Title

Encoding of Syntax and Phonology in a Premotor Songbird Nucleus

Author

Swartz, Kurtis Jon

Publication Date

2024

Peer reviewed|Thesis/dissertation

Encoding of Syntax and Phonology in a Premotor Songbird Nucleus

by
Kurtis Jon Swartz

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Loren Frank
9C247AE9D5474C3...

Loren Frank
Chair

DocuSigned by:
Saul Kato

Saul Kato

DocuSigned by:
Vikaas Sohal

Vikaas Sohal

DocuSigned by:
Michael Brainard
3A5C1850C2F1412...

Michael Brainard

Committee Members

Acknowledgements

The successful completion of my PhD was a grueling process but would have been impossible without the support of my friends, family, classmates, mentors, and coworkers. First, I would like to thank my advisor, Michael, who I always knew wanted the best for me and cared for me first as a person. He taught me to think critically and creatively about the data to make sense of things. Thank you to all my lab mates who were there to support and encourage me constantly. Ziggy for being a friend and fellow graduate student. Eric for always making me laugh and stopping me from being so hard on myself. Adria for her dedication to running the lab, without her, it would fall apart. Saipriya for being a great learner and dedicated worker. Finally, Hamish, who without his mentorship and support none of this project would be possible. For all my amazing friends, thank you for always stopping me from taking myself too seriously and keeping me having fun and loving life outside of lab. I can't imagine life without you. For my mom, thank you for always having a fun-loving attitude and instilling in me hard work, a humble attitude, and positivity. For my dad, thank you for showing me what a true passion for science looks like and making me fall in love with it. Finally, thank you to Emma for being my best friend and partner in life. Always having you to come home to makes whatever difficulties happened that day fall into the background. I wouldn't have made it this far without you.

Encoding of Syntax and Phonology in a Premotor Songbird Nucleus

Kurtis J. Swartz

Abstract

A major question in motor systems neuroscience is how complex actions are encoded, on the timescale of both shorter individual elements and longer sequences. Birdsong is a tractable system, where a learned complex vocal behavior combines a categorical set of shorter individual elements into longer sequences, making it well-suited to address this question. The song nucleus HVC (used as a proper name) contributes to song sequence and timing. While much has been studied about HVC in zebra finches, which sing linear, stereotyped songs through their adult lives, relatively little data has been collected from HVC in their close relatives, Bengalese finches, that sing flexible, variably sequenced songs. We built a custom microscope to record neural activity reported by the calcium indicator GCaMP from populations of neurons from HVC in awake, freely moving Bengalese finches. We analyzed how populations of neurons in HVC encode information around divergence points, where one syllable can be followed by multiple syllables, and convergence points, where one syllable can be preceded by multiple syllables. We found that HVC projection neuron bursting can encode for upcoming sequence many syllables ahead of divergence points, and prior sequence many syllables after convergence points. We also found that HVC bursting encodes variation in the acoustic structure (phonology) of the different renditions of a given syllable in different contexts. Moreover, we found that HVC has overlapping representations of distinct syllables, especially those which are acoustically similar. These results help to reveal how premotor regions can encode multiple types of sequence and phonological information simultaneously.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

RA = Robust nucleus of the arcopallium

HVC is used as a proper name

AFP = Anterior Forebrain Pathway

COM = Center of Mass

PCA = Principal Components Analysis

SVM = Support Vector Machine

Chapter 1: Introduction

1.1 Motor Control of Variable Sequential Actions

Many complex motor behaviors, such as speech or playing the piano are performed by sequencing together smaller units of action. These sequences require constant, moment to moment, decisions about which smaller action should be taken next. How does the brain represent individual actions and organize them into sequences? When these actions can be variably sequenced, how is the selection of what comes next (or what has just been produced) encoded in brain activity? Additionally, when a given action is produced in different motor contexts (i.e. different sequences), how similar is the neural representation of that action across contexts? Answers to such questions are fundamental to understanding how the brain organizes and produces complex sequences of motor actions yet remain difficult to study in humans.

Singing in oscine finches parallels these complex motor behaviors as song is formed by sequentially ordering a set of acoustically distinct vocalizations, called syllables, into larger motifs (Brainard and Doupe 2013). These birds are easily maintained in lab settings and have a well-defined neural circuitry specialized for song (Nottebohm et al. 1976; Konishi et al. 1985; **Figure 1.1 B**). Moreover, across different species of finches, there is a range in the complexity of song both in the number and structure of the individual syllables, and in the complexity of the syntax that defines how these syllables are sequenced. By understanding the neural mechanisms of song production, we can begin to gain a better understanding of how the brain strings together individual actions into more complex behaviors.

The song system is in part composed of the premotor cortical analog HVC (proper name) which projects to the motor cortical analog RA (robust nucleus of the arcopallium) as well as the basal ganglia homolog Area X. HVC contains two main classes of cells: inhibitory interneurons, and excitatory projection neurons which send axons to either Area X or RA (Reiner et al, 2004). The properties of these cell types have been most extensively studied in zebra finches. In these finches, both classes of projection neurons fire precisely timed bursts that are locked to specific vocal elements (Vu et al. 1994; Fee et al. 2004). Each RA projection neuron fires a single burst associated with one syllable while Area X projection neurons fire bursts during multiple different syllables (typically 2-3; Yu and Margoliash, 1996; Hahnloser et al., 2002). Together, these neurons exhibit a unique population level pattern of firing at each moment in song. It has therefore been hypothesized that HVC neurons encode the timing and sequence of each of the vocal elements of song (Yu et al. 1996; Lynch et al. 2016; Picardo et al. 2016; Katlowitz et al. 2018). Further experiments showed that cooling of HVC led to slowing down of song tempo without grossly altering song acoustics, further supporting this idea (Long et al. 2008 2010; Andalman et al. 2011; Zhang et al. 2017).

While much has been studied about HVC's role in song timing and sequencing, most of these studies were done in zebra finches which do not display variable song sequences. Instead, zebra finches produce syllables in a single stereotyped sequence across each song rendition. This lack of variability in sequencing makes it difficult to differentiate activity related to the ordering of syllables from the activity associated with the production of each individual syllable. In contrast, Bengalese finches, close relatives of zebra finches, can display variability in the ordering of syllables (Nakamura and Okanoya 2004; Kentaro et al. 2013; Kentaro et al. 2011). Despite these differences in song syntax, the neural circuitry responsible for song in zebra finch and Bengalese

finches is strikingly similar. Both species contain the same nuclei responsible for the production of song, which are well conserved across all passerine songbirds examined to date. Additionally, transcriptomic profiles of song nuclei across zebra and Bengalese finches demonstrate highly conserved cell types (Colquitt et al. 2021). The fact that both a stereotyped behavior as well as a more variable behavior are underpinned by very similar circuitry provides an excellent opportunity to enrich our understanding of neural control of complex motor sequences.

The sequence variability in Bengalese finches occurs at "branch points" in song and comes in three main types: divergences, convergences, and variable repeats (**Figure 1.1. C, D, E**). Divergence points are places in song where one syllable can lead to multiple other syllables (syllable "a" goes to either "f" or "b" in **Figure 1.1. C/D**). Convergence points occur when one syllable is sung after multiple different syllables (syllable "i" is preceded by "e" and "c" in **Figure 1.1. C/E**). Finally, variable repeats are syllables which are repeated a varying number of times before transitioning to another song element (syllable "c" in **Figure 1.1. A/C)**. These branch points have consistent statistics which remain stable over months, suggesting that the brain must have some representation that biases these statistics (Jin et al 2011, Kentaro et al. 2011 and 2013). Additionally, Bengalese finches can be trained to alter the probabilities of transitions at these branch points, suggesting active neural control over decisions of song sequence (Warren et al. 2012).

The sequence variability in Bengalese finch song provides opportunities to study how HVC encodes variable transitions. First, divergence points appear to be probabilistic transitions where the bird has two or more possibilities of what to sing in the future. One possibility is that the selection of which syllable to sing is made just before the syllable itself. In this case we may expect to see evidence of this selection very close to the divergence point. Another possibility is that this

selection is made a few or many syllables in advance of the divergence. Here, we might expect to see neurons which fire only to one type of transition multiple syllables in advance of the divergence point. By analyzing when activity in HVC is predictive of the upcoming sequence, we can gain insights into the timescale at which the syllable selection process is made.

Convergence points provide another key feature of song to study as one syllable can be sung in two different preceding contexts. This raises the question of whether HVC represents context as well as syllable identity. For example, is a syllable "a" sung after a "b" the same as an "a" sung after a syllable "c". These "a's" sung in different contexts could either have the same representation in HVC or have information about the preceding sequence. Such information about preceding context could enable subsequent transitions to be changed based on past sequence. Indeed, we see that birds can adjust probabilities of transitions at divergence points dependent on the preceding sequence (Warren et al. 2012). Additionally, birds can learn to adjust the pitch of a given syllable in one preceding sequence, but not another (Tian and Brainard 2017). This reflects some capacity to jointly represent the syllable and the sequence in which it is produced. Understanding how information about previous sequences is encoded within HVC will constrain models of how sequence history affects subsequent syllable production.

Two previous studies have provided significant insight into how HVC encodes variable sequences in songbirds: one in Bengalese finches and one in canaries (Fujimoto et al. 2011 and Cohen et al. 2020 respectively). Fujimoto and colleagues studied song syntax within HVC of Bengalese finches by recording single unit HVC_X projecting neurons individually. They sequentially recorded 48 individual neurons across nine Bengalese finches. They found that similar to what was described in zebra finches, most HVC_X projection neurons fired bursts during multiple distinct points in song. They defined multiple types of HVC_X neurons based on the specificity of

their bursts. Some HVC_X neurons were termed “syllable selective” because they consistently fired for one syllable no matter what context it was sung in (for example a firing for syllable “b” in both an “a→b” and a “f→b” context). Other neurons were termed “transition selective”, preferring to fire during one context vs the other (for example firing more for syllable “d” in a “b→ d” context vs a “g→d” context). They further split these “transition selective” neurons into two additional classes. They termed these transition selective neurons “all-or-none” if they fired only in one context and “intermediate” selective if they fired during multiple contexts and fired more spikes in one context. These data showed that HVC can encode information about past sequences, rather than encoding only what is happening at that exact moment in song as had been hypothesized in zebra finch HVC.

At divergence points Fujimoto and colleagues found “syllable selective” and “intermediate” transition selective neurons, but no “all-or-none” transition selective neurons (Fujimoto et al. 2011). The “intermediate” transition selective neurons encode future transitions one syllable prior to it being sung. For example, such neurons might fire more during a syllable “a” that transitions to syllable “b” than during an “a” that transitions to syllable “i”. Hence, HVC can encode information about variable transitions at least one syllable in advance. While this work laid the framework for understanding how HVC can encode syntax, it is limited in scope by the small number of neurons recorded and the fact that they were recorded sequentially rather than simultaneously. As this work found significant variability in the firing of HVC projection neurons, understanding if this variability is shared across the population may provide great insight into the mechanisms of production for variably sequenced songs.

Cohen et al. 2020 began to address this limitation by using calcium imaging to record from many HVC projection neurons (likely including both HVC_X and HVC_{RA}) in canaries. Canary song

differs from zebra finch or Bengalese finch song as it contains many different trills (similar to variable repeats) which last a few seconds each (Markowitz et al. 2013). These trills are treated similarly to individual syllables in Bengalese finch song as they can also transition to multiple other trills, creating convergence and divergence points. They found that individual projection neurons could encode information about either past or future sequences. Such neurons could reflect what was or would be sung up to 2 phrases in the past or future. As these trills last many seconds, they conclude that HVC can maintain a representation of past transitions over many seconds. Furthermore, HVC neurons can encode information about subsequent transitions several seconds before they occur. As canary song is very different from finch song, with transitions between different elements being sparse in time, there are still many questions on how past and future transitions are encoded in a song with shorter timescale transitions.

This previous work still leaves many open questions to address. While Cohen et al. recorded from large populations of neurons, they did not focus their analysis on population metric which may encode features not seen in individual neurons. As Fujimoto et al. only looked at shorter range information encoded within HVC, does Bengalese finch HVC also contain long range past and future information? To best address these questions, we recorded from large populations of HVC projection neurons using calcium imaging in Bengalese finches. We found that HVC in Bengalese finches encodes both long range information reflective of both past and future transitions in a more subtle fashion than seen in canaries. Additionally, we report that HVC neurons contain strong correlations in their firing strength across renditions. Our results indicate that the HVC population in Bengalese finches influences upcoming sequences far in advance but is not fully predictive.

1.2 Encoding of Acoustic Features

In speech, humans use a given word in many different sentences. While this word may sound different each time we say it, we understand that we are saying that same word and our brains likely maintain a similar representation of its production. Additionally, many different words are made up of similar sounds. By encoding words which sound similar with overlapping representations, the brain could more efficiently encode a large vocabulary. Learning how the brain represents individual vocal elements across contexts is an important question in understanding how the brain generally encodes distinct actions.

Similar to human speech, Bengalese finch song contains individual elements that are easily distinguishable called syllables. One fundamental question in speech is how the brain encodes variations in the production of the same word as every time we say the same word, it doesn't sound exactly the same. Slight variations in duration, pitch, entropy, and other features make each time a word is said, or a syllable is sung unique. Yet, we are easily able to distinguish which words are the same. Just as in human speech, the same syllable sung at different times has a great deal of acoustic variability. Additionally, there are many distinct syllables with similar acoustic elements. Given that Bengalese finch song is variability sequenced, context cues about which syllables were sung ahead of a given syllable can help to differentiate whether two iterations of a syllable should be given the same or different label. However, this can often be a difficult task deciding which syllables should be grouped together or split apart. We can begin to address this question by looking into the neural representation of individual syllables during song production. As we know that HVC should encode information about syllable identity, recordings from HVC can help identify which syllables are the same and which are different.

While we expect the same syllables to be encoded by similar neurons within HVC, it is unclear whether syllables which are clearly distinct but are acoustically similar may have overlapping representations. Prior work has indicated that HVC appears to operate as a clock and only encodes sequence and timing in song (Lynch et al. 2016, Picardo 2016). RA then controls the mapping of individual acoustic elements in song (Leonardo and Fee 2005; Sober et al. 2008). Yet, as HVC projects directly to RA, one might expect that small trial by trial variations in HVC firing during the same syllable could influence its overall acoustic features. Indeed, work from our lab has found that infusion of the cholinergic agonist carbachol into HVC, increases the firing rate of multi-unit sites within HVC and in turn alters syllable acoustics, such as pitch (Jaffe 2020). Other work has found infusion of drugs which alter inhibitory interneurons within HVC also influence spectral features of song (Isola et al. 2020). These findings raise the question of whether natural variations in the acoustic properties of syllables may be encoded within HVC.

Past recordings of HVC have reported little on correlations of the location of HVC bursts and the phonology of syllables. Lynch et al. 2016 found that burst rate in HVC projection neurons is correlated only with pitch goodness (metric of how harmonic a syllable is) when looking at 6 different acoustic features. However, they did not analyze whether representations of similar syllables are more similar within HVC. This may be because in zebra finches HVC_{RA} projecting neurons only fire one burst per motif, making these sorts of comparisons impossible. Nevertheless, the possibility that HVC_X projection neurons, which fire multiple bursts per motif, may encode syllable similarity has not been fully examined. Additionally, HVC coding of sequence variable Bengalese finch songs may look very different from the stereotyped zebra finch song. Sequence variability leads to many questions about which syllables are the same or different as there is no longer consistent context to guide human labels. Song often contains

multiple distinct syllables with some shared acoustic elements (stacks, sweeps, etc.), yet with enough differences in other elements (duration and frequency) to make them separable. By analyzing how HVC represents syllables which are acoustically similar, we can better understand which features of song HVC encodes.

By analyzing large populations of HVC projection neurons recorded in singing Bengalese finches, we correlated the neural activity associated with the production of each syllable with how similar each syllable combination was. We found that HVC encodes variation in the same syllable sung in different contexts. Additionally, we report that syllables which are acoustically similar are also similar in neural activity within HVC. These results show that HVC can encode phonological features of song as well as sequence and timing information, expanding our understanding of its role in vocal production.

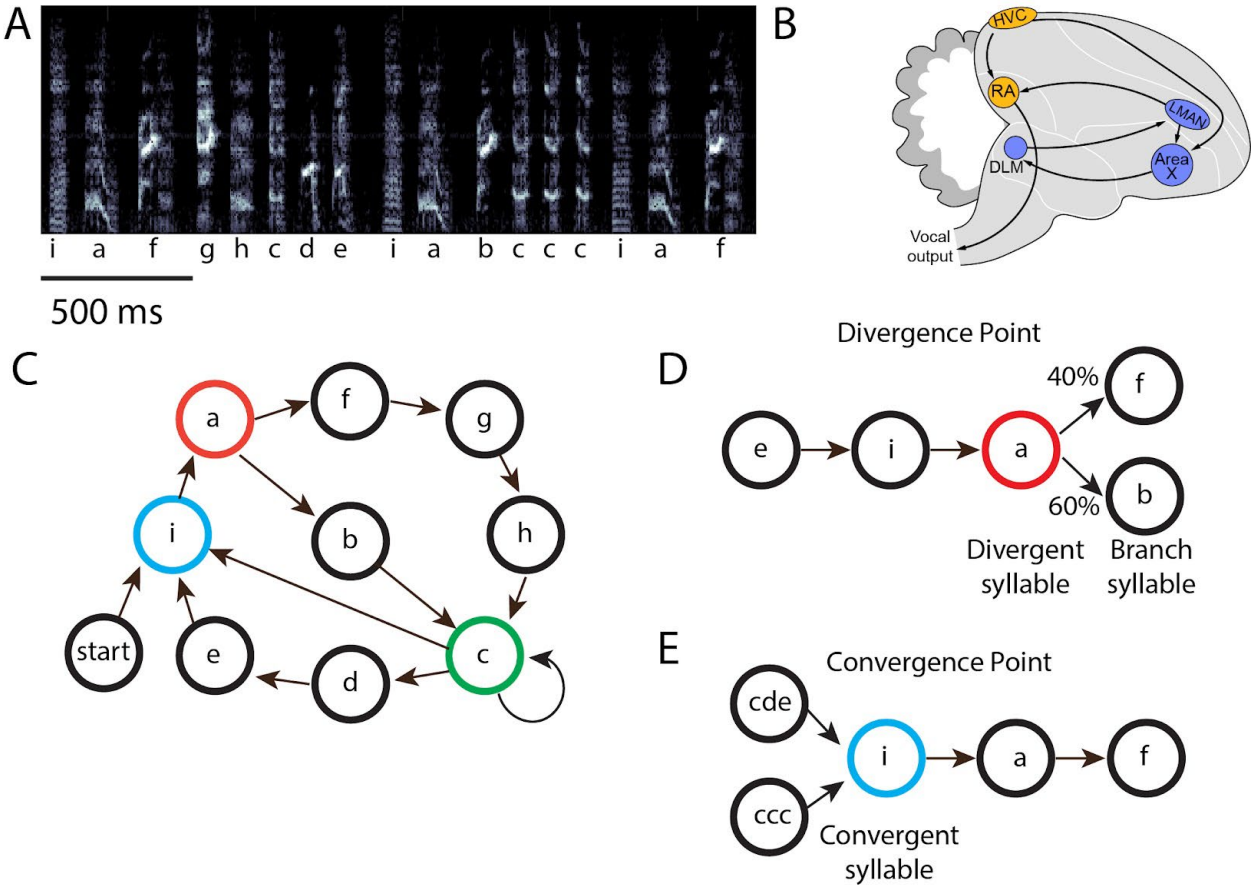


Figure 1.1: Bengalese finch song sequence and circuitry

Legend: **A.** An example of a spectrogram of Bengalese finch song **B.** Basic song circuitry of Bengalese Finches. HVC sends projections to RA and Area X. RA projects down to motor neurons while Area X is part of the Anterior Forebrain Pathway (AFP) loop which projects to RA **C.** Diagram of the song in 1a. Syllables in red are divergence points, syllables in blue are convergence points, and syllables in green are both divergent and convergent. **D.** Example of a divergence point in this song. Syllable "a" can transition to either syllable "f" or "b" with a 40:60 probability **E.** Example of a convergence point in this song. Syllable "I" can be sung after syllables "cde" or "ccc"

Chapter 2: Development of a custom 1-photon fiber optic calcium imaging microscope

2.1 Microscope design

To address questions of acoustic and sequence encoding in HVC, we need to record from large populations of individually identified neurons in singing Bengalese finches. One method to accomplish this is calcium imaging. There are currently many designs available for collecting 1-photon calcium imaging data *in-vivo* at a cellular resolution (Zhang et al. 2019; Cannon et al. 2015; Scherrer et al. 2023). Each of these requires a miniature microscope to be mounted on the head of the animal (2-3 g, Zhang et al. 2019; Scherrer et al. 2023). While this works well for larger animals such as mice and rats, for smaller animals such as finches (which weigh ~15 g), this is a significant burden and leads to animals not performing natural behaviors as easily. Previous studies in songbirds have successfully used miniaturized microscopes for freely moving 1-photon imaging, or head fixed 2-photon.

Furthermore, even for larger animals, the size of these microscopes allows for only imaging one area at a time. Additionally, having the microscope on the head of the animal can lead to significant movement artifacts which increases the inaccuracy of the signal (Zhang et al. 2019). One solution would have the microscope off the head of the animal. While advances in miniaturization will allow smaller and smaller animal-mounted imaging systems, moving the microscope off the head will always allow larger, more complex imaging pathways and components. This will allow easier movement of the animal, more advanced cameras and optical components, and easier customization of multiple cameras and excitation wavelengths.

To accomplish this, the use of fiber optic cables is necessary to transmit light from the brain of the animal to the microscope. While typical fiber optics do not allow for spatial localization, recent advances now allow for fiber optics which can create an image with a resolution of up to 2 μm (Toader, Regalado et al. 2023). Using these new fiber optics, we designed an easily customizable and affordable 1-photon microscope capable of imaging with cellular resolution. As the fiber optic cables are only 350 μm in diameter, they can easily be positioned to record from multiple brain areas simultaneously. The microscope was built using the components in **Table 1**.

2.2 Grin Lens Implant and Positioning

Bengalese finches were injected with either GCaMP6s or GCaMP8s and implanted with 1 mm diameter GRIN lenses with a working distance of 200 μm (Inscopix PN:130-000143). The lens was positioned over HVC as visualized using retrograde DiI from Area X. The lenses allowed us to cover a large area of HVC giving us the ability to search multiple areas for the best expression level. As the fiber optic cables were only 350 μm in diameter, we could only select a smaller region of the lens to image. We designed a custom 3D printed and machined parts system to couple the fiber optic and the grin lens. This device held the fiber at the working distance above the lens to create a focused image on the camera. Additionally, the holders allowed us to easily search HVC and adjust the position of the fiber while birds were awake and behaving. We adjusted the position of the fiber and allowed the birds to sing until we found an area that showed the strong song responses. These devices held the fiber at the working distance above the lens to create a focused image on the camera.

2.3 Software for collection of song aligned videos

We designed a custom software using Python and the Spinnaker Python SDK to capture video from the microscope camera and align it to audio from a microphone recording the bird's song. We constantly record and store audio in a buffer and trigger acquisition of video based on sound, as song is a spontaneous behavior and occurs throughout the day. If there was audio greater than a certain amplitude, we would begin the recording as well as turn on the excitation LED. This sound triggered recording was used to reduce bleaching of the GCaMP signal within HVC.

2.4 Potential Future Applications

One of the advantages of removing the microscope from the head of the animal is that it allows for further customization of the optical setup. In addition to using faster, and higher resolution cameras which are larger in size, we can also add in multiple cameras to the set up. This allows for multiple different features. First, we could add in a second camera that would collect another wavelength of excitation and emission. This would allow for recording of another fluorescent reporter in a different wavelength of light. This could be used to identify different cell types with retrograde tracers, or record signals using a red shifted activity reporter. Second, by having two or more cameras, we can offset the frame acquisition of these cameras to have an effective higher frame rate, while still maintaining a longer exposure time. For example, while maintaining a 30 ms exposure time across 3 cameras, if we offset the acquisition time of each camera by 10 ms, we can make our effective frame rate 3 times faster. This could allow for better temporal resolution with the newer generation of faster GCaMP8 proteins as well as open the door for 1 photon voltage imaging.

Table 1: Custom 1-photon fiber optic microscope components

Part name	**Company**	**Part number**
Tube Lens	ThorLabs	AC254-180-A
Lens Mount	ThorLabs	CP35/M
Emission Filter	ThorLabs	MDF-GFP2
Excitation Filter	ThorLabs	MDF-GFP2
Dichroic Mirror	ThorLabs	MDF-GFP2
Filter cube	ThorLabs	DFM1/M
SMA fiber mount	ThorLabs	SM1SMA
4x Achromat Objective	Olympus	RMS4x
Z-axis objective mount	ThorLabs	SM1ZA
Objective Adapter	ThorLabs	M32RMSS
Excitation LED	ThorLabs	470L5
LED Collimator	ThorLabs	SM1U25-A
LED Driver	ThorLabs	LEDD1B
Fiber Optic Cables	Fujikura	FIGH-10-350S
Camera	FLIR	BFS-U3-32S4M-C
C-Mount Adapter	ThorLabs	SM1A9
External Thread Coupler	ThorLabs	SM1T2
Camera Cable	Edmund Optics	86-770
Mirror Mount	ThorLabs	KCB1C/M
Dielectric Mirror	ThorLabs	BB1-E02
Cage assembly rods (multiple lengths)	ThorLabs	ERxx

Chapter 3: HVC encoding of future and past sequences

3.1 Characterization of single neuron firing patterns

Burst heights vary from trial to trial

To better understand how branch points are encoded within the songbird brain, we recorded neural activity with cellular resolution from HVC using calcium imaging in Bengalese finches (**Figure 3.1. A, B, C**). While HVC is known to play an important role in song timing, little is known about how it might encode variable sequence information (Cohen et al. 2020; Fujimoto et al. 2011). We recorded activity from HVC in Bengalese finches, which display sequence variability, to investigate signals that encode for the production of the current syllable, from signals which encode upcoming or past sequence.

We began our analysis by quantifying the different properties of bursting in HVC projection neurons. While much has been studied about the timing of HVC projection neuron firing, there has been less focus on the number of spikes fired for each burst and how variable this is. Calcium indicators reflect the amount of calcium in the cell which increases upon neurons firing action potentials (Zhang et al. 2023). The greater the number and frequency of action potentials, the larger the increase in GCaMP fluorescence will be. We observed that the increase from baseline in calcium signal varies from trial to trial (**Figure 3.2 A)**. Across all recorded neurons the standard deviation of burst heights ranged from 0.15 to 1.5 with a mean of 0.58 and a median of 0.53 (**Figure 3.2 B/C)**. We concluded that HVC projection neuron firing in Bengalese finches contains a large amount of variance in its magnitude.

Correlation of burst heights

While there is clearly variation in the magnitude of firing in HVC projection neurons, this variation could either come from intrinsic properties of neurons, or from larger network correlations. To address this, we pulled out the longest consistently occurring sequence of syllables which contained all syllables for each bird. We refer to these consistent sequences of syllables as motifs. We then aligned the activity from all neurons to this motif and extracted out all which occurred during this motif. As many neurons burst at more than one location in this motif, some bursts will come from the same neuron. We first treated each burst as independent and organized them into a response matrix of size number of unique bursts by number of motifs. In this matrix, each entry represents the height of a burst during a single rendition of the motif. We then took the correlation of this matrix to get a matrix of correlation coefficients of size number of bursts by number of bursts. This matrix reflects correlation between the heights of each burst across all motifs (**Figure 3.2 D, Supplemental Figure 3.2.1)**. Thus, each pair of bursts has one correlation coefficient across all motifs. We found that there were many significant correlations across birds. Approximately 30% of all pairs of bursts had correlations that were significant (threshold p=.05). The prevalence of such correlations across bursts suggests that there are network properties that change from motif to motif that lead to changes in burst height.

Correlations between bursts were present both for bursts from different neurons ("across neurons") as well as bursts for one neuron that occurred at different locations in the motif ("within neuron") (**Supplemental Figure 3.2.1 A**). To investigate the contribution of individual vs network properties to variance in burst height, we compared the strength of within neuron versus across neuron correlations. We found that within neuron correlations were more prevalent

than across neuron correlations (**Figure 3.2 E).** 70% of within neuron comparisons were correlated across motifs. In comparison 27% of across neuron comparisons were correlated. Furthermore, significant correlations of bursts from within neuron were significantly greater than those across neurons ($p<10^{-10}$ Wilcoxon rank sum). The average absolute value of the correlation within the same neuron was 0.57 vs 0.36 for different neuron comparisons. We concluded that bursts from the same neuron are more likely to be correlated and are more strongly correlated than bursts from different neurons. However, there is still a large amount of shared variance across the neural population as bursts from different neurons are also strongly correlated.

Another aspect of correlations is their sign (+/-). If the correlation coefficient is positive, this means that when one burst is high in a given motif, the other burst is also likely to be high. If the correlation is negative, it means that when one burst is high, the other is more likely to be low. We found that ~80% of significant correlations were positive. We found that the correlation coefficients between bursts from the same neuron were more likely to be positive than bursts between different neurons (**Figure 3.2 E)**. Approximately 98% of significant correlations within neuron were positive while ~80% of significant correlations across neurons were positive. This means that on certain motifs, individual neurons are consistently more or less active across all bursts, rather than larger firing in one burst correlating with lower firing in another. This effect is more nuanced with bursting across different neurons, as these bursts can either be positively or negatively correlated. This smaller fraction of negative correlations might reflect coupling of excitatory projection neurons by inhibitory interneurons. We concluded that while bursting across the population is highly correlated, changes in individual neuron excitability account for a larger fraction of the burst height variance. This suggests a model

where individual neurons maintain a state of either high, or low excitability throughout a single motif.

Burst heights decrease over each successive motif

Another variable that may correlate with neural activity is the time at which a given neuron was active. For example, a neuron which fires for syllable "a" on average might fire more for syllable "a" during the first rendition in song, and then less during each successive "a" sung. To test this, we plotted the average burst height for each neuron against the motif number in song (for example if a motif is sung for the 3rd time in song vs the 5th time in song). In each bird there was a general trend that calcium signals decreased over successive motifs (**Supplemental Figure 3.2.2 A**). This correlation of motif number vs burst height was significant across all bursts in all birds ($p< 10^{-10}$ R^2=.01). While this correlation explains relatively little of the variation of burst heights it is a consistent trend. To confirm that this trend is not an artifact of the calcium imaging itself, we reanalyzed data from HVC multi-unit electrophysiology recordings previously done in our lab (Jaffe 2020). For each motif, we calculated the average firing rate during each syllable. In both birds analyzed, motif number and average firing rate were significantly correlated at a similar level to our calcium data (**Supplemental Figure 3.2.2 B;** $p<e^{-10}$, R^2 = .04). Thus, we concluded that there is a significant decrease in the activity of HVC neurons over the course of the song.

3.2 Analysis of firing around divergence points

Divergence points are preceded by changes in burst size in HVC projection neurons

In any variably sequenced motor action, the brain must prepare the muscles to act in the desired fashion. For example, when playing the piano if the next note played requires a jump to a key far away from the ones currently being played, preparations must be made for this more difficult jump. Similarly, in Bengalese finch song, vocal musculature must be prepared to sing one note after the other. Indeed, activity during syllable "a" when the bird will transition to syllable "b" looks different than a transition to syllable "c" (Fujimoto et al. 2011). Little is known about how far in advance Bengalese finches determine these upcoming divergence points. These selections may be made many syllables in advance, or just before the onset of the divergence. For example, differences in firing may be fully predictive of the upcoming variable sequence many syllables ahead, suggesting that the selection of which syllable to sing is made far in advance and is more deterministic just before the divergence point. On the other hand, there may be more subtle differences in firing leading up to the divergence point, suggesting that while there is some level of longer-range influence, the ultimate decision at the divergence point is more stochastic.

To determine how HVC might encode changes in song sequence, we asked whether we could detect changes in firing activity that predicted the chosen sequence at divergence points. We began by aligning motifs that contained long strings of stereotyped syllables leading up to a divergence point. For example, one bird sang a motif of "iafghcdeia" which then led to a divergence point where it either sang f or b. Any neuron which burst consistently during the

stereotyped motif leading to one branch but not during the same sequence in the other would be encoding the upcoming branching syllable. This “all-or-none” type firing would be evidence that the upcoming divergence point is not decided at the branch itself, but during the stereotyped motif beforehand. Aligning the neural data as such made it clear that there are none of these “all-or-none” type neurons (**Figure 3.3A**). Across all neurons recorded in all birds during divergence points, if a neuron displayed a burst in a sequence leading up to one branch, it also burst in that same preceding sequence for the other branch (n=309 bursts in 131 ROIs). This aligns with previous results from Fujimoto et al. as they also did not find any “all-or-none” predictive neurons at divergence points. Thus, we concluded that HVC projection neurons do not encode future information in an “all-or-none” fashion, different from observations in canaries (Cohen 2020).

While there are not any neurons which are fully predictive of the upcoming branch syllable, we observed that burst size in HVC projection neurons is varied, leading us to analyze if any of these smaller differences in firing could predict the upcoming branch. Using a Wilcoxon rank-sum test, we compared all bursts leading up to the divergence point in these stereotyped sequences (**Figure 3.3B/C**). We found that across 4 sequences in 4 birds 55 of 309 bursts prior to the divergence point were significantly different (threshold of $p < .05$). These bursts were then assigned to a syllable allowing us to compare the percentage of bursts that differed in each syllable window prior to the divergence point (see methods). **Figure 3.3D** shows the percentage of bursts that differed for each syllable leading up to the divergence point across all birds.

To determine which differences were greater than expected by chance, we shuffled the labels of the post branch identities 1000 times and determined the 95th percentile of branch selective neurons (see methods).Compared to a shuffled distribution we found that the

percentage of bursts in the 1st, 2nd, 4th, 9th, and 10th syllables prior to the branch which were selective for the upcoming sequence were significantly greater than expected by chance. This reveals that while there are not "all-or-none" type selective bursts which predict the upcoming divergence points, there are more subtle changes in burst heights that do. This led us to conclude that there is evidence that signals within HVC can reflect the upcoming sequence selection. These more subtle differences suggest a model where activity in HVC influences the upcoming selection far in advance, but still allows for a more stochastic process at the divergence point.

The observation that there are significant differences in neural activity more than 8 syllables prior to the divergence was surprising. While only the differences during the 9th and 10th syllables were greater than the 95% shuffled distributions, many of the other syllables were at the 95% chance level. We sought to gain a better understanding of the dynamics of these neurons as they may provide insight into how decisions at divergence points are made. To investigate this, we looked at all of the neurons that were divergence selective more than 7 syllables prior to the divergence point (n=17). The z-scored fluorescence averages of some examples of these neurons are plotted in **Figure 3.3E**. 10 out of 17 of the neurons burst again either during the branch syllable or the syllable before it. Another 4 bursts occurred 2 syllables prior to the divergence. Of the neurons that burst in the 1 or 2 syllables before the divergence, 5 out of 8 had significantly different firing during this syllable as well. Thus, while some bursts are far from the divergence point in time, the neurons responsible for these bursts are active again close to the divergence. This aligns with our finding that different bursts from the same neurons are positively correlated across trials. Thus, the past activity of the neurons that are predictive and fire around the divergence point play a role over the upcoming sequence decision. We

concluded that the firing of individual neurons in the past can influence the upcoming branching decision.

Influence of motif number on divergence points

While we found many differences in neural activity leading up to divergence points, we know that neural activity in our dataset is also correlated with a general decrease in activity over the course of the song. While this trend only accounts for a small portion of the variation in burst heights ($R^2 = .01$), it is still possible this is partially responsible for differences across branches. If the differences we observe are due to the general decrease in neural activity, we would expect transition probabilities to also change accordingly over the course of the song. Thus, we plotted transition probabilities vs the motif number in which they were sung (**Supplemental Figure 3.2.1).** We observed that in 3 out of 4 birds, the divergence point probability changed over the course of the song, but did not consistently increase or decrease. However, in 1 bird, we observed the transition probabilities of the divergence point changed over the course of the motif. Thus, in at least one bird, changes in HVC activity may reflect a general decrease in activity which could be related to the change in probability.

To further investigate the relationship of motif number in song, divergence point probabilities and burst height we fit a linear mixed effects model using both the upcoming branch identity and motif number as fixed effects for the burst height of neurons. In the 3 birds which did not show a clear trend in divergence probability as a function of motif number, the number of significant neurons that were selective for the upcoming sequence did not significantly decrease. In the bird that did show a change in transition probability as a function of motif number, 7 neurons remained significantly transition selective out of the previously observed 25. Overall, there were now 40 of 309 bursts that were significantly transition selective

prior to divergence points). This number [MB1] exceeds the 99.99th percentile of the null distribution based on shuffled models (out of 1000 shuffles the maximum number of transition selective neurons was 30). While more analysis is needed to disambiguate the general decrease over motifs with differences in bursting around divergence points, we conclude that the observation of transition selective bursts cannot be explained exclusively by the general decrease of burst heights over the course of song.

Decoding upcoming branch syllables from neural firing

Given that we found many bursts which were transition selective during syllables prior to divergence points, we wanted to test if we can decode the upcoming branch identity from the calcium signal prior to the divergence point. Using the calcium signal increases per syllable we built a Support Vector Machine Classifier (SVM) to decode the upcoming branch syllable (**Figure 3.4).** The input to this SVM was the burst height binned by each syllable window (same data as in **Figure 3.3D).** For each additional syllable in the sequence, we used all the calcium signals that occurred before the offset of that syllable. For example, in one bird the sequence was "iafghcdeia" which diverged to either a "b" or an "f". As there are 10 syllables leading up to the divergence, the first syllable in the sequence "i" is labeled the -10 syllable and the last syllable "a" is the -1 syllable. Thus, when decoding the upcoming branch at -10, we only use calcium signals prior to the offset of the first syllable "i". For the -1 syllable, we use all the calcium signals that occurred prior to the offset of the last "a", which includes signals from the entire sequence prior to the divergence.

We found that across all birds, we could decode upcoming syllable identity with 65% accuracy using all the calcium signals prior to the branch syllable. In one bird, we could decode the upcoming syllable with 90% accuracy at this same time point, while other birds had lower

accuracy. This suggests that in certain birds, upcoming branch identity may be decided earlier in the sequence. Additionally, for 2 out of the 3 birds which had consistent sequences over 10 syllables prior to the divergence, upcoming branch identity could be decoded up with ~60% accuracy even during the first syllable of the sequence. This confirms our observation that there are transition selective bursts many syllables in advance of the divergence. In general, information about upcoming syllable selection is available prior to the divergence but is not fully predictive. This could indicate that while activity within HVC prior to the divergence influences the upcoming sequence, it is not fully deterministic until the moment the branch syllable is sung.

3.3 Analysis of firing around convergence points

Past information is encoded within HVC

Convergence points are another aspect of sequence variability in Bengalese finch song. Convergence points are places in song where one syllable can be sung after multiple different syllables, giving that one syllable multiple preceding contexts (**Figure 1.1.E** blue circle "i" contexts "cde" and "ccc", and **Figure 3.5A**). Previous studies have revealed that HVCX projecting neurons can encode information about this preceding context (Fujimoto et al. 2011). This encoding is different than the selectivity around divergence points as there are many neurons which are "all-or-none" selective (in **Figure 1.1.E** bursting during syllable "i" in context "cde" but not in context "ccc") in addition to those which fire more spikes on average during one context than another (similar to the encoding of divergence points) (Fujimoto et al. 2011). While previous work looked mainly at selectivity during syllables leading up to convergence points, we also wanted to analyze how long after a sequence has converged can neural activity still contain information about the preceding sequence. This is a fundamentally different question as we want

to understand not only the differences in encoding during the convergent syllable, but also how syllables downstream may also reflect differences in the history of that vocalization.

We began this analysis by aligning neural activity to convergence points in song and looking for differences in burst height at each successive syllable after the convergence point (**Figure 3.5A**). As expected, we found many bursts which encoded for the prior sequence during the convergent syllable (threshold $p<.005$ Wilcoxon rank sum). Across all birds ~30% of bursts were context selective at the convergent syllable. This is less than previous findings which found that ~70% of HVC_X projection neurons were context selective at convergence points (Fujimoto et al. 2011). This result may be explained by the fact that we are likely recording from both HVC_X and HVC_{RA} projection neurons, in addition to differences in how syllables are labeled (see discussion). In addition to differences during the convergent syllable, we found many significant differences in syllables up to 11 syllables after convergence. This result shows that HVC contains long range information about the past. This may help to explain the observation that song sequence often displays long range history dependence.

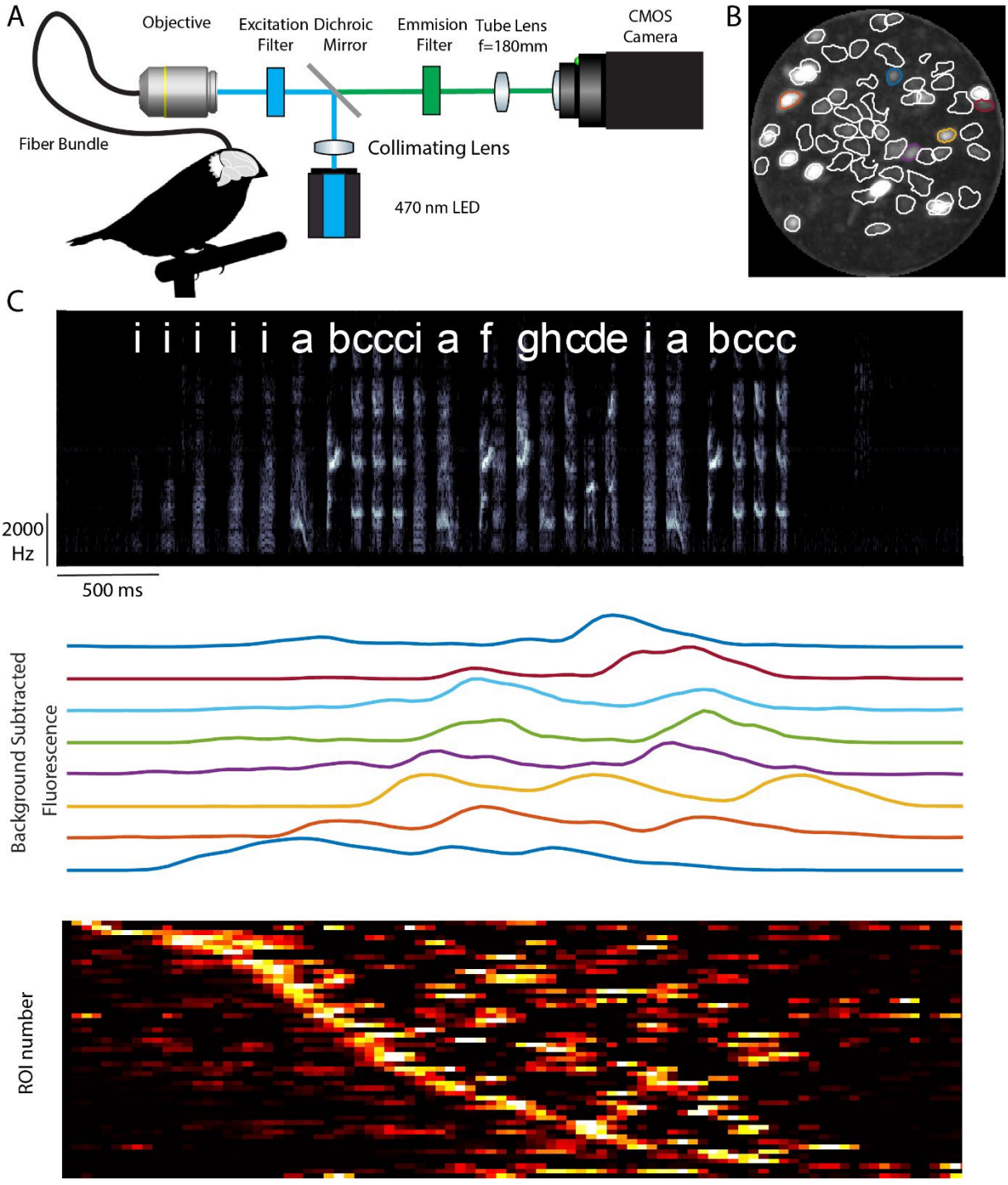


Figure 3.1: Calcium imaging in songbirds using image fiber bundles

Legend: **A.** Microscope diagram. GCaMP expressed in HVC is imaged using high-density imaging fibers capable of resolving single neurons **B.** An example of a temporal Maximum Intensity Projection of a video taken of neurons (Figure caption continued on the next page)

(Figure caption continued from the previous page) within right HVC. ROIs are shown. Colors match colors of traces in C. Note, not all colors are present as some ROIs come from fiber over left HVC **C.** Top: Example song. Middle: 14 example ROI Z-scored intensities recorded during and aligned to a single rendition of the song. Bottom: Max normalized derivatives of all ROIs recorded (n=53) aligned to song

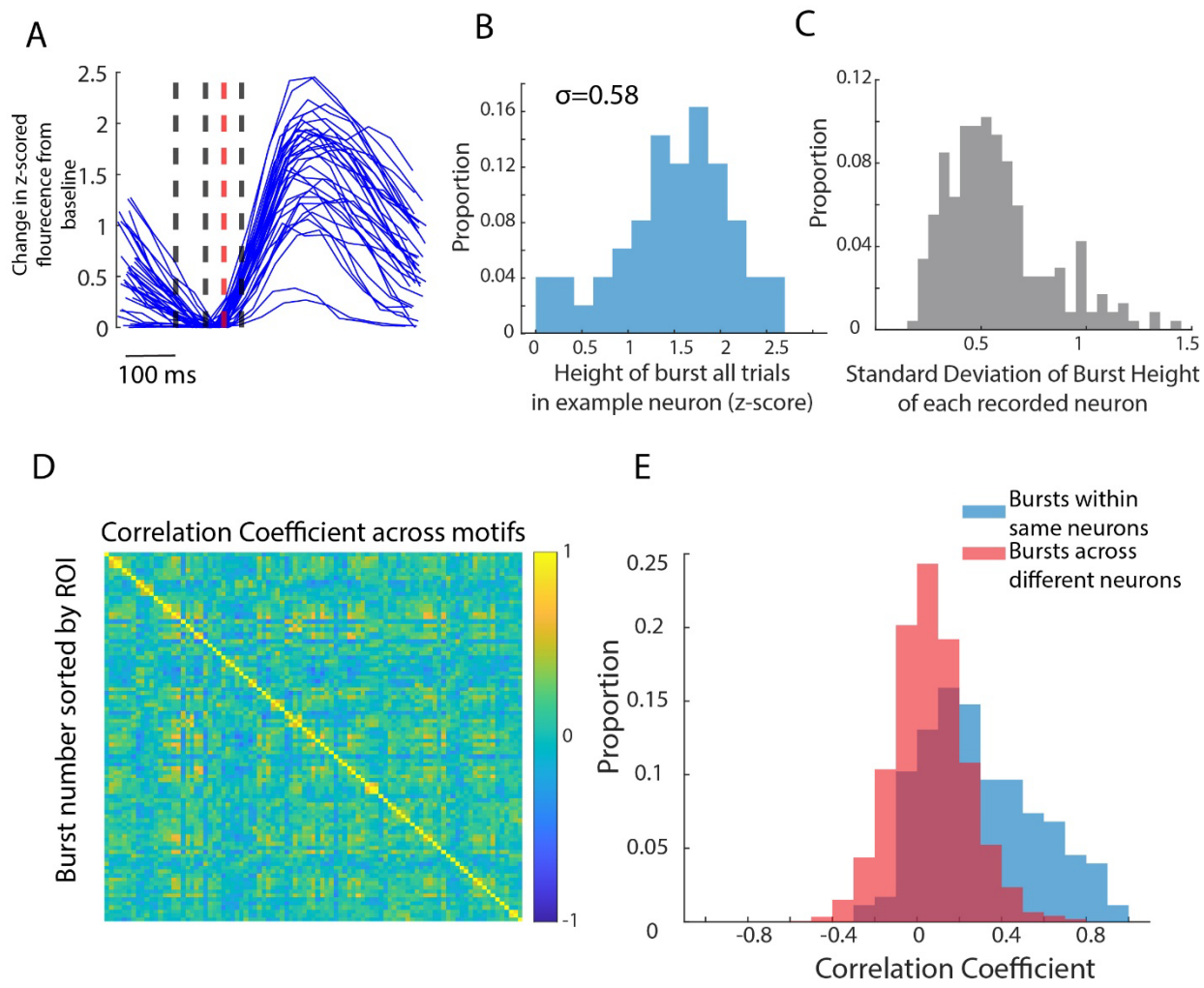
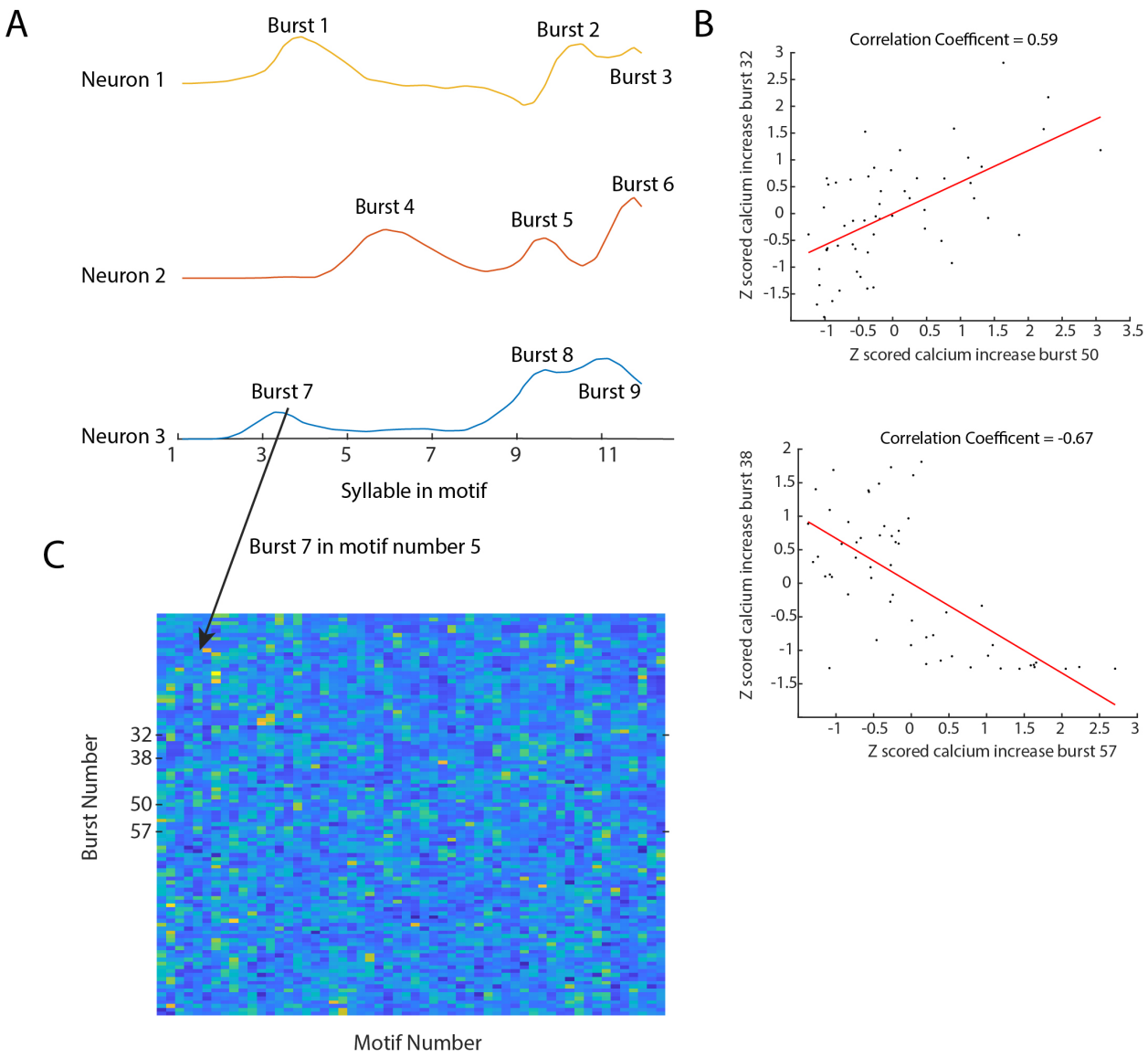


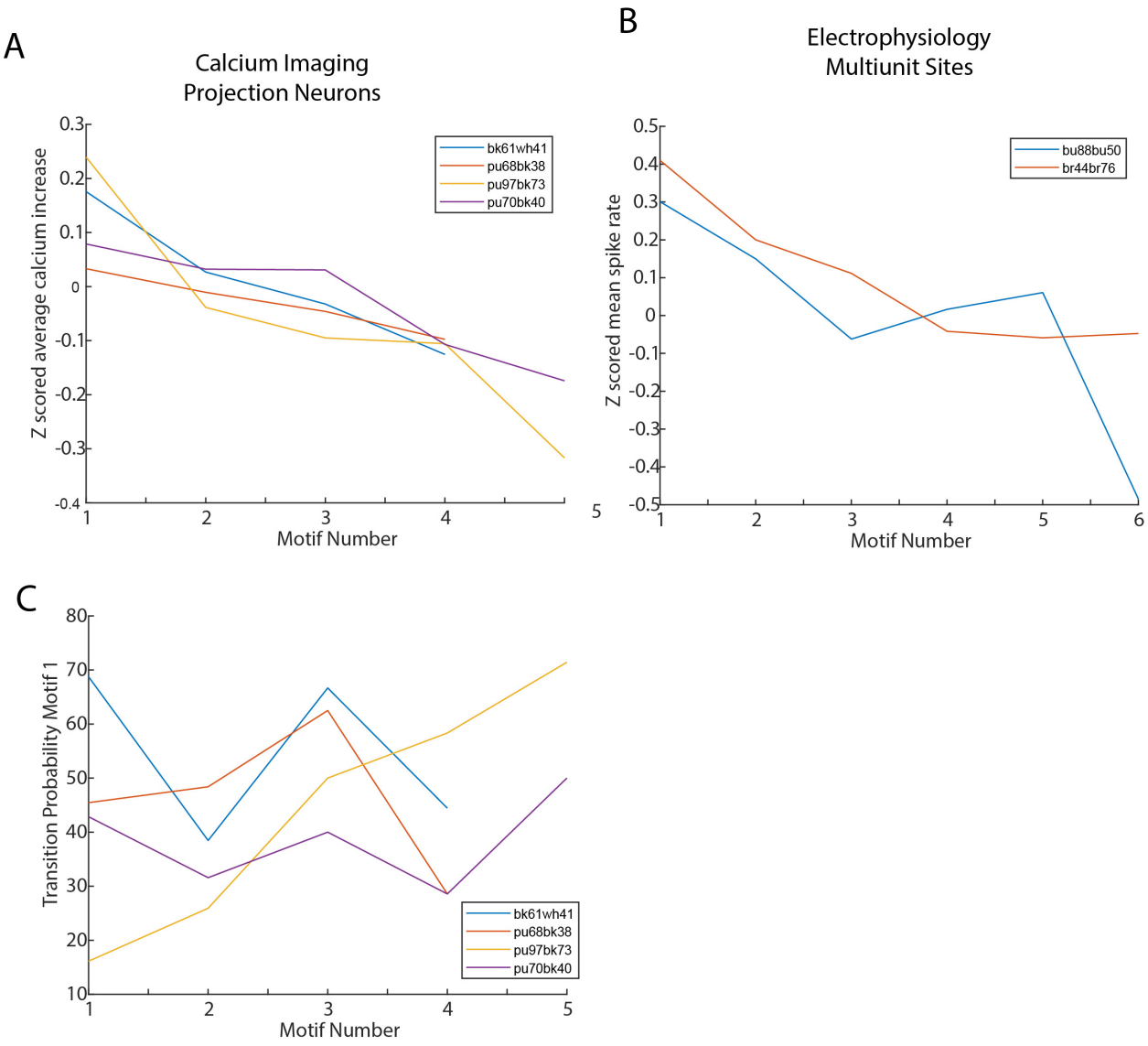
Figure 3.2: Quantification of individual and population neural properties

A. Z-scored intensity of one ROI aligned to the onset of syllable "b" in all songs recorded. Offset of previous syllable, onset and offset of "b" marked with dashed black lines. Dashed red line indicates calculated burst onset time. **B.** Histogram of the change from minimum to maximum of the traces in 1D. Standard deviation = .53 C. Histogram of the standard deviation of all recorded ROIs **D.** Correlation coefficient matrix of all individually extracted bursts in example bird in 3.1 C. Each burst is represented as an individual even if they come from the same neuron. Bursts are sorted by ROI number so bursts from the same neurons are located next to each other. **E.** Histogram of correlation coefficients of bursts from within same neurons (blue) and across different neurons (red). $p < 10e^{-15}$ that means are different Wilcoxon rank sum test.



Supplemental Figure 3.2.1: Calculation of correlation coefficients across motifs

Legend: **A.** Three example neurons from one rendition of selected motif. Each neuron contributes 3 bursts to the matrix in C **B.** Correlations of 2 pairs of example bursts from different ROIs. Top: burst 32 and burst 50 are significantly positively correlated across motif. Bottom: burst 57 and burst 38 are significantly negatively correlated **C.** Example matrix of z scored burst heights for each burst in each motif sung.



Supplemental Figure 3.2.2: Neural activity decreases over the course of each motif

Legend: **A.** Average z-scored burst height from all neurons in each motif rendition number in all 4 birds. **B.** Average z-scored mean firing rate across all multi-unit sites in two birds. **C.** Transition probability of one motif for each of the recorded in A

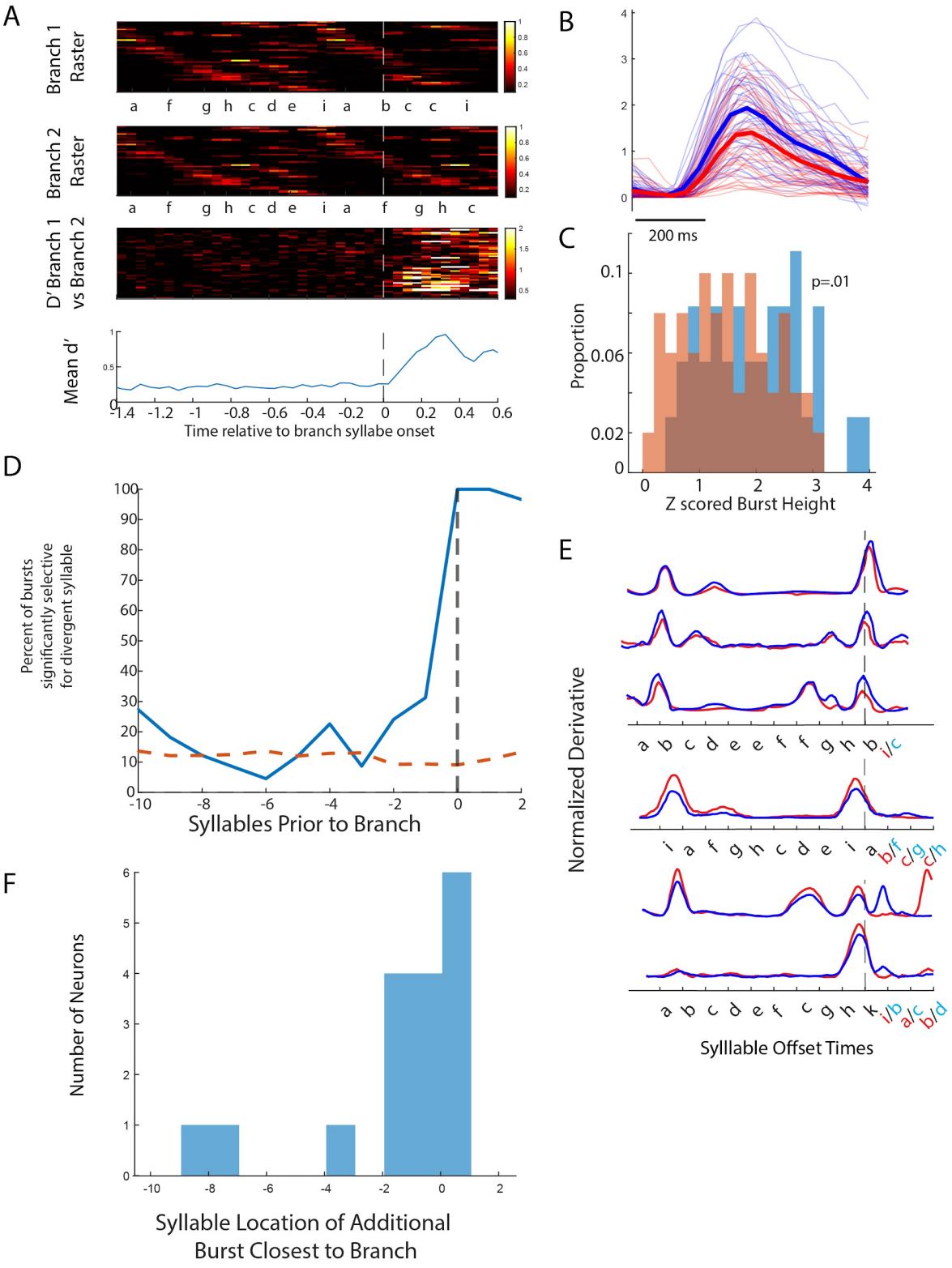


Figure 3.3: Burst heights can encode for upcoming divergence points

Legend: **A.** Top Down: 1) Raster plot of average (Figure caption continued on the next page)

(Figure caption continued from the previous page) derivative of 49 neurons recorded simultaneously in HVC aligned and linearly time warped to sequence “afghcdeiabc” 2) Same as 1 but with sequence “afghcdeiafg” 3) D’ statistic in 1 vs 2 for each neuron 4) Average D’ across all neurons. Time 0 is the onset of either syllable b or f. **B.** Z-scored fluorescence traces from all trials for an example neuron from raster in 3a, aligned to onset of syllable i. Blue traces are all trials which go to syllable b, red traces are all trials which go to syllable f. Darker traces are means for each syllable type **C.** Histogram of the peak of the traces in 3B **D.** For all bursts in all birds, percent of neurons which have significantly different bursting across divergence points in blue (as in 3c). Each burst is assigned to a syllable and plotted relative to how many syllables away from the divergence it is. Dotted orange line is 95th percentile of shuffle burst differences, showing chance level. **E.** Average fluorescence traces of example neurons in 3 birds which have significantly different bursting at least 5 syllables prior to the divergence point. Different divergence sequences are plotted in red or blue. Labeled syllables are the offset times of each syllable. Dotted line shows offset of the last same syllable across sequences. **F.** Histogram of locations of the burst closest to divergence in all ROIs that have significantly different firing at least 7 syllables prior to the divergence point.

A

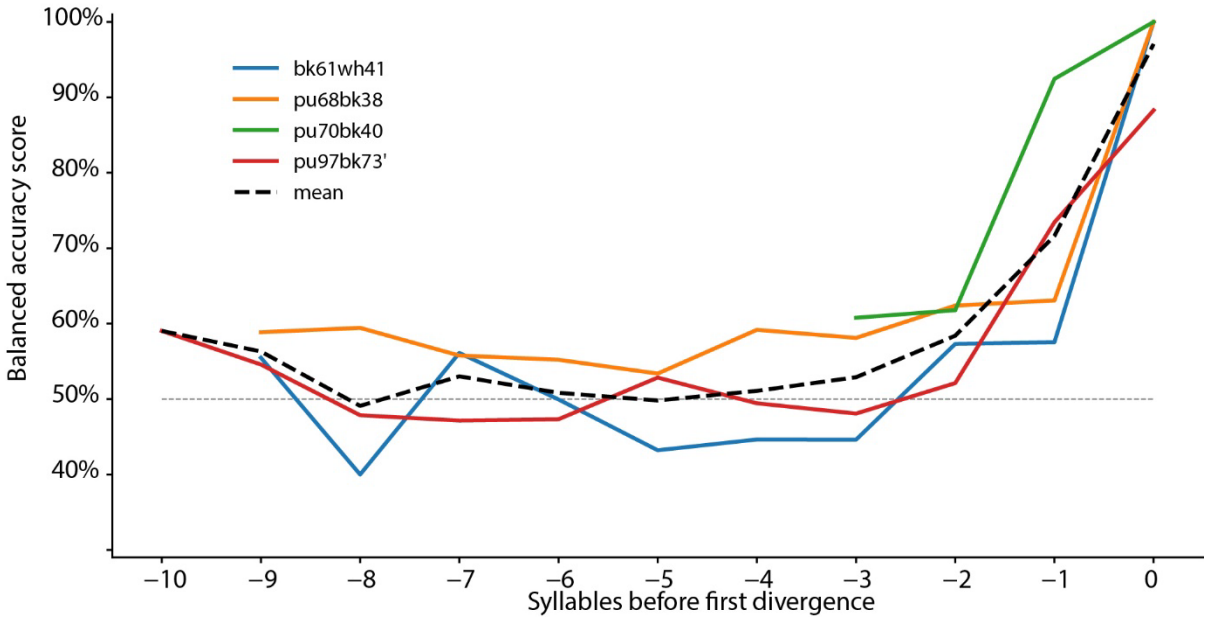


Figure 3.4: Upcoming branch identity can be decoded above chance from burst heights

Legend: **A**. Balanced accuracy score per bird with across bird means using a Support Vector Machine (SVM).

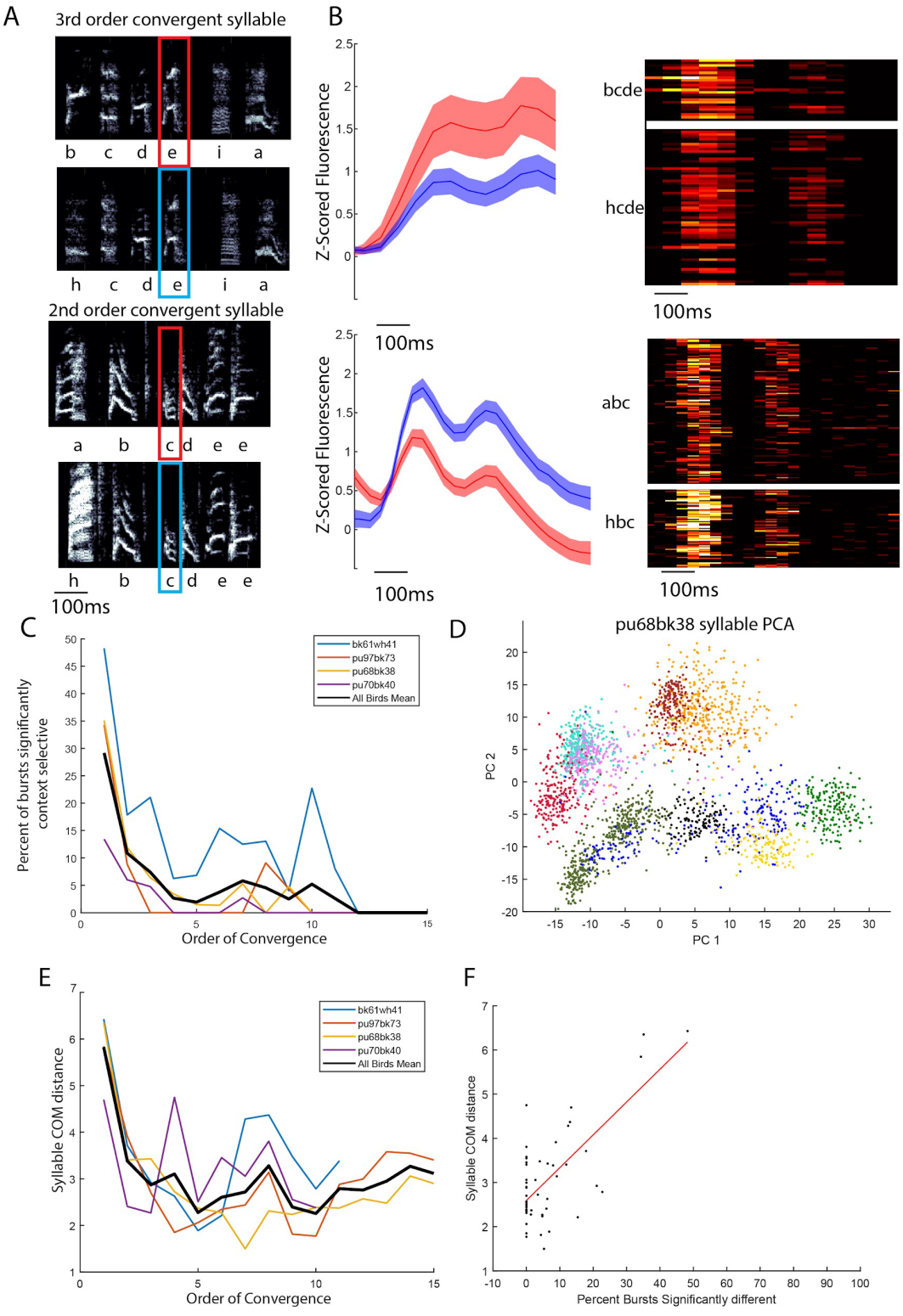


Figure 3.5: Burst heights encode for past convergence points

Legend: **A.** Examples of the spectrograms of two (Figure caption continued on the next page)

(Figure caption continued from the previous page) convergence points from two birds.3rd order convergence “bcde” vs “hcde” and 2nd order convergence “abc” vs “hbc”. The order of the convergence refers to how many syllables in the past was the last different syllable across the two sequences. Top 3rd order convergence $p<.001$. Bottom 2nd order convergence $p<10^{-8}$ B. Z-scored fluorescence traces from all trials from two example neurons which are context selective for the sequences in 3A. Left traces from different contexts plotted in red or blue. Darker trace is the mean for each context, lighter surface are 95% confidence intervals. Right, raster plot of the derivative of the traces. **C.** Per bird percent of neurons which are context selective vs the order of the convergence point. Individual birds in color, mean of all birds in black **D.** Per bird mean COM distance of syllables vs order of convergence point **E.** Correlation of percent of bursts context selective (data from 5C) vs COM distance (data from 5D) **F.** Correlation of percent of bursts context selective vs COM distance for each individual syllable. $p<10^{-8}$ $R^2 = .49$

Chapter 4: HVC encoding of phonology

4.1 HVC encodes phonology across same syllables in different contexts

Numerous studies in the Zebra Finch have contributed to a model in which HVC activity abstractly encodes timing in song, while the specific acoustic features that are produced at each moment in time depend on the pattern of connections from HVC to RA (Lynch et al. 2016; Long and Fee 2008; Long et al. 2010). However, an alternative model suggests that HVC activity encodes some acoustic features (Amador et al. 2013), and manipulations of HVC activity can influence acoustic features such as pitch (Jaffe et al. 2020). Part of the difficulty in differentiating between these models in the zebra finch arises because of the stereotypy of zebra finch song, such that any specific moment in song is highly correlated with the specific acoustic features at that moment. In contrast, in BF song the variable sequencing of syllables, and differences in phonology of individual syllables depending on when they are produced in song, provides an opportunity to examine whether HVC activity covaries with acoustic structure.

Previous work has shown that the same syllable sung in different convergent contexts ("**a**" in "xyz**a**" vs in "bcd**a**") can have significantly different acoustic structure, and that the magnitude of phonological differences are correlated with differences in RA activity (Wohlgemuth 2010). Moreover, neural activity in RA differs across acoustically distinct renditions of individual syllables and can encode acoustic features such as pitch (Sober et al. 2008; Wohlgemuth 2010). This further raises the question of whether phonological encoding first arises in RA or is also present in the inputs to RA from HVC.

Figure 4.1A shows an example song of an adult Bengalese finch which displays many of these features. We can see multiple syllables with different contexts. Syllable "b" can occur after

"a" or after "k". Syllable "c" can occur after a "b" or a "f". We wanted to test if the songs of our birds also contained phonological differences at convergent syllables. To quantify how different the same syllables sung in different contexts are, we need a metric of syllable similarity. We used a similar approach to previous work (Wohlgemuth 2010; Tchernichovski 1999). For each syllable, we measured 9 acoustic features which quantify measures of frequency, amplitude, and other spectral features (see methods; Tchernichovski 1999). These features were taken across the syllable so that each syllable had 100 data points per feature, leaving us with a 900-dimensional representation of each syllable (**Supplemental Figure 4.1B)**. This was then reduced in dimension using PCA. The top 10 PCs were used to place each syllable in a reduced dimensional space. For each syllable, we calculated the center of mass (COM) across all renditions of each syllable in each context in this space. We then measured the Euclidean distances between COMs as a way of comparing how similar two syllables are (COM distance; Wohlgemuth 2010; **Figure 4.1B**). The more similar the syllables, the lower the COM distance. This comparison can be made either across the same syllables in different contexts, or different syllables. The matrix in **Figure 4.1B** shows these COM distances for all different syllables in **Figure 4.1A**. Note that as expected this metric is smaller for syllables together which by eye appear similar. "k" and "b" are the most similar comparisons, but we can also see "a" and "e" have similar structure and have a low COM distance.

Using this metric for syllable similarity, we tested whether syllables that are more similar acoustically also have more similar neural representations. We began by analyzing if the differences in firing after convergence points can be explained in part by differences in phonology. We found that in the birds we collected calcium imaging data from, the 1st order convergent syllable had a mean COM distance of 5.9 across all birds, which was significantly

larger than the distance of any other order of convergence. This confirms the previous report that convergent syllables contain significant phonological differences dependent on their preceding sequence (Wohlgemuth 2010). The mean COM difference for same syllables in different contexts decreased as a function of the order of convergence (**Figure 3.5 D**), and this decrease had a qualitatively similar shape to that for the percent of neurons that were context selective as a function of the order of convergence. To test if there was a significant correlation between the COM distance and the percent of context selective neurons, we fit a 1st order linear model and found that they indeed significantly correlate (**Figure 3.5E;** $p < 10^{-9}$, $R^2 = .49$). We concluded that differences in same syllable phonology after convergence points are encoded within HVC projection neurons.

The observation that neural activity correlates with the mean COM distance of the same syllables in different contexts raises the question of the extent to which HVC encodes the phonology of individual variations of these syllables. To compare how similar neural activity was we used the D' statistic across the change in fluorescence during the syllable window (see methods). The lower the D', the more overlapping the two signals are. We then compared each same syllable in each different 3 syllable context (for example syllable "b" in the "kia**b**" context vs in the "ghk**b**" context. Across the 4 birds recorded the mean D' across all neurons was significantly correlated with the inter-syllable COM distance (**Figure 4.1D**, $p < 10^{-7}$ $R^2 = 0.31$). We concluded that individual variations of the same syllables in different contexts are encoded within HVC.

4.2 HVC encodes phonology across different syllables

While the finding that HVC encodes phonological differences across renditions of the same syllable is striking, we wanted to test if this would also extend to phonological differences across different syllables. Another feature of Bengalese finch song is that distinct syllables can often have similar acoustic features. For example, in **Figure 4.1A,** syllables “b” and syllable “k” contain similar frequency bands that down-sweep over the course of the syllables. However, “k” is reliably distinguishable from “b” as it has a longer duration and flatter frequency bands at the start of the syllable (**Supplemental Figure 4.1A)**. Additionally, “k” always occurs before the syllable “h” making it easier to identify. Likewise, syllables “a” and “e”, and syllables “e” and “f” have shared acoustic structure. Consistent with these qualitative observations, the COM distances of all these pairs of syllables in **Figure 4.1B** have lower than average COM distances.

To test if these syllables which by COM distance are acoustically similar are also encoded by similar firing in HVC, we analyzed bursts of single neurons across similar syllables. We observed that neurons which burst for one syllable also tend to fire for syllables which looked acoustically similar. **Figure 4.1C** shows the derivative of the fluorescence signal of 3 neurons that all burst during syllable “b” sung after syllable “a” **Figure 4.1A**. Each of these neurons also burst during syllable “k” as well as the “b” that is sung after syllable “k”. The two syllable “b’s” in different contexts were more similarly encoded than syllable “k”. There were clear differences in the two “b’s” as the top neuron was more active during the “k**b**” while the middle neuron was more active during the “a**b**”. Both neurons also had a smaller increase in calcium signal during syllable “k”. The bottom neuron burst more strongly during syllable “k”, yet reliably burst for both “b’s”. This provides a striking example of how HVC projection

neurons represent three phonologically separable elements more similarly than expected if their firing was independent of acoustic features.

To test if this trend was significant across the population of all different syllables, we fit the mean D' of the increase in fluorescence during the syllable window against the mean COM distances of all different syllable pairs in each 3 syllable context ("k" in "cgh**k**" vs "b" in "kia**b**"). We grouped each syllable in these 3 syllable contexts to account for the consistent acoustic differences across convergence points. Across 4 birds we found that COM distance and neural D' were significantly correlated (**Figure 4.1E**, $p < 10^{-8}$ $R^2 = .02$). This result shows that syllables that have more similar acoustic features have more overlapping neural representations in HVC. This correlation explains less variation than the same syllables in different context comparisons (**Figure 4.1D**). This could be due to multiple factors, including that the metric of COM distance may not accurately pull out all features that HVC represents. While within syllable similarity will remain fairly stable across different acoustic features, different syllable comparisons are more sensitive to which features are used. Further work is needed to better understand which features HVC best correlates with and which syllables HVC encodes most similarly. Overall, we conclude that HVC can represent phonology both across different syllables and within the same syllable in different contexts. This provides strong evidence that in Bengalese finches, HVC encodes more than just sequence and timing and that the previously described phonological encoding in RA derives in part from HVC.

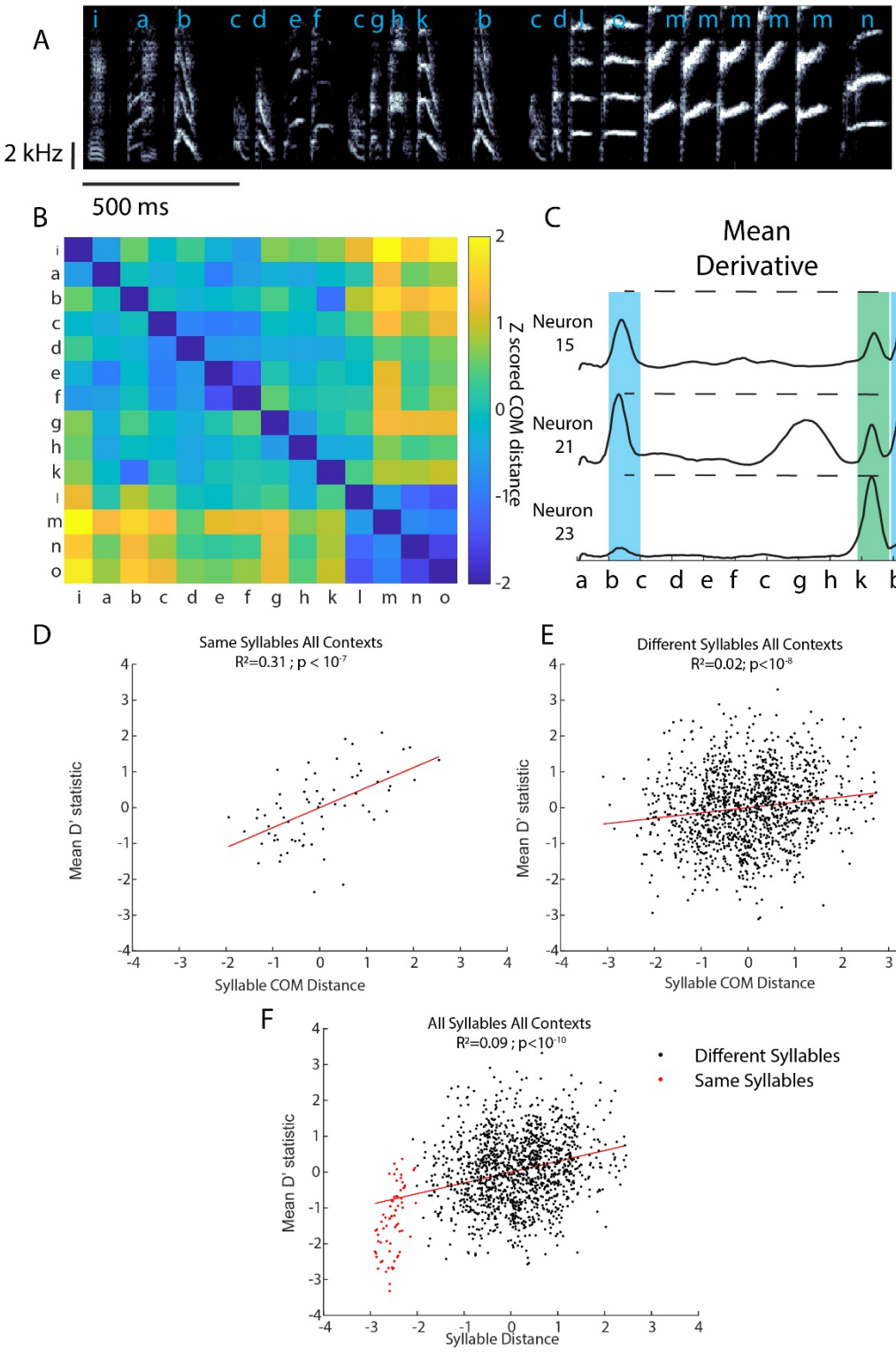
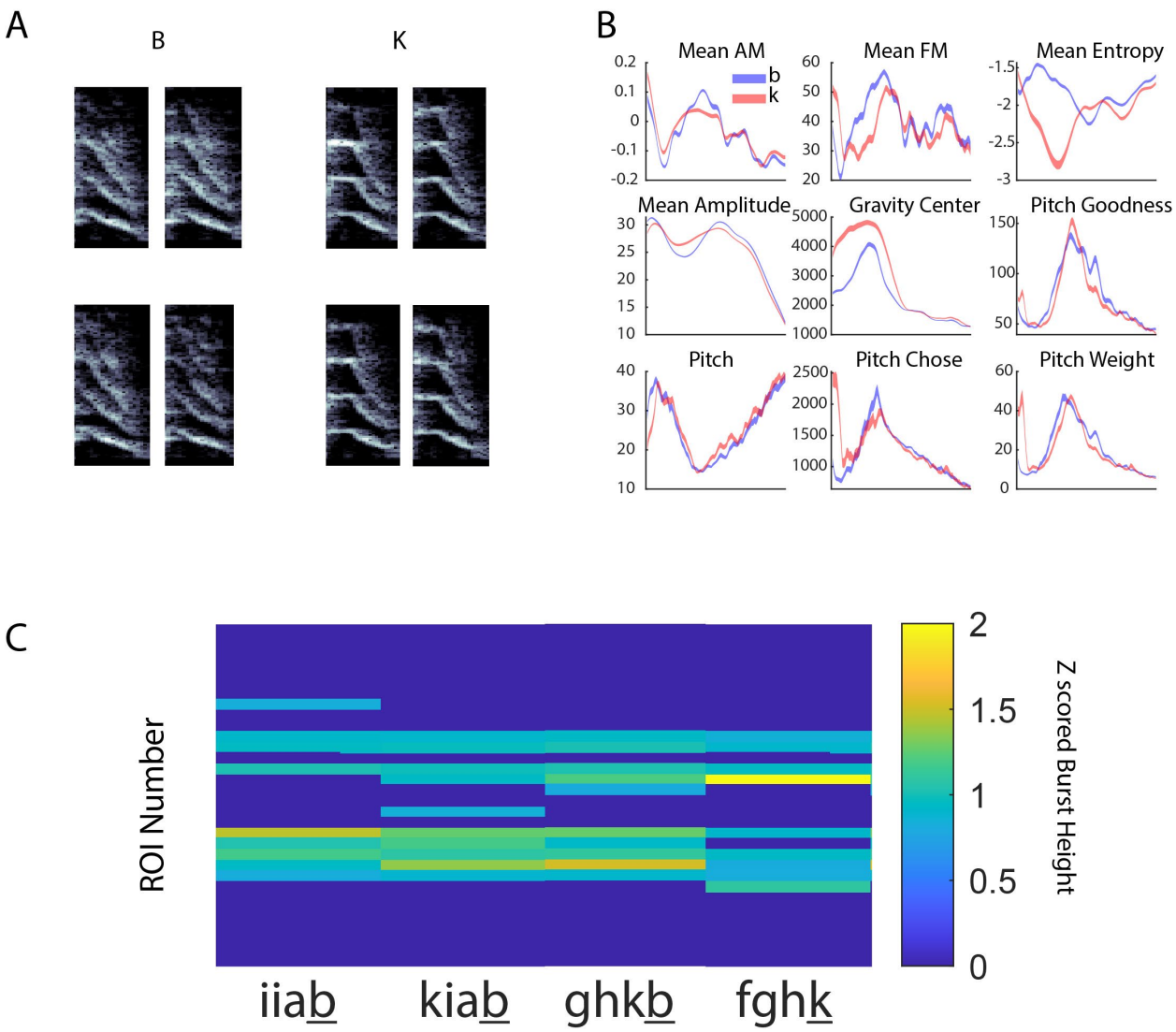


Figure 4.1: HVC encodes phonology within and across syllables

Legend: **A.** Spectrogram of song. **B.** Z-scored COM distances of syllables in spectrogram in 6A **C.** Derivative traces of 3 example neurons which fire (Figure caption continued on the next page)

(Figure caption continued from the previous page) during the song in A warped and aligned to the onset of each syllable in the motif. Dotted lines show the maximum derivative for each neuron. The z-scored D' of these neural representations is -.91 **D.** Z-scored mean D' of fluorescence increase in signal across all ROIs vs Z-scored COM distance across same syllables in different 1st-3rd order contexts **E.** Z-scored mean D' of fluorescence increase in signal across all ROIs vs Z-scored COM distance across different syllables in all 1st-3rd order contexts. **F.** All same and different syllable comparisons (combined data from D and E) z-scored and fit together



Supplemental Figure 4.1.1: SAP features extracted, and neural activity binned by syllable

Legend: **A.** Example of 4 renditions of syllable "b" and 4 renditions of syllable "k". **B**. Mean feature curves of all 9 SAP features used in PCA to calculate COM distance. Shaded area shows mean +/- 1 standard error. **C.** Neural activity for all recorded ROIs for all contexts of syllables "b" and "k"

Chapter 5: Discussion

In this study, we measured activity in populations of neurons within the HVC of Bengalese finches during the production of variably sequenced songs. Many prior studies have analyzed the role of HVC in the control of stereotyped song production in zebra finches. However, relatively little is known about how HVC contributes to birds who sing variably sequenced songs. Studying HVC activity during the variably sequenced song of Bengalese finches allows us to characterize how HVC encodes current actions, plans future actions, and keeps track of past history. Our results in the Bengalese finch show that the bursting of HVC projection neurons does not simply reflect the specific time and identity of the concurrently produced syllable, as in the zebra finch. Rather, consistent with previous studies in the Bengalese finch and canaries (Fujimoto 2011; Cohen 2020), we found that the magnitude of bursts during the production of a syllable can additionally encode information about sequence variation many syllables before and after the syllable.

We additionally took advantage of the graded acoustic similarity of many syllables in Bengalese finch songs to investigate whether HVC encodes aspects of syllable phonology. While much work has been put into automatic systems to cluster and label syllables, little has been studied about the correlation of these clusters in acoustic space vs clusters in neural space (Cohen et al. 2022; Liu et al. 2022). We found that HVC population representations were more similar across syllables which were acoustically more similar. This is true both for comparisons of activity for the same syllable produced in different sequences, and for comparisons across different syllables. These results are different from those observed in Zebra finches; indicating that HVC can encode more than just timing information about syllable production.

5.1 HVC Activity predicts upcoming divergence up to 10 syllables prior

One long standing question about divergence points in Bengalese finch song is whether the selection of which syllable to sing happens abruptly at the divergence point or evolves over a longer timescale. If these decisions appear to happen suddenly in HVC we would expect to see more "all-or-none" neurons. These could be just before the divergence point, or even further in advance. Our results did not find any of these "all-or-none" neurons until the gap after the divergent syllable. In other words, we did not observe robust differences in neural activity leading up to divergence points until just before the bird was singing different syllables. However, we found many neurons which had greater activity in one branch ("intermediate selective") up to 10 syllables prior to the divergence point. This was further in advance than previously reported and may be because a higher sample size of neurons was required to capture these types of differences (Fujimoto et al. 2011). This finding that HVC projection neurons begin to significantly bifurcate in their representation of the same syllables leading up to divergence points suggests a model where decisions are gradually made over a longer timescale. This result suggests a model where trial by trial variations of HVC activity can nudge the upcoming syllable selection process many syllables in advance. However, given the overlapping nature of these signals across branches, and the lack of "all-or-none" type neurons the ultimate selection of which syllable to sing appears to be made closer to the syllable itself.

What model of sequencing control does this suggest? Previous models of divergence points have built upon the idea of a feedforward synaptic chain driving control of song in a "branching chain" model of sequencing (Jin 2009). This model suggests that there is a "winner-take-all" competition between neural representations of the possible syllables which can be sung at the divergence points (**Figure 5.1 top)**. This competition will lead to one population of

neurons inhibiting the other through interneurons within HVC. One prediction of this model is that the relative strength of firing of neurons that are active around the divergence point may determine (and therefor correlate with) which branch of the chain "wins" during a given rendition of song. Consistent with this possibility, we found that many neurons active prior to branch points had firing rates that were predictive of which branch was subsequently followed. However, we also found such predictive differences in neural activity up to 10 syllables in advance of the divergence point. How could we account for such long-range relationships between firing and branching? One such possibility would be that there are two chains of neurons for each of the stereotyped syllables leading up to the divergence point (**Figure 5.1 bottom)**. These may inhibit each other over a longer timescale than previously thought from the predictions of the "branching chain".

We also observed that the neurons which are more active for the selected branch far in advance (> 7 syllables prior) also fire again either at the divergence point, or during one of the two branches. Thus, the past firing of neurons which are active during the variable transition to one branch or other is predictive of the upcoming decision. This may be due to intrinsic properties of the neurons which affect their firing at the divergence point. Indeed, we find the firing within the same neurons is correlated across motifs, meaning enhanced firing that is predictive of the upcoming syllable far in advance is more likely to also be predictive when that same neuron fires close to the divergence.

Finally, we often observed a significant reduction in neural activity over the course of song, including decreases in the amplitude of individual bursts over successive motifs. Such differences were not simply an artifact of calcium imaging, as we found a similar trend in multi-unit extracellular electrophysiology data and has not been previously reported. Understanding

how this decrease in activity contributes to branch selection at divergence points will require further investigation. We began this analysis by controlling for this effect by using mixed effects models. In one bird, we found that divergence point probabilities also changed over subsequent motifs. In this individual the mixed effects model significantly reduced the number of divergence selective neurons observed, especially those predictive 9 and 10 syllables prior to divergence. Thus, this effect of decreasing activity over song may play a role in the selection of upcoming syllables but does not fully explain the presence of divergence selective neurons.

5.2 HVC Maintains History of Context Many Syllables After Convergence Points

Bengalese finch song often contains long range sequence history dependence (Warren et al. 2012; Jin and Kozhevnikov 2011; Kentaro et al. 2011). For example, the statistics at divergence points are often dependent on the syllables preceding it. For example, a divergence of syllable “a” to “b” vs “c” may go to either option with a 50:50 probability when preceded by syllables “def”. This probability will change to 70:30 in favor of branch “b” if preceded by syllables “hig”. The mechanisms of these types of statistical changes dependent on past sequence are not well known. To be able to accomplish this kind of consistent statistical dependence, information of past sequence must be encoded within the brain. If HVC is important for making upcoming decisions based on past sequence, it must have access to this past information.

We found that HVC indeed does have different representations for syllables at convergence points. As previously reported, we found neurons which were selective for past sequence at the convergent syllable (Fujimoto et al. 2011). We additionally found many neurons that were selective for 2nd and 3rd order convergent syllables (syllables 1 or 2 syllables after the

convergence point. In one bird, we even found differences selective at 10th order convergent syllables. Thus, Bengalese finch HVC can encode long range information about previous sequences. This information could contribute to the previously observed history dependence in syllable sequencing.

5.3 HVC Encodes Phonological Similarity of Both Same and Different Syllables

HVC is often compared to a clock, as some view it only as a control center for timing of syllable production (Lynch et al. 2016). However, little has been studied about the encoding of phonology in HVC and whether it may influence acoustic features of syllables. The traditional view is that the mapping of specific features is encoded with RA. As HVC projects directly to RA, it follows that changes in HVC activity could also lead to changes in phonology.

Previous work has found that convergent syllables (same syllables sung in different contexts) have different acoustic properties depending on their context (Wohlgemuth 2010). This result along our current finding that HVC neurons encode past context suggests that activity in HVC could correlate with the phonology of these syllables on a trial-by-trial basis. Indeed, we found that the percent of neurons that are selective for different orders of convergent syllables correlates strongly with the COM distance of those syllables. More generally, we found that the COM distance of any pairs of syllables in different 3 syllable contexts ("fgh**c**" vs "iab**c**") correlates strongly with the mean D' of the neural activity during those syllables. Additionally, Bengalese finch song often contains syllables that are acoustically similar but are still clearly separable into two groups. We found that different syllable COM distances were strongly correlated with the similarity of their neural representations. This means that syllables which are

clearly distinct have overlapping representations in HVC if they have similar acoustic features. These results indicates that finds HVC activity can be correlated with acoustic features during unmanipulated singing.

Why has this encoding of phonology not been reported previously in recordings from Bengalese finch HVC (Fujimoto et al. 2011)? Two main reasons could be differences in labeling, as well as a lack of thorough analysis around syllable similarity. Labeling of syllables is a key decision that must be made to do any analysis of song. We found that acoustically similar syllables often had the same HVC projection neuron firing patterns. By clustering the audio features of these syllables which were similar but were labeled as distinct elements in PCA space, we observed clear separation between their clusters, indicating that they were indeed separable vocal elements. If previous analysis done in other labs had instead grouped these syllables as the same, they may not have noted this overlapping neural representation. It is important to note that even if we had labeled these different syllables as the same, our analysis would still have drawn the correlation with acoustic and neural similarity as we also compared the same syllables that were sung in different contexts. Indeed, when we correlate D' and syllable distance in both same and different syllables together, the correlation remains significant ($p < 10^{10}$; $R^2 = 0.09$; **Figure 4.1F**). As others may have not focused analysis on the acoustic variation of the same syllables in different contexts, they also may not have observed this correlation. Overall, our analysis of acoustic similarity indicates that HVC represents phonological information in Bengalese finches.

Why have these observations not been made in recordings of zebra finch HVC? One main reason is that each HVC_{RA} neuron bursts at most once per motif in zebra finch song. Any analysis focused on these neurons would be unable to compare overlapping HVC representations

as there are none. Additionally, many previous studies have treated all HVC bursts as independent and thus would not even be able to make this comparison in HVC_X projection neurons which fire during multiple acoustic elements. In our dataset, we do not know which types of HVC projection neurons we are recording from. Previous studies which express GCaMP using the same virus and CAG promoter that we used successfully labeled HVC_{RA} projection neurons (Daliparthi et al. 2019). Using retrograde DiI injections from Area X, we have confirmed expression of GCaMP in HVC_X projection neurons in our dataset. Thus, it is likely that we are also recording from both HVC_{RA} and HVC_X projection neurons. This leaves open the door that HVC_{RA} neurons may fire bursts for multiple syllables in Bengalese finches, which would be a significant discovery and show distinct coding from zebra finch HVC. This type of one-to-many encoding might be useful for building more syllables with fewer neurons. If certain populations of HVC_{RA} neurons already encode for a specific acoustic output, it would be more efficient to reuse these neurons for representations of other syllables rather than building distinct representations. Recordings from identified HVC_{RA} neurons are necessary to directly test this hypothesis.

If HVC and RA maintain more similar encodings for similar syllables, we would expect that when the motor program of one syllable is adjusted, the other one is also affected. One common paradigm in the birdsong field is to use an aversive white noise stimulus to shift the pitch of a targeted syllable (Tumer and Brainard 2007). This learning is mediated through the anterior forebrain pathway (AFP), which receives information from HVC through HVC_X neurons. Some of this learning is eventually consolidated in the motor pathway, as when AFP inputs to RA are blocked, the pitch of trained syllables is still shifted (Andalman and Fee 2009; Warren et al. 2011). If the motor pathway maintains a similar representation for similar syllables,

we would expect that if the pitch was shifted for one syllable, any syllables that are similar to it would also be shifted. In fact, if the pitch of a syllable in one context, but not the other context is targeted, there is a generalization of learning. (Hoffman and Sober 2014; Tian and Brainard 2017). For example, if the pitch of syllable "a" is targeted in the "bcd**a**" context but not the "efg**a**" context, the pitch of both syllables increases. Additionally, when AFP inputs to RA are blocked, the pitch of the non-targeted syllable remains unchanged, meaning this shift is fully encoded in the motor pathway (Tian and Brainard 2017). This suggests that the representation of the same syllable in different contexts is overlapping, as predicted by our neural observations. Further analysis is needed to find whether different syllables which are acoustically similar also have this type of generalized learning.

It is also possible the overlapping encoding of similar syllables is accomplished entirely through HVC_X neurons. HVC_X neurons can affect the acoustics of syllables through the AFP loop. However, the delay from HVC to RA through this loop is on the order of 50-100 ms (Kao et al. 2005), so the correlation of HVC with RA observed here would not affect acoustic outputs on this timescale. Thus, if HVC directly controls acoustic features of similar syllables, it is likely this is accomplished by the HVC_{RA} population. However, if the HVC_X population is responsible for encoding syllable similarity, its function may be less about more efficiently encoding syllables, and more about allowing the bird to keep track of which syllables are most similar. This could be useful for generalizing learning of one syllable to another similar one, which may help with initial vocal acquisition by co-learning similar vocal elements.

Impact

This work offers a population level look into population activity of singly identified HVC neurons in the Bengalese finch. Additionally, it provides a new method for recording from multiple brain regions simultaneously in awake freely moving songbirds. This work provides multiple insights which drive forward our understanding of variable motor sequencing in songbirds. First, we found that divergence points evolve over multiple syllables leading up to the divergence. While we can decode the upcoming syllable identity above chance, there is still a large amount of error in this signal. This suggests that the ultimate control over what syllable is sung at a divergence point does not occur until just before the branch syllable is sung. Next, we found that at convergence points, HVC encodes past sequences. This is strongly correlated with differences in phonology of convergent syllables. Finally, we found that HVC not only encodes phonological variation in the same syllables in different contexts, but also has overlapping representations for different syllables which are acoustically similar. This finding was unexpected as previous results from zebra finches have not reported any encoding of phonology within HVC. This result expands our understanding of the types of features which HVC encodes, and that different species may encode song differently, which reflects the many differences in their singing behaviors.

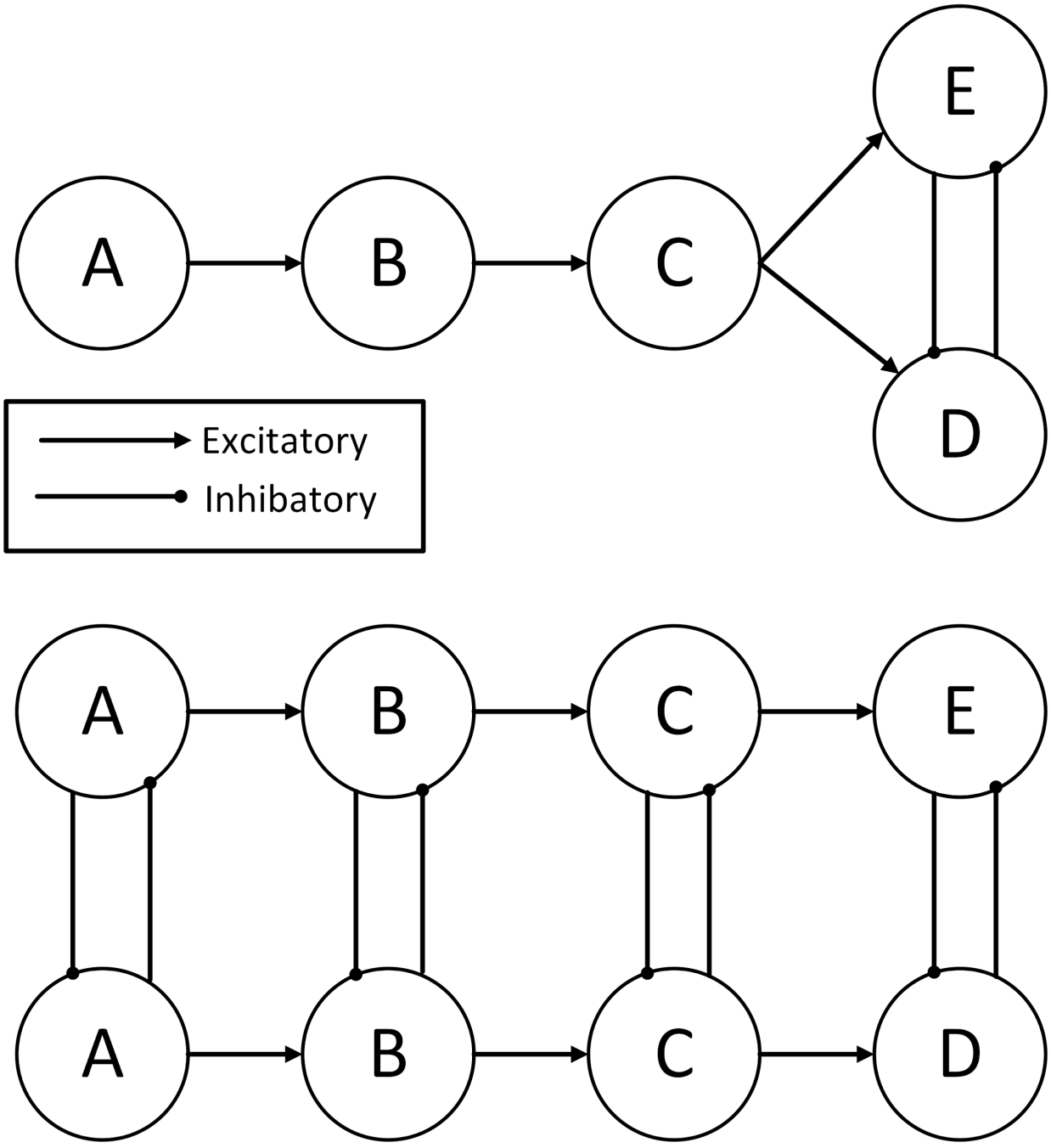


Figure 5.1: Models of divergence point branch selection

Legend Figure: Two models of divergence point syllable selection. Top: branching chain model proposed by Jin 2009 where neurons selective for the branching syllable mutually inhibit one another leading to one branch winning out. Bottom: parallel chains are simultaneously active leading to one branch or another. These chains inhibit one another leading to one ultimately winning out at the branch syllable.

Methods

Calcium Imaging

Camera settings

Calcium traces were collected using monochrome CMOS camera (FLIR BFS-U3-32S4M-C). We used the Spinnaker software from FLIR to adjust camera settings for data acquisition. We used a frame rate of either 20 or 33 Hz and binned at 2 pixels.

Raw data to calcium traces

After videos during singing were collected, we concatenated all frames into one larger video. Videos were then band pass filtered to remove both high frequency noise and low frequency background increases. To accomplish this, we used a 3-pixel gaussian kernel to smooth the video (high pass filtered) and made a separate version of the video smoothed with a 20-pixel gaussian kernel (low pass filtered). We then subtracted the low pass filtered video from the high pass filtered video. This band pass filtered video was then fed into CNMF using a custom version of the EZ calcium software package, which automatically detected ROIs (Cantu et al. 2020). Every ROI was manually checked to see if it fit the expected behavior of song aligned activity. We also confirmed activity automatically extracted for many ROIs activity by drawing manual ROIs using the maximum intensity projection of the concatenated video. Neurons were classified as either projection neurons or other based on whether the song aligned activity displayed the expected projection neuron behavior of sparse bursts. Only those ROIs that were classified as projection neurons were used for the analysis in this paper. To control for

bleaching, traces were normalized by the mean intensity of the video they were recorded from. The extracted intensity traces from ROIs were z-scored across all the songs they were recorded during to normalize across neurons.

Data analysis

Calculating burst time and magnitude

For analysis of singing related activity, traces of calcium activity were aligned to the onsets of individual syllables. Traces were then interpolated so that the new time between frames was 5ms. Traces were then averaged across 10 ms bins sliding at 1 ms to calculate the average syllable aligned activity. We used the derivative of the z-scored ROI intensities to determine whether neurons were bursting, based on the presence of well-defined peaks. The derivative was calculated by subtracting subsequent frames of ROI intensity dividing by the sampling rate. Bursts were identified as points in song where the derivative exceeded an intensity increase of 1 standard deviation per 100 ms. The time of the burst based on the first bin where the derivative was greater than 0.1 SD/100 ms prior to the burst. We additionally assumed a pre-motor delay of 25 ms, along with a delay in calcium increase after spike to be 25 ms for GCaMP6s and 0 ms for GCaMP8s, based on studies simultaneously recording calcium signal and electrophysiology (Zhang 2024). Bursts were associated with a syllable if the time of the burst occurred after the offset of the previous syllable and before the offset of the current syllable (**Figure 3.2 A**). For bursts assigned to syllables, burst magnitude was calculated for individual trials as the difference of the minimum and the maximum of z-scored trace during the burst. The minimum was the minimum intensity during the bins 50 ms prior and bin of the burst onset time. The maximum was the maximum intensity of bins up to 150 ms after the burst onset.

Significance across divergence and convergence points

To determine if a ROI had significantly different bursting across either divergence or convergence points, burst heights were calculated for each either past or future context and then compared using a Wilcoxon rank sum test.

For analysis of divergence points done in **Figure 3.3,** we identified the longest stereotyped sequence leading up to a divergence point in each bird (length of 4 to 11 syllables). This was done to control for other sequence variability leading up to a divergence in order to compare activity associated specifically with differences in divergences. We calculated where bursts occurred in the sequence and their trial-by-trial magnitude for each ROI and attributed each to a syllable (see calculating burst time and magnitude). We then used a Wilcoxon rank sum test to compare the trials of each branch type. Bursts were considered significantly different with $p < 0.05$. To control for differences expected by chance, we created a shuffled distribution of 1000 iterations where the labels for the divergence context were swapped. We calculated the percent of bursts that were significant in each of the 1000 iterations and then built a distribution. We then used the 95th percentile of shuffled bursts to determine the chance level (plotted as orange dotted line **Figure 3.3D**). For display purposes in **Figure 3.3E** we linearly time warped calcium traces relative to the time of offsets of successive syllables in the sequence. Traces were interpolated to 10 ms prior to warping. and then average across each branch.

For convergence points, we extracted out all 4 syllable chunks that occurred in song and found which of these chunks had 2 different syllable contexts preceding them, up to 15 syllables back. This allowed us to find convergent syllables up to the 15th order. For example, the 3rd order convergent syllable seen in **Figure 3.5 A** is syllable "e" sung in a "**e**iaf" context. It is a 3rd order convergence as it can be preceded by "bcd" or "hcd". Thus, the two compared sequences

are “bcd**e**iaf” and “hcd**e**iaf”. This ensures we compare convergent sequences that do not diverge until at least 3 syllables in the future. We then determined if a burst occurred in at least one of the contexts during the 1st syllable of the 4-syllable chunk. If a burst occurred, we compared the activity during this burst in the two different contexts using a Wilcoxon rank sum test.

Decoding of upcoming branch identity

We used the same data used for detecting differences leading up to divergence points (**Figure 3.3D**). We fit a Support Vector Machine (SVM) on the cumulative calcium signals prior to each syllable offset in the sequence. For example, if we have 10 syllables leading up to a divergence, at the 5th syllable before the branch syllable, we used all data which was assigned a burst time prior to the offset of the 5th syllable. If the sequence extended to 10 syllables prior to the branch this included bursts for the 10th through 5th syllables.

Data was split into 5 cross validation sets of equal size. We then fit the data 5 times using an SVM holding out 1 cross fold each time and predicting on the left-out fold. Thus, each subset of the data predicted once. To determine an overall accuracy score we took the average of the 5 cross folds. We did a grid search of various SVM parameters and select k-best features to determine which parameters we best at decoding the upcoming identity of a branch.

Correlations of burst heights

To determine if burst heights were correlated across motifs, we first identified the longest consistently occurring motif from each bird that ended in a divergence point. We then aligned, averaged, and extracted bursts in the same way as for the divergence point analysis. These bursts were organized into a matrix of size number of bursts by number of motifs. We computed the Pearson correlation of this matrix in to quantify correlations for each pair of bursts across motifs.

COM Distance

To determine acoustic similarity, we calculated 9 different Sound Analysis pro features: amplitude modulation, frequency modulation, wiener entropy, momentary amplitude, gravity center, pitch goodness, pitch, pitch chosen, pitch weight (Tchernichovski et al. 2000). These features are as described below:

Mean AM: amplitude modulation.

Mean FM: frequency modulation (0-90 degrees).

Mean Entropy: Wiener entropy or spectral flatness which describes how tonal a sound is.

Mean Amplitude: momentary sound amplitude in Db.

Gravity Center: mean frequency of the spectrogram.

Pitch Goodness: estimate of harmonic pitch periodicity. High goodness of pitch can be used as a detector of harmonic stack.

Pitch: estimate of the perceived pitch of a sound (lowest common denominator of frequency peaks).

Pitch Chosen: Similar to fundamental frequency. Derived from pitch and pitch weight.

Pitch Weight: Weight by which pitch chosen is derived from pitch. Based on pitch goodness.

These features were calculated on each bin of the spectrograms for each syllable. We then interpolated each feature so that each syllable had 100 samples across each of these 9 features. This 900-dimensional space was then reduced down to 10 dimensions using PCA. This PCA space was calculated across all syllables for each bird separately. Thus, each bird has its own PCA space. To compare how similar two syllables are in this space we calculated their center of

mass (COM) in this 10-dimensional space. We then found the Euclidean distance between these two COMs and used this as the COM distance of the two syllables.

D' statistic for neural similarity

To calculate how similar neural activity was for each pair of syllables, we determined the increase in intensity for all ROIs over each syllable window. We then calculated the D' statistics for each neuron during one context of each syllable vs. another (for example D' of syllable "c" in the "b" context vs the "h" context). The formula for D' is as follows:

$$D' = \frac{abs(\,mean\ context\ 1 - mean\ context\ 2)}{\sqrt{\frac{variance\ context\ 1 - variance\ context\ 2}{2}}}$$

This gave us a D' for each ROI. We then defined the neural similarity for two syllables as the mean D' across all ROIs.

Correlation between D' and COM distance

To compare whether the D' and COM distance of syllable pairs was related in **Figure 4.1** we fit a simple linear model using the MATLAB fitlm function. As we noted significant differences in firing during the same syllables in different contexts, we treated each syllable in all 3 syllable contexts as distinct. For example, "b" in a "ghkb" context was extracted separately from "b" in a "kiab" context.

We first compared all the same syllables in these different 3 syllable contexts. Both D' and COM distance we z-scored for each bird so that they were normalized and more comparable across birds. Next, all different syllables in 3 syllable contexts were compared. These were separately z-scored for each bird, not including the D' or COM distances of the same syllable

comparisons. As noted, if all D' statistics from same and different syllables are z-scored together, the correlation remains significant ($p < 10^{10}$; $R^2 = 0.14$).

Surgical Procedures and Bird Husbandry

Subjects

Imaging data was collected from n=4 adult (>150 dph) male Bengalese finches raised and housed at our bird colony located in the Sandler Neuroscience building on the University of California San Francisco campus. Birds were housed singly in cages and were isolated during collection of imaging data. Otherwise, birds were housed in sound boxes with other birds in isolated cages. Birds were maintained on a 14:10 light dark cycle.

Song Recording

Audio was recorded in a custom-written LabVIEW program (National Instruments; digitized at 44.1 kHz) using an omnidirectional microphone (Countryman; PN: B3P4FF05B) fed through an inline 100 Hz high pass filter (Shure; PN: A15HP) preamplifier (Symetrix; PN: 302) and 10 kHz low pass filter (Krohn-hite; PN: FMB300) into a National Instruments BNC 2090 terminal block. Songs were then digitally filtered between 500 and 10,000Hz. Onsets and offsets of syllables were determined using the rectified envelope of the amplitude smooth with a 2 ms square window. All labeling was done manually using custom MATLAB software.

Anesthesia and analgesia

Prior to all surgeries birds were injected with an appropriate amount of ketamine (20 mg/ml) and midazolam (1 mg/ml). Birds were placed on a custom stereotaxic and anesthetized using 0.1-1% isoflurane.

Virus injections and grin lens implants

HVC was targeted as reported previously using injection of retrograde lipophilic tracer DiI into Area X and using a fluorescent microscope to locate HVC after at least 1 week (Cohen 2020). After locating HVC, birds were injected either with AAV9-CAG-jGCaMP8s-WPRE or AAV9-CAG-jGCaMP8s-WPRE (Addgene 179256-AAV9 and 100844-AAV9). We injected ~ 750 ηL of virus into multiple depths around HVC ranging from 600- 200 μm below the surface of the brain. We then covered the craniotomy in silicon and waited at least 2 weeks for viral expression. For grin lens implantation we either left the hippocampus intact or removed it depending on how visible HVC was. Additionally, we used a fluorescent microscope to locate where GCaMP expression best overlapped with DiI. Grin lenses were lowered to the surface of the brain and then a further 100 μm to make sure they maintained contact with the brain and were then held in place with dental cement. We used a 1 mm diameter grin lens with a working distance of 200 μm (Inscopix 1050-004595). We implanted 61 birds with grin lens bilaterally over HVC. Of these, 14 birds had some level of song aligned increases in activity, and 4 birds which contained greater than 20 clear individual ROIs. The final dataset included n=1 bird in both left and right HVC, n=3 birds right HVC only. This is comparable to the sample size of other studies in the field.

References

1. Amador, Ana, Yonatan Sanz Perl, Gabriel B. Mindlin, and Daniel Margoliash. "Elemental Gesture Dynamics Are Encoded by Song Premotor Cortical Neurons." Nature 495, no. 7439 (March 2013): 59–64. https://doi.org/10.1038/nature11967.
2. Andalman, Aaron S., and Michale S. Fee. "A Basal Ganglia-Forebrain Circuit in the Songbird Biases Motor Output to Avoid Vocal Errors." *Proceedings of the National Academy of Sciences* 106, no. 30 (July 28, 2009): 12518–23. https://doi.org/10.1073/pnas.0903214106.
3. Andalman, Aaron S., Jakob N. Foerster, and Michale S. Fee. "Control of Vocal and Respiratory Patterns in Birdsong: Dissection of Forebrain and Brainstem Mechanisms Using Temperature." Edited by Stefan J. Kiebel. *PLoS ONE* 6, no. 9 (September 28, 2011): e25461. https://doi.org/10.1371/journal.pone.0025461.
4. Brainard, Michael S., and Allison J. Doupe. "Translating Birdsong: Songbirds as a Model for Basic and Applied Medical Research." *Annual Review of Neuroscience* 36, no. 1 (July 8, 2013): 489–517. https://doi.org/10.1146/annurev-neuro-060909-152826.
5. Cannon, Jonathan, Nancy Kopell, Timothy Gardner, and Jeffrey Markowitz. "Neural Sequence Generation Using Spatiotemporal Patterns of Inhibition." *PLoS Computational Biology* 11, no. 11 (November 2015): e1004581. https://doi.org/10.1371/journal.pcbi.1004581.
6. Cantu, Daniel A., Bo Wang, Michael W. Gongwer, Cynthia X. He, Anubhuti Goel, Anand Suresh, Nazim Kourdougli, Erica D. Arroyo, William Zeiger, and Carlos Portera-Cailliau. "EZcalcium: Open-Source Toolbox for Analysis of Calcium Imaging Data."

Frontiers in Neural Circuits 14 (May 15, 2020): 25. https://doi.org/10.3389/fncir.2020.00025.

7. Cohen, Yarden, David Aaron Nicholson, Alexa Sanchioni, Emily K Mallaber, Viktoriya Skidanova, and Timothy J Gardner. "Automated Annotation of Birdsong with a Neural Network That Segments Spectrograms." *eLife* 11 (January 20, 2022): e63853. https://doi.org/10.7554/eLife.63853.
8. Cohen, Yarden, Jun Shen, Dawit Semu, Daniel P. Leman, William A. Liberti, L. Nathan Perkins, Derek C. Liberti, Darrell N. Kotton, and Timothy J. Gardner. "Hidden Neural States Underlie Canary Song Syntax." *Nature*, June 17, 2020. https://doi.org/10.1038/s41586-020-2397-3.
9. Colquitt, Bradley M., Devin P. Merullo, Genevieve Konopka, Todd F. Roberts, and Michael S. Brainard. "Cellular Transcriptomics Reveals Evolutionary Identities of Songbird Vocal Circuits." *Science* 371, no. 6530 (February 12, 2021): eabd9704. https://doi.org/10.1126/science.abd9704.
10. Daliparthi, Vamsi K., Ryosuke O. Tachibana, Brenton G. Cooper, Richard Hr Hahnloser, Satoshi Kojima, Samuel J. Sober, and Todd F. Roberts. "Transitioning between Preparatory and Precisely Sequenced Neuronal Activity in Production of a Skilled Behavior." *eLife* 8 (11 2019). https://doi.org/10.7554/eLife.43732.
11. Egger, Robert, Yevhen Tupikov, Margot Elmaleh, Kalman A. Katlowitz, Sam E. Benezra, Michel A. Picardo, Felix Moll, Jörgen Kornfeld, Dezhe Z. Jin, and Michael A. Long. "Local Axonal Conduction Shapes the Spatiotemporal Properties of Neural Sequences." *Cell* 183, no. 2 (October 2020): 537-548.e12. https://doi.org/10.1016/j.cell.2020.09.019.

12. Fee, Michale S., Alexay A. Kozhevnikov, and Richard H.R. Hahnloser. “Neural Mechanisms of Vocal Sequence Generation in the Songbird.” *Annals of the New York Academy of Sciences* 1016, no. 1 (June 2004): 153–70. https://doi.org/10.1196/annals.1298.022.
13. Fetterman, Graham C., and Daniel Margoliash. “Rhythmically Bursting Songbird Vocomotor Neurons Are Organized into Multiple Sequences, Suggesting a Network/Intrinsic Properties Model Encoding Song and Error, Not Time.” Neuroscience, January 23, 2023. https://doi.org/10.1101/2023.01.23.525213.
14. Fiete, Ila R., Michale S. Fee, and H. Sebastian Seung. “Model of Birdsong Learning Based on Gradient Estimation by Dynamic Perturbation of Neural Conductances.” *Journal of Neurophysiology* 98, no. 4 (October 2007): 2038–57. https://doi.org/10.1152/jn.01311.2006.
15. Fujimoto, Hisataka, Taku Hasegawa, and Dai Watanabe. “Neural Coding of Syntactic Structure in Learned Vocalizations in the Songbird.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31, no. 27 (July 6, 2011): 10023–33. https://doi.org/10.1523/JNEUROSCI.1606-11.2011.
16. Hahnloser, Richard H. R., Alexay A. Kozhevnikov, and Michale S. Fee. “An Ultra-Sparse Code Underlies the Generation of Neural Sequences in a Songbird.” *Nature* 419, no. 6902 (September 5, 2002): 65–70. https://doi.org/10.1038/nature00974.
17. Hoffmann, Lukas A., and Samuel J. Sober. “Vocal Generalization Depends on Gesture Identity and Sequence.” *The Journal of Neuroscience* 34, no. 16 (April 16, 2014): 5564–74. https://doi.org/10.1523/JNEUROSCI.5169-13.2014.

18. Ikegaya, Yuji, Gloster Aaron, Rosa Cossart, Dmitriy Aronov, Ilan Lampl, David Ferster, and Rafael Yuste. “Synfire Chains and Cortical Songs: Temporal Modules of Cortical Activity.” *Science (New York, N.Y.)* 304, no. 5670 (April 23, 2004): 559–64. https://doi.org/10.1126/science.1093173.

19. Isola, Gaurav Ravindra, Anca Vochin, and Jon T. Sakata. “Manipulations of Inhibition in Cortical Circuitry Differentially Affect Spectral and Temporal Features of Bengalese Finch Song.” *Journal of Neurophysiology*, January 22, 2020. https://doi.org/10.1152/jn.00142.2019.

20. Jaffe, Paul I, and Michael S Brainard. “Acetylcholine Acts on Songbird Premotor Circuitry to Invigorate Vocal Output.” *eLife* 9 (May 19, 2020): e53288. https://doi.org/10.7554/eLife.53288.

21. Jin, Dezhe Z. “Generating Variable Birdsong Syllable Sequences with Branching Chain Networks in Avian Premotor Nucleus HVC.” *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 80, no. 5 Pt 1 (November 2009): 051902. https://doi.org/10.1103/PhysRevE.80.051902.

22. Jin, Dezhe Z., and Alexay A. Kozhevnikov. “A Compact Statistical Model of the Song Syntax in Bengalese Finch.” *PLoS Computational Biology* 7, no. 3 (March 2011): e1001108. https://doi.org/10.1371/journal.pcbi.1001108.

23. Kao, Mimi H., Allison J. Doupe, and Michael S. Brainard. “Contributions of an Avian Basal Ganglia–Forebrain Circuit to Real-Time Modulation of Song.” *Nature* 433, no. 7026 (February 2005): 638–43. https://doi.org/10.1038/nature03127.

24. Katahira, Kentaro, Kenta Suzuki, Hiroko Kagawa, and Kazuo Okanoya. “A Simple Explanation for the Evolution of Complex Song Syntax in Bengalese Finches.” *Biology Letters* 9, no. 6 (2013): 20130842. https://doi.org/10.1098/rsbl.2013.0842.

25. Katahira, Kentaro, Kenta Suzuki, Kazuo Okanoya, and Masato Okada. “Complex Sequencing Rules of Birdsong Can Be Explained by Simple Hidden Markov Processes.” *PloS One* 6, no. 9 (2011): e24516. https://doi.org/10.1371/journal.pone.0024516.

26. Katlowitz, Kalman A., Michel A. Picardo, and Michael A. Long. “Stable Sequential Activity Underlying the Maintenance of a Precisely Executed Skilled Behavior.” *Neuron* 98, no. 6 (27 2018): 1133-1140.e3. https://doi.org/10.1016/j.neuron.2018.05.017.

27. Konishi, M. “Birdsong: From Behavior to Neuron.” *Annual Review of Neuroscience* 8, no. 1 (March 1985): 125–70. https://doi.org/10.1146/annurev.ne.08.030185.001013.

28. Leonardo, Anthony, and Michale S. Fee. “Ensemble Coding of Vocal Control in Birdsong.” *The Journal of Neuroscience* 25, no. 3 (January 19, 2005): 652–61. https://doi.org/10.1523/JNEUROSCI.3036-04.2005.

29. Liberti, William A., Jeffrey E. Markowitz, L. Nathan Perkins, Derek C. Liberti, Daniel P. Leman, Grigori Guitchounts, Tarciso Velho, Darrell N. Kotton, Carlos Lois, and Timothy J. Gardner. “Unstable Neurons Underlie a Stable Learned Behavior.” *Nature Neuroscience* 19, no. 12 (2016): 1665–71. https://doi.org/10.1038/nn.4405.

30. Liu, Jiang, Yan Zhang, Danjv Lv, Jing Lu, Shanshan Xie, Jiali Zi, Yue Yin, and Haifeng Xu. “Birdsong Classification Based on Ensemble Multi-Scale Convolutional Neural Network.” *Scientific Reports* 12, no. 1 (May 23, 2022): 8636. https://doi.org/10.1038/s41598-022-12121-8.

31. Long, Michael A., and Michale S. Fee. “Using Temperature to Analyse Temporal Dynamics in the Songbird Motor Pathway.” *Nature* 456, no. 7219 (November 2008): 189–94. https://doi.org/10.1038/nature07448.

32. Long, Michael A., Dezhe Z. Jin, and Michale S. Fee. “Support for a Synaptic Chain Model of Neuronal Sequence Generation.” *Nature* 468, no. 7322 (November 18, 2010): 394–99. https://doi.org/10.1038/nature09514.

33. Lynch, Galen F., Tatsuo S. Okubo, Alexander Hanuschkin, Richard H. R. Hahnloser, and Michale S. Fee. “Rhythmic Continuous-Time Coding in the Songbird Analog of Vocal Motor Cortex.” *Neuron* 90, no. 4 (18 2016): 877–92. https://doi.org/10.1016/j.neuron.2016.04.021.

34. Markowitz, Jeffrey E., Elizabeth Ivie, Laura Kligler, and Timothy J. Gardner. “Long-Range Order in Canary Song.” Edited by Dezhe Jin. *PLoS Computational Biology* 9, no. 5 (May 2, 2013): e1003052. https://doi.org/10.1371/journal.pcbi.1003052.

35. Medina, Nelson D., and Daniel Margoliash. “Bursts from the Past: Intrinsic Properties Link a Network Model to Zebra Finch Song.” Neuroscience, May 19, 2024. https://doi.org/10.1101/2024.05.18.594825.

36. Nottebohm, Fernando, Tegner M. Stokes, and Christiana M. Leonard. “Central Control of Song in the Canary, *Serinus Canarius*.” *Journal of Comparative Neurology* 165, no. 4 (February 15, 1976): 457–86. https://doi.org/10.1002/cne.901650405.

37. Okanoya, Kazuo. “The Bengalese Finch: A Window on the Behavioral Neurobiology of Birdsong Syntax.” *Annals of the New York Academy of Sciences* 1016, no. 1 (June 2004): 724–35. https://doi.org/10.1196/annals.1298.026.

38. Okubo, Tatsuo S., Emily L. Mackevicius, Hannah L. Payne, Galen F. Lynch, and Michale S. Fee. “Growth and Splitting of Neural Sequences in Songbird Vocal Development.” *Nature* 528, no. 7582 (December 17, 2015): 352–57. https://doi.org/10.1038/nature15741.

39. Picardo, Michel A., Josh Merel, Kalman A. Katlowitz, Daniela Vallentin, Daniel E. Okobi, Sam E. Benezra, Rachel C. Clary, Eftychios A. Pnevmatikakis, Liam Paninski, and Michael A. Long. “Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch.” *Neuron* 90, no. 4 (18 2016): 866–76. https://doi.org/10.1016/j.neuron.2016.02.016.

40. Roberts, Todd F, Erin Hisey, Masashi Tanaka, Matthew G Kearney, Gaurav Chattree, Cindy F Yang, Nirao M Shah, and Richard Mooney. “Identification of a Motor-to-Auditory Pathway Important for Vocal Learning.” *Nature Neuroscience* 20, no. 7 (July 2017): 978–86. https://doi.org/10.1038/nn.4563.

41. Scherrer, Joseph R., Galen F. Lynch, Jie J. Zhang, and Michale S. Fee. “An Optical Design Enabling Lightweight and Large Field-of-View Head-Mounted Microscopes.” *Nature Methods* 20, no. 4 (April 2023): 546–49. https://doi.org/10.1038/s41592-023-01806-1.

42. Sober, Samuel J., Melville J. Wohlgemuth, and Michael S. Brainard. “Central Contributions to Acoustic Variation in Birdsong.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28, no. 41 (October 8, 2008): 10370–79. https://doi.org/10.1523/JNEUROSCI.2448-08.2008.

43. Tchernichovski, Ofer, Fernando Nottebohm, Ching Elizabeth Ho, Bijan Pesaran, and Partha Pratim Mitra. “A Procedure for an Automated Measurement of Song Similarity.”

Animal Behaviour 59, no. 6 (June 2000): 1167–76. https://doi.org/10.1006/anbe.1999.1416.

44. Tian, Lucas Y., and Michael S. Brainard. "Discrete Circuits Support Generalized versus Context-Specific Vocal Learning in the Songbird." *Neuron* 96, no. 5 (December 2017): 1168-1177.e5. https://doi.org/10.1016/j.neuron.2017.10.019.
45. Toader, Andrew C., Josue M. Regalado, Yan Ran Li, Andrea Terceros, Nakul Yadav, Suraj Kumar, Sloane Satow, Florian Hollunder, Alessandra Bonito-Oliva, and Priya Rajasethupathy. "Anteromedial Thalamus Gates the Selection and Stabilization of Long-Term Memories." *Cell* 186, no. 7 (March 2023): 1369-1381.e17. https://doi.org/10.1016/j.cell.2023.02.024.
46. Tumer, Evren C., and Michael S. Brainard. "Performance Variability Enables Adaptive Plasticity of 'Crystallized' Adult Birdsong." *Nature* 450, no. 7173 (December 2007): 1240–44. https://doi.org/10.1038/nature06390.
47. Vu, Et, Me Mazurek, and Yc Kuo. "Identification of a Forebrain Motor Programming Network for the Learned Song of Zebra Finches." *The Journal of Neuroscience* 14, no. 11 (November 1, 1994): 6924–34. https://doi.org/10.1523/JNEUROSCI.14-11-06924.1994.
48. Warren, Timothy L., Jonathan D. Charlesworth, Evren C. Tumer, and Michael S. Brainard. "Variable Sequencing Is Actively Maintained in a Well Learned Motor Skill." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32, no. 44 (October 31, 2012): 15414–25. https://doi.org/10.1523/JNEUROSCI.1254-12.2012.

49. Warren, Timothy L., Evren C. Tumer, Jonathan D. Charlesworth, and Michael S. Brainard. “Mechanisms and Time Course of Vocal Learning and Consolidation in the Adult Songbird.” *Journal of Neurophysiology* 106, no. 4 (October 2011): 1806–21. https://doi.org/10.1152/jn.00311.2011.

50. Wohlgemuth, Melville J., Samuel J. Sober, and Michael S. Brainard. “Linked Control of Syllable Sequence and Phonology in Birdsong.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30, no. 39 (September 29, 2010): 12936–49. https://doi.org/10.1523/JNEUROSCI.2690-10.2010.

51. Yu, A. C., and D. Margoliash. “Temporal Hierarchical Control of Singing in Birds.” *Science (New York, N.Y.)* 273, no. 5283 (September 27, 1996): 1871–75. https://doi.org/10.1126/science.273.5283.1871.

52. Zhang, Lifeng, Bo Liang, Giovanni Barbera, Sarah Hawes, Yan Zhang, Kyle Stump, Ira Baum, Yupeng Yang, Yun Li, and Da-Ting Lin. “Miniscope GRIN Lens System for Calcium Imaging of Neuronal Activity from Deep Brain Structures in Behaving Animals.” *Current Protocols in Neuroscience* 86, no. 1 (January 2019): e56. https://doi.org/10.1002/cpns.56.

53. Zhang, Yan, Márton Rózsa, Yajie Liang, Daniel Bushey, Ziqiang Wei, Jihong Zheng, Daniel Reep, et al. “Fast and Sensitive GCaMP Calcium Indicators for Imaging Neural Populations.” *Nature* 615, no. 7954 (March 30, 2023): 884–91. https://doi.org/10.1038/s41586-023-05828-9.

54. Zhang, Yisi S., Jason D. Wittenbach, Dezhe Z. Jin, and Alexay A. Kozhevnikov. “Temperature Manipulation in Songbird Brain Implicates the Premotor Nucleus HVC in

Birdsong Syntax." *The Journal of Neuroscience* 37, no. 10 (March 8, 2017): 2600–2611. https://doi.org/10.1523/JNEUROSCI.1827-16.2017.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Kurtis Swartz
10145D1ED51745B...
Author Signature

11/26/2024
Date