zeros.

For each bird, we gathered song during 4 age periods: 65-70 dph, 85-90 dph, 125-135 dph, and 365-375 dph. We will call a collection of songs from a given bird and a given age range a "bird-by-age sample." It has already been shown that during the late juvenile period and young adulthood, songs continue to develop in subtle ways (e. g. Brainard & Doupe, 2001). Thus, all song analysis treated each time period independently, *e.g.* song matching was based on templates that were constructed for that period only.

5.1.2 Song selection and template matching

Recordings were attributed to a target bird if the total power was greatest for the microphone directed toward the side of the recording chamber where that bird was stationed. For the template matching, recordings were initially analyzed using the log-amplitude of the fast-Fourier transform (FFT) with a 256-point (10.49 msec) window moved forward in 128-point steps. Frequency bins outside the 0.5-8.6 kHz range were excluded from all subsequent analysis because song structure is less reliable at the highest and lowest frequencies. An automated template matching algorithm (detailed below) was used to identify individual song syllables.

Each recording was composed of a series of "clips," periods of sound separated by at least 10 msec of silence. To determine if the sound in these clips matched syllables in the bird's song, syllable templates were formed by aligning and averaging 4-5 manually chosen clips corresponding to each syllable in the repertoire for that bird-by-age sample; exemplars were time-aligned by finding the peaks in a standard cross-correlation of syllable spectrograms. Each clip was then matched against each template using a sliding algorithm (Glaze & Troyer, 2007). For each template and each time point (t) in the clip, a match score ($c(t)$) was computed as the reciprocal of the mean squared difference between template and song log-amplitudes at matched time-frequency points:

$$c(t) = n \times m / \sum_{i=1}^{n} \sum_{j=1}^{m} (s(i+t, j) - s'(i, j))^2$$

where s is the song spectrogram, s' is the template spectrogram, n is the number of time bins in the template, m is the number of frequency bins, i indexes time and j indexes frequency.

The peak match score was determined as the maximum of $c(t)$ over time alignments t. Matches were only considered valid if the peak match score exceeded an initial fixed threshold of 0.5 (manually chosen based on visual inspection), and if the onset and offset for the clip and template differed by 20 msec or less. If a clip had multiple syllable matches, the template with the highest match score was chosen. This yielded a median of $14,720$, range $1686 - 120,873$ matches per syllable in each bird-by-age sample.

Template matching thresholds

After all syllables were identified, template-matching thresholds were recalculated for each syllable, based on a simplified minimum error-rate classification scheme (Duda et al., 2000): Match scores were grouped into 10 0.1-long bins from the 0.5 minimum through a value of 1.5+. A random sample of 20 syllables per score-bin were then gathered. If fewer than 20 matches existed in that bin, then it was grouped with the next highest bin; this was necessary in a few syllables which had relatively few low match scores due to a particularly high average match.

Within each bin, syllables were randomly sorted, and each syllable was manually judged as being either correctly classified (a "hit") or a false positive. The total number hits and false positives for that bin was estimated by multiplying the correct-hit probability calculated from the sample of 20 by the total number of syllable matches in that bin. An optimal threshold was then set at a value that minimized the estimated combined number of false hits and false positives for the entire sample (thresholds set by this method necessarily occur at bin boundaries). Across syllables, this yielded a median 1.8% false positive rate, range $0 - 34.2\%$, median 0.9% false negative, $0 - 34.3\%$. Approximately 86 and 87% of syllables respectively had false positive and rejection rates $< 10\%$. The highest error rates occurred for brief, noisy syllables that resemble introductory notes; these templates frequently yielded modestly high match scores even for vocalization not directly involved in song, such as tets (Zann, 1996). We found no systematic changes in these rates by age.

5.1.3 Transition probability calculations

We defined the "forward transition probability" between syllables X and Y as the total number of valid X-Y transitions divided by the total number of valid transitions from X to another syllable. Following previous analysis of sequence variability (e. g. Foster & Bottjer, 2001; Kao & Brainard, 2006), we excluded song stops (*i.e.* transitions to silence) from all probabilities. We considered a transition between two syllables to be valid if (1) both syllables had a match score above optimal threshold, and (2) the time between the offset of one syllable and the onset of the next was no more than 100 msec. Based on a previous analysis of gap-length distributions, a gap longer than 100 msec was considered an outlier and might be more related to song stops rather than a typical transition between syllables. Increasing the minimum distance criterion to 200 msec had no appreciable influence on transition probabilities. While varying match threshold yielded slightly different probabilities, we found no systematic bias that would influence our basic conclusions.

Motif and sequence definition

Song motifs were defined as the most frequent syllable sequence found across recordings. Per bird-by-age sample, $1972 - 14,057$ recordings were made (median 7132); of these $12.6 - 80.4\%$ (median 41.9%) contained at least one motif, yielding $439-9927$ recordings with song. All of the bird-by-age samples with $< 33\%$ matched recordings came from birds in the first 3 age groups (between the ages of 65 and 135 dph) that produced song sequences that were either especially variable or frequently interrupted.

Our previous studies showed that song elements repeated across multiple motifs in a song bout have especially strong correlations in timing variation. One part of the analysis focused on these repeated elements, and here we analyzed song sequences with two motifs. We omitted inter-motif gaps from all analysis because we did not restrict their lengths. We also omitted two birds from the two-motif analysis due our inability to find a sample of > 300 across all age periods that contained two or more motifs (mostly due to variable syllable sequencing and frequent interruptions of motifs at 65 dph). Across bird-by-age samples in the remaining 5 birds, this yielded a total of $439 - 7674$ recordings with at least 2 motifs.

As a further screen to obtain uncorrupted data, we identified song intervals whose length was more than 5 standard deviations from the mean of the length distribution for that interval. Intervals included syllables and gaps, and in one analysis sub-syllabic notes (see below). The vast majority of these outliers can be attributed to acoustic interference from other noises in the case (*e.g.* wing flaps, vocalization from the tutor). We omitted sequences having any such outliers. Across 1-motif syllable sequences, we omitted a median 3.1% of each bird-by-age sample for this reason, range $0.4 - 11.4\%$. Across 2-motif syllable and note sequences, we omitted a median 10.1%, range $0.7 - 22.8\%$, and median 12.1%, range $0.6 - 25.3\%$ respectively. Two-motif samples had much higher rejection rates due to a greater number of intervals that could lead to the omission of any given sequence. Without omitting outliers, we have found no significant change in basic statistics such as mean length and change in standard deviation across age periods; however, the accuracy of our statistical modeling is sensitive to the presence of these outliers.

Across birds-by-age samples, the final sample sizes had a median 3423, range 424-9876 for single-motif syllable sequences; a median 1359, range 414-7620 among two-motif syllable sequences; and a median 1330, range 392-7526 among two-motif note sequences.

Across all 7 birds, the single-motif analysis sample included a total of 33 unique syllables and 26 gaps of silence. Across the 5 birds investigated for repeated motifs, analysis included 19 syllables and 18 gaps. Some of these syllables can be divided into clearly defined sub-syllabic "notes" based on sudden changes in spectral structure (see below). Of our two-motif sample, the 10 syllables that could be reliably divided into more than one note yielded a total of 35 identified notes.

5.1.4 Song timing calculations

After syllables and syllable sequences were identified, timing variability was analyzed with a more fine-grained algorithm. Timing analysis was restricted to syllables found within identified motifs; thus, all references to syllables and gaps hereafter refer to data samples that include full song motifs. First, fine-grained spectrograms were calculated for all syllables using FFTs with a 128-point window slid forward in 4-point steps, yielding 0.16 msec time bins. Although previous research (Glaze & Troyer, 2006, 2007) had been based on log-amplitudes, for this research we used raw amplitudes, which we have found to be more reliable for timing measurements. The resulting spectrograms were then smoothed in time with a 64-point Gaussian window that had a 25.6-point ($\sim$5 msec) standard deviation. Time derivative spectrograms (TDSs), calculated as differences in amplitude in time-adjacent bins, were computed and used in the rest of the analysis.

Syllable templates were reconstructed for a fine-grained analysis in a two step process; distinct templates were constructed for each bird-by-age sample. First, we used the same sliding algorithm used previously, but now applied to the fine-grained TDSs. For each syllable type, we selected a random sample of 200 syllable TDSs. The syllable with the highest (coarse-grained) template-matching score was selected to act as an initial template, and each of the 200 TDSs were aligned to this template using our sliding algorithm. We updated the template by re-averaging the aligned TDSs, and repeated the alignment and averaging process once more. In the second step, we repeated this strategy but aligned TDSs to the template using a dynamic time-warping (DTW) algorithm (Anderson et al., 1996; Glaze & Troyer, 2006, 2007, see Appendix C). The final template was based on the average of these warped syllable TDSs. This final step in the template construction proved to be crucial for the younger developmental periods (65-90 dph), in which timing within syllables is quite variable; without the last step, many of the resulting templates exhibited degraded resolution at younger ages. For timing measurements, each syllable within a given bird-by-age sample was then mapped to its corresponding template via the same DTW algorithm.

Measuring the variability of interval lengths requires that song features be identified in a consistent manner across syllable renditions. To accomplish this, we identified feature-based markers within the TDS templates, and then used DTW to determine the precise times at which these features were observed in individual recordings of those song syllables. Markers were first identified manually in templates extracted from oldest age periods (365-375 dph). Syllable onsets/offsets were generally identified as salient time-derivative peaks/troughs in the template TDS. These correspond to "inflection points" in the rise and fall of energy at the beginning and ends of syllables, and result in syllables that are slightly shorter than would be expected from measurements based on exceeding a threshold of power. However, the differences are generally only several msec and we find that our method gives much more reliable measurements. Boundaries between notes within syllables were determined by sudden changes in spectral properties. To measure the lengths of song intervals across development, we mapped syllable templates from younger ages onto the adult templates using DTW. We then used the resulting mapping to determine the timing of the onset/offset markers and note boundaries within the templates from younger ages.

5.1.5 Timing variability analysis

Throughout this analysis we will focus on two basic kinds of timing: tempo-based variability that is shared by all song intervals, which we term "global variability," and timing variability that is independent of tempo and uncorrelated across song intervals of different identities, which we term "local variability." To separate out these two components, we used a factor analysis (FA) model (Bishop, 2006; Tresch et al., 2006) for each bird-by-age sample. FA is similar to principal components analysis in that it factors a covariance matrix with a limited set of basis vectors. We present the details of this algorithm in Appendix D.

In brief, we model global variability with a latent variable z and allow each interval to have a unique sensitivity to tempo, which we term the "global weight" and denote as $\mathbf{W}_{\mathrm{i}}$ for ith interval of the song motif (estimated in units of msec). While we are also interested in measuring independent variability that accumulates over song motifs, we have also observed jitter in feature times that induces significant negative covariances between adjacent intervals that share a feature. In order to separate jitter from accumulated variability, we explicitly modeled jitter with latent factor $\mathbf{u}$. There is jitter in both the onsets and offsets of song intervals, which we denote as $\mathbf{u}_{\mathrm{i}}^{+}$ and $\mathbf{u}_{\mathrm{i}}^{-}$ respectively. Importantly, the offset jitter of a given interval is the same as the onset jitter for the next, *i.e.* $\mathbf{u}_{\mathrm{i}}^{-} = \mathbf{u}_{\mathrm{i+1}}^{+}$.

We also sought to measure repeated local variability (*i.e.* same-id covariances from Chapters 2 and 3), which we model with latent factor $\mathbf{v}$; $\mathbf{v}$ has M dimensions, where M is the number of unique intervals. For example, in each sample of the sequence "ABCABC," $\mathbf{v}$ would have 3 entries corresponding to each unique element, and each value of $\mathbf{v}$ would be shared by two repeated elements.

With jitter and repeated variability now included in the FA model, we write interval i in sequence α as

$$\mathbf{x}_{\alpha\mathrm{i}} = \bar{\mathbf{x}}_{\mathrm{i}} + \mathbf{W}_{\mathrm{i}}\mathrm{z}_{\alpha} + \mathbf{v}_{\alpha\mathrm{k}} + \eta_{\alpha\mathrm{i}} + \mathbf{u}_{\alpha\mathrm{i}}^{+} - \mathbf{u}_{\alpha\mathrm{i+1}}^{-} \tag{5.1}$$

where $\bar{\mathbf{x}}_{i}$ is average interval length and k indexes the repeated variability source linked with that interval. Tempo-based global changes are indicated by $\mathbf{W}_{\mathrm{i}}\mathbf{z}_{\alpha}$, while the total amount of local variability is represented by the sum of repeated and independent (*i.e.* non-repeated) components, $\mathbf{v}_{\alpha\mathrm{k}} + \eta_{\alpha\mathrm{i}}$. In one part of the analysis we separate out the repeated and independent components, which are $\mathbf{v}_{\alpha\mathrm{k}}$ and $\eta_{\alpha i}$ respectively.

Parameters were fit using an expectation-maximization (EM) algorithm (Bishop, 2006, see Appendix D for details). The algorithm fits parameters to the measured covariance matrix $\mathbf{C}$, so we now write the parameters in matrix form. We introduce a matrix $\mathbf{D}$ that performs the differencing operation through multiplication (see Appendix D), so $\mathbf{Du}_{\alpha}$ represents a vector of deviations due to timing jitter in sequence α; these deviations do not accumulate so they are not expected to influence total motif length (with the exception of the first and last features). For repeated variability we include a NxM fixed weight matrix $\mathbf{Q}$ where N is the total number of intervals. For variability source k, $\mathbf{Q}_{\mathrm{ik}} = 1$ if element i shares that source and $\mathbf{Q}_{\mathrm{ik}} = 0$ otherwise. Using $\mathbf{\Sigma}$ to denote a diagonal matrix of jitter variance, $\boldsymbol{\Phi}$ to denote the diagonal covariance matrix of repeated noise, and $\boldsymbol{\Psi}$ to denote the diagonal matrix of non-repeated noise, the total covariance can now be written as

$$\mathbf{C} = \mathbf{W}\mathbf{W}^{\mathrm{T}} + \mathbf{Q}\boldsymbol{\Phi}\mathbf{Q}^{\mathrm{T}} + \boldsymbol{\Psi} + \mathbf{D}\mathbf{\Sigma}\mathbf{D}^{\mathrm{T}} \tag{5.2}$$

Interval variances due to tempo, local variability, and repeated and independent variability can be found in the diagonals of $\mathbf{W}\mathbf{W}^{\mathrm{T}}$, $\mathbf{Q}\boldsymbol{\Phi}\mathbf{Q}^{\mathrm{T}} + \boldsymbol{\Psi}$, $\mathbf{Q}\boldsymbol{\Phi}\mathbf{Q}^{\mathrm{T}}$, and $\boldsymbol{\Psi}$ respectively.

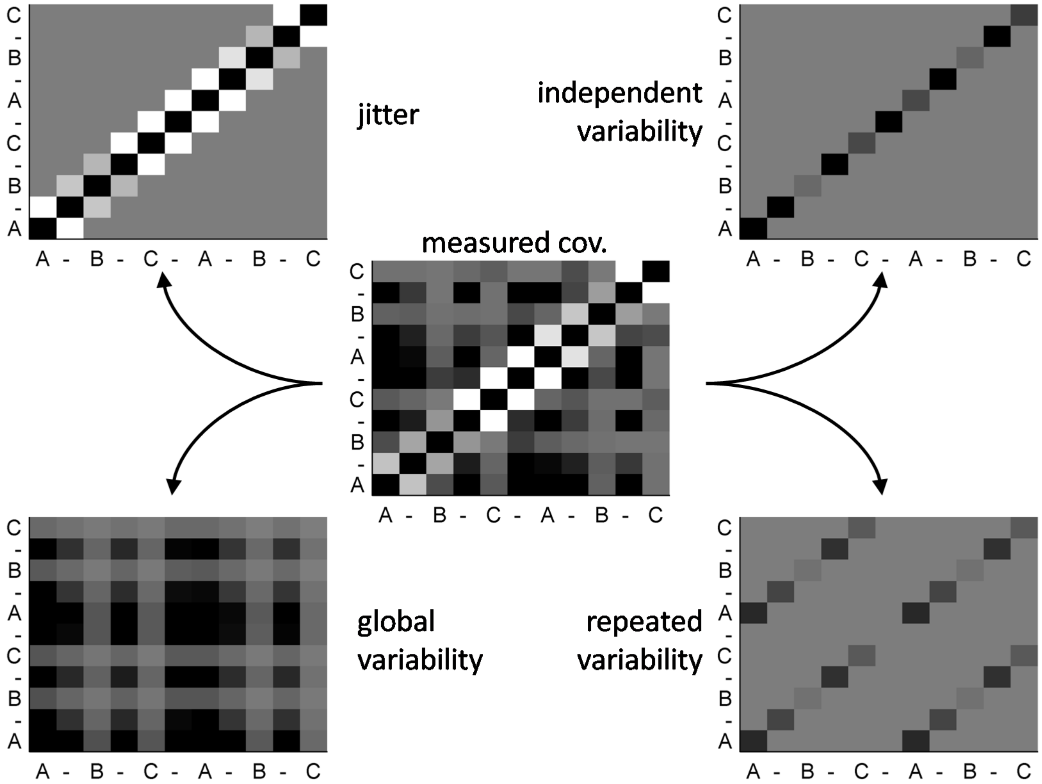

Figure 5.1: Example of the factor analysis model applied to one bird at 65 dph. See Methods and Materials and Appendix D for details of the algorithm. Center, measured covariance matrix of all song intervals, with darker shades representing stronger covariances and whiter shades more negative covariances. Other matrices, results of the factor analysis model for each of the factors discussed in Materials and Methods.

For the EM algorithm, 100 randomly chosen initial conditions were used, and the result linked with the highest log-probability was used. Figure 5.1 has an example of the FA algorithm from one bird at 65 dph.

By age, mean estimated jitter was 2.640 ± 0.238, 1.818 ± 0.159, 1.538 ± 0.156, and 1.271 ± 0.128 msec at 65-70, 85-90, 125-135 and 365+ dph respectively. The factor accounted for $65.9 \pm 2.7\%$, $63.6 \pm 2.7\%$, $61.4 \pm 2.8\%$ and $55.2 \pm 3.1\%$ of modeled variance across respective age groups. Because we have no means for determining how much of this jitter is measurement error vs. true timing variability, we discard this component for the rest of the analysis.

In order to compare global and local timing variability, we took the square root of each element of $\mathbf{\Psi}$ and $\mathbf{\Phi}$ to convert those measurements into the same units as the global weights (msec). Throughout most of the analysis we focused on single-motif sequences and compared global with local variability. Here, we did not distinguish repeated from independent variability; however, in 3 birds one or two syllables were repeated within a stereotyped motif, so we calculated the total amount of uncorrelated variability by taking the square root of the sum of both variances. In one part of the analysis we do compare repeated with independent variability; here we applied the FA model to the two-motif sequences described above.

Most of the analysis compares distributions of FA parameters, which we performed with nonparametric tests. For age-dependent changes in a given parameter we evaluated pairwise differences across adjacent ages ranges using the Wilcoxon signed-rank test. We also compared distributions across different interval and parameter groups, which we evaluated using the Wilcoxon rank-sum test. For relationships among parameters themselves, we used Spearman's correlation.

5.2 Results

We analyzed the development of temporal structure in zebra finch song from 7 males 65-375 dph. Individual song production is organized into song bouts, which generally consists of several motifs, defined as stereotyped sequences of syllables.

Syllables, distinct vocalizations separated by gaps of silence, can in turn be divided into notes, segments with distinct acoustic structure. For each bird, we gathered all recordings with at least one motif from 4 periods: 65-70 dph, 85-90 dph, 125-135 dph and 365-375 dph; for convenience, we will occasionally denote each period using the first day only (*e.g.* "65 dph" for the first period). Final bird-by-age samples contained 414-9876 sequences and a total of 33 unique syllables and 26 gaps of silence. Across the 5 birds investigated for repeated motifs, analysis included 19 syllables and 18 gaps; across the 10 syllables that could be reliably divided into more than one note we analyzed a total of 35 unique notes.

We had previously investigated rendition-to-rendition variability in adult song timing, and found several patterns that suggested how song representations are organized (Glaze & Troyer, 2006, 2007): First, song intervals share a common source of length variance that we term "global variability"; syllables are proportionally less sensitive to this variability than gaps, *i.e.* they are less elastic. Second, after factoring out tempo, we found variability on subsyllabic timescales that was uncorrelated across different notes in the same syllable, but repeated seconds later with multiple renditions of the same motif in the same song bout; we term this "repeated variability." Finally subsequent analysis and a modeling study both suggested that this repeated variability may be distinguished from what we term "independent variability," timing variations that are unique to individual song notes and not repeated across multiple motif renditions in the same bout.

Here, we use a factor analysis model (see Methods) to tease apart these different sources of timing variability and examine changes in each parameter from

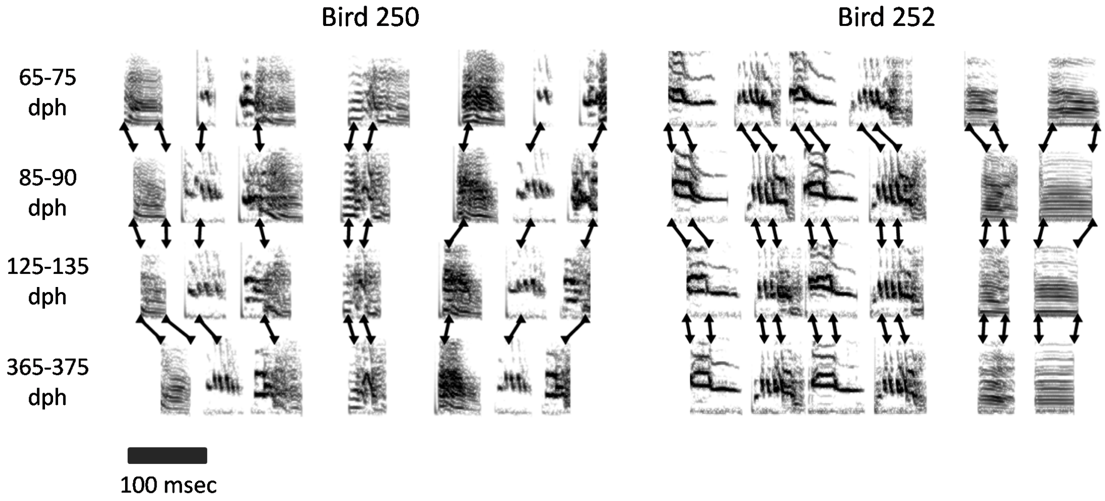

Figure 5.2: Examples of tempo changes from two birds, 65-365 dph. Each song spectrogram is from an actual recording with syllable and gaps lengths that are closest to the mean of that bird-by-age period (distance calculated as sum of squared deviations across song intervals). Arrows connect selected song features across each adjacent age period. Examples illustrate a more general developmental pattern of songs becoming faster over development.

the late plastic period, 65-70 dph, through ∼1 year of age. Throughout most of the analysis we focus on the difference between global variability and both repeated and independent variations in song timing, which together we term "local variability." Unless indicated otherwise, all values reported below include standard errors (mean±SE). However, since many of the parameter distributions showed significant skew, statistical significance was assessed using the Wilcoxon signed-rank (WSR) test for pairwise comparisons.

5.2.1 Tempo increases

We began by asking whether average song tempo changes systematically as a function of age (Figs. 5.2, 5.3). In fact, between 65 and 365 dph, motif length

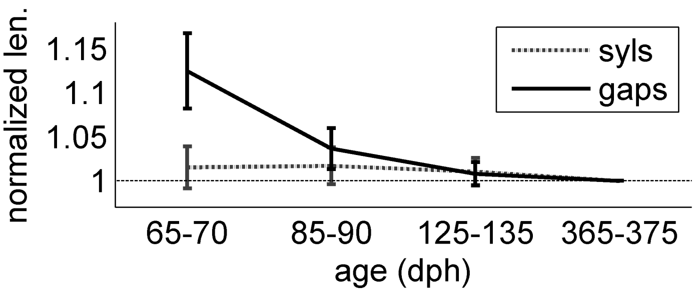

Figure 5.3: Changes in the mean length of syllables and gaps, normalized by 365-dph mean length. Errorbars indicate standard error.

decreased by 34.071 ± 11.001 msec, or 7.0 ± 2.6 % of adulthood length (WSR, $p < 0.05$), with all 7 birds showing a significant decrease. Across birds, about half of this decrease happened between 65 and 85 dph: over this period, motif length decreased by 17.699 ± 2.903 msec, or 3.1 ± 0.5 % (WSR, $p < 0.05$); again, all 7 birds showed a decrease in motif length over these periods.

Given the distinction between syllable- and gap-based timing in adult song, we then investigated the extent to which these developmental changes occurred across both types of segments (Fig. 5.3). In fact, on average, all of the dominant increases in tempo occurred during gaps. Here, length decreased by an average of 7.130±2.152 msec, or 12.5 ± 4.3 % of adulthood length (WSR, $p < 0.005$). Overall, 19 out of the 26 gaps we tracked showed significant decreases over this period, with average gap length decreasing in all 7 birds. Over half of this change occurred between 65-85 dph, during which gaps decreased in length by 4.741 ± 1.562, or 8.4 ± 3.0 % of 85-dph length (WSR, $p < 0.01$). By contrast, between 85-125 dph and 125-365 dph, respective gap decreases were 1.984 ± 0.969 msec and 0.405 ± 0.960 msec, and these changes failed to reach statistical significance.

In contrast to gaps, syllables as a whole failed to show any significant length changes, with decreases in length of 0.019 ± 0.837 msec from 65-85 dph (WSR, $p = 0.550$), and of 1.215 ± 1.631% from 85 dph through 1 year of age (WSR, $p = 0.601$). Overall, only 16 of the 35 syllables we tracked demonstrated a significant decrease in length over late development, with 4 of 7 birds showing an average decrease in syllable length.

5.2.2 Motif length variability and noise sources

We then asked whether motif length variability itself changed as a function of age. In fact, motif length was about twice as variable at 65 dph as it was 365 dph, with respective values of 17.367 ± 2.227 msec and 8.524 ± 0.646 msec standard deviation, or 2.9% and 1.5 % CV. As with changes in average tempo, the decrease in tempo variability was very robust, occurring in all 7 birds both 65-365 dph and more specifically 65-85 dph.

As above, we then examined how timing variability was expressed among syllables and gaps separately (Fig. 5.4). Here, we found decreases across all 26 gaps and 32 of 33 syllables 65-365 dph, with an almost threefold decrease among gaps, twofold among syllables: Across gaps, average variability decreased from 6.951 ± 0.827 msec to 2.427 ± 0.186 msec standard deviation, or 9.4 % to 4.3 % CV, while syllables decreased from 4.635 ± 0.407 to 2.427 ± 0.235 msec standard deviation, or 8.7 % to 4.7 % CV (WSR, $p < 0.0001$ in all changes). Across both syllables and gaps, a little under half of this decrease occurred 65-85 dph.

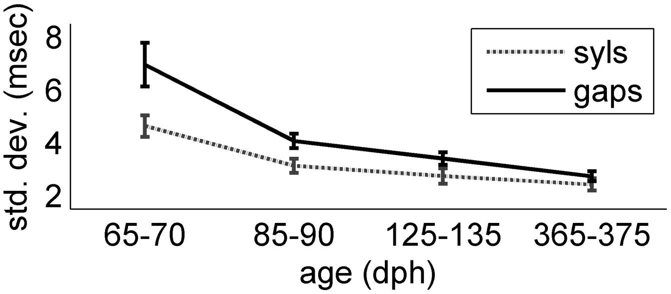

Figure 5.4: Changes in syllable and gap standard deviation across development; errorbars indicate standard error.

Variability decreases among local parameters only

We had previously found two kinds of rendition-to-rendition timing variability in adult song (Glaze & Troyer, 2006, 2007): tempo changes that are shared by all song intervals, and length changes within individual intervals that are independent of tempo and uncorrelated across the song motif, which we term local variability (see Methods for details). Importantly, motif length variability reflects a combination of both to the extent that both global and local timing changes accumulate over song rather than being offset (by contrast, measurement error is expected to be offset and should not significantly influence motif length; see Methods).

Does the decrease in timing variability over development occur in both tempo changes and local variability? This might be observed, if, for example, synaptic strength increases across song representations, yielding a decreased sensitivity in burst-onset times to both background noise and neuromodulation (Chapter 4). On the other hand, a difference in developmental trajectories might be observed if some of the decrease is due to a decline in only one of the variability sources; for example, one would find a decrease in tempo changes only if synaptic strength remained constant but fluctuations in neuromodulatory input became smaller. We investigated this question by separating out tempo changes from the accumulation of local variability using factor analysis (FA) and measuring developmental changes in each parameter (see Materials and Methods).

In fact, we found average decreases in local variability only (Fig. 5.5): Over development, this source decreased among gaps by 3.639 ± 0.782 msec, among syllables, by 0.994 ± 0.16 msec (WSR, $p < 0.0001$). As might be expected from the raw standard deviation data, roughly 42% of this decrease occurred 65-85 dph among syllables, while 60% occurred during this period among gaps. Local variability decreased across 25 of 26 gaps and 30 of 33 syllables, and averaged a decrease among syllables and gaps across all 7 birds.

By contrast, while at all ages, syllables and gaps as a whole had global weights that were significantly positive (mean 0.771 ± 0.082msec, WSR $p < 0.0005$), we did not find any significant differences by age: Across syllables, global variability decreased by 0.251 ± 0.234msec (WSR, $p = 0.131$), while gaps showed an inconsistent trajectory that actually included an increase in tempo-based variability among 13 of 26 between 65 and 85 dph; gap-based tempo variability increases across development averaged 0.559 ± 0.623msec but failed to reach significance (WSR, $p = 0.409$). Across the sample, we found decreases in 20 of 30 syllables, with average decreases in 5 of 7 birds; and 11 of 26 gaps, averaging decreases in 4 of 7 Thus, there appears to be significant tempo-based variability as early as 65-70 dph, and the magnitude of this variability remains fairly constant over late development. Although there may be trends, they are much smaller than what we observed among local timing

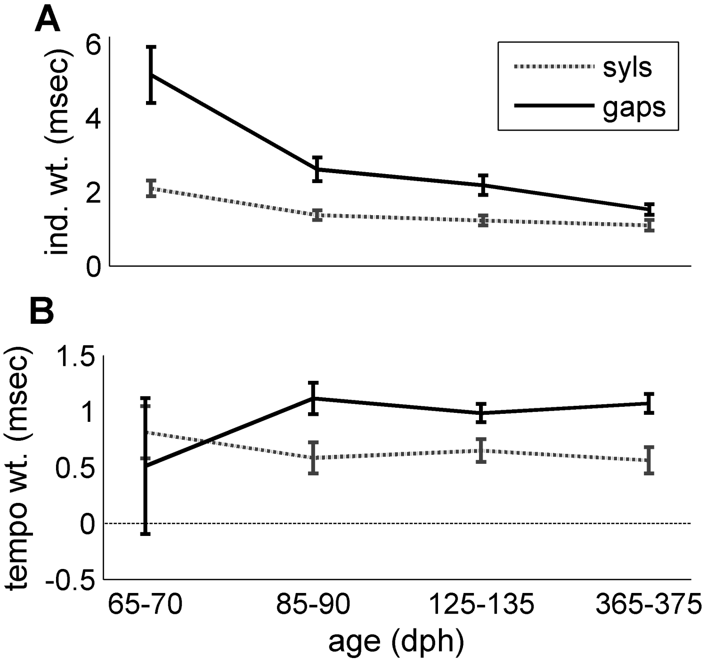

Figure 5.5: Development of global and local timing weights measured from the factor analysis model (see Materials and Methods); errorbars indicate standard error.

parameters.

Independent vs. repeated timing variability

One modeling study (Chapter 4) and previous analysis (Appendix E) have suggested that while local variability is unique to individual song intervals, each song interval may have unique variance on multiple timescales. Here we distinguish between two kinds: “Repeated variability” is unique to a song interval of a particular identity and repeated across multiple renditions of the same interval within the same song bout. By contrast, “independent variability” is unique to a song interval and not repeated at all, so this type is completely independent across all intervals in the same song bout.

Do both sources of local variance decline with age? We examined this question

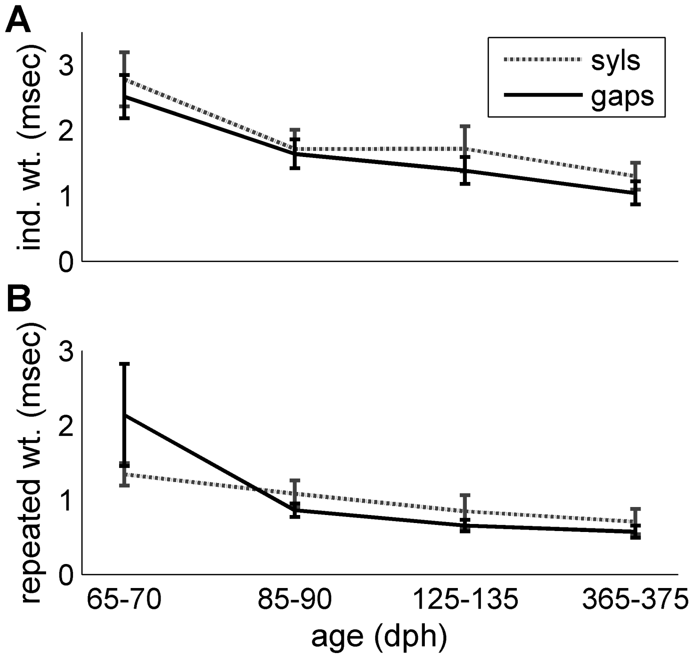

Figure 5.6: Changes in the mangitude of repeated and independent variability over development, as measured by the factor analysis model (see Materials and Methods). Errorbars indicate standard error.

using the factor analysis model and explicitly including repeated and independent latent factors that were independent of jitter (see Methods). Because repeated variability depends on multiple motif renditions in the same song bout, we gathered recordings with at least 2 motifs across all bird-by-age samples. In two birds this failed to yield sample sizes > 300 at 65 dph, so we excluded them from analysis.

In the remaining 5 birds, both kinds of local variability declined with age (Fig. 5.6): Syllable independent variability decreased from 2.776 ± 0.415 to 1.298 ± 0.203 msec over the 65-365 dph range, while among gaps this variability declined from 2.512 ± 0.331 to 1.043 ± 0.178 msec (WSR, $p < 0.001$). Repeated variability declined among syllables from 1.342 ± 0.150 to 0.708 ± 0.168, among gaps, from 2.137 ± 0.686 to 0.572 ± 0.081 msec (WSR, $p < 0.001$).

Thus, the data indicate that the decrease in local variability is due to decreases in both independent and repeated timing.

5.2.3 Links among tempo, noise and transition probability

Previous chain-based modeling (Chapter 4) suggested that the magnitude of tempo variability and local timing variability in a given song interval may be associated with the strength of synaptic connections that subserve that interval. We reasoned that transition probabilities could likewise be linked with the strength of synapses subserving the production of associated silent gaps during those transitions. If synaptic strength increases over development, this would in turn imply increases in transition probabilities over the same period, and a direct correlation between increases in tempo, increases in transition probability and decreases in noise on a gap-by-gap basis.

Indeed, transition probabilities showed developmental trajectory that was similar to timing parameters (Fig. 5.7). Specifically, we found significant increases 65-365 dph, from an average of 71.5% to 87.9% (WSR, $p < 0.0001$). Of the 20 gaps with transition probabilities < 95% at 65 dph, 17 increased, and 8 by over 10%. In contrast, from 85-125 dph and 125-365 dph, only 2 increased by more than 10%, and overall changes failed to reach statistical significance. Thus, while increases in some transition probabilities may continue after 85 dph, these increases are much smaller than the jumps observed 65-85 dph.

The data also indicate the direct link between tempo increases and both tran-

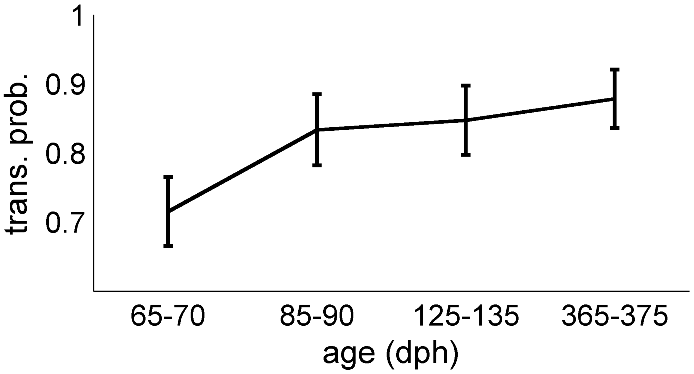

Figure 5.7: Change in average transition probability across development; errorbars indicate standard error.

sition probabilities and local timing variability: Between 65 and 85 dph, decreases in gap length were positively correlated with both increases in the corresponding transition probability and decreases in both the fast and slow noise parameters (Fig. 5.8) The correlation between length and transition probability changes was 0.679, between decreases in length and local variability, 0.683 (Spearman's correlation, $p < 0.0005$ for both). We also found a correlation between decreases in local variability and increases in transition probability (Spearman's correlation was 0.395, $p = 0.46$); however, this relationship was sensitive to outliers, and controlling for gap length (by dividing local variability by length) yielded a weak correlation that failed to each significance (Spearman's correlation was 0.198, $p = 0.331$).

Thus, the data in fact indicate, on a gap-by-gap basis, links between decreases in average tempo and both local variability and transition probability. The lack of direct relationship between the latter two parameters suggests that any measured correlation was due to the shared influence of gap tempo.

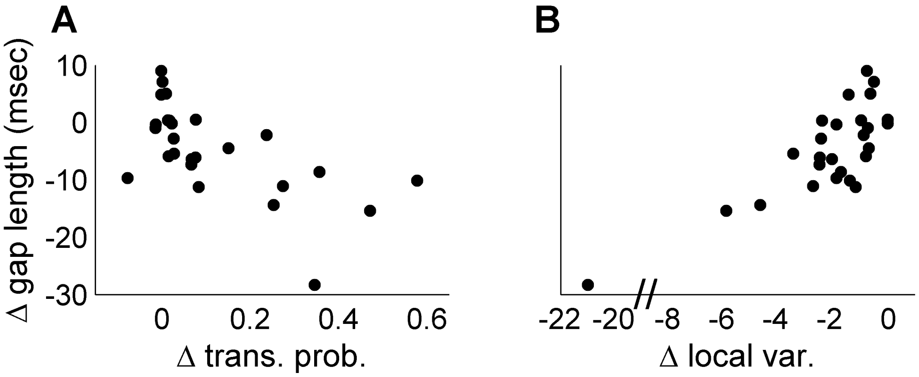

Figure 5.8: Links between changes in gap tempo and changes in both local variability and transition probability across the first 2 developmental periods, from 65-70 dph to 85-90 dph. A, change in gap length by change in transition probability; B, change in gap length by change in local timing variability.

5.2.4 Global structure

Although we failed to find significant changes in the magnitude of tempo variability across development, it is still possible that the previously reported differences between syllables and gaps (Glaze & Troyer, 2006) did change. We thus examined syllable and gap elasticity by dividing global weights from the factor analysis by mean length and multiplying by 100 to yield a percentage of mean length. This measure is analogous with the elasticity parameters derived from the linear regression with sequence length (see Methods), and may be thought of as a partial coefficient of variation; *e.g.* a 1% elasticity from the factor analysis indicates that the corresponding interval varies with global tempo by 1% of its mean length.

At 65-70 dph, syllable elasticity actually trended higher than gaps, but this difference did not reach statistical significance (Wilcoxon rank-sum, $p = 0.51$), averaging $1.482 \pm 0.307\%$ vs. $0.618 \pm 0.735\%$. However, by 85-90 dph, syllables were indeed less elastic than gaps (Wilcoxon rank-sum, $p < .01$), averaging $0.730 \pm 0.108\%$

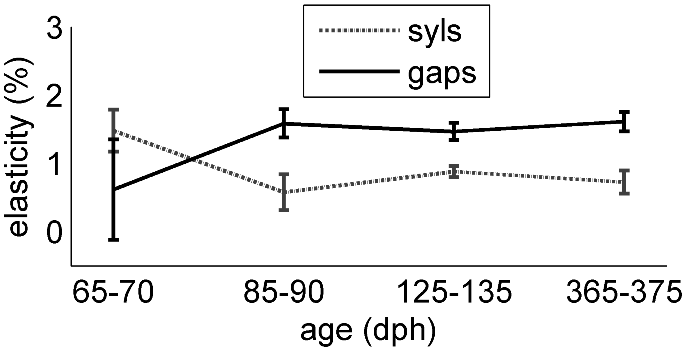

Figure 5.9: Syllable-gap elasticity averages across development. See text for definition of elasticity. Errorbars indicate standard error.

vs. $1.557 \pm 0.093\%$, and this difference persisted through adulthood (Fig. 5.9).

It is also worth noting that the elasticity measurements assume that global weights scaled with mean length. This was true 85+ dph (by age period 85-90, 125-135, 365-375 dph, $r = 0.616, 0.670, 0.684$, Spearman's correlation, $p < 0.0001$), but at 65-70 dph we only found a weak relationship that failed to reach significance ($r = 0.158$, $p = 0.23$). It is thus possible that the lack of elasticity pattern during this age period is due to a violation of that assumption.

5.2.5 Local variability is dominated by subsyllabic timescales across development

We had previously demonstrated in adult song a subsyllabic timescale of repeated variability (Glaze & Troyer, 2007). Specifically, repeated notes in the same song bout have significantly correlated rendition-to-rendition length variability while different notes in the same syllable are no more correlated than notes from different syllables. We had found analogous patterns on the 10 msec scale by dividing amplitude-modulated (AM) notes into 10-msec song slices; here, individual 10-msec song slices in the same AM note were independent, yet each had a unique source of timing variability that repeated across motifs.

While subsyllabic timescales dominate local variability in adult song, it is possible that syllable-based variability exists at younger ages. This might occur if, for example, birds learn to produce individual syllables independently of song sequence, yielding a neural network that initially groups together neurons dedicated to the same syllable. This in turn would suggest that notes and song slices in the same syllable in fact share a unique source of variability. On the other hand, it is possible that subsyllabic timescales dominate variability across development if, for example, songs are driven by synfire chains with independent chain links at all ages under examination.

We thus investigated whether subsyllabic units shared a unique source of variability at any age range. We found that most AM notes failed to show temporal structure at 65-70 dph that was reliable enough to examine the 10-msec timescale across development. Thus, we focused this part of the analysis on the note level, and restricted analysis to syllables with > 1 note, yielding a final sample of 35 unique notes in 10 unique syllables across 5 birds. We used the factor analysis model to test for syllable-based timing. To the extent that notes are independent, the sum of the square of note weights will be equal to the sum of the square of syllable weights. If notes in the same syllable have shared variability, then the note sum will be greater and the ratio of the two values will be > 1.

The data suggest a lack of syllable-based timing at all age periods analyzed. Specifically, we found syllable-note ratios were actually slightly < 1 at all ages (mean

0.884±0.026), and we found no difference by age (Fig. 5.10; Kruskal-Wallis $p = 0.41$). With ratios slightly < 1 we considered the possibility that parameter estimates were not accurate enough for the question being investigated, so for comparison we analyzed a similar ratio for global weights, with the expectation that here we would find ratios > 1 to the extent that all notes share the same global source of tempo. Indeed, we found that the global weight ratio was significantly > 1 across ages (mean 1.367 ± 0.084, WSR $p < 0.0001$), and again no difference by age (Kruskal-Wallis $p = 0.54$).

Overall, the data suggest that that local variability is dominated by song timing at subsyllabic timescales across all age periods analyzed.

5.3 Discussion

We have investigated the development of zebra finch song stereotypy from the late plastic period through 1 year of age. We collected hundreds to thousands of songs from each of four age periods extending from late plasticity through 1 year of age, and measured song timing and syllable transition probabilities within each age period to probe how the motor code changes over this age range. We found increases in song tempo that occur almost entirely during the gaps of silence between syllables. Most of the tempo increase occurs 65-85 dph, and on a gap-by-gap basis increased speed is linked with increases in the reliability of corresponding syllable transitions, as well as decreases in local timing variability. We also find interesting changes in timing variability itself: while the magnitude of local variability decreases through

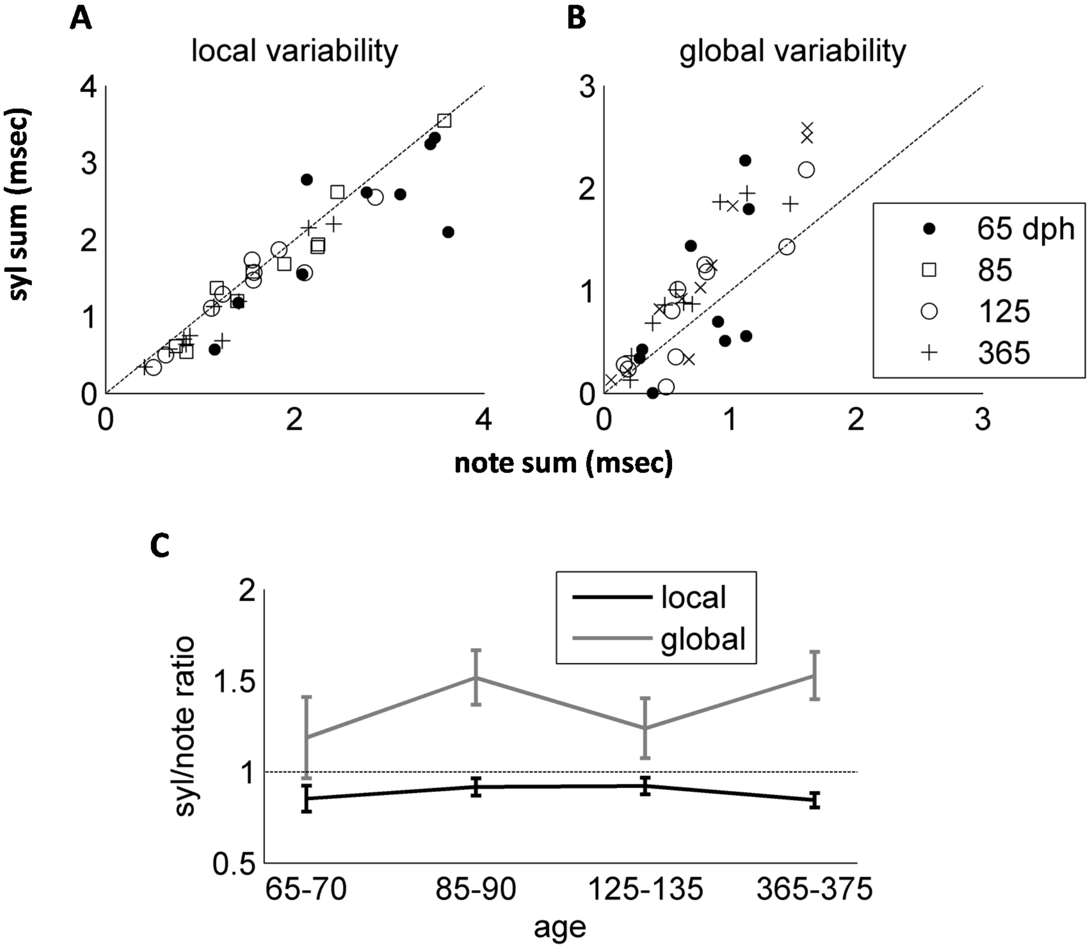

Figure 5.10: Comparison of syllable-based and note-based timing parameters from the factor analysis model (see Materials and Methods). A & B, y-axis represents the syllable-based measurements while the x-axis represents note-based measurements from the same samples; see Results for how sums were calculated. A, comparison across local timing parameters, B, across global. C, ratio of syllable- to note-based parameters as a function of age; errorbars indicate standard error.

365 dph, we found no significant changes in rendition-to-rendition global timing variability that is spread across the song bout. Furthermore, we did not find any changes in elasticity patterns as far back as 85 dph, nor changes in the timescale of local variability, a parameter which had previously been linked with proposed chaining mechanisms over the premotor pathway in the adult song system.

Overall, the data suggest a phase of late development in which song production becomes faster and more automated, while the magnitude of fine-grained timing variability decreases two- to threefold. Such a process of motor consolidation has been previously suggested for zebra finch song (Brainard & Doupe, 2001); the current study provides a rich set of behavioral constraints for models of consolidation.

Gap-based developmental processes

One of the most basic physiological parameters that might change during a developmental process is synaptic strength. For example, syllable transition probabilities might increase if the synapses subserving those links strengthen, increasing the reliability of the neural process subserving the transition and preventing transitions to other syllables. In this scenario, during late plastic song, syllable-based chains may be loosely organized into a network capable of producing a small variety of song sequences with a bias towards sequences that match song templates. Those biased sequences could completely "win out" by adulthood via spike-time dependent plasticity (STDP), in which changes in synaptic strength depend on the order of pre- and post-synaptic spikes (Nowotny et al., 2003; Melamed et al., 2004; Song

& Abbott, 2001; Froemke & Dan, 2002; Dan & Poo, 2004).

Previous modeling of premotor chaining mechanisms (Chapter 4) suggests that increased synaptic strength would also increase the tempo of the song intervals subserved by those synapses. This might explain why increased tempo among gaps is linked with increased transition probabilities–if both sequence transitions and the lengths of corresponding gaps are subserved by the same mechanisms, a process such as STDP would induce correlated changes in both parameters over late development.

LMAN-induced local timing variability?

While increased synaptic strength is an obvious hypothesis for explaining developmental changes, it cannot alone account for the changes in timing variability we have found. First, while average tempo increases among gaps only, decreases in local timing variability occur across both syllables and gaps. This contradicts one basic prediction of the chaining model, *i.e.* that changes in chain speed are coupled with sensitivity to neuromodulation and other sources of timing variability. How can syllable tempo remain the same while timing variability decreases? Second, increases in synaptic strength decreases the sensitivity of song intervals to perturbations from all possible noise sources, including neuromodulatory mechanisms that induce shared, global timing variability as well as background noise and inputs from areas such as LMAN which induce variability that varies on a local timescale (Ölveczky et al., 2005). Thus, any chain-based model of development has the burden of simultaneously accounting for both decreases in local timing variability and no changes in global variability.

The most likely mechanism to reconcile these differences may be an active decrease in input from LMAN, the output nucleus of a basal ganglia circuit that is strongly implicated in song learning (Bottjer et al., 1984; Scharff & Nottebohn, 1991), and has been demonstrated to produce spike sequences that are highly variable in both juvenile and undirected adult song (Hessler & Doupe, 1999; Ölveczky et al., 2005). It has been proposed that LMAN more generally acts as a noise generator involved in trial-and-error learning (Ölveczky et al., 2005) as well as rendition-to-rendition variability in adult song (Hessler & Doupe, 1999; Kao & Brainard, 2006; Kao et al., 2005). Furthermore, while LMAN influences song variability throughout a bird's lifetime, the degree of this influence on auditory feedback has previously been shown to progressively decline over the same age range we have analyzed, (Brainard & Doupe, 2001); it may be that this reflects a more general decline of influence across a variety of perturbations, experimental and natural. Because increased synaptic strength among gaps would also decrease the expression of noise in song timing, this scenario would suggest proportionally an even greater decrease in gap-related noise. Indeed, we found a roughly threefold decrease in accumulated variability across gaps, in contrast to a twofold decrease across syllables.

If LMAN is in fact responsible for the local timing variability we have analyzed, it may conceivably explain the simultaneous decline in both the independent component, and the slower, repeated variability that is correlated across motifs. This suggests the possibility that LMAN activity has a slower component that is repeated across multiple renditions of the same syllable (or motif) in the same song bout.

It is also possible to reconcile gap-based tempo increases with noise decreases across the motif in a scenario in which synaptic strength increases across the motif. Inhibitory HVC interneurons have been demonstrated to be somewhat more active during syllable-based activity (Kozhevnikov & Fee, 2007). Our modeling (Chapter 4) has shown that increased inhibitory feedback can simultaneously slow down activity and decrease variability; it is thus possible that synapses within syllable-based chains in fact get stronger, but potential tempo increases are offset by increased inhibitory feedback. However, inhibitory feedback can also suppress the expression of local variability in timing, so in this scenario we might expect greater decreases in noise across syllables than across gaps, which we do not find.

Overall, the data suggest a phase of development in which syllables are eventually consolidated into a longer, precisely-timed chaining mechanism that can run automatically from beginning to end in a highly reliable order. A similar process of linking simpler chains to form more functional activity patterns has been proposed for neocortex (Bienenstock, 1995), while sequence learning on multiple timescales has been proposed in behaviors such as rodent maze navigation (Melamed et al., 2004). Tempo and trial-to-trial timing variability may thus provide a window into how sequences are learned and represented across a variety of systems.

Chapter 6

Conclusion

Birdsong is frequently held up as an excellent model of speech learning and production in humans. Although both speech and birdsong contain obvious timescales of organization in the behavior, the two areas of research also share the fundamental problem that the neural bases for this organization remain elusive. In neurolinguistics, the classical neuroanatomical Geschwind model of language processing has been seriously questioned by inconsistent imaging and lesion data (Sidtis, 2006; Damasio et al., 2004; Poeppel & Hickok, 2004), and the neurophysiological processes which represent speech units have always been ill-specified (Poeppel & Embick, 2005). Similarly, the song system has yielded conflicting data on how or even whether the timescales of song organization have any neural representation. Is the acoustic hierarchy represented over neural pathways or is song produced by a single clock-like mechanism? Are individual song syllables and syllable sequence learned with distinct processes or together as part of the same underlying motor program?

We have used naturally occurring timing variability in zebra finch song to investigate these questions. Taken as a whole, the data are compatible with elements of each conflicting model and suggest a unifying framework in which the motor code may be further investigated. As predicted by the clock model, song segments stretch and compress together and have variability independent of tempo that varies on a $\leq$5-10 msec scale. However, the tempo changes are highly nonuniform and the elasticity patterns correspond to elements of the acoustic hierarchy, distinguishing syllables from gaps and aligning syllable onsets with global tempo changes. The data as a whole suggest a model that incorporates elements of both the hierarchical and clock-based models: Song may be driven by a single chain of activity on a fine timescale, but the chain may have physiological parameters that do correspond to the acoustic hierarchy by distinguishing syllable-based from gap-based chain segments.

Further analysis of timing during late development suggests a phase in which syllables are consolidated into the longer chaining mechanisms proposed for adult song. If, in adult song, the acoustic hierarchy corresponds to system parameters, it may be because different levels of the hierarchy develop at different rates. For example, synapses subserving the gaps of silence may be weaker than those corresponding to syllables because gap lengths and transitions develop at a slower pace and never "catch up" to syllables. A similar process of linking simpler chains to form more functional activity patterns has been proposed for cortex (Bienenstock, 1995), and sequence learning on multiple timescales has been proposed in behaviors such as maze navigation (Melamed et al., 2004).

Overall, the timing variability we have uncovered shows interesting similarities to other studies on sequence learning, in which learning is accompanied by increased production speed and decreased timing variability (Rhodes et al., 2004; Keele et al., 2003; Hikosaka et al., 2002). Furthermore, the adult song timing patterns presented here may be analogous with patterns of proportional scaling that have been analyzed in speech and human movement to discern different levels of hierarchy in those systems (de Jong, 2001; Heuer, 1988; Gentner, 1987).

Ethological implications

We have used temporal structure to investigate the neural mechanisms of song learning and production. However, it remains unclear whether song timing serves an important role in communication. There is evidence that the structure we have analyzed can at least be perceived by the bird producing the song as well as other birds: tests of auditory discrimination have shown that zebra finches can perceive temporal changes on the millisecond scale (Lohr et al., 2006; Dooling et al., 2002), while HVC neurons in anesthetized birds shows similar auditory sensitivity (Theunissen & Doupe, 1998). Do such perceived differences actually serve a communicative purpose? One recent investigation found that female birds prefer recordings of directed song versus the more variable, undirected version of the same songs (Woolley & Doupe, 2008); while in that study females appeared to be specifically sensitive to variability in spectral characteristics and not average song speed, timing variability itself was not tested, so it remains an open possibility that this property would show a similar functional difference. Average song speed may also serve a purpose outside the female-directed context: The links that we find between tempo and both time of day and general level of cage activity suggest the influence of neuromodulatory factors such as circadian rhythm. If so, tempo may serve a more general function of communicating the affective state of the bird, *e.g.* perhaps faster song versions indicate a very excited mood to other birds.

Do elasticity differences across song matter to other birds? The relative inelasticity of syllables might suggest that syllable lengths are more important identifiers of song than global song rhythm, *i.e.* the relative durations of syllables and gaps. However, it will be important for future studies to test just how variable gap lengths can be made before song recognition is significantly impaired. Indeed, while syllables are less elastic than gaps, it is worth noting the more basic finding that all song segments stretch and compress together to begin with, preserving song rhythm to within ∼1-2 milliseconds of difference in relative segment durations. Zebra finch birds are very social (Zann, 1996), so such preservation of song rhythm may in fact be advantageous in identifying songs in a noisy acoustic environment in which the fine structure of syllables is difficult to discern. In this respect the importance of gap lengths may depend on social context and more specifically the level of noise in the surrounding environment.

Beyond the forebrain

The analysis presented in these chapters was motivated by models of feedforward pattern generation over the HVC-RA network. Indeed, we have relied on the premise that timing variability which accumulates over relatively long timescales must be derived from the central pattern generator; noise from peripheral mechanisms must be offset by incoming signals upstream and will thus fail to influence overall sequence length. However, there is evidence for feedback loops that span the forebrain and brainstem (Vates et al., 1997; Ashmore et al., 2005). Brainstem nuclei DM and PAm are innervated by dRA and are involved in control of the respiratory pattern corresponding with syllables and gaps (Wild et al., 1997, 1998; Sturdy et al., 2003). DM and PAm also send bilateral projections back up to the forebrain through thalamic nucleus Uva (Striedter & Vu, 1998; Vates et al., 1997); these projections are important because the avian brain lacks a corpus callosum, yet both hemispheres control song and thus require coordination. Importantly, unilateral lesions to Uva cause long-term changes to syllable sequencing while leaving the basic acoustic structure of individual syllables intact (Coleman & Vu, 2005); because Uva is the recipient of bilateral feedback from the brainstem, this suggests that the brainstem-forebrain circuits may in part encode sequence transitions and thus participate in a system of pattern generation that goes well beyond HVC and RA (Ashmore et al., 2005).

Similar models of pattern generation have been proposed in other systems: For example, recent investigations suggest that the primate oculomotor system requires corollary discharge from the brainstem to produce a learned sequence of multiple visual saccades, but not a single saccade (Sommer & Wurtz, 2002). A similar anatomical segregation of timescales has also been proposed for human speech processing (Friederici, 2002; Poeppel, 2003; Luo & Poeppel, 2007). Moreover, prosody may play a role in chunking syllables and phonemes into commonly produced phrases (Byrd & Saltzman, 2003) that may be analogous to stereotyped birdsong motifs. Recent modeling and imaging data implicate a role of distributed, bilateral circuits that include the brainstem (specifically, the cerebellum) in human syllable sequencing

(Bohland & Guenther, 2006; Guenther et al., 2006). Moreover, prosody has been implicated in a range of aphasias (Baum & Pell, 1999) and lesions to key areas such as the cerebellum (Ackermann & Hertrich, 2000).

How can timing variability over such feedback circuits accumulate over sequence production? Without a central timing mechanism such accumulation requires that noise from feedback loops at least propagate throughout the song system; otherwise, activity across different nuclei would progressively become less synchronized as a song bout (or other sequenced behavior) proceeded. However, this scenario faces a challenge that confronts all of the aforementioned models of feedback and corollary discharge: neural signaling involves transmission delays (particularly over synapses) and thus can take a significant amount of time to transfer information across the system. The type of data we have analyzed may thus provide an important set of clues to investigations into how sequence learning and production is regulated by the interaction between feedforward and feedback circuitry.

More generally, birdsong has offered an excellent opportunity to develop a comprehensive, mechanistic understanding of the neural basis for vocal communication. The timing variability we have analyzed provides a common language for synthesizing results from behavioral and electrophysiological studies, and also provides strong constraints for computational models that attempt to connect the two levels of analysis.

Appendix A

Dynamic time warping algorithm for amplitude-based measurements

Following is the Dynamic Time Warping (DTW) we developed for the fine-grained measurements of syllable onsets and offsets. Both templates and actual waveforms were truncated to include only data from the beginning of the first peak to the end of the last peak. Peak beginnings and endings were defined as inflection points in the waveform on either side of the peak. In order to minimize discrepancies in the lengths of templates and waveforms to be matched, endpoint peaks used in the template were the very first and last and not necessarily the ones chosen based on height and regularity to define onsets and offsets.

Details of the DTW process are as follows (Rabiner & Juang, 1993). Paths were globally constrained by a Sakoe-Chiba band defined by the midpoints of each time-axis. Thus, total length change was limited to fall between 1/2 and 2. Pathways were locally confined to

$$P_1 \longrightarrow (1,0)(2,1)(3,2)$$

$$P_2 \longrightarrow (1,1)$$

$$P_3 \longrightarrow (0,1)(1,2)(2,3)$$

Hence a local length change was limited to fall between 2/3 and 3/2. Path sets were weighted by $[2,3,2]$ which amounts to a slight bias away from a slope of 1, so the cumulative product matrix D was computed as

$$D(i,j) = max \begin{cases} 2[d(i,j) + d(i-1,j) + d(i-2,j-1)] + D(i-3,j-2) \\ 3d(i,j) + D(i-1,j-1) \\ 2[d(i,j) + d(i,j-1) + d(i-1,j-2)] + D(i-2,j-3) \end{cases}$$

where d was the outer product, i was the template index and j the syllable index.

Endpoint constraints were relaxed because the correspondence between syllable and template boundaries was itself an object of this DTW. The default region where the path could end in time was delimited by the last 25% of the time axis for either the template or the candidate waveform. This was stretched accordingly if either length exceeded the other by more than 25%.

Each syllable's onset and offset was determined by finding the point on the syllable time axis corresponding to the onset and offset times originally defined on the template time axis.

Appendix B

Qualitative model of tempo changes

We qualitatively analyzed the relationship between interval variances and sequence length variance using a simple model set to match the variance of overall sequence length σ^2 and the mean length of the intervals in the sequence, $\bar{x}_1, \bar{x}_2, ..., \bar{x}_m$. The model was constructed so that by changing a single parameter, we could consider the case where intervals are independent and the case in which the gross covariance measure g matches the experimental data, *i.e.* interval length changes are dominated by changes in tempo. In both versions of the model, the standard deviation of all intervals was proportional to mean interval length. The length of each interval was given by $x_i = \bar{x}_i + \alpha \bar{x}_i(\gamma \eta_i + \sqrt{1-\gamma^2}\eta_0)$, where η_i and η_0 are zero-mean independent Gaussian variables with standard deviation equal to 1 (η_0 is the common variation shared by all intervals), the parameter α controls the overall variance and γ controls the degree of independence in the different intervals. Overall sequence length variance is given by

$$\sigma^2 = \alpha^2[\gamma^2 \sum \bar{x}_i^2 + (1-\gamma^2)(\sum \bar{x}_i)^2]$$

while gross covariance is given by

$$g = \frac{\sigma^2}{\alpha^2 \sum \bar{x}_i^2}$$

To obtain independent intervals, γ and g were set equal to 1 and α was solved for, while for the model with tempo changes g was set equal to the gross covariance of the bird being analyzed and both α and γ were obtained from the two equations above. The model contains a number of assumptions regarding the nature of interval length variability and so did not form the basis for any of the statistical conclusions reported.

Appendix C

Dynamic time warping algorithm for spectral-based measurements

Following is the modified dynamic time warping (DTW) we used to map syllable timepoints to mean spectrograms see (see Rabiner & Juang, 1993, for a general introduction to DTW and basic terminology). The algorithm was similar to Glaze & Troyer (2006) with several important modifications.

First, the similarity metric between each syllable and the mean was based on the time-derivative of full spectrograms rather than summed amplitude envelopes. Let the matrix m denote the mean (template) time-derivative spectrogram and the matrix s denote the time-derivative spectrogram of a particular song syllable. (A time-derivative spectrogram is computed by subtracting the frequency vectors obtained from adjacent time bins in the raw spectrogram.) The match $d(i, j)$ between time bin i of a particular syllable and time bin j of the template was equal to the overlap of the corresponding time-derivative vectors, $d(i, j) = \sum_k s_{ik} m_{jk}$, where frequency is indexed by k. With this modification, the algorithm allowed us to track changes in a syllable's spectral profile that are not always evident in the amplitude envelope.

The local path constraints and weighting were also different from the previous algorithm. At each point in the algorithm, three possible paths were available: $P_1 \rightarrow (2, 1)$, $P_2 \rightarrow (1, 1)$ or $P_3 \rightarrow (1, 2)$. The cumulative product matrix D was computed as

$$D(i,j) = max \begin{cases} 3/2\ [\frac{1}{2}d(i,j) + \frac{1}{4}d(i-1,j) + \frac{1}{4}d(i-1,j-1)] + D(i-2,j-1) \\ d(i,j) + D(i-1,j-1) \\ 3/2\ [\frac{1}{2}d(i,j) + \frac{1}{4}d(i,j-1) + \frac{1}{4}d(i-1,j-1)] + D(i-1,j-2) \end{cases}$$

This differs from the previous version in several important ways. First, the scheme allows the slope of local length changes to fall between 1/2 and 2, while the older version had more restrictive limits of 2/3 and 3/2. Second, unlike the previous algorithm, $d(i,j)$ was included in all three path calculations to determine $D(i,j)$, and its value was divided equally with $d(i-1,j)$ in path P_1 and $d(i,j-1)$ in path P_3. This allows for the fact that, geometrically, the path joining $(0,0)$ and $(2,1)$ (*i.e.* P_1) is actually equidistant from points $(1,0)$ and $(1,1)$; similarly, P_3 is equidistant from $(0,1)$ and $(1,1)$. Finally, it can be shown that with this weighting scheme, if $d(i,j)$ has the same match strength for all time bins i and j, then the algorithm accumulates the same amount between any two points for all choices of path, so there is no bias against any of the paths; in the previous version there was a slight bias against a slope of 1.

Appendix D

Factor analysis algorithm

Following are the details and derivation of the factor analysis algorithm used to model covariance matrices for Chapter 5 and Appendix E. We begin with traditional factor analysis (Bishop, 2006), in which all measurements share a global tempo factor that explains the positive timing covariances across different song intervals. Let $\mathbf{x}_{\mathrm{i}\alpha}$ denote the mean-subtracted length of interval j in sequence α, $\mathbf{z}_\alpha$ denote a 0-mean, unit-variance latent tempo factor shared by all intervals, $\mathbf{W}$ denote the matrix of tempo weights. We thus write $\mathbf{x}_{\mathrm{i}\alpha} = \mathbf{W}_{\mathrm{i}}\mathbf{z}_\alpha + \eta_{\mathrm{i}\alpha}$, where η is residual interval deviation after factoring out tempo. We write the residual covariance matrix as $\mathbf{\Psi}$, a diagonal matrix of interval variances that are independent of tempo. This allows us to model covariance matrix $\mathbf{C}$ as

$$\mathbf{C} = \mathbf{W}\mathbf{W}^{\mathrm{T}} + \Psi \tag{D.1}$$

Birdsong poses two basic challenges to this model. First, intervals are measured as pairwise differences across a sequence of feature times. Thus, feature jitter will induce a negative covariance between adjacent intervals that share that feature, yielding a combination of jitter and accumulated independent variability in η. We are interested in accumulated, independent variability, so we separate out jitter from η by including vector $\mathbf{u}$ to represent the 0-mean timing of features (*e.g.* a syllable onset or offset) that is independent of $\mathbf{z}$ and η. We include a corresponding weight matrix $\mathbf{D}$, where

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix} \tag{D.2}$$

Multiplication by $\mathbf{D}$ is equivalent to the differencing operation on a vector, so the interval deviations due to jitter can be written simply as $\mathbf{Du}$. By explicitly including a jitter factor, η can be expected to represent variability that is not offset and thus accumulates over song.

Second, birdsong often contains repeated elements. We have hypothesized that such repetitions share a common source of variability, but do not differ in their sensitivity to this source. In order to model this source directly, we include a third latent factor $\mathbf{v}$ that is shared by all elements of the same identity. Here, we include a $d\text{x}m$ fixed weight matrix $\mathbf{Q}$ where d is the total number of intervals and m is the number of unique sources contributing to this repeated variability. For variability source j, $\mathbf{Q}_{\text{ij}} = 1$ if element i shares that source and $\mathbf{Q}_{\text{ij}} = 0$ otherwise.

With the jitter and same-id sources included, we now write the vector of interval deviations as $\mathbf{x}_{\text{j}} = \mathbf{W}\mathbf{z}_{\text{j}} + \mathbf{D}\mathbf{u}_{\text{j}} + \mathbf{Q}\mathbf{v}_{\text{j}} + \eta_{\text{j}}$. Using $\mathbf{\Sigma}$ to denote a diagonal matrix of jitter variance and $\mathbf{\Phi}$ to denote the diagonal covariance matrix of slow noise, the total covariance can now be written as

$$\mathbf{C} = \mathbf{W}\mathbf{W}^{\mathrm{T}} + \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^{\mathrm{T}} + \mathbf{Q}\boldsymbol{\Phi}\mathbf{Q}^{\mathrm{T}} + \boldsymbol{\Psi} \tag{D.3}$$

It is important to note that, unlike $\mathbf{W}$, weights $\mathbf{D}$ and $\mathbf{Q}$ are held constant and diagonal covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$ are estimated ($\boldsymbol{\Psi}$ is estimated both here and in traditional factor analysis).

We used an expectation-maximization (EM) algorithm to estimate latent variables and parameters; details are as follows (Bishop, 2006): We write the complete-data log likelihood function using the joint distribution $p(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{v})$. We assume $\mathbf{z}$, $\mathbf{u}$ and $\mathbf{v}$ to be independent, so this distribution can be rewritten as $p(\mathbf{x}|\mathbf{z}, \mathbf{u}, \mathbf{v})\mathrm{p}(\mathbf{z})\mathrm{p}(\mathbf{u})\mathrm{p}(\mathbf{v})$. We can now write the likelihood function $\mathcal{L}$ as

$$\mathcal{L} = \sum_{j} \log p(\mathbf{x}_{\mathrm{j}}|\mathbf{z}_{\mathrm{j}}, \mathbf{u}_{\mathrm{j}}, \mathbf{v}_{\mathrm{j}}) + \log p(\mathbf{z}_{\mathrm{j}}) + \log p(\mathbf{u}_{\mathrm{j}}) + \log p(\mathbf{v}_{\mathrm{j}}) \tag{D.4}$$

Factor analysis assumes that all observed and latent variables have a (multivariate) Gaussian distribution, which we denote using $\mathcal{N}(\mathbf{y}|\mu, \boldsymbol{\Omega})$, where $\mathbf{y}$, μ and $\boldsymbol{\Omega}$ are the variable, mean and covariance, respectively. We thus write each component of the likelihood function as:

$$p(\mathbf{x}|\mathbf{z},\mathbf{u},\mathbf{v}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \mathbf{D}\mathbf{u} + \mathbf{Q}\mathbf{v}, \mathbf{\Psi}) \tag{D.5}$$

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|0, \mathbf{I}) \tag{D.6}$$

$$p(\mathbf{u}) = \mathcal{N}(\mathbf{u}|0, \mathbf{\Sigma}) \tag{D.7}$$

$$p(\mathbf{v}) = \mathcal{N}(\mathbf{v}|0, \mathbf{\Phi}) \tag{D.8}$$

The EM algorithm alternately estimates the latent variables in the E step and parameters (weight matrices) in the M step.

For the E step, we generate expectations for the mean and covariance of the latent variables, which we denote $\mathcal{E}\{[\mathbf{z},\mathbf{u},\mathbf{v}]\}$ and $\mathcal{E}\{[\mathbf{z},\mathbf{u},\mathbf{v}][\mathbf{z},\mathbf{u},\mathbf{v}]^{\mathrm{T}}\}$. We do so using the joint conditional distribution $p(\mathbf{z},\mathbf{u},\mathbf{v}|\mathbf{x})$, and begin by defining the joint covariance matrix

$$\mathbf{C}_{\mathrm{z,u,v}} = \begin{bmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{\Sigma} & 0 \\ 0 & 0 & \mathbf{\Phi} \end{bmatrix} \tag{D.9}$$

We solve for the expectations of the marginal mean and covariance, ($\mathcal{E}\{[\mathbf{z},\mathbf{u},\mathbf{v}]\}$ and $\mathcal{E}\{[\mathbf{z},\mathbf{u},\mathbf{v}][\mathbf{z},\mathbf{u},\mathbf{v}]^{\mathrm{T}}\}$) using Bayes' Theorem for Gaussian variables. This yields the conditional covariance matrix,

$$\mathbf{G}_{\mathrm{z,u,v}} = (\mathbf{C}_{\mathrm{z,u,v}}^{-1} + [\mathbf{W},\mathbf{D},\mathbf{Q}]^{\mathrm{T}}\mathbf{\Psi}^{-1}[\mathbf{W},\mathbf{D},\mathbf{Q}])^{-1} \tag{D.10}$$

so the joint expectation of the jitter, repeated variability and global factors w.r.t. $\mathbf{x}$ can be written as

$$\mathcal{E}\{\mathbf{z},\mathbf{u},\mathbf{v}\} = \mathbf{G}_{\mathrm{z,u,v}}[\mathbf{W},\mathbf{D},\mathbf{Q}]^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{x} \tag{D.11}$$

while

$$\mathcal{E}\{[\mathbf{z},\mathbf{u},\mathbf{v}][\mathbf{z},\mathbf{u},\mathbf{v}]^{\mathrm{T}}\} = \mathrm{N}\mathbf{G}_{\mathrm{z,u,v}} + \mathcal{E}\{\mathbf{z},\mathbf{u},\mathbf{v}\}\mathcal{E}\{\mathbf{z},\mathbf{u},\mathbf{v}\}^{\mathrm{T}} \tag{D.12}$$

In the M step, we seek to maximize parameters $\mathbf{W}$, Φ, Ψ and Σ w.r.t. the new expectations of our latent variables described above. We can now partition out $\boldsymbol{\Sigma}_{new}$ and $\boldsymbol{\Phi}_{new}$ from $\mathcal{E}\{[\mathbf{z},\mathbf{u},\mathbf{v}][\mathbf{z},\mathbf{u},\mathbf{v}]^{\mathrm{T}}\}$ by writing

$$\mathbf{C}'_{\mathrm{z,u,v}} = \frac{1}{\mathrm{N}}\mathcal{E}\{[\mathbf{z},\mathbf{u},\mathbf{v}][\mathbf{z},\mathbf{u},\mathbf{v}]^{\mathrm{T}}\} \tag{D.13}$$

$$\boldsymbol{\Sigma}_{new} = \mathbf{C}'_{\mathrm{u}} \tag{D.14}$$

$$\boldsymbol{\Phi}_{new} = \mathbf{C}'_{\mathrm{v}} \tag{D.15}$$

where $\mathbf{C}'_{\mathrm{u}}$ and $\mathbf{C}'_{\mathrm{v}}$ are the respective components of $\mathbf{C}'_{\mathrm{z,u,v}}$ derived from $\mathbf{C}_{\mathrm{u}}$ and $\mathbf{C}_{\mathrm{v}}$ in $\mathbf{C}_{\mathrm{z,u,v}}$. We discard $\mathbf{C}'_{\mathrm{z}}$ because we have fixed this portion as $\mathbf{I}$.

We solve for $\mathbf{W}_{\mathrm{new}}$ in the M step using a similar strategy. The maximum likelihood solution to the joint parameters in each iteration of the EM algorithm $[\mathbf{W},\mathbf{D},\mathbf{Q}]$ can be written as

$$[\mathbf{W}_{\mathrm{new}}, \mathbf{D}', \mathbf{Q}'] = \mathbf{X}^{\mathrm{T}} \mathcal{E}\{[\mathbf{z}, \mathbf{u}, \mathbf{v}]\} \mathcal{E}\{[\mathbf{z}, \mathbf{u}, \mathbf{v}][\mathbf{z}, \mathbf{u}, \mathbf{v}]^{\mathrm{T}}\}^{-1} \tag{D.16}$$

where $\mathbf{X}$ is the matrix of all interval length deviations across song samples. Here, we can simply partition out $\mathbf{W}_{\mathrm{new}}$ and discard $\mathbf{D}'$ and $\mathbf{Q}'$ because we have fixed those weight matrices as described above.

It can be shown that these solutions to $\mathbf{W}$, $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ are precisely what one would find via the maxima of the partial derivatives of $\mathcal{L}$ w.r.t. to each parameter. However, the joint framework naturally accounts for the conditional dependencies among $\mathbf{z}$, $\mathbf{u}$ and $\mathbf{v}$.

Finally, our generation of $\mathbf{\Psi}_{new}$ in the M step can be found by letting $\partial\mathcal{L}/\partial\mathbf{\Psi} = 0$, which yields

$$\begin{aligned}
&\mathbf{\Psi}_{new} = \frac{1}{N}\mathrm{diag}\{ \\
&\mathbf{X}\mathbf{X}^{\mathrm{T}} + \\
&[\mathbf{W}, \mathbf{D}, \mathbf{Q}]\mathcal{E}\{[\mathbf{z}, \mathbf{u}, \mathbf{v}][\mathbf{z}, \mathbf{u}, \mathbf{v}]^{\mathrm{T}}\}[\mathbf{W}, \mathbf{D}, \mathbf{Q}]^{\mathrm{T}} - \\
&2\mathbf{X}^{\mathrm{T}}\ \mathcal{E}\{[\mathbf{z}, \mathbf{u}, \mathbf{v}]\}[\mathbf{W}, \mathbf{D}, \mathbf{Q}]^{\mathrm{T}}\}
\end{aligned} \tag{D.17}$$

We defined convergence for the EM algorithm when all $\mathbf{W}$ estimates started to change by < 0.1 msec, while remaining parameters changed by < 0.01 msec^2. For each covariance matrix we ran the EM algorithm 100 times with a randomly chosen set of initial conditions, in which each initial estimate of W was the corresponding value of the measured standard deviation scaled by a random number between

0 and 1, while initial estimates of remaining parameters were scaled variances of corresponding elements. If an algorithm failed to converge we discarded those parameters estimates and continued until 100 were reached for each covariance matrix. Across the 100 estimates per 8 covariance matrices fit for the control analysis, the algorithm never failed to converge. Across all 68 covariance matrices in the data sample for developmental analysis (3 covariance matrices for each bird-by-age sample, see Materials and Methods in Chapter 5), 62 never failed to converge. For 5 of the remaining covariance matrices, the EM algorithm failed to converge 2-8 times, while for remaining 2 the algorithm failed 69 and 98 times. These last two covariance matrices came from 65-dph samples with a relatively large amount of variance in at least one song interval.

Across all 100 parameter estimates per covariance matrix, we chose for analysis those estimates corresponding with the highest log-probability, calculated as $-N/2[d\log(2\pi) + \log(\det(\mathbf{C})) + \mathrm{Tr}((\mathbf{C}\tilde{\mathbf{C}})^{-1})]$, where $\tilde{\mathbf{C}}$ is the measured covariance matrix. Visual inspection of 10-15 different fits indicated convergence of parameters.

Appendix E

Tests of the measurement algorithms and acoustic chamber

We have presented an analysis of adult and juvenile zebra finch song that focuses almost entirely on naturally occurring timing variability measured from song recordings. However, observation suggests that birds frequently shift positions in the recording chamber, which in turn will yield variability in the angle and distance between the bird and microphone. It is possible that the spectral and temporal measurements of recorded song vary systematically as a bird changes position. How much of this variability explains the timing patterns we have analyzed?

We investigated this question by analyzing recordings of invariant song played in a recording chamber at different angles and distances with the microphone. Specifically, we compiled an audio signal consisting of two renditions of the same motif from 4 different birds (from Chapters 2 and 3), yielding a total of 60 syllables and gaps (31 unique). We then played the test signal through a speaker with a frequency response that covered the range of song acoustics we analyzed, and a $\sim 1/2$ in. diaphragm to approximate beak aperture. Signal playback was recorded through a directional microphone and processed with a real time digital signal processor that was coupled to the preamplifier to ensure that recording would be triggered upon playback with a fixed delay. In order to test the maximum amount of position variability possible, we recorded signals at 32 systematic combinations of speaker-

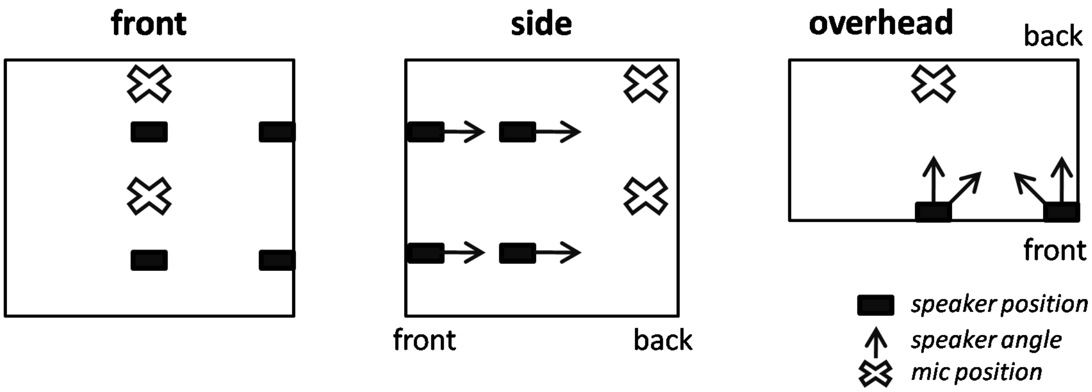

Figure E.1: Schematic of the different speaker-microphone distances and angles used for control data collection. Three separate views are shown in order to represent how positions were varied in all 3 dimensions.

microphone distance and angle (Fig. E.1), with 2 microphone heights, 8 speaker positions and 2 speaker angles per position.

We then analyzed the resulting recordings of syllables and silent gaps using the same procedure for constructing fine-grained templates detailed in Chapters 2 and 3 (Glaze & Troyer, 2006, 2007), and measured song timing with the spectral-based dynamic time warping (DTW) algorithm (Appendix C). We used the same set of song features in the control data as we had in the real data by mapping control templates to those generated for the real data using DTW.

All measurements were based on sound amplitude as opposed to log-amplitudes commonly used for analyzing song spectrograms. We have found that log-amplitudes tend yield more measurement error due to fluctuations in low-amplitude noise that often stems from both song production and background noise in the chamber. However, we have found that this added timing variability is not sufficiently large to alter our basic conclusions (see below and Chapter 5).

After compiling timing measurements across all 32 recordings, we compared the resulting variability and covariance patterns with what we have found in real data sets from the same birds using the same techniques (Chapter 5). If our previous findings on accumulated timing variability are not explained by changes in bird position, then the dominant source of variability we expect in the control data is jitter, and song length variance should be small since it depends only on jitter in the onset of the first syllable and offset of the last. The lack of appreciable song length variation in the control data in turn would preclude use of linear regression (see Methods in Chapters 2 and 3) to separate out any possible artificial tempo factor.

Thus, in order to compare both data sets using the same statistical model, we applied the factor analysis algorithm used to measure timing parameters in the developmental data (Chapter 5, Appendix D). Any acoustic variability will cause slight errors in the timing assignments of some features, so we expected some degree of timing jitter in the control data. However, to the extent that the FA algorithm successfully separates out this factor, there should be no appreciable tempo change, repeated timing variability or independent timing variability that accumulates over the length of each song (it should be noted that variability will always be at least slightly > 0 since the FA model constrains all matrices to be positive semi-definite). Below we report control variability as a percentage of real variability as well as raw timing measurements, all as means ± standard error.

On average, control timing standard deviation was $58.0 \pm 4.3\%$ of the real data, with respective averages of 1.172 ± 0.062 and 2.184 ± 0.138 msec, and the difference was statistically significant (Wilcoxon signed-rank test, $p < 0.0001$). As expected,

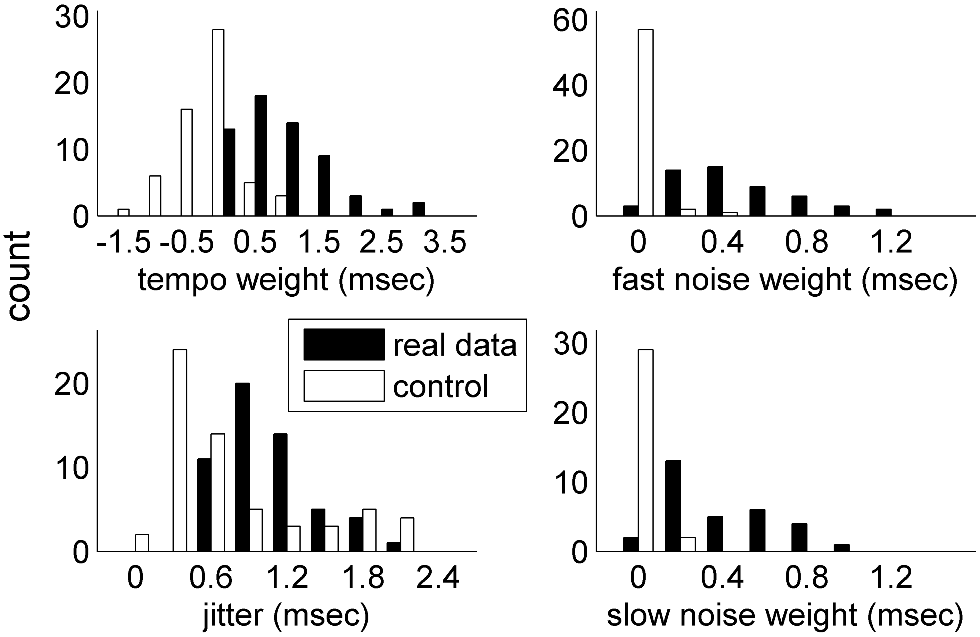

Figure E.2: Distributions of real and control timing parameters measured by the factor analysis algorithm. Differences among the distributions indicate significantly greater variability in the real data, especially among non-jitter parameters.

control song length CV was considerably smaller, averaging 9.1 ± 2.3% of real song length CV, with respective CVs of 0.1 ± 0.02% and 1.4 ± 0.03%.

In fact, the analysis suggests that jitter is the only variability source in which the real and control syllable-gap data are comparable (Fig. E.2). Here, average variability due to jitter in the control data was 74.9 ± 12.2% of corresponding real data, with values of 1.215 ± 0.206msec and 2.357 ± 0.411 respectively. By contrast, control global variability was 11.4 ± 12.8% of the real data, averaging 0.049 ± 0.075 vs. 1.059 ± 0.087 msec; control repeated variability was 23.3 ± 3.1% of the real data, with averages of 0.100 ± 0.011 vs. 0.494 ± 0.049 msec; while control independent variability was 9.0 ± 1.3% of the real data, averaging 0.059 ± 0.011vs. 0.817 ± 0.099 msec. Pairwise differences between control and real data were statistically significant across all timing parameters (Wilcoxon signed-rank test, $p < 0.0001$).

Thus, the control data as a whole suggest that the timing patterns we have investigated are not due to artifact from variability in the distance and angle between the bird and microphone.

Bibliography

Abarbanel, H. D., Gibb, L., Mindlin, G. B., Rabinovich, M. I., & Talathi, S. (2004). Spike timing and synaptic plasticity in the premotor pathway of birdsong. *Biol Cybern*, *91*(3), 159–67.

Abeles, M. (1991). *Corticonics.* Cambridge: Cambridge Univ Press.

Abeles, M., Prut, Y., Bergman, H., & Vaadia, E. (1994). Synchronization in neuronal transmission and its importance for information-processing. In *Progress in Brain Research 102: The self-organizing brain: From growth cones to functional networks* (pp. 395–404). Amsterdam: Elsevier.

Ackermann, H. & Hertrich, I. (2000). The contribution of the cerebellum to speech processing. *J Neuroling*, *13*, 95–116.

Aldridge, J. W., Berridge, K. C., & Rosen, A. R. (2004). Basal ganglia neural mechanisms of natural movement sequences. *Can J Physiol Pharmacol*, *82*(8-9), 732–9.

Anderson, S. E., Dave, A. S., & Margoliash, D. (1996). Template-based automatic recognition of birdsong syllables from continuous recordings. *J Acoust Soc Am*, *100*(2 Pt 1), 1209–19.

Arnoldi, H. M. & Brauer, W. (1996). Synchronization without oscillatory neurons. *Biol Cybern*, *74*(3), 209–23.

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

Baker, S. & Lemon, R. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J Neurobiol*, *84*(4), 1770–1780.

Barnes, T., Kubota, Y., Hu, D., Jin, D., & Graybiel, A. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, *437*, 1158–1161.

Baum, S. & Pell, M. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology*, *13*(8), 581–608.

Beggs, J. & Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J Neurosci*, *23*(35), 11167–11177.

Beggs, J. & Plenz, D. (2004). Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures. *J Neurosci*, *24*(22), 5216–5229.

Bienenstock, L. (1995). A model of neocortex. *Network: Comp Neural Sys*, *6*, 179–224.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Bohland, J. & Guenther, F. (2006). An fMRI investigation of syllable sequence production. *Neuroimage*, *32*, 821–841.

Bottjer, S., Miesner, E., & Arnold, A. (1984). Lesions disrupt development but not maintenance of song in passerine birds. *Science*, *224*(4651), 901–903.

Brainard, M. & Doupe, A. (2000). Auditory feedback in learning and maintenance of vocal behavior. *Nature Rev*, *1*, 31–40.

Brainard, M. S. & Doupe, A. J. (2001). Postlearning consolidation of birdsong: Stabilizing effects of age and anterior forebrain lesions. *J Neurosci*, *21*(7), 2501–2517.

Byrd, D. & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J Phoenetics*, *31*, 149–180.

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

Coleman, M. & Vu, E. (2005). Recovery of impaired songs following unilateral but not bilateral lesions of nucleus uvaeformis of adult zebra finches. *J Neurobiol*, *63*(1), 70–89.

Cooper, B. G. & Goller, F. (2006). Physiological insights into the social-context-dependent changes in the rhythm of the song motor program. *J Neurophysiol*, *95*, 3798–3809.

Crandall, S. R., Aoki, N., & Nick, T. A. (2007). Developmental modulation of the temporal relationship between brain and behavior. *J Neurophysiol*, *97*(1), 806–16.

Cynx, J. (1990). Experimental determination of a unit of song production in the zebra finch (taeniopygia guttata). *J Comp Psychol*, *104*(1), 3–10.

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*, 179–229.

Dan, Y. & Poo, M. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron*, *44*, 23–30.

de Jong, K. J. (2001). Effects of syllable affiliation and consonant voicing on temporal adjustment in a repetitive speech-production task. *J Speech Lang Hear Res*, (4), 826–840.

Deregnaucourt, S., Mitra, P. P., Feher, O., Pytte, C., & Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature*, *433*(7027), 710–6.

Diesmann, M., Gewaltig, M., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, *402*, 529–533.

Dooling, R., Leek, M., Gleich, O., & Dent, M. (2002). Auditory temporal resolution in birds: discrimination of harmonic complexes. *J. Acoust. Soc. Am.*, *112*(2), 748–59.

Doupe, A. & Kuhl, P. (1999). Birdsong and human speech: Common themes and mechanisms. *Annu Rev Neurosci*, *22*, 567–631.

Duda, R., Hart, P., & Stork, D. (2000). *Pattern Classification.* New York, NY: Wiley-Interscience, 2nd ed.

Fee, M. S., Kozhevnikov, A. A., & Hahnloser, R. H. (2004). Neural mechanisms of vocal sequence generation in the songbird. *Ann N Y Acad Sci*, *1016*, 153–70.

Fiete, I. R., Hahnloser, R. H. R., Fee, M. S., & Seung, H. S. (2004). Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *J Neurophysiol*, *92*(4), 2274–2282.

Foster, E. & Bottjer, S. (2001). Lesions of a telencephalic nucleus in male zebra finches: Influences on vocal behavior in juveniles and adults. *J Neurobiol*, *46*(2), 142–165.

Franz, M. & Goller, F. (2002). Respiratory units of motor production and song imitation in the zebra finch. *J Neurobiol*, *51*(2), 129–41.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cog Sci*, *6*(2), 78–84.

Froemke, R. & Dan, Y. (2002). Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature*, *416*, 433–438.

Fujii, N. & Graybiel, A. M. (2005). Time-varying covariance of neural activities recorded in striatum and frontal cortex as monkeys perform sequential-saccade tasks. *Proc Natl Acad Sci U S A*, *102*(25), 9032–7.

Gentner, D. R. (1987). Timing of skilled motor performance: Tests of the proportional duration model. *Psychol Rev*, (2), 255–276.

Gerstein, G. (2004). Searching for significance in spatio-temporal firing patterns. *Acta Neurobiol Exp*, *64*(2), 203–207.

Gilden, D. (2001). Cognitive emissions of 1/f noise. *Psychol Rev*, *108*(1), 33–56.

Glaze, C. M. & Troyer, T. W. (2006). Temporal structure in zebra finch song: Implications for motor coding. *J Neurosci*, *26*(3), 991–1005.

Glaze, C. M. & Troyer, T. W. (2007). Behavioral measurements of a temporally precise motor code for birdsong. *J Neurosci*, *27*(29), 7631–7639.

Goller, F. & Cooper, B. G. (2004). Peripheral motor dynamics of song production in the zebra finch. *Ann N Y Acad Sci*, *1016*, 130–52.

Graybiel, A. (2005). The basal ganglia: learning new tricks and loving it. *Curr Opin Neurobiol*, *15*, 638644.

Guenther, F., Ghosh, S., & Tourville, J. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Lang*, *86*, 280–301.

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

Hessler, N. A. & Doupe, A. J. (1999). Social context modulates singing-related neural activity in the songbird forebrain. *Nat Neurosci*, *2*(3), 209–11.

Heuer, H. (1988). Testing the invariance of relative timing: comment on gentner (1987). *Psychol Rev*, *95*(4), 552–8.

Hikosaka, O., Nakamura, K., Sakai, K., & Nakahara, H. (2002). Central mechanisms of motor skill learning. *Curr Opin Neurobiol*, *12*(2), 217–22.

Hough, G. & Volman, S. (2002). Short-term and long-term effects of vocal distortion on song maintenance in zebra finches. *J Neurosci*, *22*(3), 1177–1186.

Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., & Yuste, R. (2004). Synfire chains and cortical songs: Temporal modules of cortical activity. *Science*, *304*(5670), 559–564.

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

Jansen, R., Metzdorf, R., van der Roest, M., Fusani, L., ter Maat, A., & Gahr, M. (2005). Melatonin affects the temporal organization of the song of the zebra finch. *Faseb J*, *19*(7), 848–50.

Kao, M. H. & Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J Neurophysiol*, *96*, 1441–1455.

Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature*, *433*(7026), 638–43.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychol Rev*, *110*(2), 316–339.

Kimpo, R. R., Theunissen, F. E., & Doupe, A. J. (2003). Propagation of correlated activity through multiple stages of a neural circuit. *J Neurosci*, *23*(13), 5750–61.

Kobayashi, K., Uno, H., & Okanoya, K. (2001). Partial lesions in the anterior forebrain pathway affect song production in adult bengalese finches. *Neuroreport*, *12*(2), 353–8.

Konishi, M. (1965). The role of auditory feedback in the control of vocalizations in the white-crowned sparrow. *Z Tierpsychol*, *22*, 770–783.

Konishi, M. (1985). Birdsong: from behavior to neuron. *Annu Rev Neurosci*, *8*, 125–170.

Kozhevnikov, A. & Fee, M. (2007). Singing-related activity of identified HVC neurons in the zebra finch. *J Neurophysiol*, *97*, 42714283.

Lashley, K. (1951). The problem of serial order in behavior. In L. Jeffress (Ed.), *Cerebral Mechanisms in Behavior* (pp. 112–146). New York: Wiley.

Leonardo, A. (2004). Experimental test of the birdsong error-correction model. *Proc Natl Acad Sci U S A*, *101*(48), 16935–16940.

Leonardo, A. & Fee, M. S. (2005). Ensemble coding of vocal control in birdsong. *J Neurosci*, *25*(3), 652–61.

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

Liu, W., Gardner, T., & Nottebohm, F. (2003). Juvenile zebra finches can use multiple strategies to learn the same song. *Proc. Natl. Acad. Sci. USA*, *101*(52), 1817718182.

Lohr, B., Dooling, R., & Bartone, S. (2006). The discrimination of temporal fine structure in call-like harmonic sounds by birds. *J. Comp. Psychol.*, *120*(3), 239–51.

Luo, H. & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*, 1001–1010.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav Brain Sci*, *21*(4), 499–511; discussion 511–46.

Marler, P. (1970). A comparative approach to vocal learning: song development in white-crowned sparrows. *J Comp Physiol Psychol*, *71*, 1–25.

McCasland, J. (1987). Neuronal control of bird song production. *J Neurosci*, *7*(1), 23–39.

Melamed, O., Gerstner, W., Maass, W., Tsodyks, M., & Markram, H. (2004). Coding and learning of behavioral sequences. *Trends Neurosci*, *27*(1), 11–14.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). The unit of analysis. In *Plans and the Structure of Behavior* (pp. 21–39). New York: Henry Holt and Company.

Mooney, R. (2000). Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the zebra finch. *J Neurosci*, *20*(14), 54205436.

Mooney, R. & Prather, J. F. (2005). The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *J Neurosci*, *25*(8), 1952–64.

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, *Serinus canarius*. *J Comp Neurol*, *165*(4), 457–86.

Nowotny, T., Rabinovich, M., & Abarbanel, H. (2003). Spatial representation of temporal information through spike-timing-dependent plasticity. *Phys Rev E Stat Nonlin Soft Matter Phys*, *68*(011908), 1–11.

Ollason, J. & Slater, P. J. B. (1973). Changes in the behaviour of the male zebra finch during a 12-hr day. *Anim Behav*, *21*, 191–196.

Ölveczky, B. P., Andalman, A. S., & Fee, M. S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol*, *3*(5), 902–9.

Oram, M., Wiener, M., Lestienne, R., & Richmond, B. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J Neurophysiol*, *81*(6), 3021–3033.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time”. *Speech Comm*, *41*, 245–255.

Poeppel, D. & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones*. Philadelphia: Lawrence Erlbaum.

Poeppel, D. & Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, *92*, 1–12.

Rabiner, L. & Juang, B. (1993). Time alignment and normalization. In *Fundamentals of Speech Recognition*, Prenctice Hall Signal Processing Series (pp. 200–241). Englewood Cliffs, NJ: Prentice-Hall PTR.

Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., & Page, M. P. (2004). Learning and production of movement sequences: behavioral, neurophysiological, and modeling perspectives. *Hum Mov Sci*, *23*(5), 699–746.

Riehle, A., Grun, S., Diesmann, M., & Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, *278*(5345), 1950–1953.

Scharff, C. & Nottebohn, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: Implications for vocal learning. *J Neurosci*, *11*(9), 2896–2913.

Schmidt, M. F. (2003). Pattern of interhemispheric synchronization in HVC during singing correlates with key transitions in the song pattern. *J Neurophysiol*, *90*(6), 3931–49.

Sidtis, D. (2006). Does functional neuroimaging solve the questions of neurolinguistics? *Brain Lang*, *98*(3), 276–290.

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

Solis, M. M. & Perkel, D. J. (2005). Rhythmic activity in a forebrain vocal control nucleus in vitro. *J Neurosci*, *25*(11), 2811–22.

Sommer, F. T. & Wennekers, T. (2005). Synfire chains with conductance-based neurons: internal timing and coordination with timed input. *Neurocomputing*, *65*, 449–454.

Sommer, M. & Wurtz, R. (2002). A pathway in primate brain for internal monitoring of movements. *Science*, *296*, 1480–1482.

Song, S. & Abbott, L. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, *32*, 339–350.

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

Spiro, J. E., Dalva, M. B., & Mooney, R. (1999). Long-range inhibition within the zebra finch song nucleus ra can coordinate the firing of multiple projection neurons. *J Neurophysiol*, *81*(6), 3007–20.

Striedter, G. & Vu, E. (1998). Bilateral feedback projections to the forebrain in the premotor network for singing in zebra finches. *J Neurobiol*, *34*, 27–40.

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

Sturdy, C. B., Wild, J. M., & Mooney, R. (2003). Respiratory and telencephalic modulation of vocal motor neurons in the zebra finch. *J Neurosci*, *23*(3), 1072–86.

Suthers, R. A. & Margoliash, D. (2002). Motor control of birdsong. *Curr Opin Neurobiol*, *12*(6), 684–90.

Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science*, *291*(5513), 2564–9.

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

Thompson, J. & Johnson, F. (2007). HVC microlesions do not destabilize the vocal patterns of adult male zebra finches with prior ablation of LMAN. *Devel Neurobiol*, *67*(2), 205 – 218.

Tresch, M., Cheung, V., & d'Avella, A. (2006). Matrix factorization algorithms for the identification of muscle synergies: Evaluation on simulated and experimental data sets. *J Neurophysiol*, *95*, 2199–2212.

Troyer, T. W. & Doupe, A. J. (2000a). An associational model of birdsong sensorimotor learning i. efference copy and the learning of song syllables. *J Neurophysiol*, *84*, 1204–1223.

Troyer, T. W. & Doupe, A. J. (2000b). An associational model of birdsong sensorimotor learning ii. temporal hierarchies and the learning of song sequence. *J Neurophysiol*, *84*, 1224–1239.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., & Aersten, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioral event. *Nature*, *373*(6514), 515–518.

Vates, G., Vicario, D., & Nottebohm, F. (1997). Reafferent thalamo-"cortical" loops in the song system of oscine songbirds. *J Comp Neurol*, *380*(2), 275–290.

Vu, E. T., Mazurek, M. E., & Kuo, Y. C. (1994). Identification of a forebrain motor programming network for the learned song of zebra finches. *J Neurosci*, *14*(11 Pt 2), 6924–34.

Wennekers, T. & Palm, G. (1996). Controlling the speed of synfire chains. In *Artificial Neural Networks ICANN 96*, Lecture Notes in Computer Science (pp. 451–456). Heidelberg: Springer Berlin.

Wild, J. M., Goller, F., & Suthers, R. A. (1998). Inspiratory muscle activity during bird song. *J Neurobiol*, *36*(3), 441–53.

Wild, J. M., Li, D., & Eagleton, C. (1997). Projections of the dorsomedial nucleus of the intercollicular complex (dm) in relation to respiratory-vocal nuclei in the brainstem of pigeon (*Columba livia*) and zebra finch (*Taeniopygia guttata*). *J Comp Neurol*, *377*(3), 392–413.

Williams, H. (2004). Birdsong and singing behavior. *Ann N Y Acad Sci*, *1016*, 1–30.

Williams, H., Cynx, J., & Nottebohm, F. (1989). Timbre control in zebra finch (*Taeniopygia guttata*) song syllables. *J Comp Psychol*, *103*(4), 366–380.

Williams, H. & Mehta, N. (1999). Changes in adult zebra finch song require a forebrain nucleus that is not necessary for song production. *J Neurobiol*, *39*, 14–28.

Williams, H. & Staples, K. (1992). Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *J Comp Psychol*, *106*(3), 278–286.

Williams, H. & Vicario, D. S. (1993). Temporal patterning of song production: participation of nucleus uvaeformis of the thalamus. *J Neurobiol*, *24*(7), 903–12.

Woolley, S. C. & Doupe, A. J. (2008). Social context - induced song variation affects female behavior and gene expression *PLoS Biol*, *6*(3), 525–537.

Yu, A. C. & Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science*, *273*(5283), 1871–5.

Zann, R. (1993). Structure, sequence and evolution of song elements in wild australian zebra finches. *Auk*, (4), 702–715.

Zann, R. (1996). *The Zebra Finch: A Synthesis of Field and Laboratory Studies.* New York: Oxford University Press.