UNIVERSITY OF CALIFORNIA SAN DIEGO

Establishing Songbirds as an Animal Model for the Development of Human Speech Prosthesis

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Electrical Engineering

(Medical Devices & Systems)

by

Daril EVan Brown II

Committee in charge:

Professor Vikash Gilja, Chair
Professor Timothy Gentner, Co-Chair
Professor Shadi Dayeh
Professor James Friend
Professor William Hodgkiss
Professor Bradley Voytek

2022

The Dissertation of Daril EVan Brown II is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my Grandparents.
You each fought and sacrificed so that I may reach this summit.

EPIGRAPH

A life is not important except in the impact it has on other lives.

— Jackie Robinson

Not everything that is faced can be changed;
but nothing can be changed until it is faced.

—James Baldwin

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Look, you tell me I can't, I tell you I can. You tell me I can't, I do it again
—Tyler, The Creator

Graduate school has been anything *but* easy for me. However, I was blessed to have nothing but support from the people who mattered most. First and foremost I must thank God (I would not hear the end of it from my family otherwise). It is a miracle, honestly, that I have kept my faith throughout the arduous process that is the PhD training process. This is likely due to the amazing people who have rallied around me over the past 8 years as I have navigated the academic process.

I must next thank my main PhD Advisor Prof. Vikash Gilja. There is no succinct way to thank you for everything you have done for me over the past 6 years. Your leadership and mentorship have furthered my development as a neuroscientist, engineer, leader, and most importantly as a person. You have been one of my greatest allies and my most sincere reality checker. You allowed me to bring my full self into my research training and I would not be where I am now without your unyielding, enthusiastic support. I am in awe of your work ethic, moral character, diplomatic tact, optimistic demeanor, and always well-timed improv jokes. I will do my absolute best to live up to the amazing example you have set for me (excluding the improv).

Next, I must thank my co-advisor Prof. Timothy Gentner. The freedom you granted me as a member of the Gentner Lab was unparalleled. Whether prototyping crazy mechanical systems or disappearing for a week to write analysis software, you always trusted me to get my

work done and to do it well. You never once doubted my abilities and always pushed me to do my best. For that I thank you.

Prof. Bradley Voytek, though not officially my third PI, this has been the role you have played for years. When I was at a particularly low point in my academic career, your resume of failures was the catalyst that prevented me from losing faith in my ability to contribute meaningfully to science and engineering. I've said this to you personally, but I'd prefer to pass this kernel of wisdom on to anyone else who may read this thesis: "Failure is part of science. Don't let it keep you from pushing forward." Thank you for your unyielding and enthusiastic support.

I would like to acknowledge and thank the rest of my thesis committee, Prof. Shadi Dayeh, Prof. James Friend, and Prof. William Hodgkiss. It has been my privilege and honor to have each of you on my doctoral thesis committee. You each saw my potential and have worked to help me reach it. Thank you.

To my team of undergraduate mentees: Derek Nguyen, Jairo Chavez, Ahmed Abdalsatar (Adam Kadwory), and Xavier (Xavi) Perez. Your trust in me as your mentor and leader is greatly appreciated. Watching each of you grow from fledgling scientists to respected peers pursuing your own scientific and engineering careers has been the most rewarding part of my PhD Training. This accomplishment is in large part due to your collective efforts. Thank you for your hard work, and I can't wait to see all the amazing things you each will achieve in the future!

Working across three labs means that I have a ton of labmates that I must thank. However, I must first thank my labmates in the Translational Neural Engineering Lab (TNEL). Years ago, Paolo Gabriel first introduced me to TNEL when I was looking to transfer PhD

programs, and when I eventually joined the lab he took me under his wing. Thank you for being the best lab mom a fledgling grad student could have! TNEL has been my home for 6 years and our camaraderie and friendship has survived both pandemics and graduations. I want to acknowledge and thank as many of you as possible, Aashish Patel, Wahab Al-asfoor, John Hermiz, Abdullah Alothman, Pablo Tostado, Lauren Stanwicks, Paul Ngo, Sophia Huang, Nicholas Rogers, Tejaswy Pailla, Kenny Chen, Akin Omigbodun, Venkatesth Elango, Kevin Moses, Werner Jiang, Chenghao Gong and last but certainly not least Aparna Srinivasan. (This is not at all exhaustive, but I would be remiss if I didn't try) Thank you for the late night submission crunches, insightful conversations, random coffee runs (though I always drink tea) and random hangout sessions.

I want to also acknowledge and thank my labmates in both the Gentner Lab and Voytek Lab. There are way too many of you to list, however the years I have spent working and growing alongside each of you has been amazing. I especially want to acknowledge and thank Thomas Donoghue, Anna Mai, Tim Sainburg, Richard Gao, Kai Chen, Srihita Rudraraju, Sasen Cain, Dr. Natalie Schaworonkow, Trevor McPherson, Michael (MJ) Preston, Michael Turvey, Andrew Bender, Sydney Smith, Quirine van Engen, Sean Coffinger, and Ezequiel (Zeke) Arneodo. I want to thank all of you for your companionship, patience, insights, and support.

I am blessed to have a rather extensive support network outside of my research labs. This support network includes some of the best people on the planet. To my amazing friends Jim Short, Kevin White, Jordan (JoJo) Campbell, Kistine Andall, Jade Parker, Sage Aronson, Jaleesa Harrigan, Benjamin Rangel, Ollie D'Amico, Christopher Frick, Hristos Courellis, Kim McCabe, Nick Forsch, Khari Rockward, Nailah Seale, Alice Mukora, Nicole Hoffner, and Jonathan Bransomb. You guys keep me sane. Through the ups and downs you have always seen

the best in me even when I couldn't. Thank you! I also want to acknowledge and thank all of the members of the UCSD BGPSA and the entire UCSD Black Community. I got to thank all of you in person during my Graduate Keynote Speech, but it is worth repeating.

I want to acknowledge and thank the mentors who have helped me reach this point in my life. First my elementary school teacher Mrs. Paige. You were my first black teacher and although you were only in my life for an extremely short period of time, none of this would have been possible if you hadn't taken an interest in my education. From the bottom of my heart thank you. I want to thank both Dr. Will Liberti III and Dr. Ebonee Williams. You both took me under your wing and I can't thank you enough for all of your support. I want to thank Eugene Duvall, who, in my opinion, is a mechanical engineering genius. Working and learning with you has been a highlight of my graduate career. I look forward to collaborating on many more mechanical systems with you! I would like to thank my masters advisor Prof. Todd Coleman who mentored me as both a summer undergraduate and a graduate student. Truly, thank you. I want to also thank all my undergraduate research mentors: Prof. James Hammonds, Prof. Baratunde Cola, Prof. Tendo Foba, Prof. Wayne Patterson, and Prof. Marcus Alfred. I still use the skills each of you taught me and I thank you for taking me on when I was just beginning my science journey. Finally, I would like to thank Emily Bass for her expert editing to ensure that my research manuscripts and grants applications were as clear as possible. You juggle a lot of hats and I appreciate all the help you have given me over the years.

Finally, I must thank my family which is several orders of magnitude larger than the list of people already acknowledged above. I will have to thank each of you in person when I see you. I come from a long line of hard workers. Farmers, Teachers, Civil Rights Activists, and Veterans. This is as much your accomplishment as it is my own. Thank you for always having

my back, keeping me humble, and loving me unconditionally. I want to thank my amazing parents. The older I get the more I realize how lucky I am to be your son. My strength and resolve come from both of you, and I hope that I continue to make you proud. Finally, I want to acknowledge and thank my younger brother Brandyn. You keep me humble, and you are always someone I can rely on. Thank you.

Chapter 2, in full, has been adapted from material as it appears in PLOS Computational Biology, 2021, "Local Field Potentials in a Pre-motor Region Predict Learned Vocal Sequences," D. E. Brown II, J. I. Chavez, D. H. Nguyen, A. Kadwory, B. Voytek, E. Arneodo, T. Q. Gentner, and V. Gilja. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, has been adapted from material as it appears in PLOS Computational Biology, 2021, "Local Field Potentials in a Pre-motor Region Predict Learned Vocal Sequences," D. E. Brown II, J. I. Chavez, D. H. Nguyen, A. Kadwory, B. Voytek, E. Arneodo, T. Q. Gentner, and V. Gilja. The dissertation author was the primary investigator and author of this paper. Chapter 3, in part, also contains material as it adapted from Current Biology, 2021, E. M. Arneodo, S. Chen, D. E. Brown II, V. Gilja. The dissertation author was the primary author of this chapter.

Chapter 4, in full, has been submitted in a provisionary patent, 2021, "Dynamic Counterbalance to Enable Chronic Free-behaving Research with Small Animals", Daril EVan Brown II, Eugene Duval, Derek Hung Nguyen, Vikash Gilja, Timothy Q. Gentner. The dissertation author was the primary author of the white paper that informed the provisionary patent applicant and the primary author of this chapter.

VITA

2014 Bachelor of Science in Mechanical Engineering, Howard University

2016 Master of Science in Bioengineering, University of California San Diego

2022 Doctor of Philosophy in Electrical Engineering (Medical Devices & Systems), University of California San Diego

2022 Postdoctoral Fellow, University of California Davis Health, Neuroprosthetics Lab

PUBLICATIONS

D. E. Brown II, J. I. Chavez, D. H. Nguyen, A. Kadwory, B. Voytek, E. Arneodo, T. Q. Gentner, and V. Gilja, "Local Field Potentials in a Pre-motor Region Predict Learned Vocal Sequences," PLOS Computational Biology 17, no. 9, Sept. 2021. doi: 10.1101/2020.06.30.179861.

E. M. Arneodo, S. Chen, **D. E. Brown II**, V. Gilja, and T. Q. Gentner, "Neurally driven synthesis of learned, complex vocalizations," Current Biology, p. S0960982221007338, Jun. 2021, doi: 10.1016/j.cub.2021.05.035.\

D. E. Brown II, P. Tostado, J. Chavez, D. H. Nguyen, A. Kadwory, B Voytek, E. Arneodo, V. Gilja, T. Gentner. "Generalized Decoding of Vocal Elements in Zebra Finch". (2022. In Preparation)

DATASETS

Brown, Daril; Arneodo, Ezequiel; Gentner, Timothy Q; Gilja, Vikash (2021): Chronic Recording of HVC in Free Behaving Zebra Finch with Behaviors Hand Annotated. figshare. Dataset, doi: https://doi.org/10.6084/m9.figshare.15094219.v2

PATENTS

Daril EVan Brown II, Eugene Duval, Derek Hung Nguyen, Vikash Gilja, Timothy Q. Gentner. Dynamic Counterbalance to Enable Chronic Free-behaving Research with Small Animals. (*Provisional Patent submitted by UCSD OIC August 2022*)

ABSTRACT OF THE DISSERTATION

Establishing Songbirds as an Animal Model
for the Development of Human Speech Prosthesis

by

Daril EVan Brown II

Doctor of Philosophy in Electrical Engineering

(Medical Devices & Systems)

University of California San Diego, 2022

Professor Vikash Gilja, Chair
Professor Timothy Gentner, Co-Chair

Losing the ability to speak—whether from stroke, traumatic brain injury, or other neurological disorders— significantly reduces a person's quality of life. Research studies demonstrate proof-of-concept Speech-Synthesis Brain-Machine Interface (BMI) systems, but several limitations impede their clinical viability. A major rate limiting factor impeding progress in developing speech prosthesis is the lack of an established animal model to ask basic science questions regarding the neural encoding of vocal communication. This dissertation aims to address this gap by establishing songbirds as an animal model for a human speech prosthesis. Songbirds are a well-established model for vocal learning, and their motor nuclei are

homologous to the human motor cortex with respect to function and gene transcription. This work builds upon this basis to demonstrate songbirds' suitability as a preclinical model to accelerate the development of speech-prosthesis technology.

First, we answer basic science questions regarding nuclei important for the song system by characterizing neurophysiological similarities and differences with respect to human motor areas during vocalization. In analyses of data recorded with electrodes chronically implanted in the premotor region HVC of awake free-behaving zebra finches, we detail novel Local Field Potential (LFP) signatures correlated to vocal behavior. These LFP signatures are decomposed using signal processing methods to characterize their relation to vocal production. This work found that HVC exhibits many remarkably similar spectral characteristics to LFP in human motor cortex during speech.

Next, we developed proof-of-concept systems that demonstrate algorithms feasible for real time vocal BMIs. Utilizing simple algorithms, we show that HVC LFP features can be leveraged to predict vocal activity. These algorithms can be run in real time to predict both the identity and onset of syllable production. Leveraging these simple algorithms, we analyze preliminary system requirements necessary for decoding vocal elements. The methods developed to leverage these LFP features to predict vocal behavior can be implemented in real-time and suggest a path for developing a similar system for humans.

Finally, this thesis details both software and hardware designs to enable reproduction and wider adoption of the songbird animal model by the speech prosthesis research field. We developed novel methods to partition freely produced vocal behavior data based on the subjects' behavior, which are provided to the field as open-source

software. We also designed an integrated counterweight and tether management system that dramatically lowers the stress on chronically implanted small animal subjects.

Collectively, these works enrich the literature connecting human and avian vocal-motor production, and we believe strengthen the argument for utilizing songbirds to supplement human speech prosthesis research.

Chapter 1 : Introduction

1.1 Background

Vocal communication is a fundamental tool necessary to survive and thrive. However, each year, hundreds of thousands of people will experience a total loss or reduction in their ability to communicate [1]. Losing the ability to speak—whether from stroke, traumatic brain injury, or other neurological disorders— can have a considerable negative impact on someone's quality of life. It can stress interpersonal relationships, impair one's ability to perform at work, and become a major barrier to receiving adequate medical care [2–4].

Neuroprostheses — electronic devices that directly decode the neural correlates of a person's desired actions, thereby bypassing the injured part of the nervous system — stand to radically improve the quality of life for individuals with lost motor, speech, and language function. Current state of the art Motor Brain-Computer Interfaces (BCIs) have enabled paralyzed individuals to communicate through handwriting; with the current highest BCI performance to date being 19 words per minute[1] with 94.1% raw accuracy online [5]. However, the speed at which information can be conveyed through handwriting—which is about 15 to 40 words per minute for an average adult [6]—pales in comparison to natural speech; with an average speaker able to say 100 to 300 words per minute [7–9] as compared to writing. Vocal communication focused neural prosthesis aim to eventually return naturalistic, or close to naturalistic, speech rates to individuals who have lost the ability to speak.

[1] The system developed in this study was originally benchmarked at 90 characters per minute. This is roughly equivalent to 19.14 words per minute; given the average word length in the English language is 4.7 characters.

1.2 Overview of the Vocal Communication Field

Vocal communication requires dynamic coordination of the vocal tract articulator muscles. Remarkably, these synchronized sequences of motor movements produce a complex vocal output that is mutually comprehensible to both the communicator and their target. Speech focused BCIs are designed to decode the motor commands of a user's intended speech from neural signals recorded directly from motor regions of the brain. This is accomplished by applying statistical signal processing techniques to extract useful features from electrical signals recorded from speech related brain regions during vocal production. These features are then used to develop mappings from the neural activity to the message vocalized by the subject. Broadly, the communication output of all speech focused BCIs fall under two main types; (1) *Acoustic Synthesis* where the BCI tries to reproduce the intended sounds of the decoded communication [10–13], and (2) *Symbolic Decoding* where the BCI tries to decode symbolic components of the decoded communication (i.e. phonemes, vowels, words, etc) [14–16].

To date major progress has been made utilizing both approaches. One such recent breakthrough allowed a patient with anarthria—where someone loses all or significant amount of control of the speech musculature—was once again able to communicate [14]. However, current Speech BCIs are still significantly slower than natural speech. Increases in data quantity (many hours), scale (hundreds of parallel views of the brain), and diversity (different behavioral conditions) would help accelerate performance, however data from humans during speech is limited and no animal model supplements human work.

1.3 Motivating the Need for the Songbird Animal Model

While neural prostheses to restore limb movement have improved significantly over the years, speech and communication focused prostheses have lagged significantly behind. This is in large part due to the availability of a translational animal model, namely the rhesus macaque, to answer basic science questions regarding the neural encoding of motor movements. These basic science advances are then applied to proof-of-concept prototypes that are first validated in lab settings with the macaques before being translated to devices that are able to be used by humans in clinical settings. There is a long history of this translation of basic science advances in motor control in macaques [12,17–24] into advances in clinical motor BCIs in humans [5,25–28]. Unfortunately, macaques are a poor model for human speech as their capacity for learning complex vocalizations does not approach the abilities of humans. Nevertheless the speech BCI field would benefit from having a translatable animal model to supplement work currently done in humans.

A viable candidate to fulfill this need is the songbird. There is extensive research, spanning over 50 years [29,30], focused on the song system of songbirds. As a member of a short list of complex vocal learners—which includes humans, bats, elephants, whales, seals, and dolphins—they were first studied for their vocal learning capacity. However they have recently been studied for vocal motor control [31–33] and vocal memory [34]. Songbirds as a whole have the benefit of numerous species with varied levels of complexity ranging from simple (e.g. zebra finch) to complex (e.g. Starlings) with multiple species in between these levels (e.g. Bengalese finch and Canary).

1.4 Requirements of a Translatable Animal Model

To fulfill the role of a viable animal model for speech BCI development, songbirds must meet some minimum requirements. They must have similar physiology for vocal production, comparable neurophysiology for vocal motor control, enable invasive chronic recordings over long time periods (weeks to months), and have validated scientific tools to facilitate in-depth study. Although the underlying physiology for producing vocalizations differs between songbirds and humans they have been found to have similar auditory feedback requirements while producing vocalizations [35–37]. In addition, the nuclei crucial for song production have been found to be both functionally [38–41] and genetically [42,43] analogous to the premotor cortex of humans and non-human primates.

The birdsong neuroscience field, taking advantage of the accessibility of the song system, has largely focused on fundamental physiology. For this reason most songbird research has largely focused on carefully identified neural units and their spiking activity. As a result much is known about the spiking activity of nuclei crucial for song production, such as HVC and RA. However, while spiking activity from these nuclei has been well documented, local field potential activity (LFP) from these regions has only just recently begun to be studied [32,44]. This is significant because volume conductance signals, such as LFP, have been the most commonly studied features for most state-of-the-art speech BCIs [11–14]. It is important then to validate this proposed animal model in terms of its similarities and differences in regards to this important neural feature; as well as to demonstrate that these signals can be leveraged for decoding and predicting vocal elements.

Finally to enable widespread adoption of the animal model by the broader vocal communication BCI field there must be accessible tools to facilitate its study. This will be crucial as an animal model is only as reliable as the tools available to study it.

1.5 Chapter Previews

This thesis aims to address many of the unaddressed hurdles to validating and developing the songbird animal model as a suitable preclinical model to accelerate the development of speech-prosthesis technology. I first analyze LFP activity of HVC during vocal production using neural engineering approaches. Finding many remarkably similar spectral characteristics to LFP in human motor cortex during speech. I then extract useful neural features to develop simple proof-of-concept algorithms that can be leveraged to decode which syllable the bird will sing and predict when they will sing the. Finally I develop research tools to help facilitate reproduction of my findings by other labs and help lower the barriers for the songbird animal model to be adopted by the wider vocal communication prosthesis research field.

Chapter 2 is an in-depth analysis of HVC LFP from a neuroengineering perspective. It gives a brief introduction of the simplest songbird animal model, the zebra finch, then elucidates the network neural activity that proceeds vocal production.

Chapter 3 leverages the features documented in Chapter 2 to develop proof-of-concept BCIs for birdsong. It characterizes which component of the LFP, phase or power, is the most salient for decoder performance. It also asks many preliminary questions regarding the minimum system requirements for adequate decoder performance.

Chapter 4 documents tools developed to facilitate long term chronic invasive neural engineering research in small animals. Originally developed for songbirds, this system can be easily translated to other small animal models commonly used in neuroscience research.

Chapter 5 concludes this work with a brief description of future directions of this work as well as a discussion on where the songbird animal model fits into the larger Vocal Communication BCI field

1.6 References

1. Quick Statistics About Voice, Speech, Language. In: NIH National Institute on Deafness and Other Communication Disorders (NIDCD) [Internet]. 19 May 2016. Available: https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language
2. Beukelman DR, Fager S, Ball L, Dietz A. AAC for adults with acquired neurological conditions: A review. Augment Altern Commun. 2007;23: 230–242. doi:10.1080/07434610701553668
3. Morris MA, Meier SK, Griffin JM, Branda ME, Phelan SM. Prevalence and etiologies of adult communication disabilities in the United States: Results from the 2012 National Health Interview Survey. Disabil Health J. 2016;9: 140–144. doi:10.1016/j.dhjo.2015.07.004
4. Stransky ML, Jensen KM, Morris MA. Adults with Communication Disabilities Experience Poorer Health and Healthcare Outcomes Compared to Persons Without Communication Disabilities. J Gen Intern Med. 2018;33: 2147–2155. doi:10.1007/s11606-018-4625-1
5. Willett FR, Avansino DT, Hochberg LR, Henderson JM, Shenoy KV. High-performance brain-to-text communication via handwriting. Nature. 2021;593: 249–254. doi:10.1038/s41586-021-03506-2
6. van Drempt N, McCluskey A, Lannin NA. A review of factors that influence adult handwriting performance: FACTORS THAT INFLUENCE ADULT HANDWRITING PERFORMANCE. Aust Occup Ther J. 2011;58: 321–328. doi:10.1111/j.1440-1630.2011.00960.x
7. Jacewicz E, Fox RA, O'Neill C, Salmons J. Articulation rate across dialect, age, and gender. Lang Var Change. 2009;21: 233–256. doi:10.1017/S0954394509990093
8. Krause JC, Braida LD. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. J Acoust Soc Am. 2002;112: 2165–2172. doi:10.1121/1.1509432
9. Yoo J, Oh H, Jeong S, Jin I-K. Comparison of Speech Rate and Long-Term Average Speech Spectrum between Korean Clear Speech and Conversational Speech. J Audiol Otol. 2019;23: 187–192. doi:10.7874/jao.2019.00115
10. Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, Panko M, et al. A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. Ben-Jacob E, editor. PLoS ONE. 2009;4: e8218. doi:10.1371/journal.pone.0008218
11. Herff C, Diener L, Angrick M, Mugler E, Tate MC, Goldrick MA, et al. Generating Natural, Intelligible Speech From Brain Activity in Motor, Premotor, and Inferior Frontal Cortices. Front Neurosci. 2019;13: 1267. doi:10.3389/fnins.2019.01267
12. Angrick M, Ottenhoff MC, Diener L, Ivucic D, Ivucic G, Goulis S, et al. Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. Commun Biol. 2021;4: 1055. doi:10.1038/s42003-021-02578-0
13. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. Nature. 2019;568: 493–498. doi:10.1038/s41586-019-1119-1
14. Moses DA, Metzger SL, Liu JR, Anumanchipalli GK, Makin JG, Sun PF, et al. Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. N Engl J Med. 2021;385: 217–227. doi:10.1056/NEJMoa2027540
15. Moses DA, Leonard MK, Makin JG, Chang EF. Real-time decoding of question-and-answer speech dialogue using human cortical activity. Nat Commun. 2019;10: 3096. doi:10.1038/s41467-019-10994-4
16. Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic Feature Encoding in Human Superior Temporal Gyrus. Science. 2014;343: 1006–1010. doi:10.1126/science.1245994
17. Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov DF, et al. Learning to Control a Brain–Machine Interface for Reaching and Grasping by Primates. Idan Segev, editor. PLoS Biol. 2003;1: e42. doi:10.1371/journal.pbio.0000042

18. Ganguly K, Carmena JM. Emergence of a Stable Cortical Map for Neuroprosthetic Control. Ashe J, editor. PLoS Biol. 2009;7: e1000153. doi:10.1371/journal.pbio.1000153
19. Mulliken GH, Musallam S, Andersen RA. Decoding Trajectories from Posterior Parietal Cortex Ensembles. J Neurosci. 2008;28: 12913–12926. doi:10.1523/JNEUROSCI.1463-08.2008
20. Nuyujukian P, Kao JC, Ryu SI, Shenoy KV. A Nonhuman Primate Brain–Computer Typing Interface. Proc IEEE. 2017;105: 66–72. doi:10.1109/JPROC.2016.2586967
21. Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP. Instant neural control of a movement signal. Nature. 2002;416: 141–142. doi:10.1038/416141a
22. Suminski AJ, Tkach DC, Fagg AH, Hatsopoulos NG. Incorporating Feedback from Multiple Sensory Modalities Enhances Brain-Machine Interface Control. J Neurosci. 2010;30: 16777–16787. doi:10.1523/JNEUROSCI.3967-10.2010
23. Taylor DM, Tillery SIH, Schwartz AB. Direct Cortical Control of 3D Neuroprosthetic Devices. Science. 2002;296: 1829–1832. doi:10.1126/science.1070291
24. Velliste M, Perel S, Spalding MC, Whitford AS, Schwartz AB. Cortical control of a prosthetic arm for self-feeding. Nature. 2008;453: 1098–1101. doi:10.1038/nature06996
25. Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, et al. High-performance neuroprosthetic control by an individual with tetraplegia. The Lancet. 2013;381: 557–564. doi:10.1016/S0140-6736(12)61816-9
26. Hochberg LR, Bacher D, Jarosiewicz B, Masse NY, Simeral JD, Vogel J, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature. 2012;485: 372–375. doi:https://doi.org/10.1038/nature11076
27. Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature. 2006;442: 164–171. doi:10.1038/nature04970
28. Pandarinath C, Nuyujukian P, Blabe CH, Sorice BL, Saab J, Willett FR, et al. High performance communication by people with paralysis using an intracortical brain-computer interface. eLife. 2017;6: e18554. doi:10.7554/eLife.18554
29. Brainard MS, Doupe AJ. What songbirds teach us about learning. Nature. 2002;417: 351–358. doi:10.1038/417351a
30. Nottebohm F. The Neural Basis of Birdsong. PLoS Biol. 2005;3: e164. doi:10.1371/journal.pbio.0030164
31. Kozhevnikov AA, Fee MS. Singing-Related Activity of Identified HVC Neurons in the Zebra Finch. J Neurophysiol. 2007;97: 4271–4283. doi:10.1152/jn.00952.2006
32. Markowitz JE, Liberti WA, Guitchounts G, Velho T, Lois C, Gardner TJ. Mesoscopic Patterns of Neural Activity Support Songbird Cortical Sequences. Ashe J, editor. PLOS Biol. 2015;13: e1002158. doi:10.1371/journal.pbio.1002158
33. Picardo MA, Merel J, Katlowitz KA, Vallentin D, Okobi DE, Benezra SE, et al. Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch. Neuron. 2016;90: 866–876. doi:10.1016/j.neuron.2016.02.016
34. Yu K, Wood WE, Theunissen FE. High-capacity auditory memory for vocal communication in a social songbird. Sci Adv. 2020;6: eabe0440. doi:10.1126/sciadv.abe0440
35. Brainard MS, Doupe AJ. Auditory feedback in learning and maintenance of vocal behaviour. Nat Rev Neurosci. 2000;1: 31–40. doi:10.1038/35036205
36. Keller GB, Hahnloser RHR. Neural processing of auditory feedback during vocal practice in a songbird. Nature. 2009;457: 187–190. doi:10.1038/nature07467
37. Sakata JT, Brainard MS. Real-Time Contributions of Auditory Feedback to Avian Vocal Motor Control. J Neurosci. 2006;26: 9619–9628. doi:10.1523/JNEUROSCI.2027-06.2006
38. Bottjer SW, Johnson F. Circuits, hormones, and learning: Vocal behavior in songbirds. : 17.
39. Luo M, Perkel DJ. A GABAergic, Strongly Inhibitory Projection to a Thalamic Nucleus in the Zebra Finch Song System. J Neurosci. 1999;19: 6700–6711. doi:10.1523/JNEUROSCI.19-15-06700.1999

40. Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary,Serinus canarius. J Comp Neurol. 1976;165: 457–486. doi:10.1002/cne.901650405
41. Vates GE, Vicario DS, Nottebohm F. Reafferent thalamo-“cortical” loops in the song system of oscine songbirds. : 16.
42. Doupe AJ, Kuhl PK. BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms. Annu Rev Neurosci. 1999;22: 567–631. doi:10.1146/annurev.neuro.22.1.567
43. Pfenning AR, Hara E, Whitney O, Rivas MV, Wang R, Roulhac PL, et al. Convergent transcriptional specializations in the brains of humans and song-learning birds. Science. 2014;346: 1256846–1256846. doi:10.1126/science.1256846
44. Liberti WA, Markowitz JE, Perkins LN, Liberti DC, Leman DP, Guitchounts G, et al. Unstable neurons underlie a stable learned behavior. Nat Neurosci. 2016;19: 1665–1671. doi:10.1038/nn.4405

Chapter 2 : Characterization of HVC LFP

2.1 Abstract

Neuronal activity within the premotor region HVC is tightly synchronized to, and crucial for, the articulate production of learned song in birds. Characterizations of this neural activity detail patterns of sequential bursting in small, carefully identified subsets of neurons in the HVC population. The dynamics of HVC are well described by these characterizations, but have not been verified beyond this scale of measurement. There is a rich history of using local field potentials (LFP) to extract information about behavior that extends beyond the contribution of individual cells. These signals have the advantage of being stable over longer periods of time, and they have been used to study and decode human speech and other complex motor behaviors. Here we characterize LFP signals presumptively from the HVC of freely behaving male zebra finches during song production to determine if population activity may yield similar insights into the mechanisms underlying complex motor-vocal behavior. Following an initial observation that structured changes in the LFP were distinct to all vocalizations during song, we conduct an in depth analysis using signal processing techniques. Surprisingly, the time frequency structure of HVC LFP is qualitatively similar to well-established oscillations found in both human and non-human mammalian motor areas. This physiological similarity, despite distinct anatomical structures, may give insight into common computational principles for learning and/or generating complex motor-vocal behaviors.

2.2 Introduction

Learned vocalizations, such as speech and song, are generated by the complex coordination of multiple muscle groups that control the vocal organs[1–3]. As with other voluntary movements, this coordinated action arises from premotor brain regions [4–8] and is prepared prior to the physical initiation of movement [4,9,10]. How these behaviors are encoded during both preparation and generation remains an important topic of ongoing research. Developing an understanding for how the brain encodes complex sequential motor movements carries implications for the development of neural prostheses that aim to return or supplement lost motor function. In addition to their clinical application, such systems will help create new tools for examining the brain's mechanisms for learning and executing motor sequences.

At present, studying the motor encoding of speech and other complex motor movements in humans is challenging due to the intrinsic hurdles of conducting invasive human studies [11–15] and the complexity of human language [2,14]. Invasive studies that employ implanted electrodes are typically conducted in a clinical setting and clinical studies are inherently difficult. The clinical setting constrains experimental study duration and access to brain regions are limited. The study of other motor movements in humans is often supplemented by first studying simpler animal models such as non-human primates [16,17] and rodents [18–21]. However, these animal models fall short with respect to more complex freely generated motor sequences such as speech; this is primarily because none of the dominant models employed are capable of learning vocal behavior resembling the complexity of human speech [2,22]. For this reason, speech production studies, unlike other motor behavioral fields, have been limited exclusively to invasive [11,23,24] [25] and non-invasive [26], clinical studies in humans.

Chapter 2: Characterization of HVC LFP

Work in non-human primates and rodents have yielded physiological and engineering insights that enable the continued development of neurally driven upper limb prostheses for clinical functional restoration [4,16,27–32]. Given this track record of translation from animal models to clinical studies, despite anatomical differences in both limbs and brain structure, an animal model that exhibits similar vocal behavior, despite anatomical differences, could similarly benefit human speech prosthesis development. A natural candidate model is the songbird, which is widely used to study complex learned vocal behavior [33,34]. The zebra finch (*Taeniopygia guttata),* in particular, is known for its precisely timed sequentially structured song which is stable over the course of its adult lifetime. Experiments with this system have yielded insights into the acquisition [35,36], maintenance [37,38], and generation [8,39,40] of complex motor-vocal sequences.

Although the anatomical structure of avian and mammalian brains are divergent, relevant similarities between avian nuclei and human cortex have been identified. The premotor nucleus HVC, used as a proper name, contains neurons with sparse spiking patterns precisely aligned to vocal production [8,41]. Multiple studies suggest HVC is analogous to premotor cortex in human and non-human primates with respect to both functional [42–45] and genetic profile [46,47]. HVC provides input to two pathways that lead to neurons within nuclei that control the vocal muscles of the syrinx (Figure 2.1). The first is directly through the arcopallium (RA) in the posterior descending pathway (PDP), which is necessary for both the acquisition and production of learned motor-vocal behavior. For reference to mammalian anatomy, the PDP is analogous to a motor pathway that starts in the cerebral cortex and descends through the brain stem. The second, named the anterior forebrain pathway (AFP), plays a strong role in acquisition and adjustment of vocal output throughout life and projects indirectly through several nuclei; it is analogous to a cortical

pathway through the basal ganglia and thalamus in mammals [34]. These analogous regions share thousands of convergent genes despite their last shared ancestor being millions of years ago [47].

At a circuit level, the similarities between the neural activity in humans and songbirds is much harder to compare. Single-unit work and techniques, like those studied in songbirds, are difficult to conduct within a clinical setting. In contrast, given the accessibility of the system and a focus on fundamental physiology, songbird studies [6,10,34,35,37,38,40,44,48,49] have largely focused on carefully identified neural units and their spiking activity, with limited examination of other neural features that may correlate with vocalization behavior. Examining neural features readily accessible in both species and their relationship to motor-vocal behavior will enable direct comparison between the neural activity in birdsong and human speech production. Clarifying similarities (and differences) that may exist will bridge the gap between the two species.

Local field potentials (LFP) are thought to reflect both aggregate local postsynaptic activity and presynaptic inputs to the recording site [50]. As stable single units can be difficult to acquire and maintain in humans and non-human primates, there is a rich history of literature looking at both spiking and LFP for understanding, decoding, and predicting motor production [28,51–54]. At present, characterization of the neural dynamics of HVC and their relationship to behavior in songbirds has focused primarily on single- and multi-unit spiking activity [6,34,40,55,56], and limited work has focused on the structure of LFP and how it relates to song production. The most detailed characterization of this signal in songbirds is the relationship between LFP and interneuron synchrony [39]. This leaves a gap in the literature regarding the structure of LFP activity in HVC and whether its characteristics have any similarities to LFP in human, non-human primates, or mammalian premotor and motor regions.

We address this gap by chronically implanting and recording from freely behaving male zebra finch (Figure 2.1B) and analyzing LFP, presumptively from HVC, in relation to each bird's performed song (Figure 2.1C and D). We identify narrow-band oscillations [57–59] similar to those reported in human, non-human primate, and rodent motor electrophysiology literature. Further we provide evidence that phase and amplitude modulation within these frequency bands is predictive of vocalization behavior.

2.2 Data Acquisition

2.2.1 Overview of Recording Methods and Data Collected

Adult male zebra finches (n=3) were chronically implanted with laminar silicone microelectrode probes (Figure 2.1A). Local field potentials from these probes and vocal behavior from a microphone were simultaneously recorded (Figure 2.1B-D) (see section 2.2.4 Electrode implantation procedure). All references to LFP localization are recognized to be presumptively from HVC. The male zebra finches' learned song is structured and was annotated (Figure 2.2). The song consists of 1-7 repetitions of a motif, each of which is composed of a precisely ordered sequence of 3-10 discrete units called syllables. Beyond the syllables of the motif, male zebra finches may have an additional syllable or sequence of syllables that they will optionally insert between successive repetitions of motifs. These are called "connector" [60] or intra-motif notes. Song motifs are also grouped into larger structures called bouts, which consist of multiple repetitions of motifs. The bout is typically preceded by a variable number of repetitions of the same note, called Introductory Notes. Syllables of each bird's learned song, other non-learned vocalizations, and tokens of non-vocal intervals were segmented and annotated within periods of vocal behavior, referred to as vocally active periods (VAP), using acoustic landmarks (Figure 2.2B,C). Figure 2.3 provides state diagrams that describe the specific transitions between syllables

and other vocalization behaviors observed for all three of the subjects' VAP (colloquially, their song's "grammar"). Temporal boundaries of VAPs were used to define behaviorally relevant epochs in the simultaneously recorded neural signal. Subjects were recorded for 1-10 hours per day; statistics of recorded behavior data are documented in Table 2.1 and Table 2.2. For statistical power, we selected the two recording session days with the most vocal activity from each bird for the analyses reported here (Table 2.3, Table 2.4, and Table 2.5). Results from days with fewer recorded motifs were qualitatively similar.

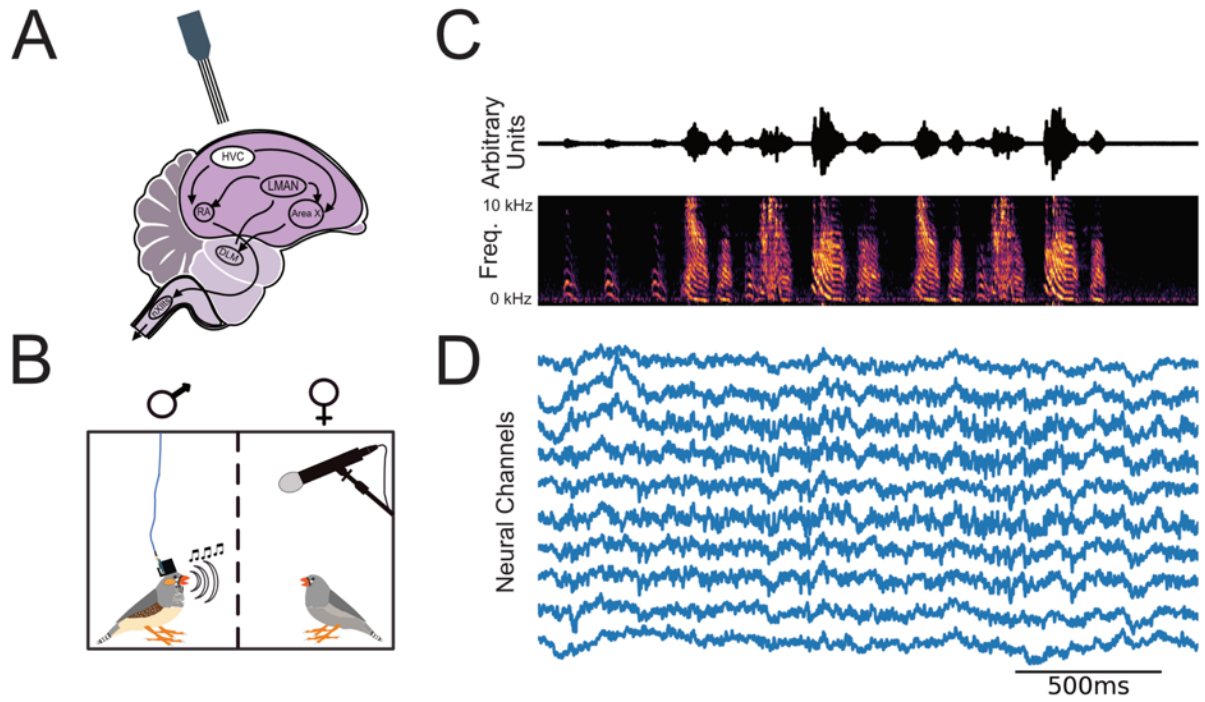


Figure 2.1: Continuous electrical and audio recording of chronically implanted freely behaving male zebra finch.

(A) Four-Shank multi-site Si-Probes were chronically implanted targeting nucleus HVC. HVC provides input to two motor pathways that innervate neurons in the tracheosyringeal half of the hypoglosssal nucleus (nXIIts) that projects to vocal muscles. The posterior descending pathway (PDP) comprises a direct projection to the robust nucleus of the arcopallium (RA) and is necessary for both the acquisition and production of learned vocal behavior (song). The anterior forebrain pathway (AFP), provides an indirect pathway to RA, through Area X, the dorsolateral anterior thalamic nucleus (DLM), and LMAN, and is necessary for song acquisition. The PDP is homologous to a motor pathway in mammals that starts in the cerebral cortex and descends through the brain stem, while the AFP is homologous to a cortical pathway through the basal ganglia and thalamus. (B) Neural and audio recording apparatus. We recorded LFP and vocal activity from male zebra finches stimulated to sing by the presence of a female conspecific. (C) Exemplar sound pressure waveform (1.3 seconds in duration, top) from bird z007 above the corresponding spectrogram. (D) Voltage traces (μV) of ten randomly selected channels of simultaneously recorded neural activity aligned to the song during audio recording, after a 400 Hz low-pass filter.

2.2.2 Subjects

All procedures were approved by the Institutional Animal Care and Use Committee of the University of California (protocol number S15027). Electrophysiology data were collected from three adult male zebra finches. Birds were individually housed for the entire duration of the experiment and kept on a 14-hour light-dark cycle (Figure 2.1). Each day had one session that lasted multiple hours: recording days were unstructured, as they depended on the subject's willingness to sing (Figure 2.2). All available days were analyzed; however, the two highest motif yielding days, hereafter referred to as *high-yield days*, were reported and used for statistical analysis. The full duration of chronic recordings ranged from 5 to 19 days. The birds were not used in any other experiments.

Table 2.1: Overall Description of Subject Data

Overall Description					
bird_id	# of Days Recorded Total	# of uniquely Labeled Syllables in Song (Including Intra Motif Notes)	# of Intra Motif Notes in Song	Presence of Unique Introductory Note	Electrode Geometry (Neuronexus Probes)
z020	5	4	1	Yes	A4x1-tet-3mm-150-121
z007	9	5	0	Yes	A4x4-tet-5mm-150-200-121
z017	19	7	2	No	A4x4-3mm-50-125-177

2.2.3 Electrophysiology and audio recording

Electrophysiological recordings were gathered from 3 subjects, in which two were implanted with 16-channel probes and one with a 32-channel probe (Figure 2.1A). We used 4-shank, 16/32 site Si-Probes (Neuronexus), which were PEDOT-coated in-house[2]. The probes were mounted on a custom designed printable microdrive [61] and implanted targeting nucleus HVC

[2]Protocol is available at https://www.protocols.io/view/EDOT-PSS-c2syed

(Figure 2.1A & B). Audio (vocalizations) was recorded through a microphone (Earthworks M30) connected to a preamplifier (ART Tube MP) and registered temporally to ongoing neural activity (Figure 2.1)]. Extracellular voltage waveforms and pre-amplified audio were amplified and digitized at 30 kHz using an intan RHD2000 acquisition system, Open Ephys and custom software (Figure 2.1D) [62].

2.2.4 Electrode implantation procedure

Animals were anesthetized with a gaseous mixture of Isoflurane/oxygen (1-2.5%, 0.7 lpm) and placed in a stereotaxic frame. Analgesia was provided by means of a 2mg/kg dose of carprofen (Rimadyl) administered I.M. The scalp was partially removed and the upper layer of the skull over the y-sinus was uncovered. The probe was attached to the shaft of a microdrive of our design[3], which was printed in-house using a b9 Creator printer and the BR-9 resin. A craniotomy site was open 2400 μm medial to the y-sinus (right/left hemispheres). The dura was removed and the electrode array was lowered to a 300-500μm depth. The opening was then covered with artificial dura (DOWSIL 3-4680 Silicone Gel Kit) and the microdrive was cemented to the skull using dental cement (C&B Metabond). A reference wire was made with a 0.5 mm segment of platinum-iridium wire (0.002") soldered to a silver wire lead and inserted between the dura and the skull through a craniotomy roughly 3mm medial (contralateral to the hemisphere where the electrode was inserted) and 1.5 mm anterior to the y-sinus. The reference electrode was cemented to the skull and the silver lead was soldered to the ref and gnd leads of the neuronexus probe. The craniotomy, the electrode, and the Microdrive were then covered with a chamber designed and 3D printed in-house, and which was subsequently cemented to the skull. The skin incision was sutured and adhered to the chamber with adhesive. The mass of the probe, Microdrive, and protective chamber

[3] Design Files for microdrive are available on Github.
https://github.com/singingfinch/bernardo/tree/master/hardware/printable_microdrive

were measured to be 1.2-1.4g. Upon the finches returning to a single-housing cage, a weight reliever mechanism was attached using the end of a thin nylon wire that was attached to an ad-hoc pin in the chamber; the other end routed through a set of pulleys and attached to a counterweight mass of ~1g [62].

2.2.5 Analysis of electrophysiology data

Extracellular voltage traces were multiplexed and digitized at 30kHz on the headstage, and stored for offline analysis. –**offline–** They were then low-passed filtered at 400 Hz using a Hamming finite impulse response (FIR) filter and downsampled to 1 kHz. The group delay introduced by the filter is compensated by introducing a temporal shift to the filter output [63].

2.3 Annotation and alignment of behavioral data

The song of an adult male zebra finch can be partitioned and labeled in multiple ways. However, the most fundamental characteristic of their song that is agreed upon is their stereotyped motif. The motif is a consistent sequence of sounds interleaved with silent periods that is learned from a tutor while they are young [10,35,36]. Song motifs are also grouped into larger structures called bouts, which consist of multiple repetitions of motifs [10,35,60]. Depending on the definition used, bouts can include a short, repeated vocalization that typically precedes the first motif of the bout. For the purpose of the analysis done, the start of the bout is explicitly defined to be the start of the first motif in that bout. This is due to the focus of understanding the correlation between LFP and learned motor sequences. Beyond the syllables of the motif male zebra finch may have an additional syllable or sequence of syllables that they will optionally insert between successive repetitions of motifs. This is called a “connector” [60] or intra-motif note. In our recordings, both z017 and z020 had intra-motif notes. As z007 did not have any intra-motif notes,

and therefore had a more stereotyped song, this bird will be used for all of the empirical figures shown.

A custom template matching algorithm written in Python was used to find potential instances of vocal activity (using an exemplar motif as the template); these instances were then curated manually to rule out false positives [62]. The curated motif start times were grouped into larger time segments that ensured that the gap between each motif was no greater than 20 seconds (Figure 2.2B). These chunks of time are subsequently referred to as vocally active periods (VAP). These VAPs were then manually labeled using the Praat speech analysis software [64]. Vocal events were segmented by hand (Figure 2.2C) and labeled based on identity. There is a unique label for each individual syllable and introductory note (if the bird demonstrated one). All other male vocalizations were grouped together and labeled as a call. Calls are short, simple vocalizations that are produced by both sexes and which mostly do not have learned components [36]. There were also two labels for silence in these VAPs: one for the gaps between song syllables, and another for silence not during song in the VAP.

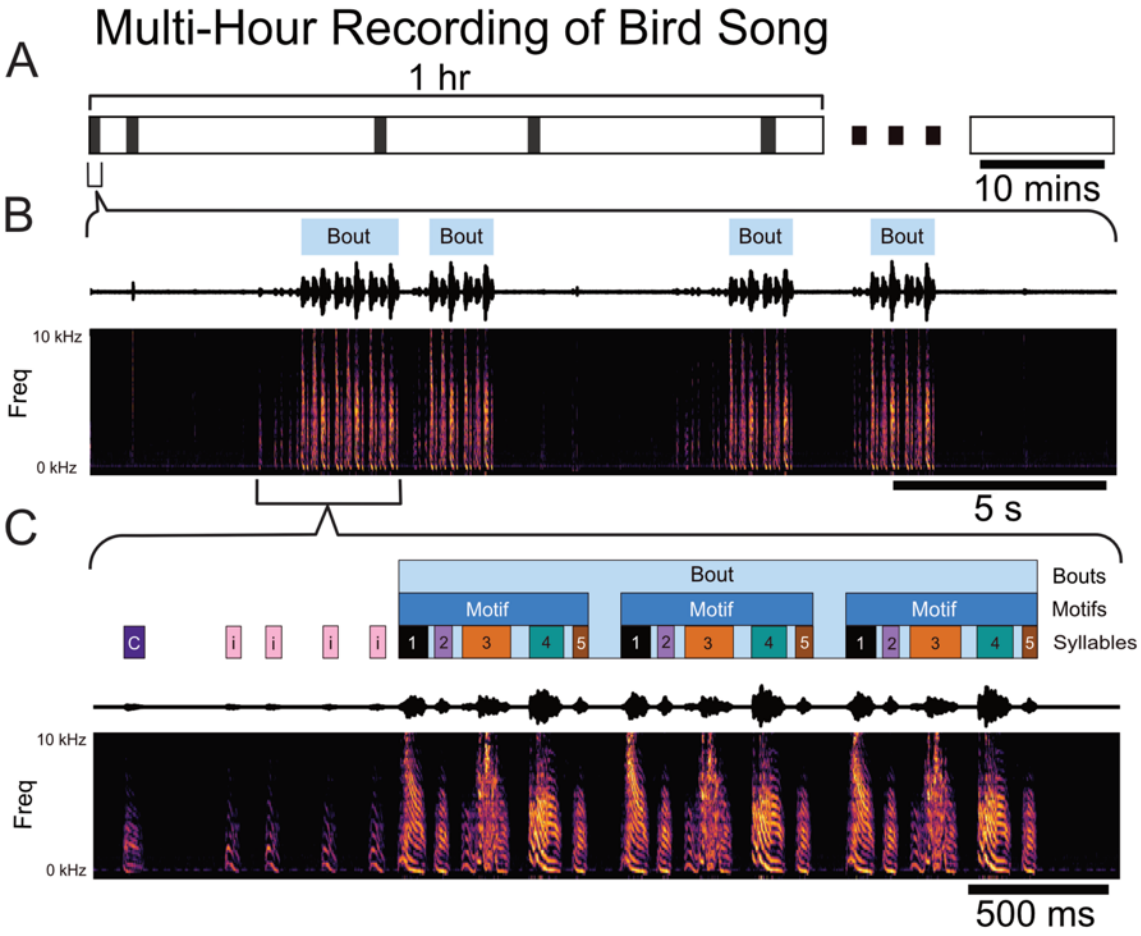


Figure 2.2: Temporal structure of vocal behavior.

(A) Schematic showing the intermittent occurrence of vocally active periods (VAPs, black bar) at sporadic intervals throughout the first hour of a multiple-hours-long continuous recording session of one male zebra finch (z007). (B) Zoomed-in view of one 25-second-long VAP comprising several bouts, denoted by light-blue rectangles above the sound pressure waveform and corresponding spectrogram. (C) Zoomed-in view of the first bout in segment (B) showing introductory notes, typically repeated a variable number of times prior to the start of the bout and labeled as ‘’'; other vocalizations not part of the courtship song, labeled ‘C’; and syllables comprising the courtship song, labeled ‘1’, ’2’, ‘3’, ‘4’, and ‘5’, based on their sequential order in the motif.

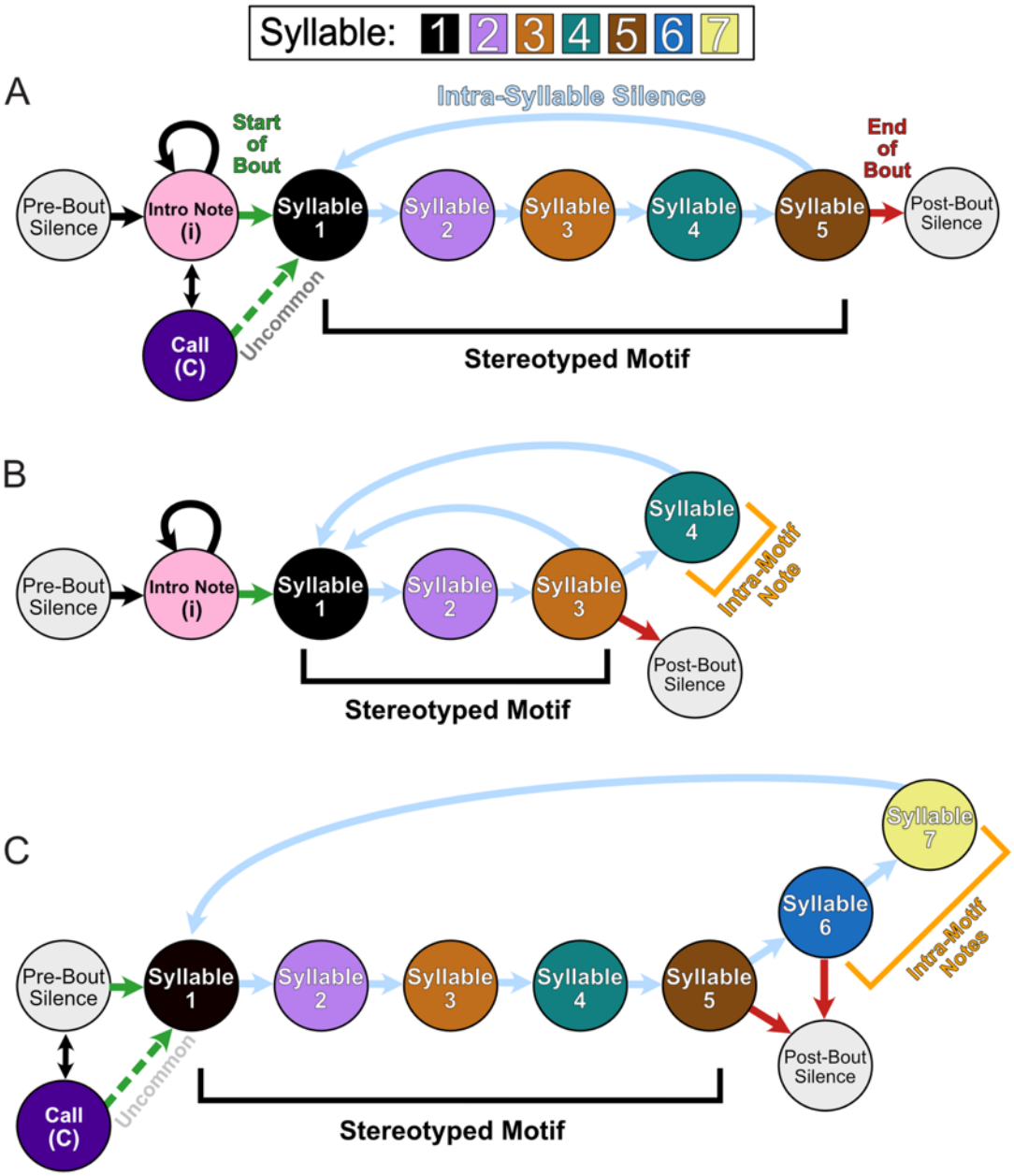


Figure 2.3: State Diagram of All Subjects Observed Song Behavior during High Yield Days.

(A) Schematic showing the vocalization structure of subject z007's song. (B) Schematic showing the vocalization structure of subject z020's song, including its intra-motif note syllable 4. (C) Schematic showing the vocalization structure of subject z017's song, including both of its intra-motif notes, syllable 6 and 7. For all three diagrams the green arrow represents the transition into the stereotyped motif song behavior; this sequence is underlined by a black bracket. The red arrow indicates the termination of the repeated motif sequence and the end of the bout.

Table 2.2: Detailed Description of available labeled data

Bird ID	Date	# of VAPs	Duration (hr. min. sec.)	# of Autodetected Motifs
z020	day-2016-06-02	6	5 hr. 13 min 9.092 sec	34
z020	day-2016-06-03	27	7 hr. 32 min 18.526 sec	113
z020	day-2016-06-04	14	7 hr. 43 min 43.603 sec	24
z020	day-2016-06-05	15	3 hr. 9 min 32.552 sec	85
z020	day-2016-06-06	9	2 hr. 40 min 39.42 sec	44
z007	day-2016-09-07	10	2 hr. 32 mins 46.24 sec	37
z007	day-2016-09-09	5	2 hr. 19 min 5.739 sec	19
z007	day-2016-09-10	14	2 hr. 13 min 26.45 sec	71
z007	day-2016-09-11	20	5 hr. 14 min 40.959 sec	92
z007	day-2016-09-12	11	5 hr. 27 min 7.001 sec	44
z007	day-2016-09-13	5	1 hr. 18 min 5.539 sec	60
z017	day-2016-06-19	18	8 hr. 38 min 29.42 sec	97
z017	day-2016-06-21	16	10 hr. 8 min 10.959 sec	76

Table 2.3: Behavior of Subject z007 during High Yield Days

z007			
Day 1			
Syllable	# of Instances	Duration (ms.) +/- std	Start within Motif relative to First Syllable (ms.) +/- std
1	72	100.084 +/- 3.558	-
2	72	55.406 +/- 2.877	125.694 +/- 3.101
3	71	164.915 +/- 3.296	221.830 +/- 4.087
4	71	120.870 +/- 2.654	448.069 +/- 6.111
5	71	53.958+/- 2.620	604.796 +/- 7.754
I	157	53.520+/- 4.662	-
Day 2 (Highest Yield Day)			
Syllable	# of Instances	Duration (ms.) +/- std	Relative Start within Motif (ms.)
1	100	102.532 +/- 6.561	-
2	100	57.612 +/- 7.598	127.509 +/- 6.525
3	100	167.799 +/- 8.020	224.254 +/- 8.082
4	100	124.418 +/- 5.822	452.354 +/- 9.369
5	98	55.548 +/- 4.489	611.310 +/- 9.780
I	178	53.621 +/- 6.870	-

Table 2.4: Behavior of Subject z020 during High Yield Days

z020			
Day 1 (Highest Yield Day)			
Syllable	# of Instances	Duration (ms) +/- std	Start within Motif relative to First Syllable (ms) +/- std
1	160	38.882 +/- 4.716	-
2	160	132.902 +/- 8.230	65.039 +/- 5.053
3	160	49.861 +/- 5.099	203.551 +/- 10.469
4	91	110.964 +/- 3.476	347.219 +/- 23.193
I	235	42.081 +/- 5.792	-
Day 2			
Syllable	# of Instances	Duration (ms) +/- std	Relative Start within Motif (ms) +/- std
1	109	39.887 +/- 6.139	-
2	109	132.925 +/- 4.865	67.670 +/- 5.01
3	109	50.257 +/- 5.011	206.310 +/- 4.985
4	75	111.741 +/- 6.112	348.720 +/- 12.329
I	191	46.922 +/- 9.378	-

Table 2.5: Behavior of Subject z017 during High Yield Days

z017			
Day 1 (Highest Yield Day)			
Syllable	# of Instances	Duration (ms) +/- std	Start within Motif relative to First Syllable (ms) +/- std
1	97	50.692 +/- 5.341	-
2	97	57.389 +/- 6.629	73.109 +/- 5.061
3	97	117.752 +/- 7.757	157.192 +/- 7.368
4	97	106.304 +/- 5.832	307.916 +/- 9.093
5	97	127.465 +/- 4.870	420.749 +/- 9.343
6	97	154.392 +/- 6.617	591.343 +/- 13.956
7	52	72.894 +/- 8.067	809.046 +/- 21.541
Day 2			
Syllable	# of Instances	Duration (ms) +/- std	Relative Start within Motif (ms) +/- std
1	84	49.162 +/- 6.525	-
2	84	51.217+/- 9.207	75.058 +/- 6.447
3	86	114.807 +/- 15.198	155.762 +/- 12.082
4	83	106.109 +/- 8.912	306.191 +/- 18.28
5	82	127.825 +/- 11.710	420.302 +/- 22.832
6	66	154.634+/- 11.803	584.710 +/- 14.729
7	41	68.696 +/- 8.431	803.838 +/- 26.629

2.3.1 BirdSongToolbox

To leverage the behavioral annotations for contextual analysis of their synchronously recorded neural activity, custom Python software was developed [BirdSongToolbox][4]. This software added additional labels to the hand labels based on their sequence and context within the larger bout structure (e.g., first or last motif in bout). Labeled vocalization segmentation times were adjusted to the proper sampling rate for the downsampled neural data using this same software. Finally, the software used these additional contextual labels to select particular event times to align both vocal and neural data that fit certain criteria matching the particular vocalization of interest.

BirdSongToolbox is designed to take in a dictionary that describes all annotated vocal units for a particular subject. This dictionary includes information about what type of vocalization the annotation refers to; i.e. 'part of song', 'call', or 'introductory note'. For vocalizations that are part of the song there is an additional indication of whether they are part of the stereotyped motif or an intra-motif note. The software then reviews the sequence of vocalizations and groups them into hierarchical labels based on the song sequence rules inferred from the subject specific dictionary. The base class for these annotations are designed specifically for the song structure of Zebra Finches, however it can be expanded to work with songbirds whose song structure is more complicated. This agnostic approach allows for a flexible software that can quickly be adapted for multiple subjects with little user intervention and can be expanded as the burgeoning songbird BCI field expands its repertoire of songbird species studied.

Although originally intended to be applied to annotated birdsong data, the species agnostic design of BirdSongToolbox can potentially be expanded to work with human speech research to

[4] BirdSongToolbox and its documentation are available on Github.
https://github.com/Darilbii/BirdSongToolbox

enable neurophysiology and neural engineering studies. This would be accomplished by using a dictionary derived from the language of focus and using techniques from natural language processing to create a class that would create relevant hierarchical labels. BirdSongToolbox's indexing functionality would be the real utility to researchers in this approach. Human Speech Researchers would be able to study free vocal behavior in humans yet have the capability to index both the recorded audio and neural activity data to conduct traditional neural engineering analysis.

2.4 Time series power spectral analysis of LFP during song production

2.4.1 Time series power spectrum

Spectrograms were calculated by first filtering the data with 100 partially overlapping narrow frequency bands that were equal width in log-space. These filtered time series were Hilbert-transformed to extract the analytic amplitude at each time point. Each narrowband amplitude time series was then normalized by the mean of that frequency band over the entire duration of its VAP.

2.4.2 Cross trial z-scored ratings of averaged spectrograms

The cross trial z-scored ratings of averaged spectrograms was calculated by first mean centering each frequency by its trial-wise mean

$$\hat{s}(f,t,n) = s(f,t,n) - \frac{1}{T*N}\sum_{m=1}^{N}\sum_{\tau=1}^{t} s(f,\tau,m) \quad (1)$$

Where s(f,t,n) represents every time, t, frequency, f, sample for each of the trials, n, spectrogram which were averaged in the manner used to create Fig 4. The mean of each time-frequency sample was then divided by the standard deviation of its time-frequency sample across trials.

$$S(f,t) = \frac{\frac{1}{N}\sum_{n=1}^{N}\hat{s}(f,t,n)}{stdev_n(\hat{s}(f,t,n))} \quad (2)$$

This approach calculates a z-score of each time-frequency mean to evaluate how well it represents that result across all trials. The closer to zero the value the more accurately it reflects the spectral change across all renditions of the motif. The greater in absolute magnitude the value the more it is above or below the norm across trials. A value of +/-2 is a approximately to within a 95% confidence interval for all trials values.

2.4.3 Song-related LFP spectral changes in HVC

Figure 2.4A illustrates the spectrotemporal characteristics of the LFP amplitude during the preparation and production of bird song, with an exemplar single trial time-varying power spectral density (PSD) during the course of three separate bouts. For each frequency, power is scaled to percentage of the mean amplitude during the respective VAP (see section 2.4.1). Through visual inspection of individual trials, relative to the post-bout period we noted a broadband increase in power in all frequencies above roughly 50 Hz prior-to and during vocal production (Figure 2.3). This occurs simultaneously with rhythmic changes in frequencies below roughly 50Hz (i.e. bout aligned increases and decreases in amplitude). When assessing single trials, we noted that changes in LFP amplitude appeared to correspond with each instance of vocal behavior, including non-learned portions such as introductory notes.

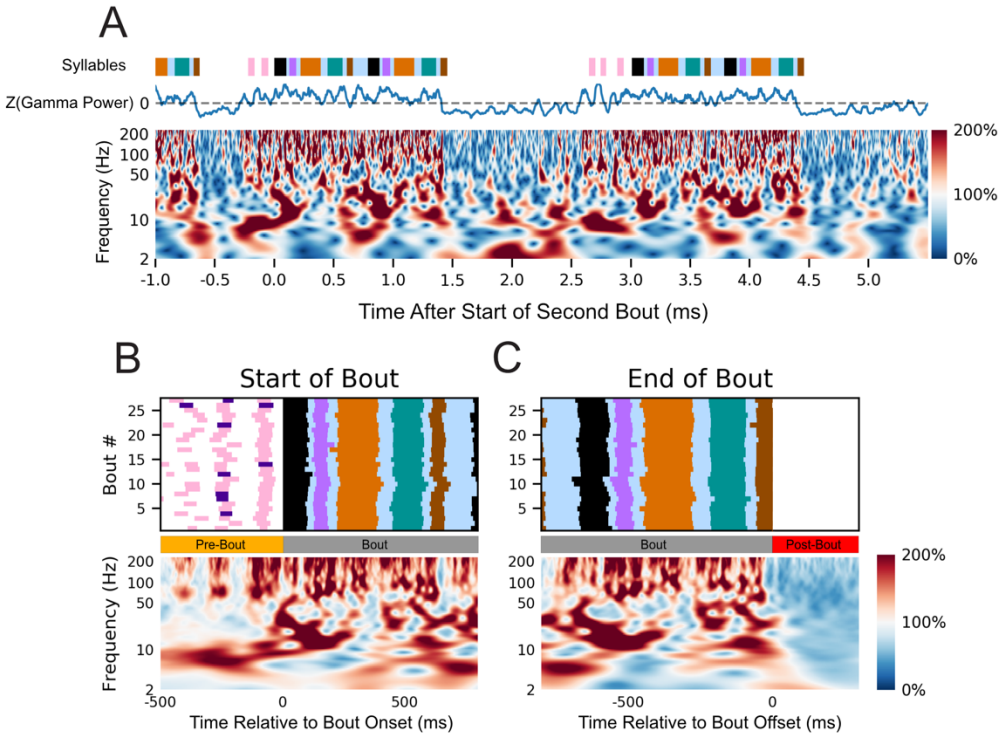


Figure 2.4: Variation in HVC LFP power correlate with vocal behavior.
(A) Normalized spectrogram (bottom) of LFP activity from one representative channel (see section 2.4.1 **Time series power spectrum**) aligned to the free vocal performance of a male zebra finch (z007) during the production of three separate bouts. Each frequency is scaled to its percentage of the mean amplitude during the respective VAP. Above the spectrogram is the z-scored power of the 50-200 Hz band smoothed with a 50ms rolling mean window aligned to the same behaviors. Colored bars above the spectrum annotate the vocal behavior, with color representing a distinct syllable with the same color-coding as Figure 2.2. (B) Normalized spectrogram (bottom) from the same representative channel as (A) averaged across multiple renditions of similar bouts (n=27, top) aligned to the start of the first motif in the bout. (C) As in (B) but aligned to the end of the last motif in the bout. Above each is a behavioral raster showing the time course of the behavior being averaged (n=27 bouts). No dynamic-time warping was used, and all motifs were from the highest yield day for the subject. To ensure the start and end of the bout are unique time periods, only bouts with more than one motif in duration were used. Behaviorally inconsistent bouts were excluded for clarity of visualization; including them does not qualitatively alter the result.

To evaluate the consistency of these changes in amplitude, we used the one-sided z-test to compare the distributions of power values during vocalizations to the distribution of power values during silence. A statistically significant increase in power for frequency bands above 50 Hz was seen across most, if not all, channels for all but one of the high yield days for z020 (Table 2.6 and Table 2.7). Changes in amplitude for frequency bands below 50 Hz were more nuanced, with no

consistent trend across subjects; however, frequency changes tended to be consistent across days for two of the birds, particularly z017 and z007 (Table 2.6 and Table 2.7).

Table 2.6: Characteristic Increases in Power in Narrowband Frequencies during Song Across Channels.

The percentage of channels whose distribution of power values were greater during periods of vocal activity than during silence. To determine statistical significance between the two distributions the one-sided z-test was used with each frequency and channel pair. The percentages shown are the number of good channels that still were significant after using the Benjamini-Hochberg False Discovery Rate ($p<0.05$ and $q<0.05$).

		Percentage of Channels with Increased Power During Vocally Active Periods (After FDR Correction)					
Bird_ID	Date	4-8 Hz	8-12 Hz	25-35 Hz	35-50 Hz	50-70 Hz	80-200 Hz
z020	day-2016-06-03	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
z020	day-2016-06-05	53.33%	26.66%	26.66%	26.66%	40.00%	26.66%
z007	day-2016-09-10	70.00%	76.66%	86.66%	100.00%	100.00%	100.00%
z007	day-2016-09-11	73.33%	80.00%	100.00%	100.00%	100.00%	100.00%
z017	day-2016-06-19	0.00%	6.25%	62.50%	68.75%	87.50%	100.00%
z017	day-2016-06-21	0.00%	6.25%	81.25%	81.25%	100.00%	100.00%

Table 2.7: Characteristic Decrease in Power in Narrowband Frequencies during Song Across Channels.

The percentage of channels whose distribution of power values were smaller during periods of vocal activity than during silence. To determine statistical significance between the two distributions, the one-sided z-test was used with each frequency and channel pair. The percentages shown are the number of good channels that still were significant after using the Benjamini-Hochberg False Discovery Rate ($p<0.05$ and $q<0.05$).

		Percentage of Channels with Decreased Power During Vocally Active Periods (After FDR Correction)					
Bird_ID	Date	4-8 Hz	8-12 Hz	25-35 Hz	35-50 Hz	50-70 Hz	80-200 Hz
z020	day-2016-06-03	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
z020	day-2016-06-05	40.00%	53.33%	40.00%	60.00%	46.60%	60.00%
z007	day-2016-09-10	30.00%	19.99%	3.33%	0.00%	0.00%	0.00%
z007	day-2016-09-11	26.66%	19.99%	0.00%	0.00%	0.00%	0.00%
z017	day-2016-06-19	100.00%	93.75%	31.25%	31.25%	0.00%	0.00%
z017	day-2016-06-21	100.00%	87.75%	6.25%	0.00%	0.00%	0.00%

The spectral structure of the LFP activity immediately prior to and during the production of learned song is highly consistent across bouts, as suggested when viewed across time for multiple bouts (Figure 2.4A) and when they are aligned to their initiation and termination (Figure 2.4B,C and Figure 2.5). Visualizations of activity around the times of bout initiation and termination were calculated by aligning to the start of the first syllable of the first motif in the bout (initiation) and the start of the last syllable of the last motif of the bout (termination) for all bouts recorded over the course of a single day. The aligned neural activity was then averaged for each time-frequency sample. These spectral structures were validated as a strong summary representation of the activity during the transition into and out of the stereotyped motif, despite trial-by-trial variability, by a modified z-score method (Figure 2.6) (see section 2.4.2 Cross trial z-scored ratings of averaged spectrograms). As with the single trial time-varying PSD, there is an increase in amplitude for all frequencies above roughly 50 Hz (Figure 2.5 and Figure 2.7).

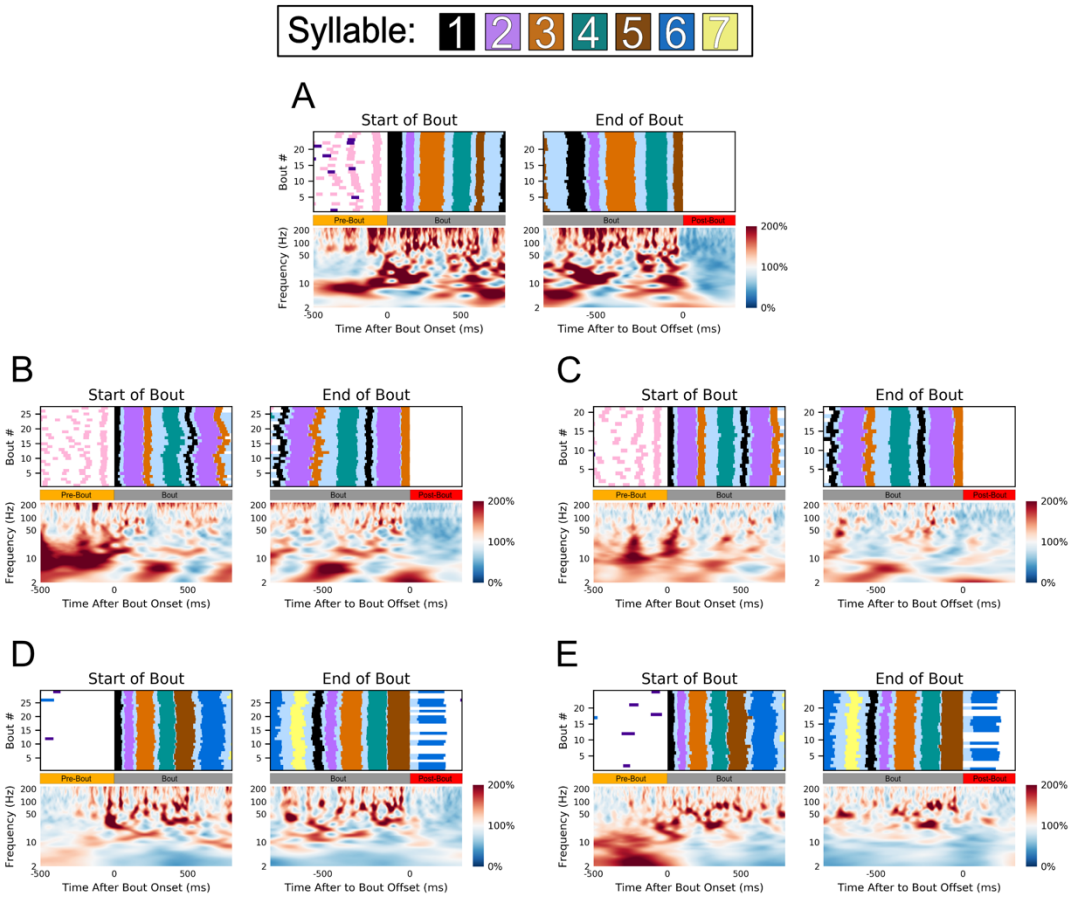


Figure 2.5: Song-correlated modulation of LFP power across subjects and days.
Averaged spectrotemporal power activity (see section 2.4.1 **Time series power spectrum**) aligned to the start of the first motif in the bout, left, and the last motif in the bout, right, for each recording day not shown in Fig 4. Shown above all results is a behavioral raster showing the time course of the behavior being averaged. (A) The averaged results for the second highest-yielding day, designated Day 1, for z007 (n=25 Bouts). The other subjects' results are show as follows; (B) z020's first high yield day (n=29 Bouts), (C) z020's second high-yield day (n=25 Bouts), (D) z017's first high-yield day (n=27 Bouts), and (E) z017's second high-yield day (n=21 Bouts). As z017 would end its bout on either syllable '5' or '6', the end of the bout was aligned to syllable '5'. No dynamic-time warping was used. To ensure the start and end of the bout are unique time periods, only bouts with more than one motif in duration were used. Behaviorally inconsistent bouts were excluded for clarity of visualization; however, results are consistent when including them in the analysis.

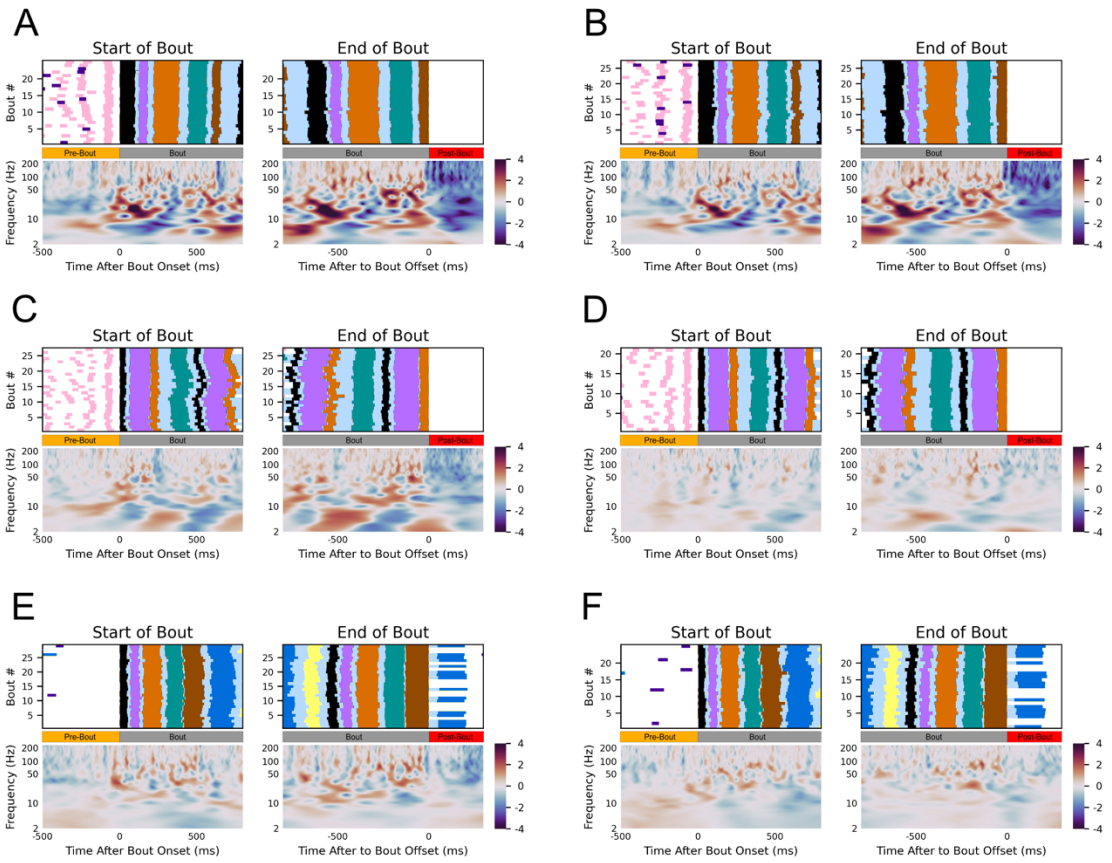


Figure 2.6: Evaluation of consistency across trials of song correlated modulation of LFP power across subjects and days.

Cross trial z-scored ratings of average spectrograms (see section 2.4.2 **Cross trial z-scored ratings of averaged spectrograms)** aligned to the start of the first motif in the bout, left, and the last motif in the bout, right, for each recording day. The z-score metric normalizes the modulation in LFP power at each frequency and time point by its standard deviation across trials. Thus the metric quantifies the number of standard deviations, as measured across trials, between the mean LFP power at each timepoint/frequency and the mean LFP power across the VAP. Shown above all results is a behavioral raster showing the time course of the behavior being evaluated. (A) The z-scored results for the first high yield day, designated Day 1, for z007 (n=25 Bouts) (B) The z-scored results for the second high-yield day, designated Day 2, for z007 (n=27 Bouts). The other subjects' results are shown as follows; (C) z020's first high yield day (n=29 Bouts), (D) z020's second high-yield day (n=25 Bouts), (E) z017's first high-yield day (n=27 Bouts), and (F) z017's second high-yield day (n=21 Bouts). As z017 would end its bout on either syllable '5' or '6', the end of the bout was aligned to syllable '5'. No dynamic-time warping was used. To ensure the start and end of the bout are unique time periods, only bouts with more than one motif in duration were used. Behaviorally inconsistent bouts were excluded for clarity of visualization; however, results are consistent when including them in the analysis

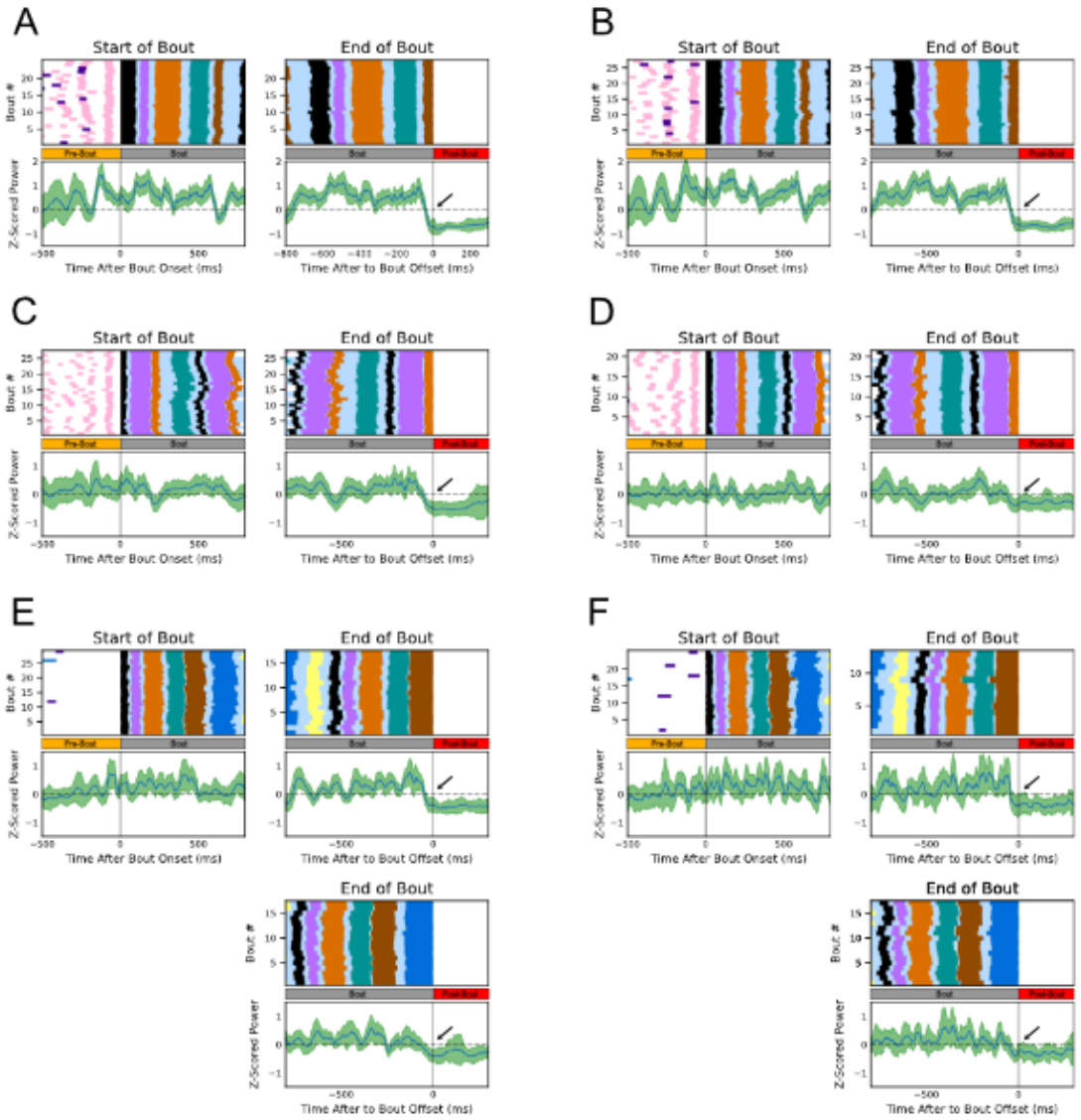


Figure 2.7: Song correlated changes in power of 50-200 Hz band across days and subjects.
Z-scored changes in power of the 50-200 Hz band aligned to the start of the first motif in the bout, left, and the last motif in the bout, right, for each high-yield recording day. Black traces in each subpanel show the mean, and the green shading is the standard deviation. The end of the bout, and the subsequent drop in power are annotated by a black arrow. Above all results is a behavioral raster showing the time course of the behavior being averaged. (A) The results of the first-high yield day of z007 (n=27 Bouts). (B) The results for the second high-yield day for z007 (n=25 Bouts). The other subjects results are shown as follows; (C) z020's first high-yield day (n=29 Bouts), (D) z020's second high-yield day (n=25 Bouts), (E) z017's first high yield day (n=27 Bouts), and (F) z017's second high-yield day (n=21 Bouts). As z017 ends its bout on either syllable '5' or '6', both types of bout endings are shown separately. No dynamic-time warping is used. To ensure that the start and end of the bout are unique time periods, only bouts with more than one motif in duration are used. Behaviorally inconsistent bouts are excluded for clarity of visualization; however, results are consistent when including them in the analysis.

The amplitude of these high frequencies were found to significantly decrease during the 100 millisecond window immediately after the end of the bout for most if not all channels for all days; with the second high-yield day of z020 being the only exception (Table 2.8). Within subject, there was strong stereotyped structure for frequency bands below 50 Hz arising from co-occurrences in changes from both phase and amplitude. Figure 2.8 and Figure 2.9 Fig illustrate this stereotypy of time correlated spectral structure in HVC as it relates to the preparation and production of bird song. No dynamic time warping was used to align the behavior.

As in previous work [37,39], we found a strongly stereotyped 25–35 Hz oscillation that was consistent across renditions of motifs (Figure 2.8 and Figure 2.9). Consistent with common usage in neuroscience literature [57,58], by oscillation we refer to energy in the specified frequency range. In addition to this previously reported frequency band, structured activity was observed in several frequency ranges that are commonly studied in mammalian literature, namely 4–8 Hz, 8–12 Hz, and 35–50 Hz. This structure, observed to be stable across trials, arises from fine time-alignment of both phase and amplitude, and occurred over times-scales longer than a single cycle for a given frequency below approximately 50 Hz (Figure 2.8 and Figure 2.9). These oscillations, with consistency in both phase and amplitude, were observed to start just prior to song onset and to end immediately after vocal production stopped (Figure 2.9). Similar patterns emerged in each of the birds and within birds across days (Figure 2.8 and Figure 2.9) [39]. This finding led us to ask: What frequencies in the LFP might carry information about the preparation and production of song through either phase or amplitude changes?

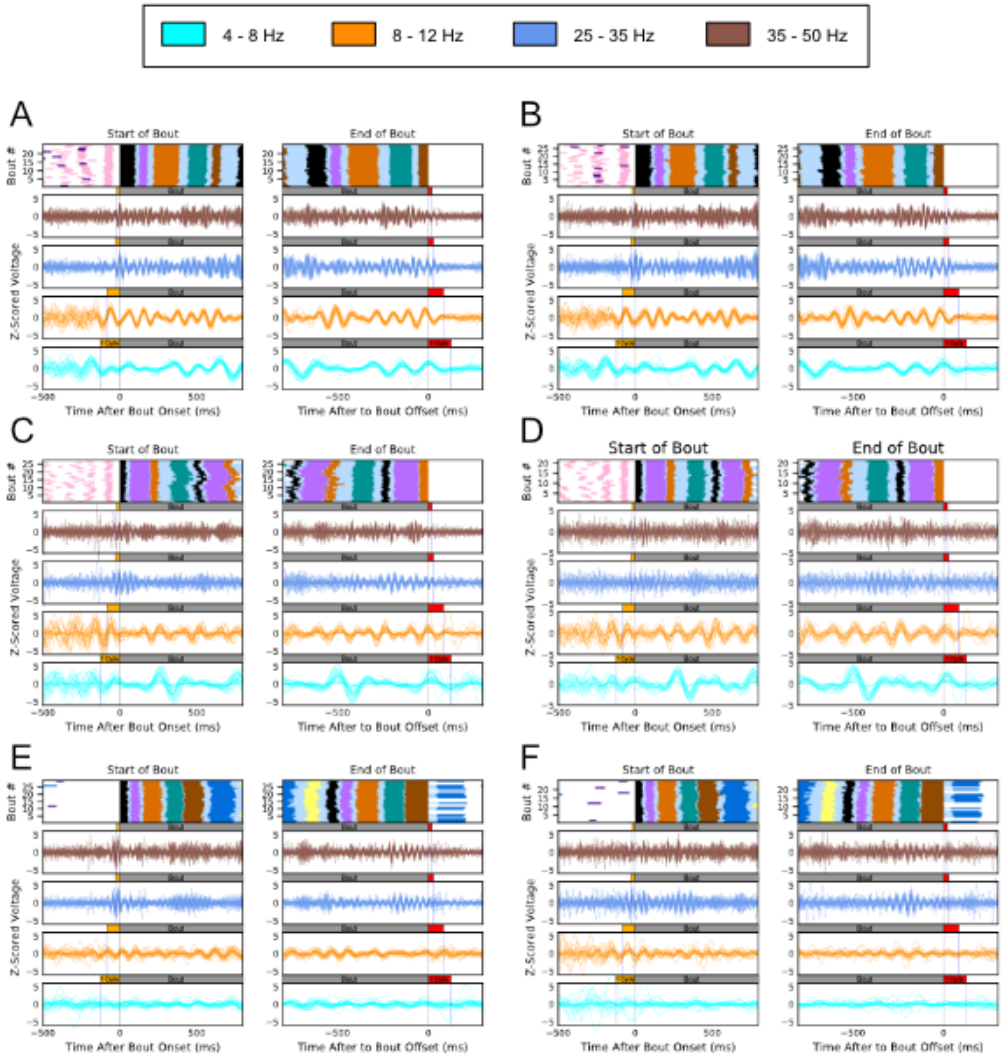


Figure 2.8: Stereotyped song correlated rhythmic changes in LFP across days and subjects (single trial).

Single-trial z-scored LFP traces of four narrowband frequency bands aligned to the start of the first motif in the bout, left, and the last motif in the bout, right, for each high-yield recording day. Each trace is colored its respective narrowband frequency. Shown above all results are behavioral rasters showing the time course of the behaviors being shown. The black line in each row below the behavior shows the time point the trials are aligned to, and the blue line denotes the time prior to (left) or after (right) one full cycle of the highest frequency in the narrowband frequency. (A) The results of the first high-yield Day of z007 (n=27 Bouts) (B) The results for the second high-yield day for z007 (n=25 Bouts). The other subjects' results are shown as follows; (C) z020's first high-yield day (n=29 Bouts), (D) z020's second high-yield day (n=25 Bouts), (E) z017's first high-yield day (n=27 Bouts), and (F) z017's second high-yield day (n=21 Bouts). As z017 would end its bout on either syllable '5' or '6', the end of the bout was aligned to syllable '5'. No dynamic-time warping was used. To ensure that the start and end of the bout are unique time periods, only bouts with more than one motif in duration were used. Behaviorally inconsistent bouts were excluded for clarity of visualization; however, results are consistent when including them in the analysis.

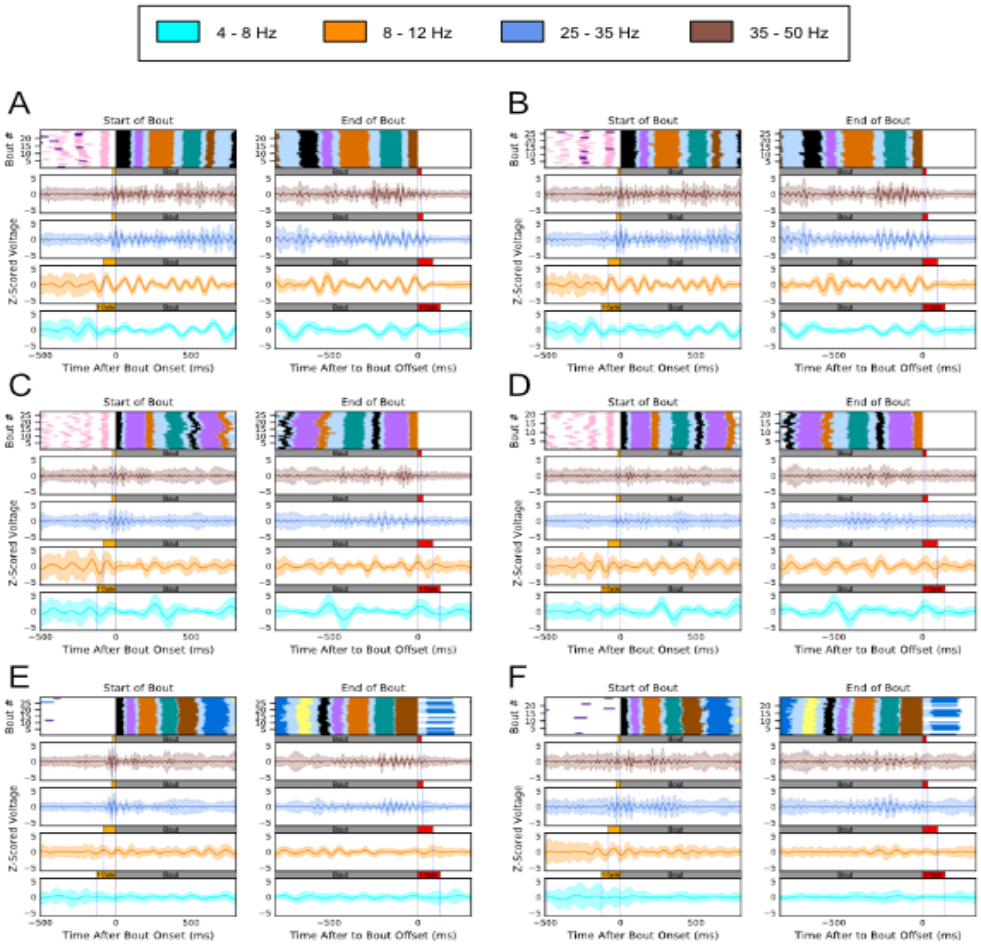


Figure 2.9: Stereotyped song correlated rhythmic changes in LFP across days and subjects (mean and standard deviation).

The mean and standard deviation of the z-scored LFP traces of four narrowband frequency bands aligned to the start of the first motif in the bout, left, and the last motif in the bout, right, for each high-yield recording day. Each row is colored its respective narrowband frequency. Above all results is a behavioral raster showing the time course of the behaviors being shown. The black line in each row below the behavior shows the time point that the trials are aligned to, and the blue line denotes the time prior to (left) or after (right) one full cycle of the highest frequency in the narrowband frequency. (A) The results of the first high yield day of z007 (n=27 Bouts) (B) The results for the second high-yield day for z007 (n=25 Bouts). The other subjects' results are shown as follows; (C) z020's first high-yield day (n=29 Bouts), (D) z020's second high-yield day (n=25 Bouts), (E) z017's first high-yield day (n=27 Bouts), and (F) z017's second high-yield day (n=21 Bouts). As z017 would end its bout on either syllable '5' or '6', the end of the bout was aligned to syllable '5'. No dynamic-time warping was used. To ensure that the start and end of the bout are unique time periods, only bouts with more than one motif in duration were used. Behaviorally inconsistent bouts were excluded for clarity of visualization, however results are consistent when including them in the analysis.

2.5 Unsupervised Decomposition of HVC LFP

2.5.1 Principal spectral decomposition

Following Human ECoG studies [65], a principal component method was applied to find consistent changes in LFP spectrum related to motor-vocal behavior. This method required calculation of the power spectral density (PSD) for windows of time during either vocal activity or inactivity. Although the behavioral changes in birdsong are on a smaller timescale (20-100 ms) than human limb movements, to keep methods consistent between studies the same window size of 1 second was used for the trial length, τ_q.

PSDs of vocally active trials centered on song activity, which included introductory notes and calls, were analyzed alongside trials of neural activity centered within larger periods of vocal inactivity longer than 2 seconds in duration. All PSDs were calculated using the multitaper method with the time_frequency.psd_array_multitaper function from the MNE-Python software package [63]. The frequency range was set from 0 to 200 Hz with 14 orthogonal slepian tapers. Each PSD, $P(f, \tau_q)$, was individually normalized using two steps: each spectral sample was elementwise divided by the mean across the ensemble, at each frequency, and then the log was taken. This centers the data around the log of the mean spectrum.

$$\underline{\boldsymbol{P}}(\boldsymbol{f}, \boldsymbol{\tau_q}) = \boldsymbol{ln\ ln}\left(\boldsymbol{P}(\boldsymbol{f}, \boldsymbol{\tau_q})\right) - \boldsymbol{ln\ ln}\left(\frac{\mathbf{1}}{\boldsymbol{N_q}}\sum_{\boldsymbol{p=1}}^{\boldsymbol{N_q}} \boldsymbol{P}(\boldsymbol{f}, \boldsymbol{\tau_p})\right) \quad \textbf{(3)}$$

The label q refers to the times centered within periods of high vocal activity and vocal inactivity (silence). The number of instances, or trials, per class were balanced to be equal; their combined total number of PSDs is denoted by N_q. The order of the trials was explicitly ignored, meaning that they together represent a balanced ensemble of N_q independent measurements of the power spectrum during the two conditions. The covariance matrix $C(f, \tilde{f})$ between frequencies of these normalized PSDs were calculated:

$$C(f, \tilde{f}) = \sum_{\tau_q} \tilde{P}(f, \tau_q)\tilde{P}(\tilde{f}, \tau_q) \quad (4)$$

The covariance measure is centered with respect to the log of the mean spectrum. The eigenvalues, λ_k, and eigenvectors, $\vec{e}_k$, of this matrix elucidate common features during song production. These eigenvectors, $\vec{e}_k$, are referred to as "principal spectral components" (PSCs) and they have been described in prior literature to reveal which frequencies vary together [65].

2.5.2 Cosine Similarity.

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space. For the purpose of comparing the principal spectral components, cosine similarity evaluates the orientation, not magnitude, of two vectors in relation to one another. To calculate this, one takes the inner product of two unit vectors, or two vectors that have each been normalized to have a length Of 1.

$$Similarity = A \cdot B = \|A\|\|B\| \cos\theta \quad (5)$$

To calculate the cosine similarity for the PSCs we must take the dot product between two unit vectors. As the PSCs are all eigenvectors calculated using PCA they are already unit vectors and their sign can thus be flipped without altering their information. A template of each PSC is calculated by taking the mean of the PSCs across all good channels after all of their signs have been aligned. This template PSC is then normalized to have a length of one by dividing it by its norm. The cosine similarity matrix is then calculated by taking the dot product of these two unit vectors. This is expressed below with the template PSC represented as T and the PSC symbolized by P from a selected channel, c.

$$Similarity = P_c \cdot \frac{T}{\|T\|} \quad (6)$$

2.5.3 Decoupling LFP power-spectrum reveals song related spectral features

A spectral decomposition technique previously applied to human motor cortex data [65] was used to determine which frequencies' amplitude changes were correlated with the production of birdsong. As suggested in Figure 2.10 and supported by Table 2.8, during times of vocal activity for all birds we observe an increase in power aggregated across higher frequencies that are in a range often referred to as "High Gamma" [66,67]. The frequency ranges for "High Gamma" vary in the literature, but Miller et. al. described it as 80-200Hz. Following the steps of the spectral decomposition approach described in the methods of Miller et. al., the principal component decomposition of these PSDs found principal spectral components (PSC) similar to those previously reported in [65] (Figure 2.9B & C). The PSC's shown in Figure 2.10 are results from one channel, which was representative of all channels for each subject. The first PSC, or the most significant principal component, in Miller et al. is characterized for having most of its element magnitudes with the same sign and being consistently non-zero, with values approaching a constant across frequencies. This was subsequently described as reflecting a broad spectrum increase in power most clearly visible in High Gamma.

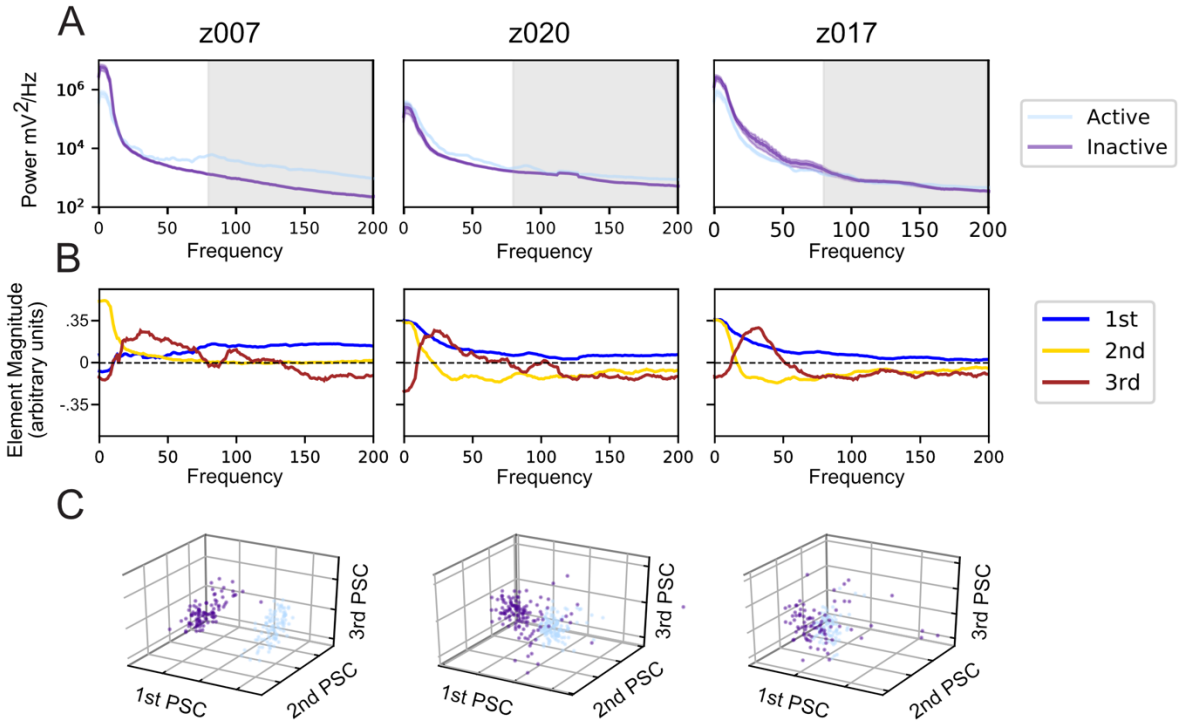


Figure 2.10: Naïve decomposition of LFP power spectra reveals song correlated features. Representative results from a single electrode channel for each of the three subjects during its highest yield day. (A) Averaged power spectra during trials centered in 1-second intervals where the bird either initiates a motif, called vocally active (light blue), or does not vocalize, called vocally inactive (purple). The 80–200 Hz, or 'High Gamma' band, is shaded in grey. (B) The power spectra in Figure 2.4A are normalized and naively decomposed into PSCs (see section 2.5.1 **Principal spectral decomposition**). The elements of the first principal spectral component (1st PSC, blue) is non-zero across all frequencies, likely due to the power law in LFP PSDs. The 2nd PSC, golden-yellow, peaks between 0 and 10 Hz. The 3rd PSC, burgundy, peaks between 10 and 30 Hz, but has variations across birds that extend into 50Hz. As the PSCs are all eigenvectors, their signs do not matter when interpreting them. This structure is largely consistent across channels and across days. Note that the other PSCs are not shown. (C) Projection of both the vocally active and vocally inactive trials onto the first three PSCs for the same channels in (A) and (B). The color coding is the same as (A).

Table 2.8: Characteristic Decrease in High Gamma Power After Bout Termination Across Channels.

The percentage of channels whose distribution of power values for 80-200 Hz, or High Gamma, were smaller during the 100ms silent period after the bout ends than the 100ms period prior to the end of the bout. To determine statistical significance between the two distributions, the one-sided Welch's t-test was used for each channel. The percentages shown are the number of good channels that still were significant after using the Benjamini-Hochberg False Discovery Rate ($p<0.05$ and $q<0.05$).

		Percentage of Channels with Decreased Power Immediately After the Bout (After FDR Correction)	
Bird_ID	Date	Primary Bout Ending Syllable	Alternative Bout Ending Syllable (z017 Syllable 6)
z020	day-2016-06-03	100% (n= 15/15 Channels)	—
z020	day-2016-06-05	19.99% (n= 3/15 Channels)	—
z007	day-2016-09-10	100% (n= 30/30 Channels)	—
z007	day-2016-09-11	100% (n= 30/30 Channels)	—
z017	day-2016-06-19	100% (n= 16/16 Channels)	93.75% (n= 15/16 Channels)
z017	day-2016-06-21	100% (n= 16/16 Channels)	93.75% (n= 15/16 Channels)

In our analysis the most consistent PSC, which is the first principal component, was a spectrum wide-amplitude change that was consistent across channels and subjects (Figure 2.11A Fig). The second PSC in Miller et al. peaked in the "alpha/low beta range"—alpha is described as 8-12 Hz and beta is described as 12-30 Hz—which mirrors PSC 3 found across subjects and channels in zebra finches (Figure 2.11C Fig) [65]. These frequency bands in humans have been proposed to reflect the resting rhythms that decrease when cortical areas are activated [68].

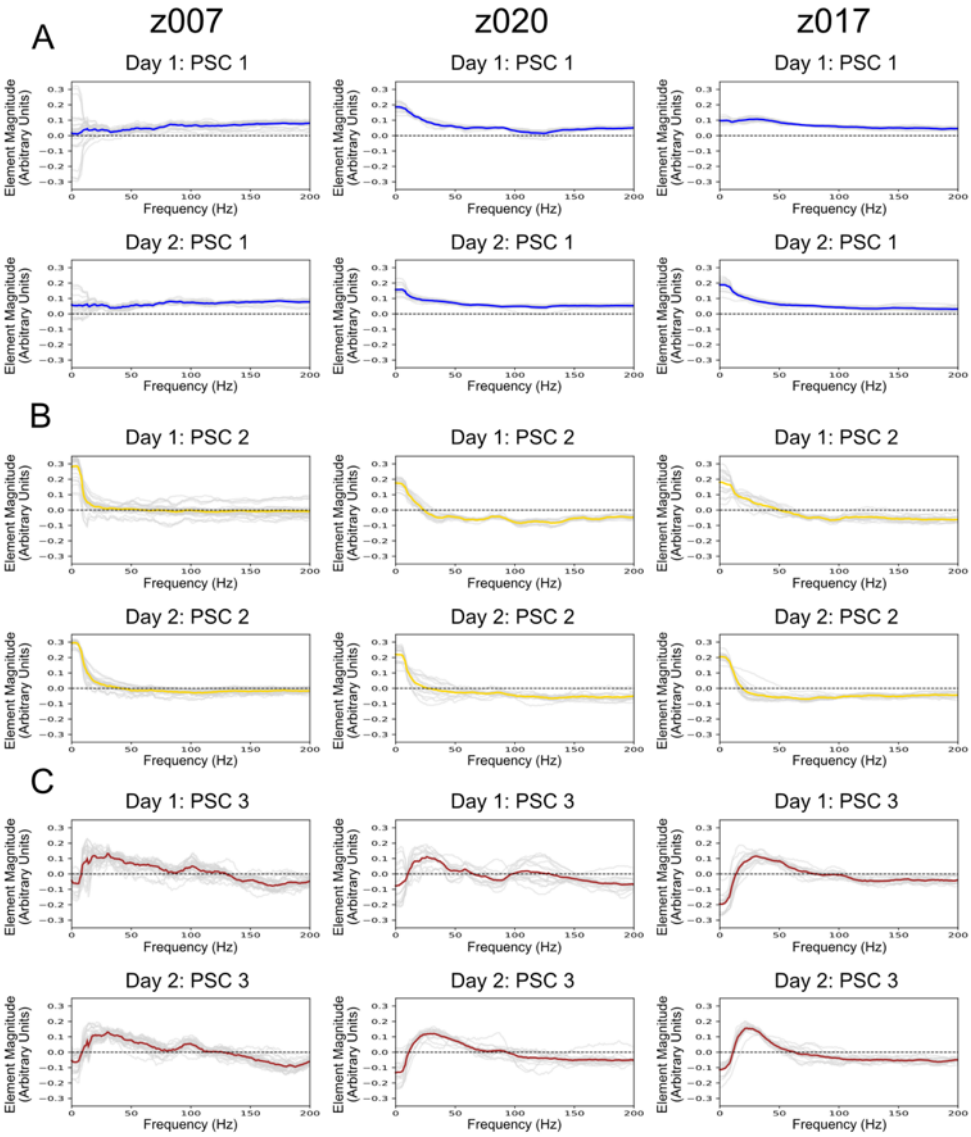


Figure 2.11: Summary of principle spectral component analysis.
(A) The 1st Principle spectral components for every channel (light grey) with the mean across channels in Blue for all High Yield Days for each subject. (B) The 2nd Principle spectral components for every channel (light grey) with the mean across channels (golden-yellow) for all High Yield Days for each subject. (C) The 3rd Principle spectral components for every channel (light grey) with the mean across channels (burgundy) for all High Yield Days for each subject.

The characteristic structure of each PSC was found to be consistent across subjects, channels, and days using the cosine similarity metric (see section 2.5.2) (Figure 2.12 Fig). Cosine similarity is a measure of the degree of similarity of two non-zero vectors of an inner product space. It evaluates their orientation—and not their magnitude—with two vectors that have the exact same orientation having a value of 1 and two vectors that are orthogonal to one another having a value of 0. In addition, with this metric we also found that the similarity of each PSC to another across subjects was greater than was the case with a different PSC (Figure 2.12E-G Fig). Although the PSCs were calculated without explicit knowledge of distinct vocally active and inactive time periods, the first three PSCs show that the power spectrum during vocal production is separable from vocally inactive periods (Figure 2.10C). This separation between behavioral states was evaluated using the sensitivity index, or d', between the two behavior states; each channel was analyzed independently (Figure 2.13A). The sensitivity index for each PSC was compared to the bootstrapped distribution of sensitivity indexes (n = 20,000) when the state labels were shuffled (Figure 2.13); the Benjamini-Hochberg False Discovery Rate was used to account for multiple comparisons across all of the channels (Table 2.9). This separation without explicit knowledge of the state labels suggests that these frequencies could be used to detect the onset of vocal production.

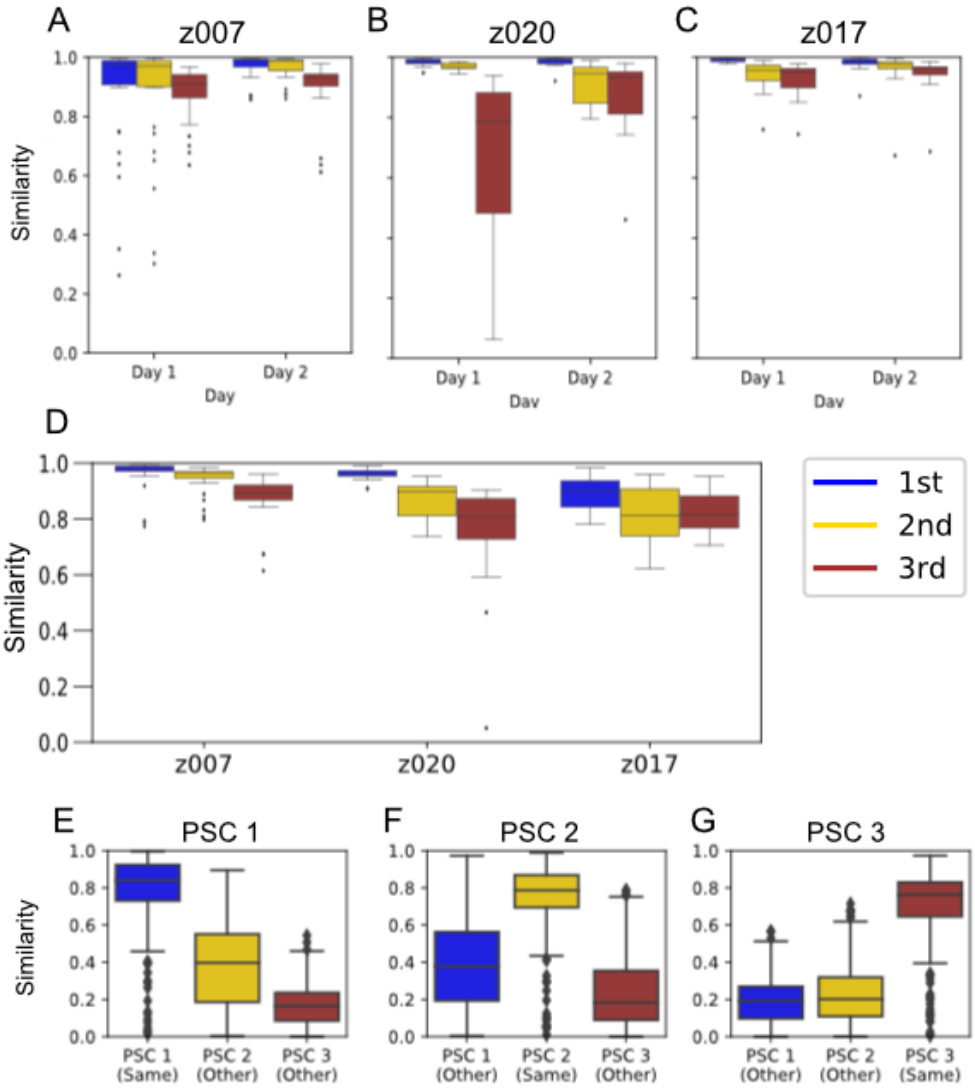


Figure 2.12: Similarity of principal spectral components across channels, days and subjects.

Boxplot of the distribution of cosine similarity metric values between a template, which is created by taking the mean across the sign-aligned PSC of all channels for a specific PSC. The cosine similarity matrix ranges from 1 and -1, however, the absolute value of the metric is shown (see section 2.5.2). (A) The cosine similarity of each channel's PSC to the same recording day's template PSC for both high-yield days for subject z007. (B) The cosine similarity of each channel's PSC to the same recording day's template PSC for both high-yield days for subject z020. (C) The cosine similarity of each channel's PSC to the same recording day's template PSC for both high-yield days for subject z017. (D) The cosine similarity of each channel's PSC from the second high-yield day with the template of the PSC from the first high-yield day. (E) All templates for PSC 1 compared either with the PSC 1 for the other two birds, same, or the PSC for one of the other PSCs for the other two birds, other. (F) All templates for PSC 2 compared either with the PSC 2 for the other two birds, same, or the PSC for one of the other PSCs for the other two birds, other. (E) All templates for PSC 3 compared either with the PSC3 for the other two birds, same, or the PSC for one of the other PSCs for the other two birds, other.

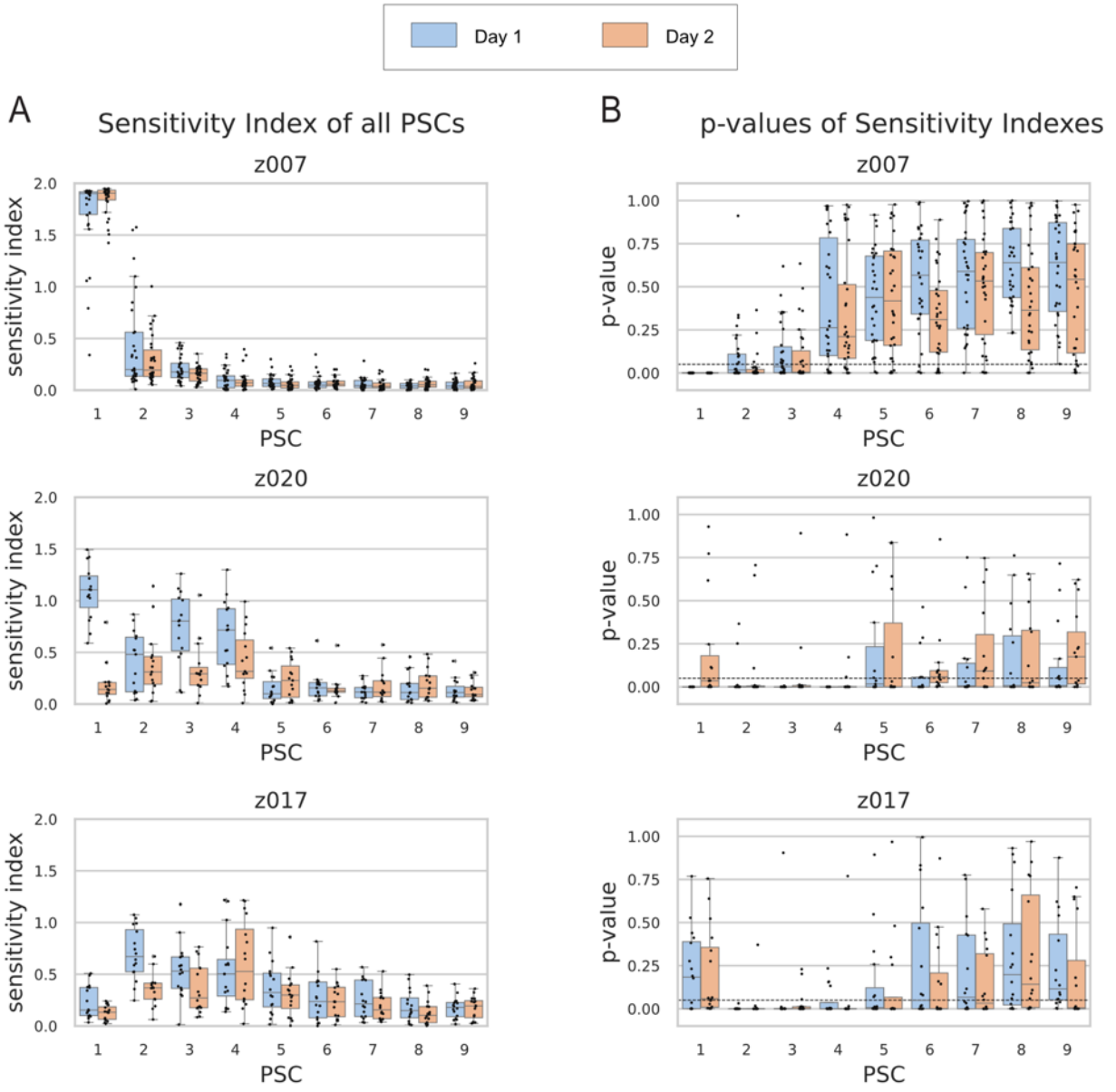


Figure 2.13: Separability of Active vs. Inactive trials using principal spectral components. (A) Box plots of the sensitivity indexes (d') for the separation of the Active and Inactive trials for all channels using the first 9 PSCs. The plots show the distribution of values for both high yield days for all three subjects. (B) Blox plots of the p-values of the sensitivity indexes shown in (A) when tested against a bootstrapped shuffle control (N=20,000 Shuffles).

Table 2.9: Sensitivity of Channels as Assessed by Permutation Test

	Number of Channels with Sensitivity Index Significantly Greater than the Permutation Test Based Control After FDR Correction (N=20,000 Shuffles)					
	z007		z020		z017	
	day-2016-09-10	day-2016-09-11	day-2016-06-03	day-2016-06-05	day-2016-06-19	day-2016-06-21
PSC 1	30 Channels (100%)	30 Channels (100%)	15 Channels (100%)	7 Channels (46.67%)	6 Channels (37.5%)	5 Channels (31.25%)
PSC 2	17 Channels (56.67%)	26 Channels (86.67%)	13 Channels (86.67%)	12 Channels (80%)	16 Channels (100%)	15 Channels (93.75%)
PSC 3	12 Channels (40%)	19 Channels (63.33%)	15 Channels (100%)	13 Channels (86.67%)	15 Channels (93.75%)	13 Channels (81.25%)
PSC 4	4 Channels (13.33%)	3 Channels (9.99%)	15 Channels (100%)	12 Channels (80%)	12 Channels (75%)	15 Channels (93.75%)
PSC 5	1 Channel (3.33%)	2 Channels (6.67%)	8 Channels (53.33%)	10 Channels (66.67%)	10 Channels (62.5%)	12 Channels (75%)
PSC 6	1 Channel (3.33%)	2 Channels (6.66%)	10 Channels (66.67%)	3 Channels (19.99%)	8 Channels (50%)	10 Channels (62.5%)
PSC 7	1 Channel (3.33%)	2 Channels (6.66%)	10 Channels (66.67%)	6 Channels (40%)	6 Channels (37.5%)	8 Channels (50%)
PSC 8	0 Channels (0%)	2 Channels (6.66%)	9 Channels (60%)	8 Channels (53.33%)	4 Channels (25%)	5 Channels (31.25%)
PSC 9	0 Channels (0%)	2 Channels (6.66%)	8 Channels (53.33%)	4 Channels (26.67%)	2 Channels (12.4%)	10 Channels (62.5%)

2.6 Phase locking of LFP during song production

2.6.1 Computation and statistical testing of phase locking across renditions

We computed the inter-trial phase coherence (ITPC) to assess trial-to-trial synchronization of LFP activity with respect to time-locked syllables within the motif. The ITPC is a measure of the phase synchronization of neural activity in a given frequency, calculated relative to critical event times that are repeated over an experimental paradigm. To calculate the ITPC across trials, for each channel, data were first filtered in 100 partially overlapping narrow frequency bands that were equal width in log-space. These filtered time series were then Hilbert-transformed to extract

the analytic phase at each time point. Then, at each time point, for each channel and frequency band, we calculated the phase-consistency across trials to estimate the ITPC. Here, 'trials' were defined as the start or end of a specific type of vocal event, which could be further refined by its context within the bout (e.g., first or last in bout). Once all instances of the event of interest were selected, a window of time centered on each was implemented. The ITPC for each frequency and time sample pair must be calculated independently. The equation for calculating the ITPC is described as below, for n trials, if, $F_k(f, t)$ is the spectral estimate of trial k at frequency f and time sample t

$$\boldsymbol{ITPC}\ (\boldsymbol{f}, \boldsymbol{t}) = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{k}=\mathbf{1}}^{\boldsymbol{n}} \frac{\boldsymbol{F_k(f,t)}}{|\boldsymbol{F_k(f,t)}|} \qquad \textbf{(7)}$$

where $|x|$ represents the complex norm of x [69]. The spectral estimate, $F_k(f, t)$, is the instantaneous phase of frequency f at time sample t. This was repeated for all narrow-band frequencies and time samples within the event window.

The ITPC value, or resultant vector length, scales from 0 to 1, with 0 being the null hypothesis where phases are uniformly distributed about the polar axis and 1 being perfect synchrony. To determine the significance of this vector length, and to determine if it could have been randomly sampled from a uniform distribution by chance, p-values were calculated for each frequency and time sample. The mean resultant vector of the instantaneous phase across trials for a specific time sample, and its corresponding bootstrapped p-value, were calculated using the pycircstats toolbox. To enable visualization of the results for comparing ITPC values while accounting for all of the frequency and time sample combinations, the Rayleigh Z statistic was calculated.

2.6.1.1 Rayleigh statistic.

The ITPC over the time length of a syllable or motif is evaluated in terms of the Rayleigh Z statistic and is defined as $Z = \frac{R^2}{n}$where R is the Rayleigh's R, $R = nr$ where r is the resultant vector length, and n is the number of trials. P-values were estimated using the following equation [70–73]:

$$\boldsymbol{P = exp\left[\sqrt{1 + 4n + 4(n^2 - R^2) - (1 + 2n)}\right]} \quad \textbf{(8)}$$

2.6.1.1 Normalized sustained Rayleigh Z statistic.

To determine which, if any, frequency bands could potentially be used as a feature to decode vocal behavior, a measure of how consistently significant a band stayed phase-locked was created. As we don't expect the Rayleigh Z statistic across both channels and frequencies to be exactly the same, we wanted to understand the relative significance across birds and channels. The measure ranks how consistently phase-locked across time a frequency was during song production overall for one recording session. The mean Rayleigh Z-statistic of samples over the course of one oscillation for a set frequency normalized by the maximum Z-statistic value for the entire time period analyzed was used and will be referred to as the normalized sustained Z statistic. This analysis informed the frequency bands used for the classification and onset detection analysis.

2.6.2 Syllable onset is phase-aligned to underlying LFP rhythms

We calculated the inter-trial phase coherence (ITPC) of the spectrum (see section 2.6.1 Computation and statistical testing of phase locking across renditions) from 2–200 Hz to determine which frequencies had stable song-aligned structured changes in phase. The ITPC is a measure of the phase synchronization of neural activity in a given frequency at a specific time sample. A

defining characteristic of zebra finch song is its stereotyped motif. Leveraging this stereotypy, we calculated the ITPC aligned to the first motif of all bouts recorded in a single day (n>28). Any LFP phase coherence occurring prior to song production can be attributed to the preparation of the subsequent vocalization. In two of the three birds, the first motif was almost always preceded by an introductory note. As each frequency and time sample are separate results with an attributed p-value, they are all visualized in terms of their Rayleigh Z statistic (see section 2.6.1.1 Rayleigh statistic.) to allow for comparisons. By scaling the p-values of each independent ITPC result, the Rayleigh Z statistic provides a measure by which multiple comparisons in circular statistics can be accounted. Based upon the number of trials and conditions tested, a threshold value can be applied to the Rayleigh Z statistic to determine statistical significance. Figure 2.11A shows long-lasting phase structure that precedes the onset of the bout by up to 100 ms in frequencies lower than 50 Hz (note that in this figure and in all subsequent ITPC Rayleigh Z statistic images we have set all non-significant times and frequencies to black based upon the appropriate threshold value for the condition evaluated). This consistency in phase continues during the course of the bout and terminates soon after the end of the bout (Figure 2.11B). Although the precise frequency ranges vary across subjects, the subject specific pattern present prior to bout onset and sustained throughout the bout was consistently observed across channels, subjects, and days [Figure 2.15].

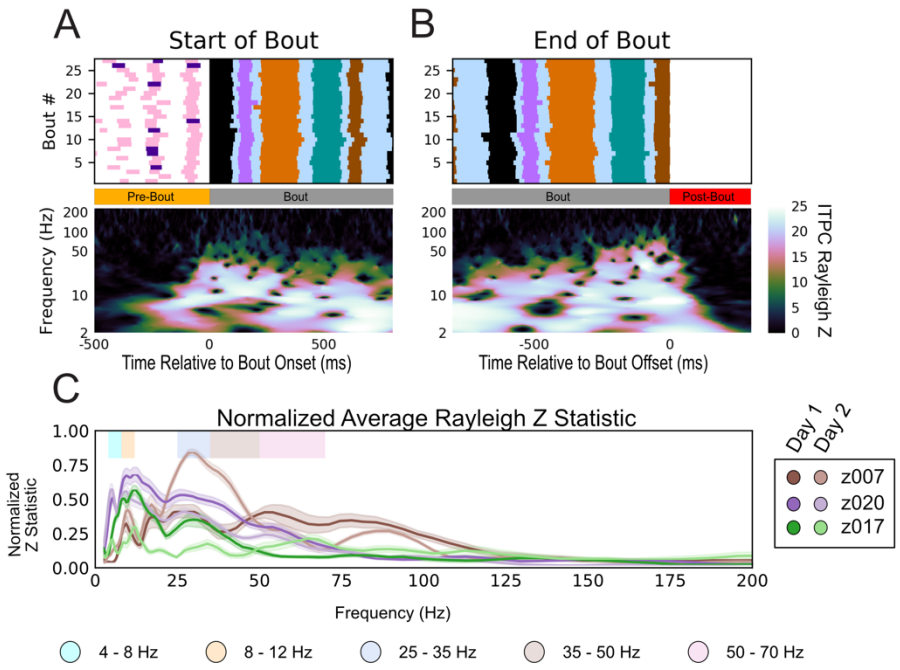


Figure 2.14: LFP Inter-trial phase coherence during production of learned sequences. (A) ITPC of LFP aligned to the start of the first motif of the bout for bird z007 (bottom) across multiple renditions of similar bouts (n=27, top) aligned to the start of the first motif in the bout. (B) the same as (a) aligned to the end of the last motif of the bout. No dynamic-time warping was used. To ensure that the start and end of the bout are unique time periods, only bouts with more than one motif in duration were used. Behaviorally inconsistent bouts were excluded for clarity of visualization; however, including these bouts does not alter the result. ($p<0.006$ for all Rayleigh Z > 5; all non-black time-frequency points in this plot are above the significance threshold). (C) Normalized sustained Z-statistic of the ITPC for samples preceding the labeled start of the bout. The steps for calculating this metric are detailed in the section 2.6.1.1 **Normalized sustained Rayleigh Z statistic.**. Data reflect all bouts from both high-yield days for three birds.

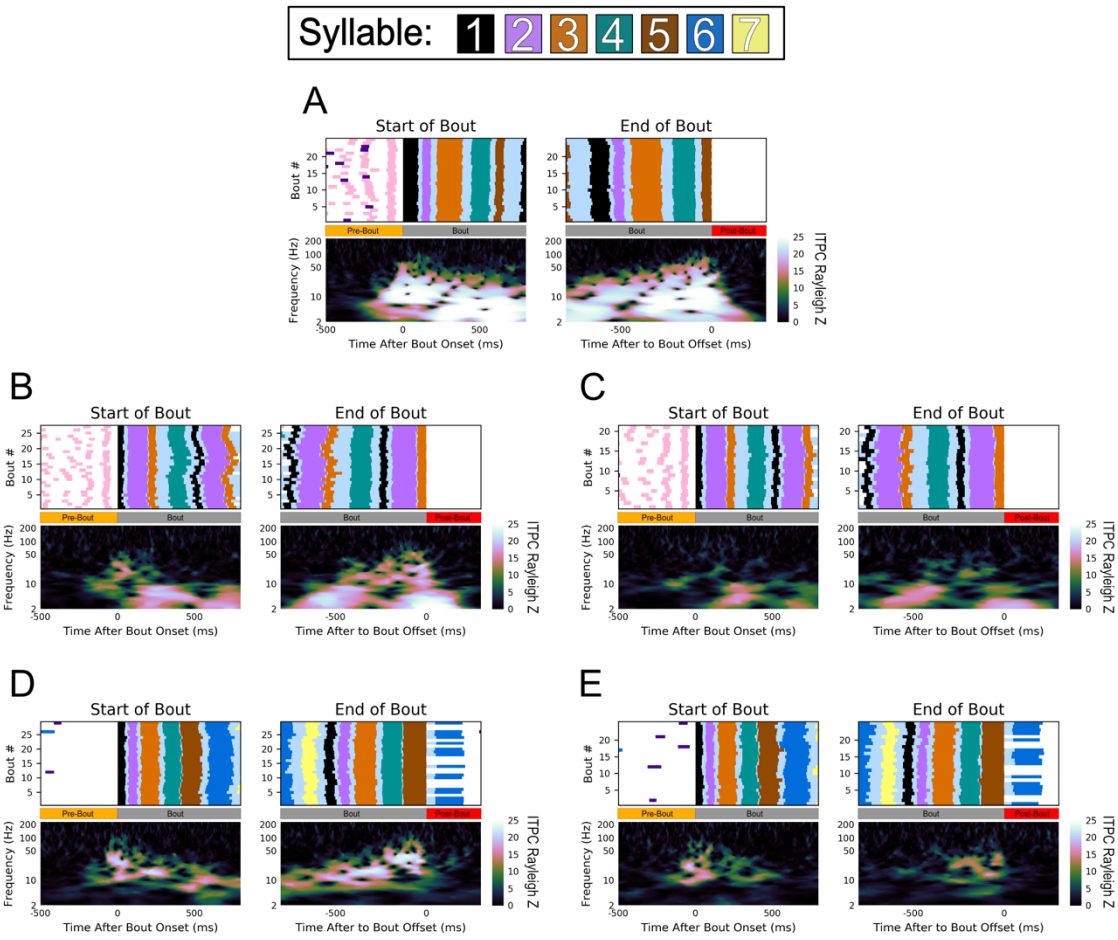


Figure 2.15: Inter-trial phase coherence of LFP phase during production of learned sequences across subjects and days.

ITPC of LFP aligned to the start of the first motif in the bout, left, and the last motif in the bout, right, for each recording day not shown in Fig 4. Shown above all results is a behavioral raster showing the time course of the behavior being averaged. (A) The averaged results for the second highest-yielding day, designated Day 1, for z007 (n=25 bouts). The other subjects' results are show as follows; (B) z020's first high-yield day (n=29 bouts), (C) z020's second high-yield day (n=25 bouts), (D) z017's first high-yield day (n=27 bouts), and (E) z017's second high-yield day (n=21 bouts). As z017 would end its bout on either syllable '5' or '6', the end of the bout was aligned to syllable '5'. No dynamic-time warping was used. To ensure that the start and end of the bout are unique time periods, only bouts with more than one motif in duration were used. Behaviorally inconsistent bouts were excluded for clarity of visualization; however, results are consistent when including them in calculating the ITPC. ($p<0.006$ for all $Z > 5$ for all subjects and days; all non-black time-frequency pints in this plot are above the significance threshold).

To determine the frequencies where phase coherence is most consistent across individual subjects, we computed the normalized sustained Rayleigh Z Statistic (Section 2.6.1.1 Normalized sustained Rayleigh Z statistic.) for the two high-yield days for all subjects (Figure 2.14C). The frequencies that contained sustained high ITPC values across animals include the 25–35 Hz range, as previously established [37,39], and several other frequencies: 4–8 Hz, 8–12 Hz, 35–50 Hz, and 50–70 Hz. These oscillations fall within well-documented frequency ranges in mammalian literature, namely theta (approximately 4–8 Hz), alpha (approximately 8–12 Hz), and low gamma (approximately 30–70 Hz) [74]. The lower frequencies exhibited longer periods of phase coherence through time than did the higher frequencies. These periods occurred over several cycles, and they fell out of alignment faster than a full cycle once the song terminated (Figure 2.14B and Figure 2.15). Strong phase coherence was observed throughout the production of the motif without the need for dynamic time warping to force motifs into alignment on a warped time scale (Figure 2.14 and Fig. Figure 2.15). However, when viewing the ITPC in Figure 2.14 across the longer time scale of the bout, it is important to note that there is considerable variation in the timing of individual renditions of the bout. If the dynamics of LFP phase are aligned to the time scale of underlying behavior, then the variation in motif timing across bouts will result in a reduction in ITPC at time points further away from the point of alignment. In particular, such a reduction will be more pronounced for higher frequencies which have shorter cycle times.

To better understand the dynamics underlying HVC population activity within the bout, we examined the song-aligned LFP phase coherence in greater detail. One hypothesis, given the highly stereotyped spectro-temporal structure of the motif including the brief gaps between each syllable, is that both the vocal output and the underlying HVC dynamics are largely deterministic

(i.e., the observed coherence may reflect an initial, single alignment at the start of the motif, which endures only by virtue of the very low song-to-song variability). Alternatively, it could be possible that HVC population activity reflects a more tightly controlled timescale, in which the observed oscillations are aligned to shorter timescale vocal events such as syllables. Previous work has shown that local spiking activity in HVC is aligned to syllables not motif onset [55,56], however as LFP largely reflects postsynaptic and transmembrane currents coming into the region of the recording the site it is unknown whether LFP shares this temporal alignment.

To determine whether alignment is tied to the motifs (hypothesis one), or is unique to each syllable (hypothesis two), a smaller scale time alignment was used. We first verified that phase-locking was found when localized to shorter sub-components of vocalization. Syllable-specific polar phase analysis (Figure 2.16A) shows a strong phase preference to the onset of each syllable in all of the previously determined frequency bands of interest.

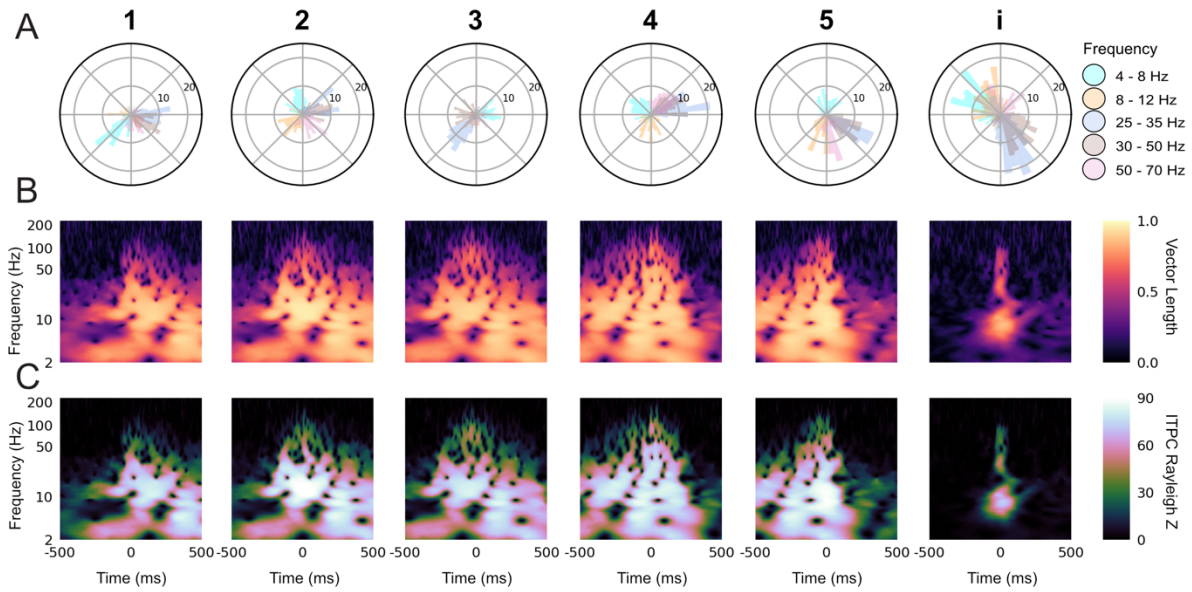


Figure 2.16 Syllable specific inter-trial phase coherence reveals phase preference to syllable onset.
(A) Polar histogram of the phase for each LFP frequency band at the labeled start of all instances of a given syllable or the introductory note over the course of one day (Day 2), for bird z007. The number of instances (n=100) are equal for all syllables and the introductory note, and is set by the syllable class with the fewest renditions. (B) ITPC resultant vector length, indicating the level of phase consistency across trials, for each frequency over time relative to the labeled start of each syllable or introductory note (0 ms) over the same instances as in (A). (C) Rayleigh Z-statistic of the ITPC over the same time and frequencies as (B) ($p<0.007$ for all $Z > 5$ for all syllables; all non-black time-frequency points in these plots are above the significance threshold).

These phases were unique to each frequency band for each syllable. Similar results were seen for all subjects across all days (Figure 2.18–Figure 2.21). We also found the same phase preference to vocalization onset for the introductory note despite its huge variability in duration and structure beyond a single utterance. Figure 2.16C demonstrates that this stereotyped phase structure occurs prior to and during each vocalization and is statistically unlikely to have occurred by chance. To directly test the two hypotheses, we compared, relative to the onset time of the first syllable, the ITPC centered on onset time of each subsequent syllable to the ITPC centered on the stereotyped time of the same syllable (Figure 2.17). Centering on the actual labeled syllable start time yielded significantly stronger ITPC compared to centering on the stereotyped, or across-bout average, start

time for each syllable (Figure 2.18–Figure 2.21). Thus, the result of this analysis supports hypothesis two as the more likely description of this phase locking characteristic.

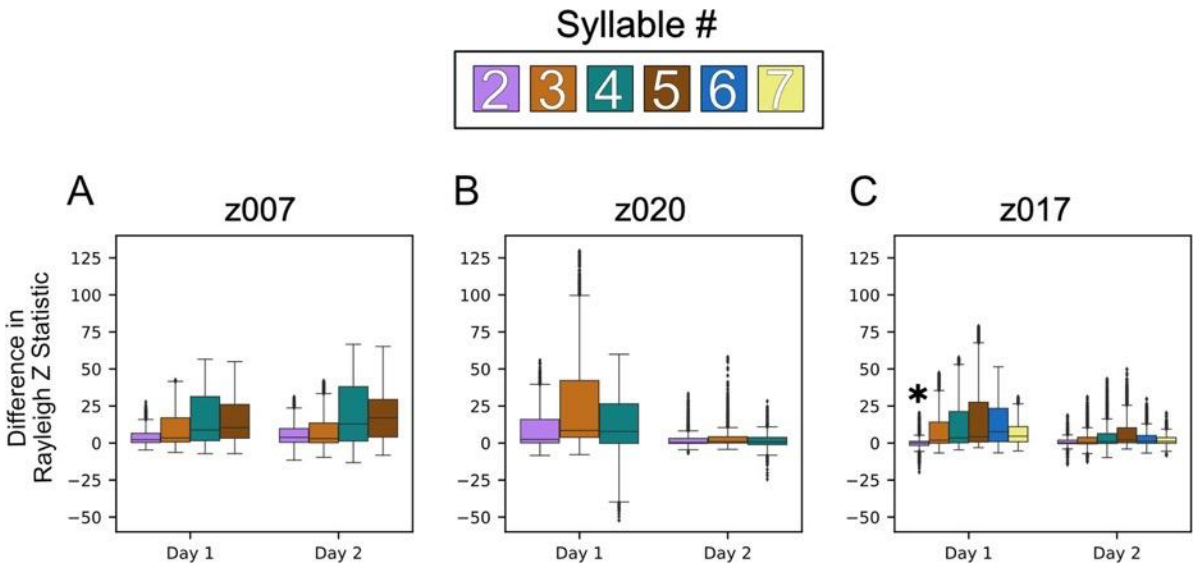


Figure 2.17: Phase preference to syllable onset is reflective of constitutive syllables rather than of the larger motif structure.

Box plots of the differences in Rayleigh Statistic between the syllable aligned ITPC and the ITPC aligned to that same syllable's stereotyped onset time within the motif. Positive differences indicate greater phase consistency at the syllable aligned time versus that of the stereotyped time across LFP frequencies (4–8 Hz, 8–12 Hz, 25–35 Hz, 35–50 Hz, and 50–70 Hz). This difference was determined by first concatenating the Z statistic for each channel centered at the designated time point to get a vector of all partially overlapping frequencies for all channels. The vector of stereotyped alignment was then subtracted from the labeled onset alignment to achieve the difference for each frequency on every channel. This was repeated for all syllables, excluding the first, with each syllable represented by a specific color, as indicated. (A) Shows the results for the two high-yield days for z007, (B) shows the results for z020, and (C) shows the results for z017. All instances of each syllable that were preceded by the first syllable were used. To determine statistical significance, we used the one-sided Wilcoxon signed-rank test, with each frequency and channel pair. * denotes that the comparison was not statistically significant when using the using the Benjamini-Hochberg False Discovery Rate. All other results $p<0.05$ and $q<0.05$.

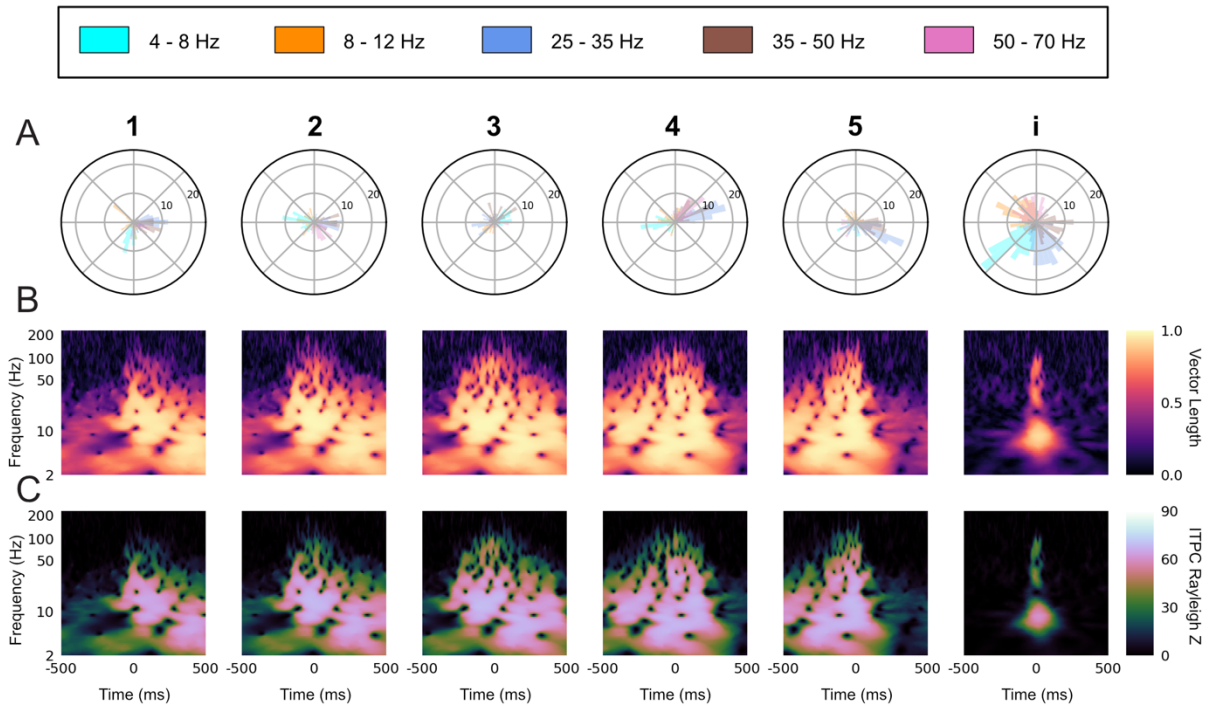


Figure 2.18: Unique LFP phase preferences at each syllable onset (z007, day 1).
(A) Polar histogram of the phase for each LFP frequency band at the labeled start of all instances of a given syllable or the introductory note over the course of one day (Day 1), for one bird (z007). The number of instances (n=71) are equal for all syllables and the introductory note, and are set by the syllable class with the fewest renditions. (B) ITPC resultant vector length for each frequency over time relative to the labeled start of each syllable or introductory note (0 ms) over the same instances as in (A). (C) Rayleigh Z-statistic of the ITPC over the same time and frequencies as (B). ($p<0.007$ for all $Z > 5$ for all syllables and the introductory note; all non-black time-frequency pints in this plot are above the significance threshold).

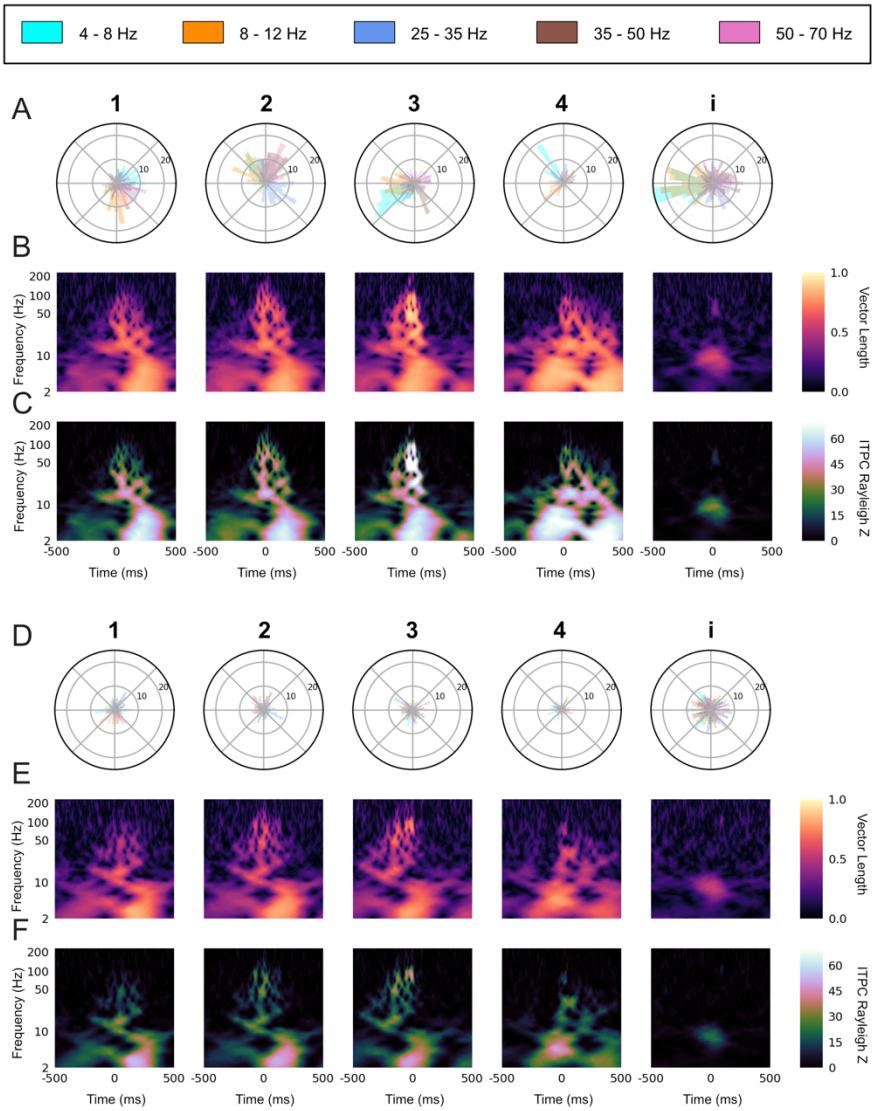


Figure 2.19: Unique LFP phase preferences at each syllable onset (z020).
(A) Polar histogram of the phase for each LFP frequency band at the labeled start of all instances of a given syllable or the introductory note over the course of one day (Day 1), for one bird (z020). The number of instances (n=91) are equal for all syllables and the introductory note, and are set by the syllable class with the fewest renditions. (B) ITPC resultant vector length for each frequency over time relative to the labeled start of each syllable or introductory note (0 ms) over the same instances as in (A). (C) Rayleigh Z-statistic of the ITPC over the same time and frequencies as (B). (D) Polar histogram of the phase for each LFP frequency band at the labeled start of all instances of a given syllable over the course of one day (Day 2), for one bird (z020). The number of instances (n=75) are equal for all syllables and the introductory note, and are set by the syllable class with the fewest renditions. (E) ITPC resultant vector length for each frequency over time relative to the labeled start of each syllable (0 ms) over the same instances as in (D). (F) Rayleigh Z-statistic of the ITPC over the same time and frequencies as (E). ($p<0.007$ for all $Z > 5$ for all syllables and the introductory note for both days; all non-black time-frequency pints in this plot are above the significance threshold).

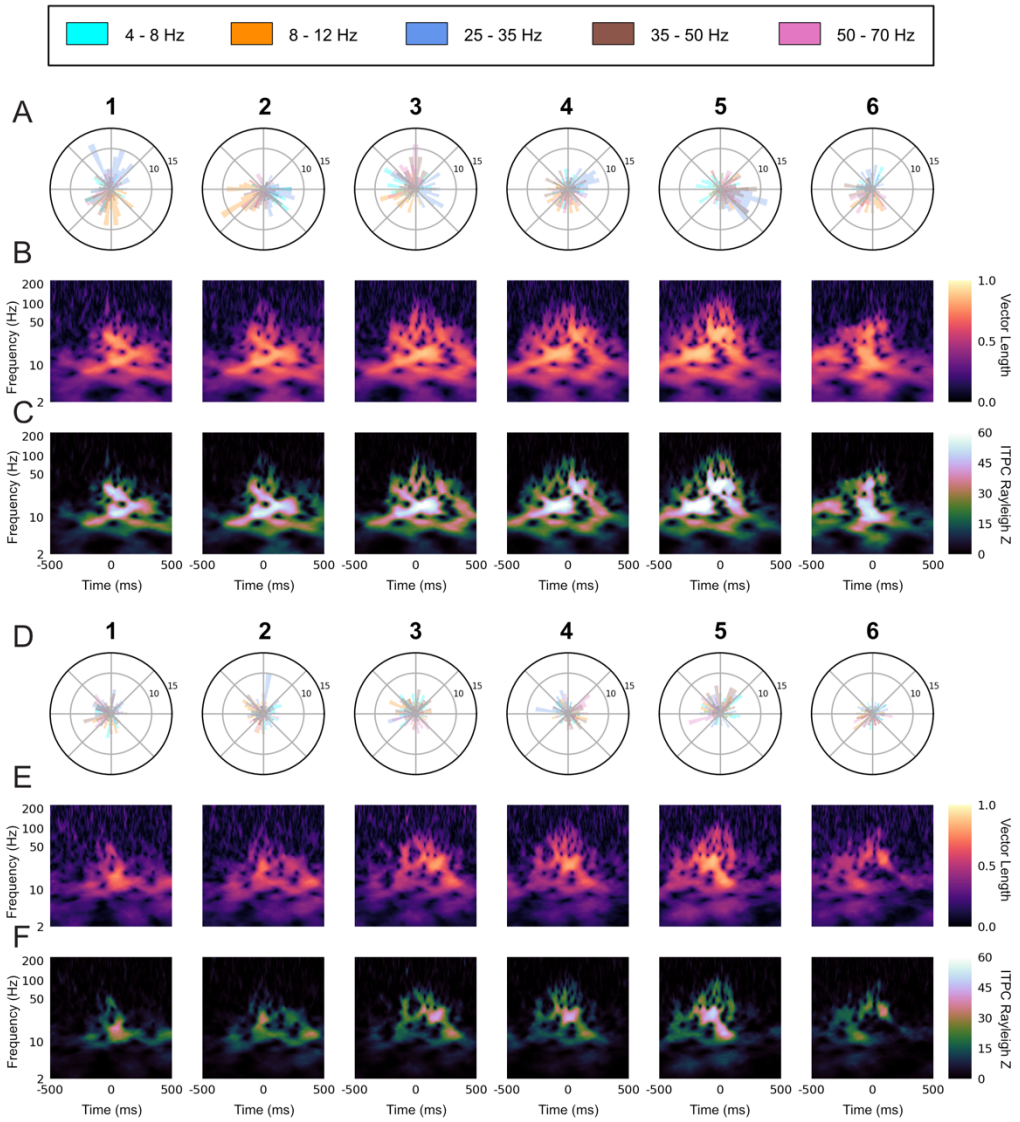


Figure 2.20: Unique LFP phase preferences at each syllable onset (z017).
(A) Polar histogram of the phase for each LFP frequency band at the labeled start of all instances of a given syllable over the course of one day (Day 1), for one bird (z020). The number of instances (n=97) are equal for all syllables, and set by the syllable class with the fewest renditions. (B) ITPC resultant vector length for each frequency over time relative to the labeled start of each syllable (0 ms) over the same instances as in (A). (C) Rayleigh Z-statistic of the ITPC over the same time and frequencies as (B). (D) Polar histogram of the phase for each LFP frequency band at the labeled start of all instances of a given syllable over the course of one day (Day 2), for one bird (z020). The number of instances (n=82) are equal for all syllables, and set by the syllable class with the fewest renditions. (E) ITPC resultant vector length for each frequency over time relative to the labeled start of each syllable (0 ms) over the same instances as in (D). (F) Rayleigh Z-statistic of the ITPC over the same time and frequencies as (E). ($p<0.007$ for all $Z > 5$ for all syllables for both days; all non-black time-frequency pints in this plot are above the significance threshold).

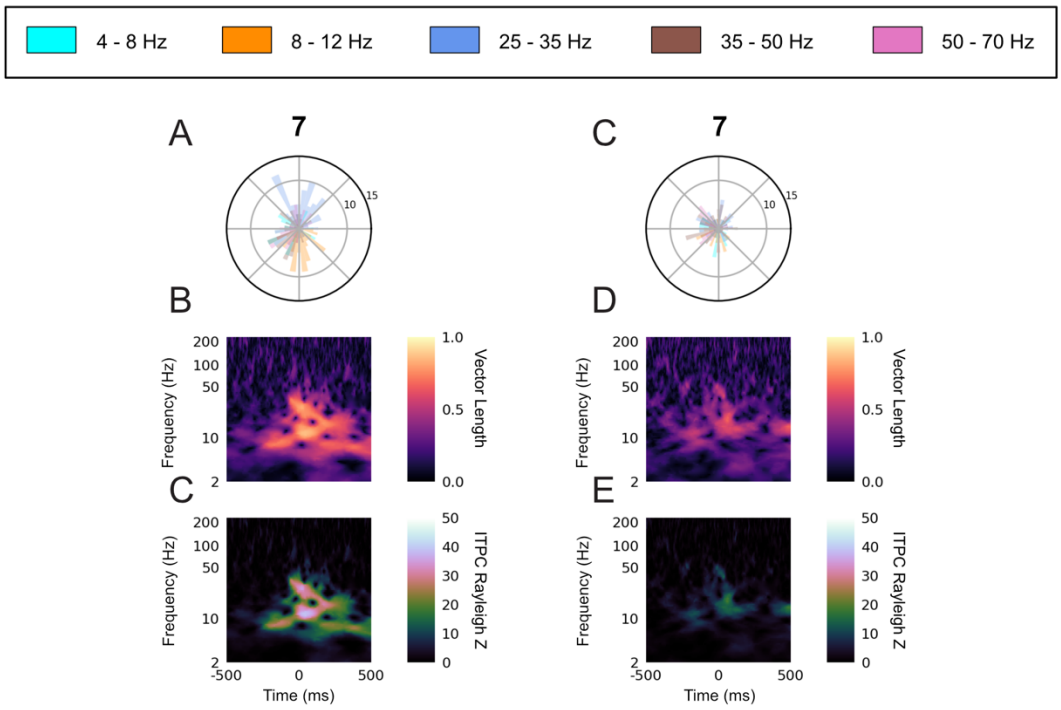


Figure 2.21: Unique LFP phase preferences for sparsely used intra-motif note onset (z017). (A) Polar histogram of the phase for each LFP frequency band at the labeled start of all instances (n=52) of syllable 7 over the course of one day (Day 1), for bird z017. (B) ITPC resultant vector length for each frequency over time relative to the labeled start of syllable 7 (0 ms) over the same instances as in (A). (C) Rayleigh Z-statistic of the ITPC over the same time and frequencies as (B). (D) Polar histogram of the phase for each LFP frequency band at the labeled start of all instances (n=41) of syllable 7 over the course of one day (Day 2), for bird z017. (E) ITPC resultant vector length for each frequency over time relative to the labeled start of syllable 7 (0 ms) over the same instances as in (D). (F) Rayleigh Z-statistic of the ITPC over the same time and frequencies as (E). ($p<0.007$ for all $Z > 5$ for both syllables; all non-black time-frequency pints in this plot are above the significance threshold).

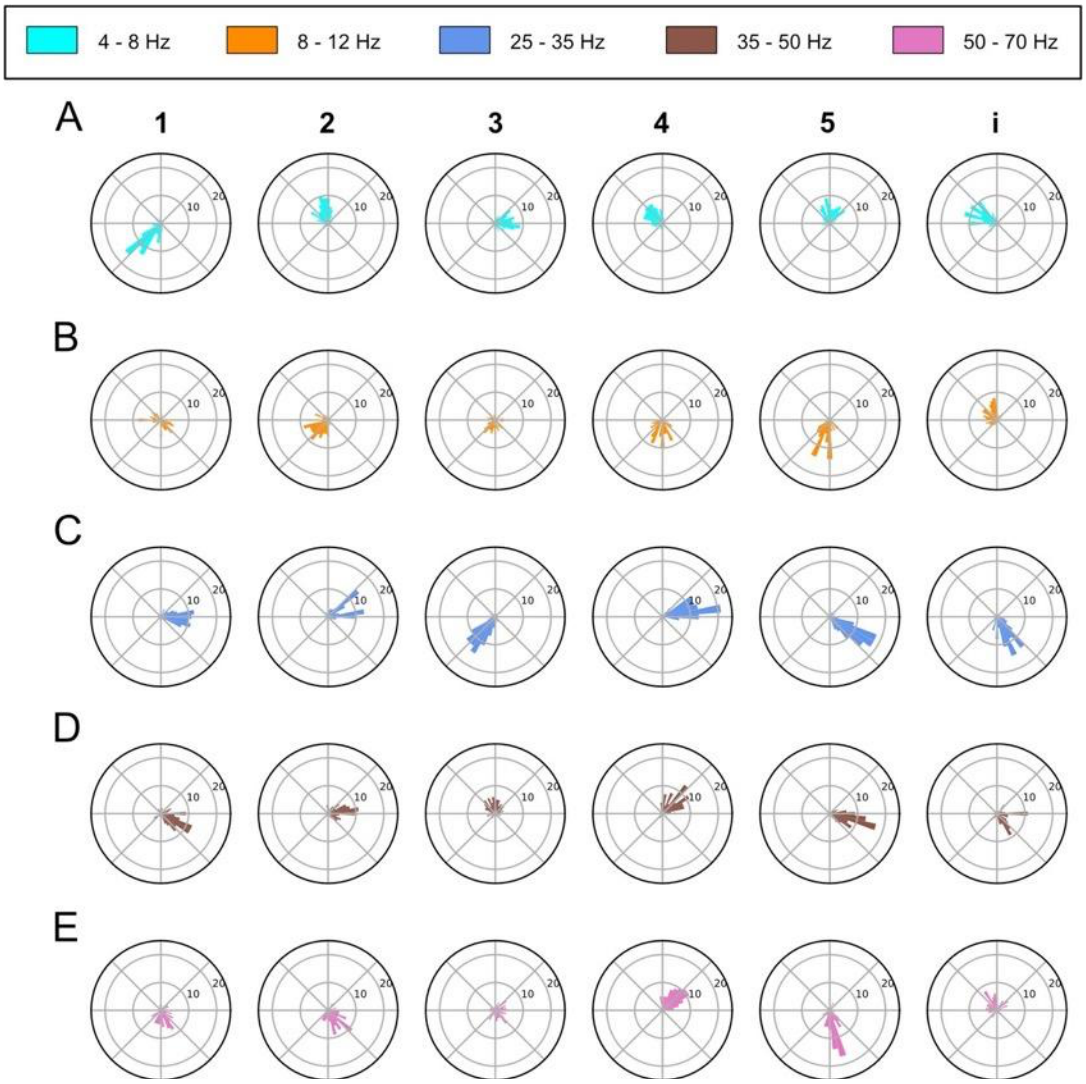


Figure 2.22: Detailed view of the syllable phase preference to syllable onset for z007.
Detailed rendering of the phase preference to syllable onset for z007 shown in Fig 7A. Each row shows a different vocalization type, which includes the five syllables of the motif and the introductory note. Each column shows a different frequency band and is organized top to bottom from least to greatest. As such they are the polar plots for the (A) 4-8 Hz band, (B) 8-12 Hz band, (C) 25-35 Hz band, (D) 35-50 Hz band, and (E) the 50-70 Hz band. The number of instances have been balanced to match the class with the least number of instances (n=100) for each class.

2.7 Discussion

In the current study we demonstrate that LFP in zebra finch HVC carry significant information tied to vocal production. More specifically, the time-varying structure of this meso-scale signal, captured by both the power and phase of various LFP frequency bands, correlates with behaviorally-relevant timepoints during this complex, learned motor behavior. Our results builds upon previous studies in zebra finch HVC [39], by establishing that LFP features can be used to detect both when the bird will sing and the specific syllable that the bird will sing. Prior to that previous study and our work here, little was known about the relationship between LFP features and vocal production in birds. This limited the generalizability of the otherwise powerful birdsong model to other systems, including non-human primates and humans, where LFP and surface potentials are broadly used to measure the dynamics of cortical motor control [12,54,65,75]. In addition, we note that the bandwidth and features of the LFP signals investigated in this paper share similarities with LFP features tied to motor control in humans, non-human primates, and other mammals. For example, the power spectrum components most closely tied to song in finches (Figure 2.10 and Figure 2.11–Figure 2.13) match those documented in the human motor cortex during finger flexion [65]. We suggest that LFP recordings can serve as useful targets to further our understanding of songbird neurophysiology, and to more closely connect the powerful birdsong model to motor control studies in mammals, including non-human primates and humans.

A striking feature of the amplitude changes in LFP during song is the increase in higher frequencies that are within the band often referred to as “high gamma”. Although changes in this frequency band in another brain region in the song bird system, named Nif, have previously been reported [76], they were not found to be premotor related. We have shown the first preliminary

evidence, to our knowledge, of premotor activity related amplitude changes in high gamma (Figure 2.7). This is significant, as this band is often the feature of choice for many state-of-the-art neural decoding studies in humans and suggests designs for neurally driven speech prostheses [12,24]. Notably, these changes empirically appeared to correspond to introductory notes in addition to song syllable. Future work should evaluate if these amplitude changes also occur during calls, which are considerably less deterministic and more syntactically complex than the zebra finch motif.

A consistent feature observed in the LFP of HVC during singing is tight phase-locking to vocal behavior. In evaluating the nature of the observed song-aligned phase-locking; we proposed two hypotheses: (1) where the phase locking is related to the larger motif sequence and (2) in which the phase-locking can be attributed to the timing of each unique vocal unit (syllables). In simpler terms, we ask if the phase-locking is a process aligned to the start of the motif (hypothesis one) or if it is a process that aligns to the timing of each syllable (hypothesis two). In our examination of syllable level timing, we aligned time based upon the syllable onset time. This mirrors the time relationship between spiking activity and syllable onset [55,56]. Further examination would be required to examine the progression of phase with respect to the temporal structure within each syllable.

When directly testing these hypotheses, the second hypothesis was supported by statistical analyses of the data. However, it is important to note that while there is noticeable non-deterministic jitter in the brief gaps of silence between the syllables that make up the motif, this may not be sufficient to completely disprove hypothesis one. On the other hand, there are additional notes, beyond the stereotyped sequences of the motif, that occur non-deterministically between motifs. These notes are referred to as intra-motif notes or “connectors” [60] and

introductory notes for whose timing before the motif is variable. Examination of the phase polar-plots of individual syllables would either show high selectivity of phase preference for all unique vocalizations if our second hypothesis is correct, or low levels of phase-locking for motif syllables and a random distribution about the polar axis (null) for the intra-motif notes and introductory notes if our first hypothesis is correct. As shown in Figure 2.16 and Figure 2.17, each syllable has phase preference that is stable across every specific vocalization type, including intra-motif notes (z017 & z020) (Figure 2.19–Figure 2.21) and introductory notes (z020 & z007) (Figure 2.16 and Figure 2.18). This finding demonstrates that precise phase preference is behaviorally relevant to both learned and non-learned vocalizations. We argue that the long template phase traces in lower frequencies that precede and persist during song production are a byproduct of the zebra finches' stereotyped song structure. This structure is composed of encoding at a smaller time scale centered on each individual syllable during song production. Similar phase preference to motor onset has been found in both human and non-human motor cortex [77].

The results of the normalized sustained Z statistic of inter-trial phase coherence suggest that there are differences in magnitude of coherence peaks both between and within birds. This is likely due to significant variation in signal statistics. This variation likely has many sources, including precise electrode placement and varying signal-to-noise ratios across days. Within days, the peaks in the Z-statistic denote consistency in phase-locking. Although the magnitude of these peaks varied, the frequency bands that contained them were empirically similar to those described previously in mammalian literature. These results, the existing literature, and our empirical inspection of the narrowbands, as shown in Figure 2.8 and Figure 2.9, informed our narrowband feature selection.

Previous work in zebra finch by Markowitz et al. demonstrated that spiking of both excitatory and inhibitory neurons were locked to the phase of the 25-35 Hz band [37,39]. Admittedly, this previously identified phase-locking, combined with the behavior-locked manner of local spiking activity to song behavior, gives further credence to some of the phase-locking we observed. However, it is important to note that local neural spiking activity and LFP are not deterministically linked. LFP reflects the postsynaptic inputs to the local neuronal population and transmembrane currents in the recorded region, which is of course different from the spiking in that region, which reflects neuronal outputs [78]. Additionally, our work establishes that other frequency bands beside the 25-35 Hz band are phase locked to the behavior. More importantly, we find that these frequencies, and the 25-35 Hz band, are also locked to behaviors beyond the syllables of the motif, namely intra-motif notes and introductory notes. Collectively these results show the utility of using LFP to study vocal production in songbirds.

2.8 Acknowledgment

Chapter 2, in part, has been adapted from material as it appears in PLOS Computational Biology, 2021, “Local Field Potentials in a Pre-motor Region Predict Learned Vocal Sequences,” D. E. Brown II, J. I. Chavez, D. H. Nguyen, A. Kadwory, B. Voytek, E. Arneodo, T. Q. Gentner, and V. Gilja. The dissertation author was the primary investigator and author of this paper.

2.9 References

1. Goller F, Cooper BG. Peripheral Motor Dynamics of Song Production in the Zebra Finch. Ann N Y Acad Sci. 2004;1016: 130–152. doi:10.1196/annals.1298.009
2. Jarvis ED. Evolution of vocal learning and spoken language. Science. 2019;366: 50–54. doi:10.1126/science.aax0287
3. Nowicki S. Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. Nature. 1987;325: 53–55. doi:10.1038/325053a0
4. Churchland MM, Cunningham JP, Kaufman MT, Ryu SI, Shenoy KV. Cortical Preparatory Activity: Representation of Movement or First Cog in a Dynamical Machine? Neuron. 2010;68: 387–400. doi:10.1016/j.neuron.2010.09.015
5. David A. Rosenbaum. Human Movement Initiation: Specification of Arm, Direction, and Extent. J Exp Psychol Gen. 1980;109: 444–474. doi:https://doi.org/10.1037/0096-3445.109.4.444
6. Fee MS, Kozhevnikov AA, Hahnloser RHR. Neural Mechanisms of Vocal Sequence Generation in the Songbird. Ann N Y Acad Sci. 2004;1016: 153–170. doi:10.1196/annals.1298.022
7. Ghez C, Hening W, Gordon J. Organization of voluntary movement. Curr Opin Neurobiol. 1991;1: 664–671. doi:10.1016/S0959-4388(05)80046-7
8. Kozhevnikov AA, Fee MS. Singing-Related Activity of Identified HVC Neurons in the Zebra Finch. J Neurophysiol. 2007;97: 4271–4283. doi:10.1152/jn.00952.2006
9. Daliparthi VK, Tachibana RO, Cooper BG, Hahnloser RH, Kojima S, Sober SJ, et al. Transitioning between preparatory and precisely sequenced neuronal activity in production of a skilled behavior. eLife. 2019;8: e43732. doi:10.7554/eLife.43732
10. Rajan R. Pre-Bout Neural Activity Changes in Premotor Nucleus HVC Correlate with Successful Initiation of Learned Song Sequence. J Neurosci. 2018;38: 5925–5938. doi:10.1523/JNEUROSCI.3003-17.2018
11. Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. Nature. 2013;495: 327–332. doi:10.1038/nature11911
12. Bouchard KE, Chang EF. Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Chicago, IL: IEEE; 2014. pp. 6782–6785. doi:10.1109/EMBC.2014.6945185
13. Martin S, Iturrate I, Millán J del R, Knight RT, Pasley BN. Decoding Inner Speech Using Electrocorticography: Progress and Challenges Toward a Speech Prosthesis. Front Neurosci. 2018;12: 422. doi:10.3389/fnins.2018.00422
14. Rabbani Q, Milsap G, Crone NE. The Potential for a Speech Brain–Computer Interface Using Chronic Electrocorticography. Neurotherapeutics. 2019;16: 144–165. doi:10.1007/s13311-018-00692-2
15. Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E. Sequential Processing of Lexical, Grammatical, and Phonological Information Within Broca's Area. Science. 2009;326: 445–449. doi:10.1126/science.1174481

16. Churchland MM, Santhanam G, Shenoy KV. Preparatory Activity in Premotor and Motor Cortex Reflects the Speed of the Upcoming Reach. J Neurophysiol. 2006;96: 3130–3146. doi:10.1152/jn.00307.2006
17. Tanji J, Evarts EV. Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. J Neurophysiol. 1976;39: 1062–1068. doi:10.1152/jn.1976.39.5.1062
18. Svoboda K, Li N. Neural mechanisms of movement planning: motor cortex and beyond. Curr Opin Neurobiol. 2018;49: 33–41. doi:10.1016/j.conb.2017.10.023
19. Erlich JC, Bialek M, Brody CD. A Cortical Substrate for Memory-Guided Orienting in the Rat. Neuron. 2011;72: 330–343. doi:10.1016/j.neuron.2011.07.010
20. Guo ZV, Li N, Huber D, Ophir E, Gutnisky D, Ting JT, et al. Flow of Cortical Activity Underlying a Tactile Decision in Mice. Neuron. 2014;81: 179–194. doi:10.1016/j.neuron.2013.10.020
21. Li N, Chen T-W, Guo ZV, Gerfen CR, Svoboda K. A motor cortex circuit for motor planning and movement. Nature. 2015;519: 51–56. doi:10.1038/nature14178
22. Sainburg T, Theilman B, Thielk M, Gentner TQ. Parallels in the sequential organization of birdsong and human speech. Nat Commun. 2019;10: 3636. doi:10.1038/s41467-019-11605-y
23. Angrick M, Herff C, Mugler E, Tate MC, Slutzky MW, Krusienski DJ, et al. Speech synthesis from ECoG using densely connected 3D convolutional neural networks. J Neural Eng. 2019;16: 036019. doi:10.1088/1741-2552/ab0c59
24. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. Nature. 2019;568: 493–498. doi:10.1038/s41586-019-1119-1
25. Stavisky SD, Willett FR, Wilson GH, Murphy BA, Rezaii P, Avansino DT, et al. Neural ensemble dynamics in dorsal motor cortex during speech in people with paralysis. eLife. 2019;8: e46015. doi:10.7554/eLife.46015
26. Herff C, Schultz T. Automatic Speech Recognition from Neural Signals: A Focused Review. Front Neurosci. 2016;10. doi:10.3389/fnins.2016.00429
27. Gilja V, Pandarinath C, Blabe CH, Nuyujukian P, Simeral JD, Sarma AA, et al. Clinical translation of a high-performance neural prosthesis. Nat Med. 2015;21: 1142–1145. doi:10.1038/nm.3953
28. Gilja V, Nuyujukian P, Chestek CA, Cunningham JP, Yu BM, Fan JM, et al. A high-performance neural prosthesis enabled by control algorithm design. Nat Neurosci. 2012;15: 1752–1757. doi:10.1038/nn.3265
29. Velliste M, Perel S, Spalding MC, Whitford AS, Schwartz AB. Cortical control of a prosthetic arm for self-feeding. Nature. 2008;453: 1098–1101. doi:10.1038/nature06996
30. Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, et al. High-performance neuroprosthetic control by an individual with tetraplegia. The Lancet. 2013;381: 557–564. doi:10.1016/S0140-6736(12)61816-9
31. Mijail D. Serruya, Nicholas G. Hatsopoulos, Liam Paninski, Matthew R. Fellows, John P. Donoghue. Instant neural control of a movement signal. Nature. 416: 141–142. doi:10.1038/416141a
32. Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature. 2006;442: 164–171. doi:10.1038/nature04970
33. Brainard MS, Doupe AJ. What songbirds teach us about learning. Nature. 2002;417: 351–358. doi:10.1038/417351a

34. Nottebohm F. The Neural Basis of Birdsong. PLoS Biol. 2005;3: e164. doi:10.1371/journal.pbio.0030164
35. Okubo TS, Mackevicius EL, Payne HL, Lynch GF, Fee MS. Growth and splitting of neural sequences in songbird vocal development. Nature. 2015;528: 352–357. doi:10.1038/nature15741
36. Woolley SMN. Early experience shapes vocal neural coding and perception in songbirds. Dev Psychobiol. 2012;54: 612–631. doi:10.1002/dev.21014
37. Liberti WA, Markowitz JE, Perkins LN, Liberti DC, Leman DP, Guitchounts G, et al. Unstable neurons underlie a stable learned behavior. Nat Neurosci. 2016;19: 1665–1671. doi:10.1038/nn.4405
38. Poole B, Markowitz JE, Gardner TJ. The Song Must Go On: Resilience of the Songbird Vocal Motor Pathway. Coleman MJ, editor. PLoS ONE. 2012;7: e38173. doi:10.1371/journal.pone.0038173
39. Markowitz JE, Liberti WA, Guitchounts G, Velho T, Lois C, Gardner TJ. Mesoscopic Patterns of Neural Activity Support Songbird Cortical Sequences. Ashe J, editor. PLOS Biol. 2015;13: e1002158. doi:10.1371/journal.pbio.1002158
40. Picardo MA, Merel J, Katlowitz KA, Vallentin D, Okobi DE, Benezra SE, et al. Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch. Neuron. 2016;90: 866–876. doi:10.1016/j.neuron.2016.02.016
41. Schmidt MF. Pattern of Interhemispheric Synchronization in HVc During Singing Correlates With Key Transitions in the Song Pattern. J Neurophysiol. 2003;90: 3931–3949. doi:10.1152/jn.00003.2003
42. Bottjer SW, Johnson F. Circuits, hormones, and learning: Vocal behavior in songbirds. : 17.
43. Luo M, Perkel DJ. A GABAergic, Strongly Inhibitory Projection to a Thalamic Nucleus in the Zebra Finch Song System. J Neurosci. 1999;19: 6700–6711. doi:10.1523/JNEUROSCI.19-15-06700.1999
44. Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary,Serinus canarius. J Comp Neurol. 1976;165: 457–486. doi:10.1002/cne.901650405
45. Vates GE, Vicario DS, Nottebohm F. Reafferent thalamo-"cortical" loops in the song system of oscine songbirds. : 16.
46. Doupe AJ, Kuhl PK. BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms. Annu Rev Neurosci. 1999;22: 567–631. doi:10.1146/annurev.neuro.22.1.567
47. Pfenning AR, Hara E, Whitney O, Rivas MV, Wang R, Roulhac PL, et al. Convergent transcriptional specializations in the brains of humans and song-learning birds. Science. 2014;346: 1256846–1256846. doi:10.1126/science.1256846
48. Kosche G, Vallentin D, Long MA. Interplay of Inhibition and Excitation Shapes a Premotor Neural Sequence. J Neurosci. 2015;35: 1217–1227. doi:10.1523/JNEUROSCI.4346-14.2015
49. Leonardo A. Ensemble Coding of Vocal Control in Birdsong. J Neurosci. 2005;25: 652–661. doi:10.1523/JNEUROSCI.3036-04.2005
50. Buzsáki G, Anastassiou CA, Koch C. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat Rev Neurosci. 2012;13: 407–420. doi:10.1038/nrn3241
51. Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov DF, et al. Learning to Control a Brain–Machine Interface for Reaching and Grasping by Primates. Idan Segev, editor. PLoS Biol. 2003;1: e42. doi:10.1371/journal.pbio.0000042
52. Fetz EE. Operant Conditioning of Cortical Unit Activity. Science. 1969;163: 955–958. doi:10.1126/science.163.3870.955

53. Kubánek J, Miller KJ, Ojemann JG, Wolpaw JR, Schalk G. Decoding flexion of individual fingers using electrocorticographic signals in humans. J Neural Eng. 2009;6: 066001. doi:10.1088/1741-2560/6/6/066001
54. Miller KJ, Leuthardt EC, Schalk G, Rao RPN, Anderson NR, Moran DW, et al. Spectral Changes in Cortical Surface Potentials during Motor Movement. J Neurosci. 2007;27: 2424–2432. doi:10.1523/JNEUROSCI.3886-06.2007
55. Yu AC, Margoliash D. Temporal Hierarchical Control of Singing in Birds. Science. 1996;273: 1871–1875. doi:10.1126/science.273.5283.1871
56. Hahnloser RHR, Kozhevnikov AA, Fee MS. An ultra-sparse code underliesthe generation of neural sequences in a songbird. Nature. 2002;419: 65–70. doi:10.1038/nature00974
57. Buzsaki G, Draguhn A. Neuronal Oscillations in Cortical Networks. Science. 2004;304: 1926–1929. doi:10.1126/science.1099745
58. Wang X-J. Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. Physiol Rev. 2010;90: 1195–1268. doi:10.1152/physrev.00035.2008
59. Lopes da Silva F. EEG and MEG: Relevance to Neuroscience. Neuron. 2013;80: 1112–1128. doi:10.1016/j.neuron.2013.10.017
60. Hyland Bruno J, Tchernichovski O. Regularities in zebra finch song beyond the repeated motif. Behav Processes. 2019;163: 53–59. doi:10.1016/j.beproc.2017.11.001
61. Ezequiel Arneodo. Software and hardware designs for chronic, high channel count electrophysiology. Biocircuits Institute; 2016. Available: https://github.com/singingfinch/bernardo.git
62. Arneodo EM, Chen S, Gilja V, Gentner TQ. A neural decoder for learned vocal behavior. Neuroscience; 2017 Sep. doi:10.1101/193987
63. Gramfort A. MEG and EEG data analysis with MNE-Python. Front Neurosci. 2013;7. doi:10.3389/fnins.2013.00267
64. Paul Boersma, David Weenink. Praat: doing phonetics by computer. 2018. Available: http://www.praat.org/
65. Miller KJ, Zanos S, Fetz EE, den Nijs M, Ojemann JG. Decoupling the Cortical Power Spectrum Reveals Real-Time Representation of Individual Finger Movements in Humans. J Neurosci. 2009;29: 3132–3137. doi:10.1523/JNEUROSCI.5506-08.2009
66. Brovelli A, Lachaux J-P, Kahane P, Boussaoud D. High gamma frequency oscillatory activity dissociates attention from intention in the human premotor cortex. NeuroImage. 2005;28: 154–164. doi:10.1016/j.neuroimage.2005.05.045
67. Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Kirsch HE, et al. High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. Science. 2006;313: 1626–1628. doi:10.1126/science.1128115
68. Pfurtscheller G, Graimann B, Huggins JE, Levine SP, Schuh LA. Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement. Clin Neurophysiol. 2003;114: 1226–1236. doi:10.1016/S1388-2457(03)00067-1
69. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004;134: 9–21. doi:10.1016/j.jneumeth.2003.10.009
70. Berens P. CircStat: A MATLAB Toolbox for Circular Statistics. J Stat Softw. 2009;31. doi:10.18637/jss.v031.i10
71. Fisher NI. Statistical Analysis of Circular Data. 1st ed. Cambridge University Press; 1993. doi:10.1017/CBO9780511564345

72. Jammalamadaka SR, Sengupta A. Topics in circular statistics. River Edge, N.J: World Scientific; 2001.
73. Zar JH. Biostatistical analysis. 5th ed. Upper Saddle River, N.J: Prentice-Hall/Pearson; 2010.
74. Buzsáki G. Rhythms of the Brain. Oxford University Press; 2006. doi:10.1093/acprof:oso/9780195301069.001.0001
75. Jiang W, Pailla T, Dichter B, Chang EF, Gilja V. Decoding speech using the timing of neural signal modulation. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Orlando, FL, USA: IEEE; 2016. pp. 1532–1535. doi:10.1109/EMBC.2016.7591002
76. Lewandowski BC, Schmidt M. Short Bouts of Vocalization Induce Long-Lasting Fast Gamma Oscillations in a Sensorimotor Nucleus. J Neurosci. 2011;31: 13936–13948. doi:10.1523/JNEUROSCI.6809-10.2011
77. Miller KJ, Hermes D, Honey CJ, Hebb AO, Ramsey NF, Knight RT, et al. Human Motor Cortical Activity Is Selectively Phase-Entrained on Underlying Rhythms. Behrens T, editor. PLoS Comput Biol. 2012;8: e1002655. doi:10.1371/journal.pcbi.1002655
78. Gilja V, Moore T. Electrical Signals Propagate Unbiased in Cortex. Neuron. 2007;55: 684–686. doi:10.1016/j.neuron.2007.08.012

Chapter 3 : Proof-of-Concept Birdsong BCIs

3.1 Abstract

The previous chapter focused on the population activity presumptively from HVC in awake free behaving zebra finches. The analyses described characterize Local Field Potentials (LFP) during song production and document several features that are qualitatively similar to well-established oscillations found in both human and non-human mammalian areas. Given these newly described physiological similarities, despite distinct, species specific anatomical structures, we next examine if these features can be leveraged to decode vocal activity using methods that are complementary to current human speech BCI studies. In this chapter we describe novel approaches and methods that show that this neural activity can be used to predict both the identity of each vocal element (syllable) and when it will occur during song. We also briefly describe contemporaneous work that uses this same data and develops novel methods for synthesizing songs directly from neural activity. All novel methods described in this chapter can be implemented in real-time using neural features that proceed the vocal behavior. Collectively these works demonstrate novel methods that could be translated to human clinical speech BCIs research and provide strong preliminary evidence for the utility of the songbird animal model for supplementing speech BCI studies.

3.2 Introduction

At a high level vocal communication involves the production of an acoustic signal that is mutually comprehensible to the communicator and the intended listener. This acoustic signal can be perceived and decomposed into symbolic components that carry the intended message. Human

speech BCI researchers aim to develop therapeutic devices that enable those who have lost or diminished ability to control the speech articulator musculature necessary to produce these communication signals. Most approaches to study and develop these systems involve using features derived from neural signals recorded directly from motor brain regions related to speech during vocal production.

Broadly, the communication output of all speech focused BCIs fall under two main types; (1) *Symbolic Decoding*, and (2) *Acoustic Synthesis:*

(1) *Symbolic Decoding* attempts to decode the intended vocal communication by approaching it using classification methods. Instead of trying to reconstruct all of the spectral features of the acoustic signal it tries to decode the components of the message itself. This approach has found success by decoding symbolic components at varying levels of complexity such as articulatory gestures [1], phonemes [2,3], phonetic features [4,5], words [6] and continuous sentences [7–9].

(2) *Acoustic Synthesis* attempts to produce all of the spectral features of the intended vocalizations. The communication target is tasked with both hearing and understanding the message in the same way they would comprehend another person speaking. This synthesis can involve synthesizing different scales of vocalizations from components such as vowels [10], spectral features of subunits of speech [11,12], and full sentences [13]. This approach has also had recent progress synthesizing both imagined and whispered speech [14].

Both approaches have their advantages and challenges, however neither approach has achieved system performance levels that would enable widespread adoption as a therapeutic device. Speech prosthesis research as a whole has several challenges that impedes scientific progress and translation. Namely limits in data quantity, experiment duration, and brain coverage. It is important then to supplement this work with a viable translatable animal model. For this reason we have sought to validate both approaches in the songbird animal model. This chapter will mostly focus on a Symbolic Decoding approach in songbirds using features described in chapter 2 [15]. We will also include a brief summary of Acoustic Synthesis results, which use the same neural recordings, so that we can discuss both approaches and their context within the large vocal prosthesis research field.

3.3 Symbolic Decoding of Bird Song: Decoding Vocalization Identity

3.3.1 Overview of Methods

We implanted male zebra finches with laminar electrodes and simultaneously recorded their vocal behavior and neural activity as described in sections 2.2.2 Subjects–2.2.4 Electrode implantation procedure. The recorded neural signals were processed with signal processing techniques that are described in detail in section 2.2.5 Analysis of electrophysiology data. Through a series of steps, explained in detail in section 2.3 Annotation and alignment of behavioral data, we found segments of the recording that contained songs and hand-annotated them using Praat (Figure 2.2). These behavioral labels were applied to analyze the neural activity in relation to specific classes of behavior to determine what frequencies, if any, correlated to the behavior. The resulting frequencies were used to classify and predict behavior onset to clarify their relationship to the vocalizations (Figure 3.1–Figure 3.8).

3.3.2 Feature extraction

Feature extraction from the preprocessed data comprised a series of steps that involved first Common Average Referencing and then narrow band-pass filtering. Common Average Referencing has been shown to minimize or remove the contributions of correlated noise sources such as 60 Hz noise and intermittent noise sources such as movement artifacts [16]. We used the Hilbert transform to compute the analytic signal. To extract oscillatory power, we used the absolute value of the analytical signal; this effectively removed all phase-related information. To extract oscillatory phase, we used the instantaneous angle of the complex argument of the analytical signal; this was further smoothed using a sine function. This effectively removed all information related to power from the signal.

These signals, whether Hilbert-transformed or not, were then sub-selected at times relative to the label onset for a syllable using two parameters: bin width and offset. The bin width is the number of samples used prior to the offset, and the offset is the number of samples, or milliseconds, prior to the labeled start time to a specific syllable. All combinations of these hyperparameters used offset that were prior to the start of the vocal behavior.

3.3.3 Band templates and Pearson correlation features.

Using the optimized bin width and offset for a particular frequency band, we calculated a template for that band by taking the mean of the LFP traces of the training set of a specific behavior class. This template represents the average LFP activity prior to and during the production of a particular vocal behavior. This template was then used to extract features from a narrow-band LFP trace by taking the Pearson correlation of the template from a segment of neural activity of the same length in samples. This correlation value is set between -1 and 1. For the behavioral

classification results, a segment of the corresponding LFP frequency band that is set by the optimized bin width and offset is used as the feature. For the onset detection analysis, the Pearson correlation for each sliding segment equal in length to the optimal bin width is used to detect behavior that corresponds to its optimal offset.

3.3.4 Linear classifier

To classify behavior using the LFP features, a linear discriminant analysis (LDA) model using singular value decomposition (SVD) was trained using the scikit-learn toolbox [17] in Python. This classifier is tasked to correctly classify examples of all vocalizations during song, both motif syllables and intra-motif notes, in addition to introductory notes and periods of silence. As the classifier must learn how to distinguish between both vocal and non-vocal events, these events are collectively referred to as classes. All priors for each class were set equal, and all classes were balanced in the datasets used for classification analyses. Detailed information regarding the number of instances in each class can be found in S3-S5 Table. No shrinkage or regularization were used; however, the SVD optimizer was used to avoid ill-conditioned covariance matrices. Results were validated using 5-fold cross-validation. Templates for feature extraction were created by taking the mean across the training set. These templates were used to extract features from both the training set and the testing set. All frequencies were trained and tested independently of one another.

3.3.5 Channel-Adding Analysis.

Channel-adding curves were calculated using a bootstrap approach to determine how many channels were needed until additive information saturation. The channel-adding curves were calculated by first training and testing a classifier with the neural features of only one channel,

using the steps described previously, then repeating this analysis after adding the features of another randomly selected channel. This repeated training and testing after adding a random channel is considered complete once a classifier that uses the features from all available good recording channels is evaluated. This channel-adding analysis was repeated 5,000 times, changing the order in which each channel was included, with 5-fold cross validation. The order in which channels were added over the 5,000 repetitions was maintained across the folds to enable fair calculation of their validated mean. These are subsequently used to calculate the mean and standard deviation across repetitions.

3.3.6 LFP features encode intended syllable identity

Having isolated potential LFP bands of interest through the decomposition of both phase and power, we next asked whether the LFP bands' spectral features were correlated to vocalization identity. If so, then these features could be used to classify vocalization identity from neural activity alone. As the dominant characteristic of these features are their consistent structure, a promising approach was to create frequency band-specific LFP templates that could be correlated with representative time traces. It was unclear, however, what the ideal time range and latency relative to song onset for this information might be. With the goal of better understanding the ideal time range and latency, we conducted a hyperparameter search using a linear discriminant analysis (LDA) model and singular value decomposition (SVD) (see section 3.3.4 Linear classifier). The classifier was trained with LFP templates for which the duration in samples (bin width) and the time between the event onset and the end of the template (offset) was varied in the hyperparameter search. All templates considered fully precede the onset of the event. For most frequencies found, multiple combinations of bin width and offset yielded classifiers that could distinguish, significantly above binomial chance, between song syllables, silent (non-vocal)

periods, and —for the subject that performed them— introductory notes (z020 & z007) (Figure 3.1 and Figure 3.2).

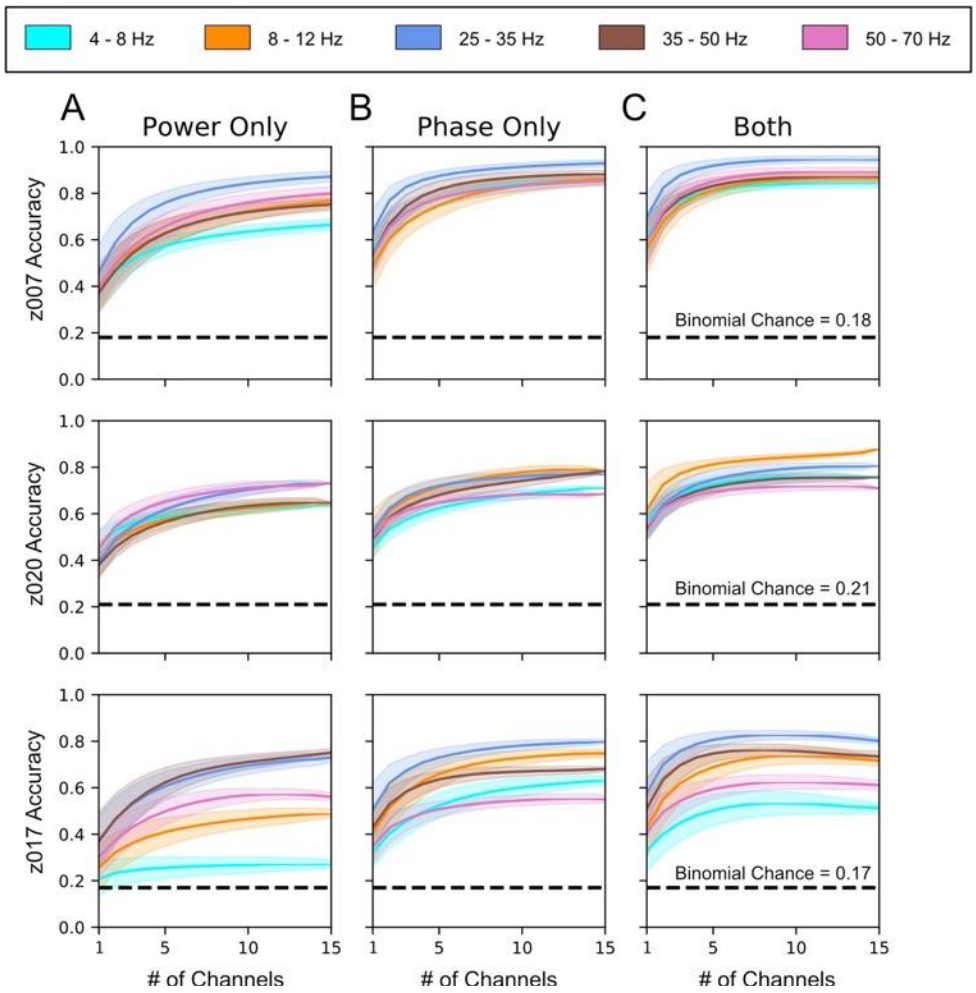


Figure 3.1: LFP phase and power provide independent and additive information for vocalization classification.

Channel-adding curves calculated by repeatedly training classifiers with an increasing number of randomly selected channels (see section 3.3.5 **Channel-Adding Analysis.**). (A-C) Channel-adding curves showing classifier performances with either (A) all LFP phase information removed, (B) all LFP power information removed, or (C) with both phase and power used as independent features. Each row corresponds to data from the highest-yield day for each bird. These vocal and non-vocal classification events are collectively referred to as classes (see section 2.3 **Annotation and alignment of behavioral data**). z007 n=98 for each of 7 classes (1, 2, 3, 4, 5, i (introductory note), Silence); z020 n=91 for each of 6 classes (1, 2, 3, 4, i (introductory note), Silence); and for z017 n=52* for each 8 classes (1, 2, 3, 4, 5, 6, 7, Silence). The results of up to 15 channels are shown to allow for direct comparison across subjects. Dark lines show the mean for each vocalization class, shaded areas give the standard deviation over the bootstrapped analysis using n=5,000 repetitions across 5 cross-validation folds. The p-value for all of the binomial chances calculated for each bird was 0.05.

*The number of instances for each class was limited by Syllable 7, which is an intra-motif note.

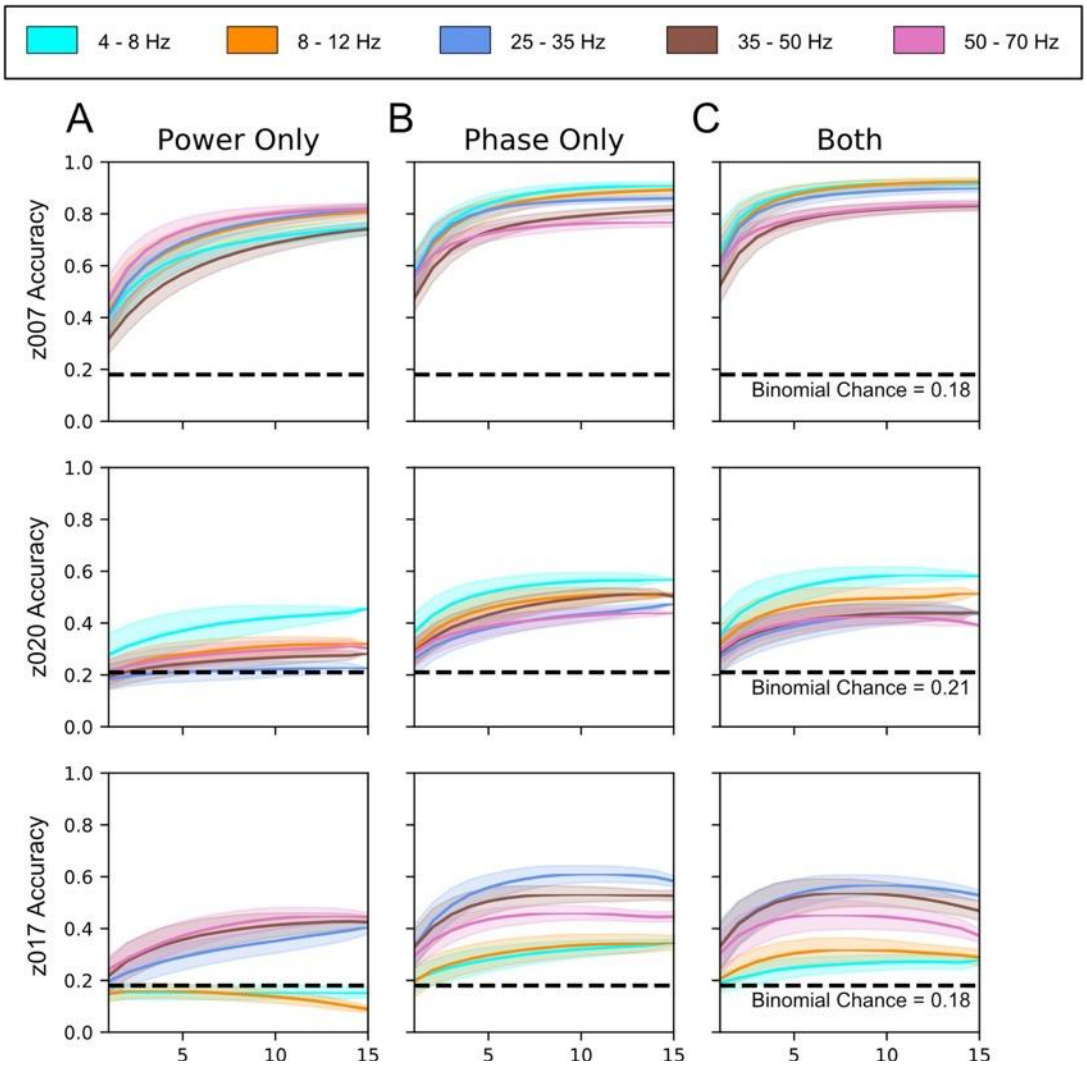


Figure 3.2: Channel-adding Curves for the second highest-yield days across subjects.
Channel-adding curves calculated by repeatedly training classifiers with an increasing number of randomly selected channels (see section 3.3.5 **Channel-Adding Analysis.**). Channel-adding curves of classifier performances with either (A) all phase related information removed, (B) all power related information removed, or (C) with both phase and power used as independent features. Each row corresponds to data from the second highest yielding day for each bird. z007 n=71 for each class n=7 (1, 2, 3, 4, 5, i, Silence), z020 n=75 for each class n = 6 (1, 2, 3, 4, i, Silence), and for z017 n=41* for each class n=8 (1, 2, 3, 4, 5, 6, 7, Silence). Error bars represent the standard deviation over the bootstrapped analysis using n=5,000 repetitions across 5 cross-validation folds. The p-value for all of the binomial chances calculated for each bird was 0.05.

*The number of instances for each class was limited by Syllable 7, which is an intra-motif note.

As classifiers were trained to distinguish both vocal and non-vocal events, these vocalization event types are collectively referred to as classes (see section 3.3.4 Linear classifier). The parameter search for each frequency was performed separately, and, although there was considerable variation, the best bin width and absolute value of offset tended to decrease as LFP frequency increased. Table 3.1 and Table 3.2 summarize the parameters in the search that yielded the highest classification accuracy.

Table 3.1: Bin Width Hyperparameter Search Results

		Best Bin Width (ms)				
Bird ID	Date	4-8 Hz	8-12 Hz	25-35 Hz	35-50 Hz	50-70 Hz
z020	day-2016-06-03	90	185	90	85	90
z020	day-2016-06-05	190	125	35	160	70
z007	day-2016-09-10	170	75	90	25	30
z007	day-2016-09-11	195	170	70	20	70
z017	day-2016-06-19	175	170	85	50	35
z017	day-2016-06-21	150	170	70	60	40

Table 3.2: Offset Hyperparameter Search Results

		Best Offset (ms)				
Bird ID	Date	4-8 Hz	8-12 Hz	25-35 Hz	35-50 Hz	50-70 Hz
z020	day-2016-06-03	-10	-5	0	0	0
z020	day-2016-06-05	-25	-5	0	-20	0
z007	day-2016-09-10	-35	-5	-10	-10	0
z007	day-2016-09-11	-5	-5	-15	0	-15
z017	day-2016-06-19	-50	0	-10	-5	-5
z017	day-2016-06-21	-10	-15	-30	0	0

Nearly all frequencies we examined were useful for classifying prior neural activity as it relates to the song it will produce downstream; however, it is not immediately clear what component of these oscillations (i.e., phase or amplitude), carries information regarding song identity.

To determine which component (phase or amplitude) of each frequency band provided information about the vocalization identity, we applied the Hilbert transform to selectively remove either phase or power, while retaining only the contributions of power or phase information, respectively. Information provided by a given component was inferred from classifier accuracy. As shown in Figure 3.1A–C, both phase- and amplitude-based classifiers had accuracies above binomial chance for all frequencies; from this, we inferred that they each carried song-related information. In general, however, phase had higher classification accuracy than power for the frequency bands below 50 Hz (Figure 3.3). This difference in accuracy between phase and power was greater for lower frequencies (i.e., those below 30 Hz). The highest frequency band, 50–70 Hz, had marginally higher performance for power-only than for phase-only. We also queried if the number of channels recorded were necessary for decoding song identity by running a channel-adding analysis for each frequency band (see 3.3.5 Channel-Adding Analysis.). This analysis evaluates improvements in classifier performance enabled by increasing the number of recording sites on the probe and provides insight into the point beyond which additional electrodes provide diminishing returns. As shown in Figure 3.1A, the LDA classifier performed well above binomial chance with only 5–10 channels of neural activity when classifying between song syllables, silence and introductory notes.

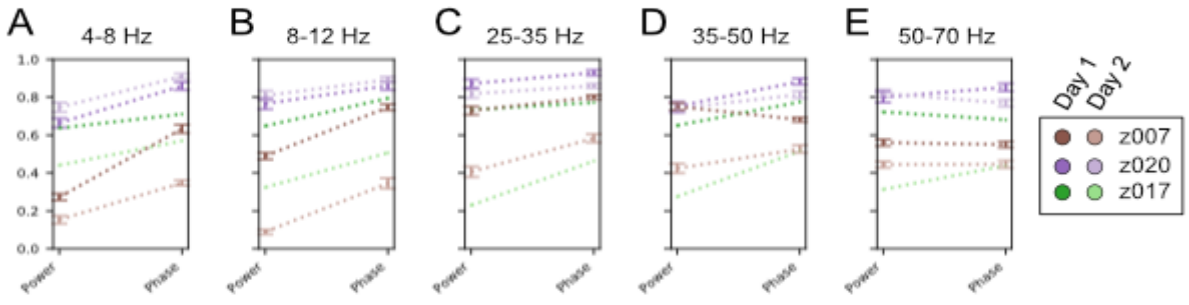


Figure 3.3: Difference in Decoding Accuracy between Phase Only and Power Only Classification.

Classification accuracy for each high-yield day for each bird with 15 channels of neural data when all phase-related information is removed, left, and all power-related information is removed, right, for (A) 4–8 Hz band, (B) 8–12 Hz band, (C) 25–35 Hz band, (D) 35–50 Hz band, and 50–70 Hz band.

3.4 Symbolic Decoding of Bird Song: Onset Detection

3.4.1 Syllable onset detection

In order to analyze the temporal relationship between LFP and song production, we used a template-matching approach to determine whether LFP can predict syllable onset. Each syllable of the motif—excluding the first—had every labeled instance recorded in the same day split into training and testing sets using a 5-fold stratified split (80% training set and 20% testing) (S3-S5 Table). Both the template and the stereotyped onset of the syllable were calculated from the training set. Templates are the mean of the LFP traces of the training set. The stereotyped onset is the average time the syllable occurs in the training set with respect to the first syllable of the motif it occurred in. These templates were then used to compute the Pearson correlation across time for each of the motifs that contain the syllables of the test set, maintaining the temporal relationship of the optimal offset and the bin width for its respective frequency.

The prediction confidence of a single frequency band was calculated by first thresholding the Pearson correlation values at zero, and taking the average of the resultant time series across channels. The maximum confidence within a 100-millisecond window centered on the behavior-derived stereotyped onset time is then used as the frequency's prediction for that instance of the syllable. We followed the same steps as previously stated in order to produce the results of a predictor that uses all of the frequencies; only the prediction confidence across all frequencies and channels is used. The prediction of the syllable onset using the birds' stereotyped behavior was determined by adding the stereotyped onset time, calculated from the training set, to the actual start of the first syllable for the motif the syllable occurred in. It is important to note two things with this approach: (1) neither approach receives information on the actual start time of the syllable in its respective motif and (2) this 100-millisecond window is significantly larger than the natural

variability of syllable onsets, providing the neural-based predictor a more difficult task than the baseline stereotyped comparison. In addition, for later syllables of the motif there are a few instances where the syllable occurs after the 100-millisecond window; meaning that the neural detector will be forced to predict the syllable in a window where it hasn't yet occurred. Due to this the neural decoder has a larger possible maximum prediction error than the behaviorally based predictor. Both predictions were then normalized by taking the difference between the predicted time and the actual labeled start time of the syllable within the same motif.

Statistical significance of the results was calculated with a one-sided Wilcoxon signed-rank test using the difference between the relative predictions of the stereotyped behavioral model and the neural features models. Two null hypotheses were tested, first that there was no difference between the stereotyped behavior and the prediction using neural features, and the second being that the predictor using the neural features was closer to the actual labeled onset time. The second null hypothesis requires that the LFP based predictor must outperform the behavior-based predictor. Results must pass both tests with a p-value less than .05 to be considered significant. The Benjamini-Hochberg procedure was used to control the False Discovery Rate to account for the multiple comparisons done. The procedure was implemented for each day treating the results of each individual frequency and the predictor which uses all frequencies as separate results (n=6, q=.05).

3.4.2 LFP Features Predict Syllable Onsets

The demonstrated ability to accurately classify syllable identity with LFP features could be a consequence of a unique LFP signal structure associated with specific syllables or could be a consequence of a motif level LFP signal structure and stereotypy in the syllable onset time within the motif. To disambiguate these two possibilities, we examined whether the template features optimized for determining syllable identity (Table 3.1 and Table 3.2) could also predict specific

onset times of each separate vocalization within the full motif. To accomplish this task, we implemented a naive pattern-matching approach (see Section **Error! Reference source not found.**). As the syllables' onset is predicted using features derived from causal bandpass filters and samples that occur prior to their start, the features are based entirely upon signals recorded prior to syllable onset. As the stereotyped structure of the zebra finch's song is well documented, we used the mean error between the stereotyped (average across trials) start time of each syllable—relative to its first syllable—and the actual time the syllable occurred as a benchmark to test the performance of a predictor that used the LFP features described previously. The observed behavioral variability in onset timing for each syllable relative to the first is detailed in Table 2.3–Table 2.5. An example of one motif and its corresponding confidences is shown in Figure 3.4A.

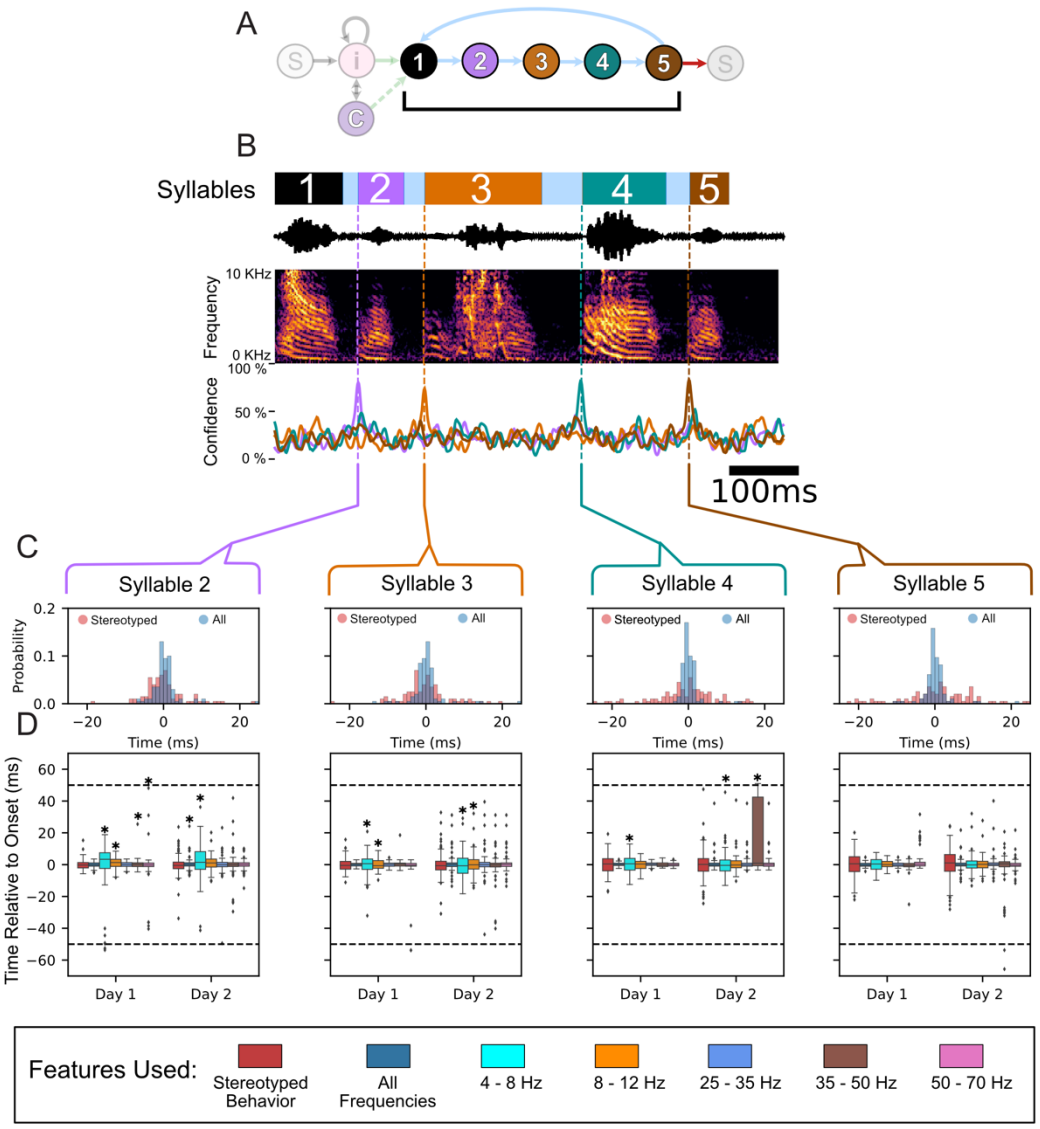


Figure 3.4: Onset detection using LFP features for Subject z007.
(A) State diagram of z007's observed song structure. Syllable colors are the same as in Fig 2 and Fig 3. (B) Example motif from the highest-yield day for subject z007. Annotated behavior (top) using the same color scheme as in Fig 2, sound pressure waveform and the corresponding time-aligned spectrogram (middle), and the time-varying naïve confidence of the onset prediction (bottom) for each syllable in this example motif. Confidence signal traces are the same color as the syllable they are meant to predict. (B) Histogram of onset prediction times for each syllable relative to its annotated start time. The annotated start times are relative to the start of the first syllable of the motif that the syllable occurred in. The histogram compares two approaches: The first, in pink, uses only behavioral information (the stereotyped onset relative to the start of the first syllable), and the second, in blue, uses all the neural features to predict the start of the syllable. (C) Boxplot of onset prediction times relative to the labeled onset time for both of the high-yield days for z007. The order of each feature used is the same, going left to right: first is the stereotyped onset time using only the deterministic behavior (colored red), next are the results using all of the neural features (colored navy blue), then each frequency band only in order from least to greatest (4–8 Hz, 8–12 Hz, 25–35 Hz, 35–50 Hz, and 50–70 Hz). The time window that the neural based predictor must make a prediction within is represented by the dotted black line (see section **Error! Reference source not found.**). Statistical significance was calculated using the o ne-sided Wilcoxon signed-rank test, and * denotes results that are not statistically significant when using the Benjamini-Hochberg False Discovery Rate. All other results $p<0.05$ and $q<0.05$.

Across all syllables for all birds, the predictor that used all of the frequencies performed better than predication based upon the stereotyped onset of the syllable, Figure 3.4B and Figure 3.5–Figure 3.7. Similar results were found for the 2 additional syllables for z017 as well (S20 Fig). Statistically significance was assessed based upon the Wilcoxon signed rank test and the Benjamini-Hochberg false discovery rate (FDR) correction ($p<0.05$; $q<0.05$). The paired test directly compared the error in onset time prediction between the neural predictor and the behaviorally based stereotyped start time to evaluated if the neural-based predictor was more accurate. When applying the FDR correction, the higher performance achieved by the all frequencies LFP predictor over the stereotyped onset time predictor was statistically significant for all syllables across all birds and high-yield days except for only three cases out of the 26 tested; these were syllables 2 & 4 for z020's 2nd high-yield day and syllable 2 for z017's 2nd high-yield day (Figure 3.7).

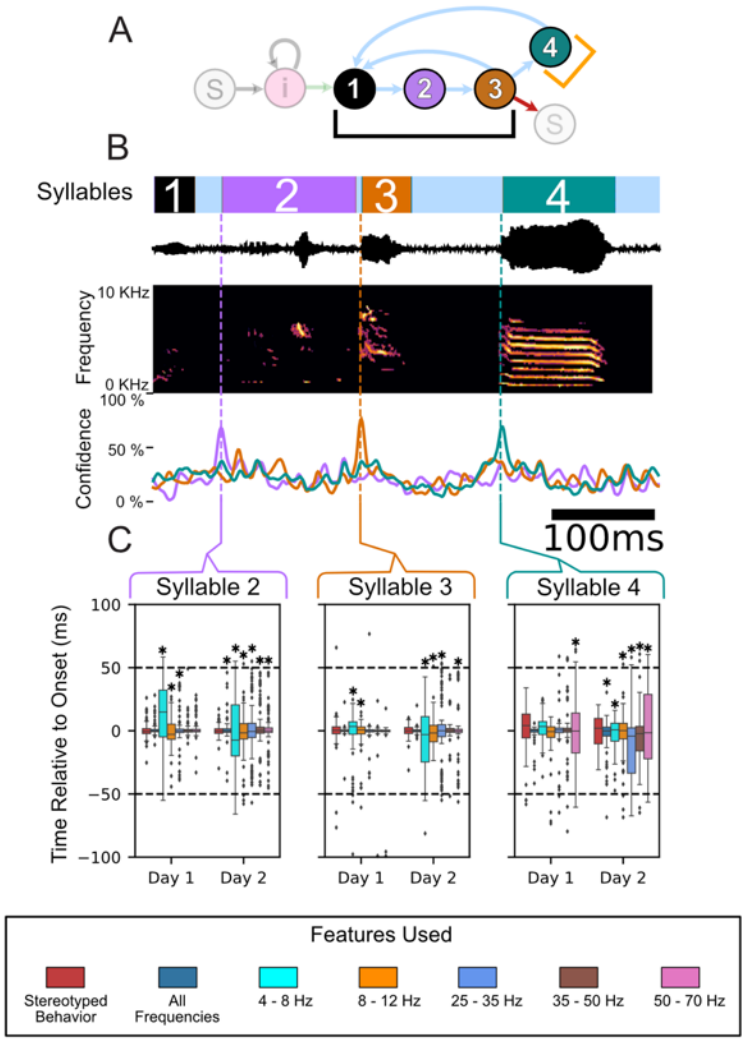


Figure 3.5: Onset detection and branch behavior analysis using LFP features for Subject z020.

(A) State diagram of z020's observed song structure. Syllable colors are the same as in Fig 3. (B) Example motif from the highest-yield day for subject z020. Annotated behavior (top) using the same color scheme as in Fig 3, sound pressure waveform and the corresponding time-aligned spectrogram (middle), and the time-varying naïve confidence of the onset prediction (bottom) for each syllable in this example motif. Confidence signal traces are the same color as the syllable they are meant to predict. (C) Boxplot of onset prediction times relative to the labeled onset time for both of the high-yield days for z020. The order of each feature used is the same, going left to right, as is shown in Fig 10. The time window that the neural based predictor must make a prediction within is represented by the dotted black line (see Section 3.4.1 **Syllable onset detection**). Statistical significance was calculated using the one-sided Wilcoxon signed-rank test, and * denotes results that are not statistically significant when using the Benjamini-Hochberg False Discovery Rate. All other results $p<0.05$ and $q<0.05$.

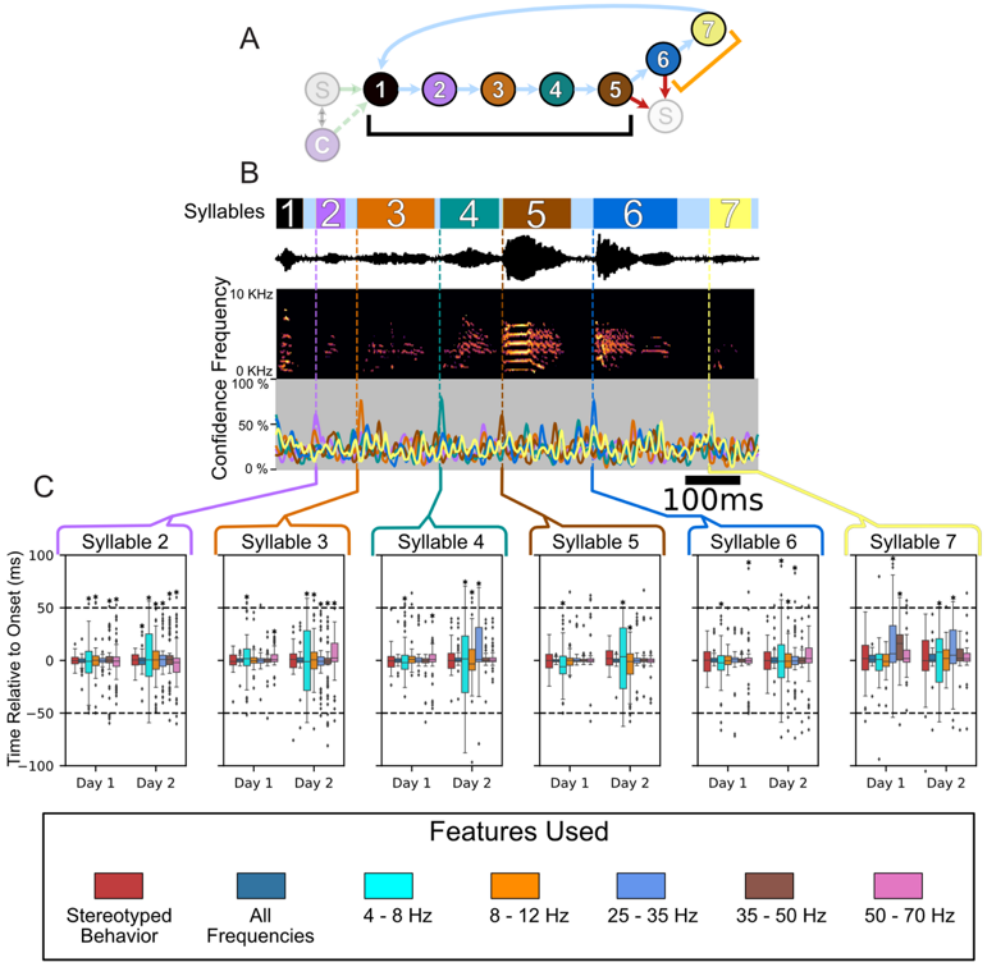


Figure 3.6: Onset detection and branch behavior analysis using LFP features for Subject z017.

(A) State diagram of z017's observed song structure. Syllable colors are the same as in Fig 3. (B) Example motif from the highest yield day for subject z017. Annotated behavior (top) using the same color scheme as in Fig 3, sound pressure waveform and the corresponding time-aligned spectrogram (middle), and the time-varying naïve confidence of the onset prediction (bottom) for each syllable in this example motif. Confidence signal traces are the same color as the syllable they are meant to predict. (C) Boxplot of onset prediction times relative to the labeled onset time for both of the high-yield days for z017. The order of each feature used is the same, going left to right, as shown in Fig 10. The time window that the neural based predictor must make a prediction within is represented by the dotted black line (see Section 3.4.1 **Syllable onset detection**). Statistical significance was calculated using the one-sided Wilcoxon signed-rank test, and * denotes results that are not statistically significant when using the Benjamini-Hochberg False Discovery Rate. All other results $p<0.05$ and $q<0.05$.

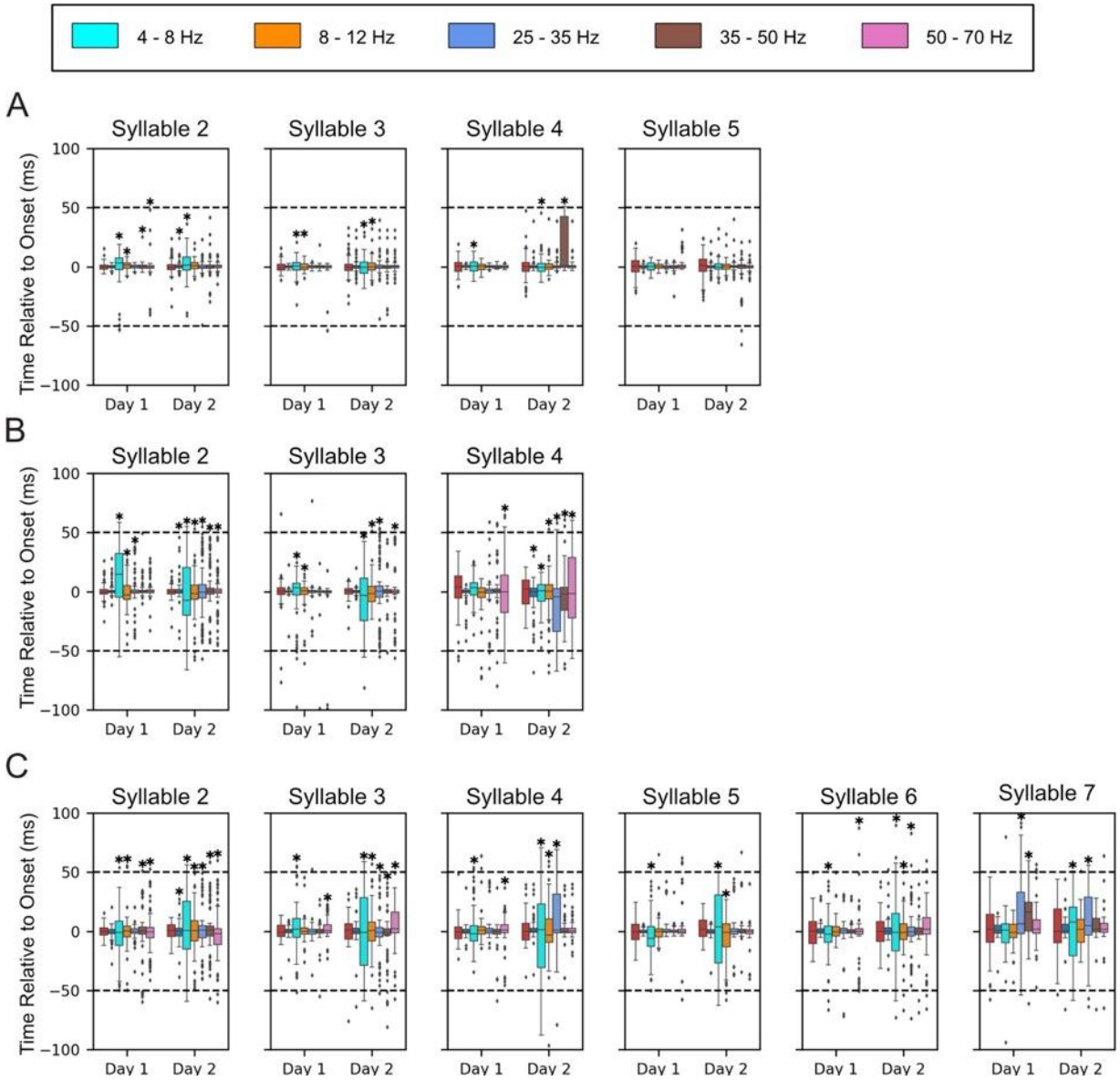


Figure 3.7: Onset detection across all high-yield days.

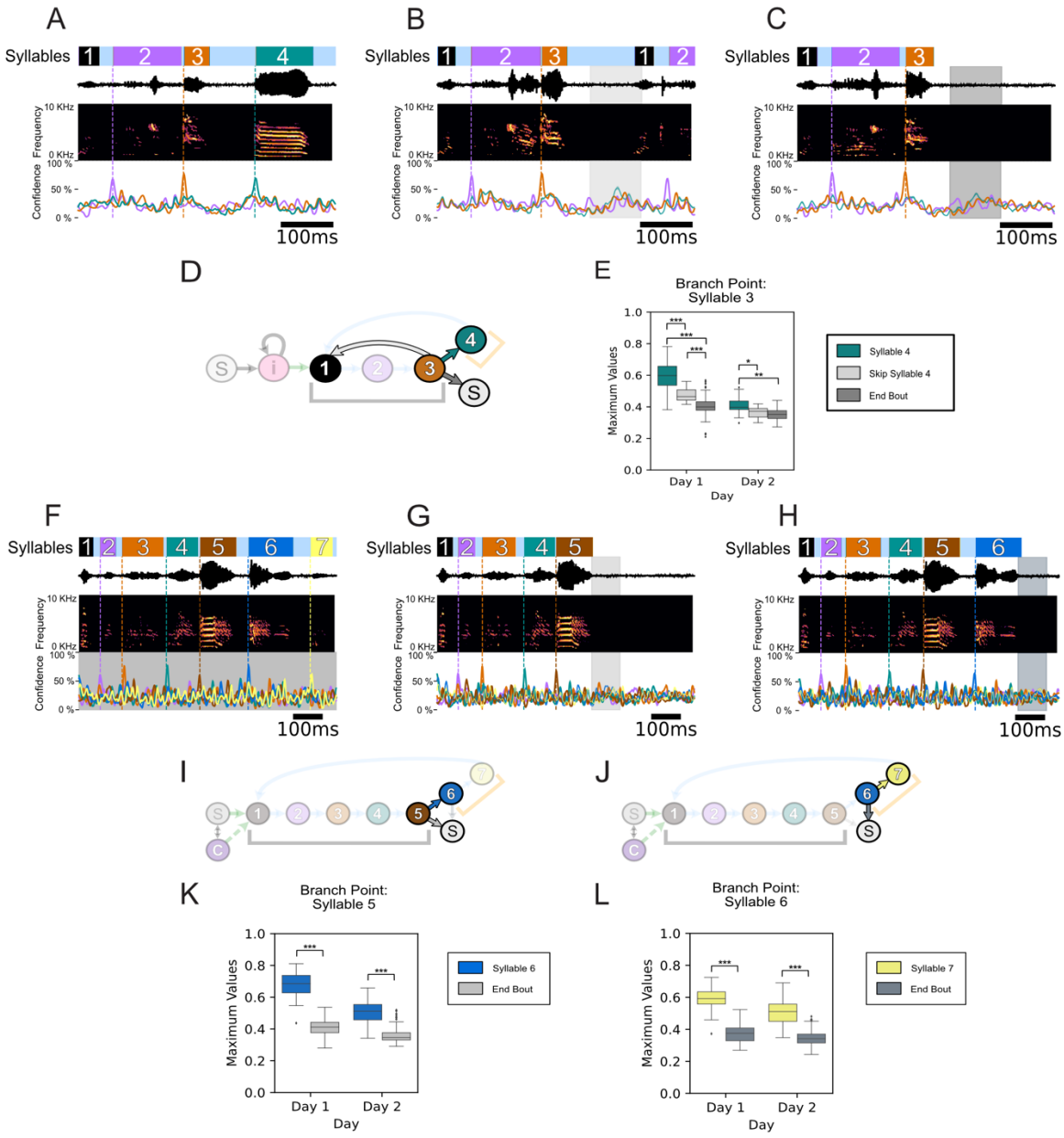
Boxplots of onset prediction times relative to the labeled onset time for each bird for every syllable for two highest-yielding days. Each column reflects the result for syllable number within the motif. Each row is for a specific bird with (A) corresponding to z007, (B) corresponding to z020, and (C) corresponding to z017. The order of each feature used is the same, going left to right: first is the stereotyped onset time using only the deterministic behavior, next is the results using all of the neural features, then each frequency band only in order from least to greatest (4- 8 Hz, 8-12 Hz, 25-35 Hz, 35-50 Hz, and finally 50-70 Hz). The recording day designation number refers to the chronological order that the recordings took place. Statistical significance was calculated using the one-sided Wilcoxon signed-rank test, and * denotes results that were not statistically significant when using the Benjamini-Hochberg False Discovery Rate. All other results $p<0.05$ and $q<0.05$.

We next asked which frequency, if any, might be best at predicting the onset time of the syllables. As demonstrated in Figure 3.4C and Figure 3.7, no single frequency performs best for predicting onset times across all frequencies and all subjects. Although the higher frequencies tend to perform better with less variance than the lower frequencies, using all of the frequencies yields better performance than any one frequency across all subjects and sessions. There were poor prediction performances for 4–8 Hz, 8–12 Hz, and the 50–70 Hz for some syllables for certain birds with multiple examples of performance lower than that of the stereotyped onset.

As much of the sequence of the zebra finch song can be viewed as deterministic when only viewing the syllables of the stereotyped motif, it could be hypothesized that our onset prediction results are not decoding the song identity and are instead only finding the warped time of the syllable within the motif. To further evaluate this possibility, we analyzed the model's performance during the non-deterministic syllables of the birds' behaviors (i.e., their intra-motif notes). As shown in Figure 2.3B, Subject z020 has one behavioral branch point immediately after the third syllable, where it can either transition to its intra-motif note (syllable 4) or omit syllable 4. There are two different behavior options when the subject omits syllable 4; he can either immediately start another motif (skip syllable 4), or terminate the bout (end bout). Thus there are two types of omissions: (1) skip syllable 4 and (2) end bout. An example of the model's single-trial performance for each case is shown in Figure 3.8A–C. As syllable 4 does not occur during every motif, when testing the model's performance on motifs that omit syllable 4 every template from the cross-validation folds is valid; because none of the motifs being tested can be in any of the training sets. Thus, we used each template to independently evaluate performance of each fold. Confidences that were derived from the same template across the different conditions were evaluated against one another. When comparing the distribution of the maximum confidence values during the time

window the model would be required to predict this divergent behavior to occur, its confidence was statistically higher when the syllable occurred versus when it was omitted, grouping the two omission types together, across all five folds. These results were evaluated using the one-sided Welch's t-test for each individual fold. In addition, repeating this test while treating the two syllable omission types as separate distributions yielded similar pair-wise results (Figure 3.8E). This same branch point analysis approach was used to evaluate the performance of subject z017, which has two intra-motif notes (Figure 2.3C) and two different branch points Figure 3.8I and Figure 3.8J. Figure 3.8F–H show exemplar single-trial results that demonstrate performance that is robust to these non-deterministic behaviors. In addition, following the same branch point evaluation approach as previously described, we found when evaluating subject z020's behavioral bifurcations the distribution of confidence values between motifs that contained the intra-motifs notes were significantly higher than when they were omitted (Figure 3.8K and Figure 3.8L).

Figure 3.8: Onset prediction of non-deterministic syllables and decoding divergent behaviors using LFP features.
(A) Example motif with all four syllables from the highest-yield day for subject z020. Annotated behavior (top) using the same color scheme as in Fig 3, sound pressure waveform and the corresponding time-aligned spectrogram (middle), and the time-varying naïve confidence of the onset prediction (bottom) for each syllable in this example motif. (B) Example motif from the same high-yield day as (A) where the subject skips syllable 4 and continues the bout. The shaded lightgray time period is the time window that the behavior-based model would predict that the omitted syllable would occur. (C) Example motif from the same high-yield day as (A) where the subject ends the bout with syllable 3. The shaded darkgray time period is the time window that the behavior-based model would predict that the omitted syllable would occur. (D) State diagram of z020's observed song structure with the three transition types highlighted; 'Syllable 4': syllable 3 to syllable 4 (teal) shown in (A); 'Skip Syllable 4': syllable 3 to syllable 1 (lightgray) shown in (B); 'End Bout': syllable 3 to silence (dark gray) shown in (C). Syllable colors are the same as in Fig 3. (E) Boxplot of the difference in maximum confidence values between the three behavioral transition types shown in (A–C) using the same transition colors used in (D) across both high-yield days. (F) Example motif with all seven syllables from the highest yield day for subject z017. (G) Example motif from the same high yield day as (F) where the subject ends the bout with syllable 5. The shaded lightgray time period is the time window that the behavior-based model would predict the omitted syllable would occur. The prediction trace of the omitted syllables, both syllables 6 and 7, shows the mean across all folds, with shading indicating the standard deviation. (H) Example motif from the same high-yield day as (F) where the subject ends the bout with syllable 6. The shaded slategray time period is the time window that the behavior-based model would predict the omitted syllable would occur. The prediction trace of the omitted syllable, syllable 7 only, shows the mean across all folds with shading indicating the standard deviation across folds. (I) State diagram of z017's observed song structure with the two transition types of the first behavioral branch point highlighted; 'Syllable 6': syllable 5 to syllable 6 (blue) shown in (F); 'End Bout': syllable 5 to silence (lightgray) shown in (G). Syllable colors are the same as in Fig 3. (J) State diagram of z017's observed song structure with the two transition types of the second behavioral branch point highlighted; 'Syllable 6': syllable 6 to syllable 7 (yellow) shown in (F); 'End Bout': syllable 6 to silence (slategray) shown in (G). Syllable colors are the same as in Fig 3. (K) Boxplot of the difference in maximum confidence values between the two transition types, shown in (F) and (G), of the first behavioral branch point using the same transition colors used in (I) across both high-yield days. (L) Boxplot of the difference in maximum confidence values between the two transition types, shown in (F) and (G), of the second behavioral branch point using the same transition colors used in (J) across both high-yield days. Statistical significance was calculated using the one-sided Welch's t-test (***$p < .02$ for all folds; **$p < .05$ for all folds; *$p < .05$ for four out of five folds).


A
B
C
Syllables
Frequency
Confidence
10 KHz
0 KHz
100 %
50 %
0 %
100ms
D
E
Branch Point:
Syllable 3
Maximum Values
Day 1
Day 2
Day
Syllable 4
Skip Syllable 4
End Bout
F
G
H
I
J
K
Branch Point:
Syllable 5
Syllable 6
End Bout
L
Branch Point:
Syllable 6
Syllable 7
End Bout

3.5 Acoustic Synthesis of Bird Song

3.5.1 Overview of Methods

In work done in collaboration with Ezequiel M. Arneodo and Shukai Chen [18], we describe methods for synthesizing realistic vocal signals from neural activity recorded in a premotor nucleus of zebra finches (Taeniopygia guttata). These methods leverage the field's current understanding of the biomechanics of birdsong production. In this work we employ a biomechanical model of the vocal organ that captures much of the spectro-temporal complexity of song in a low-dimensional parameter space [19]. This approach enables training of a shallow feedforward neural network (FFNN) that maps neural activity onto the model parameters. Although both the described symbolic decoding study and this song synthesis study decode behavior from neural activity putatively recorded in sensorimotor nucleus HVC, they use different features of this recordings. In contrast to LFP features applied for symbolic decoding, the song synthesis study employed spiking activity. All subjects described in section 2.2.2 Subjects were also used in this work, with the addition of one subject that wasn't, using recording methods described in sections 2.2.3 Electrophysiology and audio recording–2.2.5 Analysis of electrophysiology data.

To obtain ensemble HVC activity and vocal output, we implanted 16- or 32-channel Si probes in male, adult (>120-day-old) zebra finches and recorded extracellular voltages simultaneously while each bird sang (n = 4 birds, 70–120 vocal motifs per session). Neural recordings were sorted automatically using Kilosort and manually curated to exclude noise [20]. Non-noise clusters were classified as single- or multi-unit activity (SUA or MUA) based on the number of refractory period violations and putatively as projection or interneurons based on the

sparseness of the activity during singing. The recordings were dominated by MUA clusters (n = 88) and HVC interneurons (HVCI; n = 29), with relatively few putative projection neurons (HVCx/RA; n = 15). Example song-aligned neural activity histograms are shown in Figure 3.9A. Example rasters with the numbers of clusters per bird are shown in Figure 3.10.

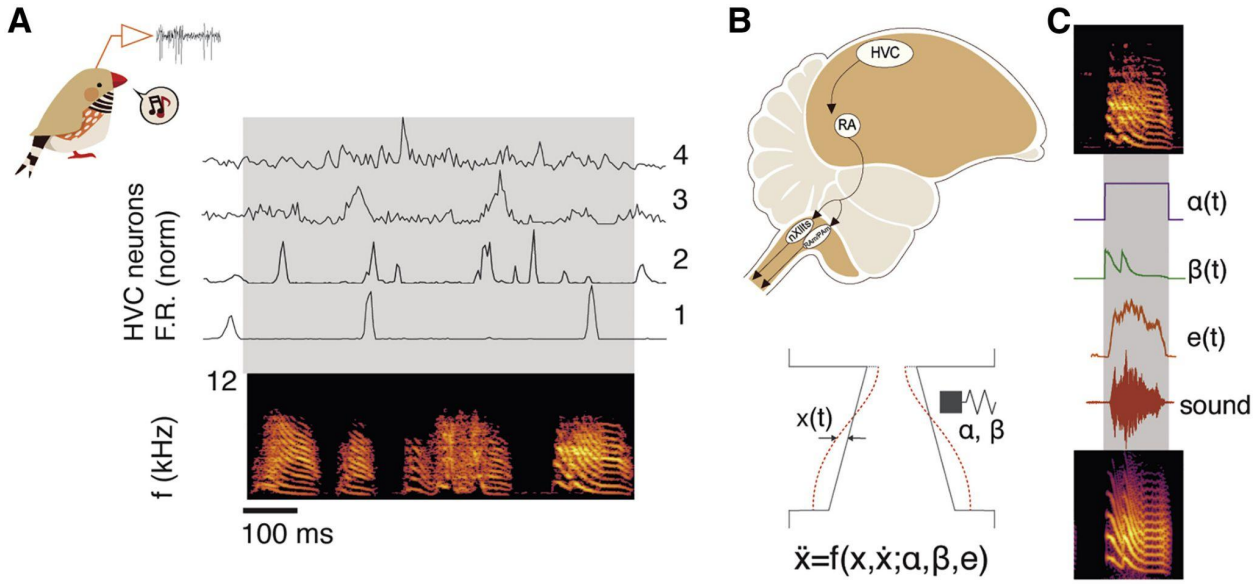


Figure 3.9: A neural-network-based decoder to synthesize birdsong from premotor neural activity

(A) Neural activity is collected from awake-singing animals. Sorted, extracellularly recorded single- and multi-units show different degrees of singing-related sparseness, robustness, and spiking precision (4 example clusters; top traces: normalized mean firing rate over 70 repetitions of the bird's motif; below: spectrogram of the motif; see also Figure 3.10). (B) Downstream of HVC, the posterior motor pathway leads into nuclei that control the muscles driving the sound production (nXII and RAm/ PAm) [21]. Syringeal and respiratory muscles act coordinately to modulate the flow of air through sets of labia and produce sound [22]. The complex labial motion is captured by the equations of a nonlinear oscillator [23]; parameters that define acoustic properties of the sounds are surrogates of the activities of syringeal and respiratory muscles [24]. (C) To reproduce a particular vocalization (top) from the biomechanical model, we fit the parameters (middle {α(t), β(t), e(t)}) such that, upon integration, the synthetic song (bottom) matches the pitch and spectral richness.

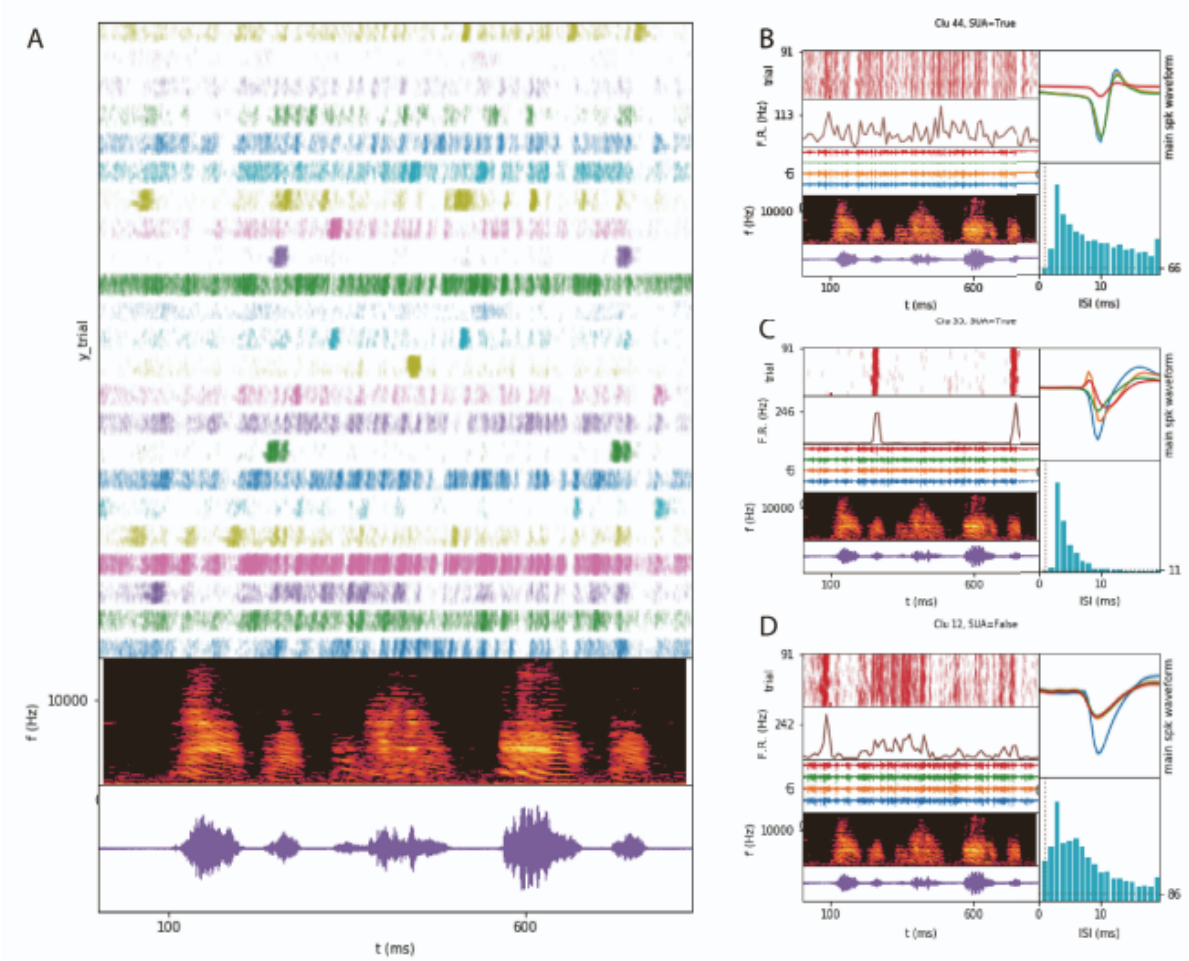


Figure 3.10: Example clusters of sorted units, Related to Figure 1.
(A) Raster plots of 23 automatically clustered units, for 91 repetitions of a motif, spanning 12 hours of recording (top), aligned to an example motif's spectrogram (middle) and waveform (bottom). (B) Example of a (putatively) single unit activity cluster (SUA), likely an interneuron (HVCI). In the left panels are, from bottom to top: the raster; the corresponding histogram (10ms bin); example traces of the 4 neural channels where the cluster's representative waveform has the largest amplitudes; spectrogram of an example motif; waveform of an example motif. In the right panels: (top) a representative waveform for the cluster (mean of 10,000 events), plotted for the 4 channels with the largest amplitude (peak to trough) and (bottom) inter-spike-interval (ISI) histogram (0.5ms bins). Vertical dotted line indicates 1ms, horizontal dotted line indicates 3% level of refractory period violations. (C) Example of a (putatively) SUA cluster, likely a projection neuron (HVCx or HVCRA). (D) Example of a (putatively) multi unit activity cluster (MUA). For this study, we used the highest yield session for each bird (z007: 29 MUA, 11 HVCI, 12 HVCX/RA; z017: 18 MUA, 4 HVCI; z020: 19 MUA, 2 HVCI; z028: 22 MUA, 4 HVCI, 11 HVCX/RA).

3.5.2 Dataset Preparation

3.5.2.1 Spike sorting

Spikes were detected and sorted using Kilosort; details of the procedure can be found in Pachitariu et al [20]. The number of clusters was initialized to 32/64 (twice the number of channels of the probe) and the algorithm was allowed to automatically merge similar clusters. In post hoc curation, we removed the clusters that were visibly noise (as per the waveform) and labeled units as putatively SUA/MUA depending on whether the fraction of refractory period (2ms) violations was below/above 3% respectively

3.5.2.2 Single Unit type classification

Single Unit Activity (SUA) clusters were classified as putatively representing sparse firing projection neurons or tonically firing interneurons, based on their base firing rate and their bursting behavior. We labeled a SUA cluster a putative projection neuron if its mean, spontaneous firing rate was below 5Hz and it produced at most 4 bursts with a frequency of 100 Hz or higher during the motif [25].

3.5.2.3 Neural activity features

With all clusters spike-sorted or supra-threshold events, we extracted spike counts within each motif and collapsed them into 1ms (30 samples at 30,000 samples/second) time bins.

3.5.2.4 Spectral features

When training the networks with spectral features, the target at each time step was a vector containing a spectrogram slice (in log power scale). We generated the spectral slices using the spectrogram function of the signal module in the scipy package [26]. We used 5ms windows (150 samples) and kept the 64 first bands above 300 Hz.

3.5.3 Biomechanical model of the vocal organ

3.5.3.1 Model

A model of the zebra finch vocal organ has been previously introduced and it is explained in detail in Perl et al. [24] and Arneodo et al [27]. This model considers mainly a sound source and a vocal tract that further shapes the acoustics of the vocalizations.

The source (syrinx) comprises two sets of tissues or labia that can oscillate induced by the sub-syringeal pressure and modulate the airflow to produce sound [22]. The motion of the labia is represented as a surface wave propagating in the direction of the airflow, that can be described in terms of the lateral displacement of the midpoint of the tissue [23]. Its mathematical form is the motion equation of a nonlinear oscillator in which two parameters that determine the acoustic features of the solutions are controlled by the bird: the sub-syringeal air sac pressure and the stiffness of the restitution (through the activity of syringeal muscles). In order to integrate the model in real time, a set of equations was found that is computationally less expensive yet capable of displaying topologically equivalent sets of solutions as the parameters are varied: [28]

$$\begin{cases} \frac{dx}{dt} = y \\ \frac{dy}{dt} = \gamma^2\alpha + \gamma^2\beta x + \gamma^2 x^2 - \gamma^2 x^3 - \gamma x y - \gamma x^2 y \end{cases}$$

where x represents the departure of the midpoint position of the oscillating labia, γ is a time scaling factor, and the parameters α and β are functions of the air sac pressure and the activity of the ventral syringeal muscle, respectively. The upper vocal tract further shapes the sound produced by the source, determining spectral properties such as the timbre. We used a model for the vocal that includes a tube, accounting for the trachea, followed by a Helmholtz resonator, accounting for the oropharyngeal-esophageal cavity (OEC) [29,30] (see also Figure 1A in Arneodo et al. [27]). The pressure at the input of the tube that represents the trachea is $P_i(t) = ax(t) - rx(t - \tau)$, where

$ax(t)$ is the contribution to the fluctuations by the modulated airflow, r is the reflection coefficient at the opposing end of the tube of length L and $\tau = 2l/c$, with c the sound velocity. The pressure fluctuations at the output of the trachea force the air at the glottis, approximated by the neck of the Helmholtz resonator that represents the OEC. The mass of air at the glottis, forced into the cavity, is subject to a restitution force exerted by the larger mass of air in it. In acoustics, it is common to write an analog electric computational model to describe a system of filters. The acoustic pressure is represented by an electric potential and the volume flow by the electric current. In this framework, short constrictions are inductors, and cavities (smaller than the wavelengths) are well represented by capacitors. The equations for the equivalent circuit of the post-tracheal part of the vocal tract, (see Figure 1B in Arneodo et al. [27]) read:

$$\begin{cases} \dfrac{di}{dt} = \Omega_1 \\ \dfrac{d\Omega_1}{dt} = \frac{1}{L_g C_h} i_1 - R_h \left(\frac{1}{L_b} + \frac{1}{L_g}\right) \Omega_1 + \; i_3 \left(\frac{1}{L_g C_h} - \frac{R_b R_h}{L_b L_g}\right) + \frac{1}{L_g} \frac{dV_{ext}}{dt} \; + \frac{R_h}{L_g L_b} V_{ext} \\ \dfrac{di_3}{dt} = \dfrac{L_g}{L_b}\, \Omega_1 - \dfrac{R_b}{L_b} i_3 + \dfrac{1}{L_b}\, V_{ext} \end{cases}$$

where the electric components relate to geometric parameters of acoustic elements, and are described in detail in Perl et al [24]. and Arneodo et al [27]. The pressure fluctuations at the glottal end of the trachea relate linearly to the electric tension V_{ext} driving the circuit. Following the same scheme, the electrical potential at the resistor standing for the beak $V_b = \; i_3 R_b$ is the analog of the pressure fluctuations at the output of the beak. In our model, this quantity is the sound radiated by the vocal organ.

3.5.3.1 Parameter fitting

In order to fit the parameter series that will lead to reconstruction of the song, we perform a procedure similar to that previously described [24,31]. Timescale parameter is set to a value of 23,500; α is set to 0.15 during vocalization and 0.15 otherwise, and β is set in order to minimize

the distance in the (pitch, spectral content) space between the synthesized and the recorded song segments [24]; the envelope $\big(e(t)\big)$ is obtained by rectifying and smoothing the recorded waveform; the parameters of the vocal tract were fixed, in the same values as in Perl et al [24]. In order to extract the pitch of the song, we follow a modification of the automatic procedure presented in Boari et al. [32], and we add a layer of manual curation. When integrating the model, we apply the extracted envelope $\big(e(t)\big)$ as an extra multiplicative factor when computing $\big(ax(t)\big)$, since it recovers the amplitude fluctuations that were discarded when reducing the model to its normal form and driving it with the bi-valued parameter α. The parameters accounting for the geometry of the vocal tract are constants and are set to the same values as in Perl et al [19].

3.5.5 Quantification and statistical analysis

3.5.5.1 Performance evaluation root mean square error (RMSE)

We used RMSE between each pair of original and predicted spectrogram magnitude as a metric to evaluate the performance of our models.

3.5.5.2 Spectral correlation

To obtain the spectral correlation across time for a pair of spectrograms, we first computed the pearson correlation coefficient between each corresponding pair of spectral slices that conform the two spectrograms (via the function pearsonr from the stats module of the scipy python package [26]). Then, we obtained the time-averaged value across the span of the motif.

3.5.5.3 Earth mover's distance

To obtain the distance across time for a pair of spectrograms, we computed the earth mover's distance (d_{EMD}) or Wasserstein metric between each pair of spectral slices that conform the two spectrograms (via the function wasserstein_distance from the stats module of the scipy python package [26]). Prior comparison, each spectral slice was normalized such that the total area under the slice be 1; for silences, a value of 1 was assigned to the first bin of the spectrogram. Then, we obtained the time-averaged value across the span of the motif.

3.5.6 Biomechanically meaningful compression enhances neurally driven synthesis

Synthesizing a complex motor sequence from neural activity requires mapping between two high-dimensional representations. To reduce the dimensionality of the problem, we leveraged a biomechanical model of the avian vocal organ that transforms neural activity to vocal output. The model accounts for the syrinx and the vocal tract [19,27,31]. The syrinx contains labial folds that oscillate when induced by the sub-syringeal air sac pressure and modulate the airflow to produce sound (Figure 3.9B) [22]. The dynamics of the labia can be modeled after the motion equations of a nonlinear oscillator, in which the features of the sounds produced are determined by two time-varying parameters [19,23,24], representing physiological motor instructions (the sub-syringeal pressure and the activity of the muscles that tense the labia) [19]. In its simplest form, the syrinx model is computable in real time to produce synthetic vocalizations [27]. We model the vocal tract (the trachea, the oropharyngeal-esophageal cavity, and the beak) as a passive acoustic filter that determines species-specific spectral traits, such as the timbre [19,31,33].

To synthesize song from neural activity via the biomechanical model, we first fit the parameters of the model to produce a synthetic version of each vocalization [19,27,31]. We

searched for the parameters that produce, upon integration of the equations of the model, the closest match in pitch, spectral richness, and amplitude of the target vocalization. This effectively compresses each segment of a bird's own song (BOS) into a time series in a 3D parameter space, which generates a corresponding segment of synthetic song (SYN) (Figure 3.9C) [19,27]. For each session, we randomly select 60% of the motifs for training, split each motif into 5-ms bins, and train a one-hidden-layer FFNN to predict the biomechanical model parameters corresponding to each bin independently from the neural activity in a 50-ms, immediately preceding time window. The neural activity was represented by the average firing rate of each cluster, split into 1-ms bins. To avoid introducing temporal correlations, we randomized the order in which each pair of neural activity window and target model parameters was presented to the network. After training, we predict the values of the biomechanical model parameters corresponding to a test set of neural activity and integrate the differential equations of the model to produce each bin of neurally driven synthetic song. This yields synthetic vocalizations that sound similar to the bird's own. An example motif from each bird is illustrated in Figure 3.11.

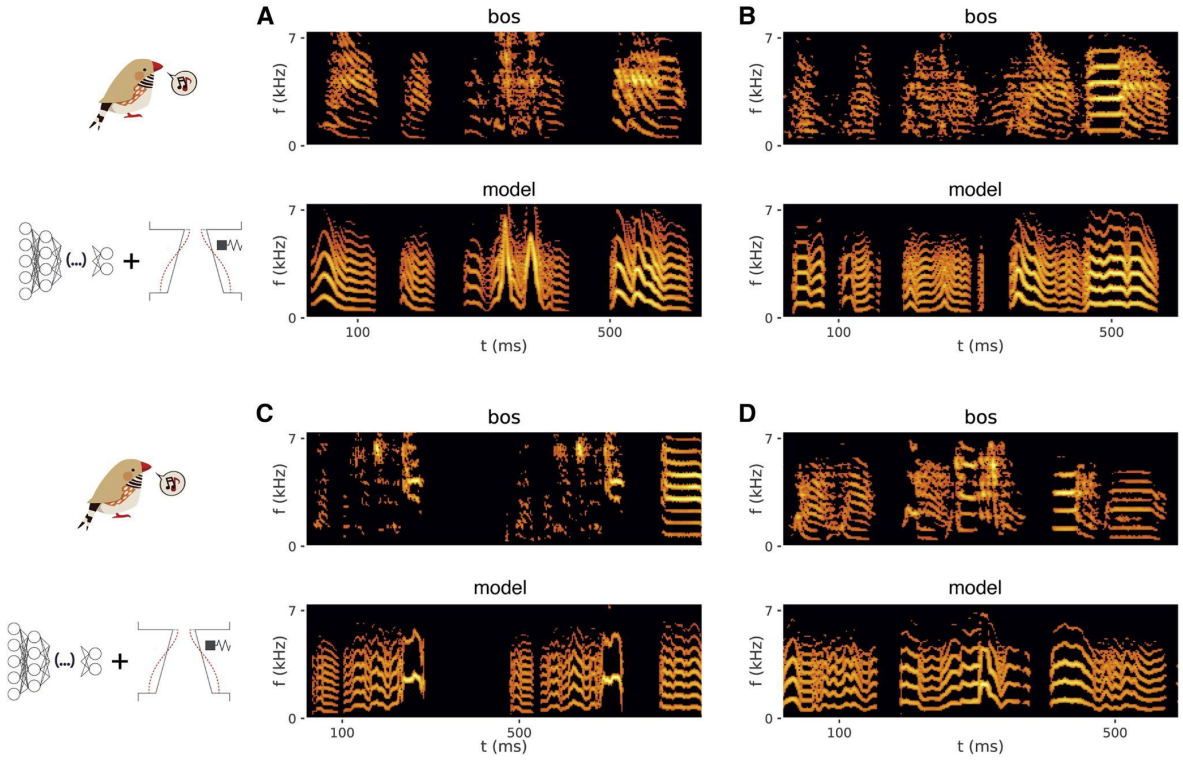


Figure 3.11: Song synthesized from premotor neural activity via a biomechanical model of the vocal organ is similar to the recorded bird's own song.
Spectrogram of one or two instances of a bird's motif (BOS; upper) and corresponding song generated by inferring the biomechanical model parameters from neural activity using a shallow FFNN and integrating the model, for four different birds (z007, z017, z020, and z028, respectively).

3.6 Discussion

3.6.1 Symbolic Decoding

Naturalistic motor behavior is intrinsically difficult to study due to its high degree of variability. Zebra finch represents a desirable model species because they mitigate this difficulty. Their song allows for repeated segments of the same sequence to be produced with two almost opposing characteristics: near perfect precision of each utterance and the natural irregularities that exist in all of nature. This provides a dataset that has both high repetitions of each syllable class and a non-deterministic structure beyond the motif level that facilitates detailed analyses. However this stereotypy alone would not be an adequate model for the complex sequential structure of human speech. Fortunately, there is an additional level of nondeterminism that arises from the intramotif notes which have been found to have various combinations of 'syntactic' rules. These nondeterministic sequences were leveraged in the results described in Figure 3.8, where we show that LFP features predict the presence or absence of intra-motif notes, as well as their onset time (Figure 3.8). These findings suggest that additional vocalizations in the zebra finch repertoire can be used to study motor-vocal control. These behaviors are significantly less deterministic than the syllables of the motif. The stereotyped structure of the zebra finch's song differs from the more complex structure of human speech. Broadening the vocal behaviors that can be used in physiological and neural engineering motivated studies can help mitigate this weakness in the zebra finch model. In addition, the methods and insights learned from these less deterministic vocalizations can be applied to songbirds with more complex song structures such as starlings, canaries, and Bengalese finches. Collectively these songbird models provide an opportunity to investigate vocal-motor encoding at varying levels of behavioral complexity.

We present a unique testbed for approaches to predict the onset of vocalizations in other animals, particularly humans. The onset prediction results in which each syllable beyond the first was predicted from LFP features, described in Figure 3.4, were computed with little-to-no parameter optimization. Better performance could be achieved with more elegant approaches with minimal increases in computational complexity. Furthermore, all computations were conducted with neural activity preceding syllable onset, suggesting that these features could be leveraged to predict syllable onset with low latency.

At present the state of the art for human speech decoding intended for neural prosthesis applications are non-causal, utilizing neural signals before and after the intended vocalization to synthesize sounds similar to those of the intended speech [34–36]. While promising, the non-causal nature of these algorithms introduces significant latency [36] and ultimately limits the quality of interaction with such a prosthesis. Even further, the limits on duration over which neural signals can be studied impedes the pace at which these methods can be established and translated into clinical application. This is where the songbird model can contribute to the development of neural prostheses: providing an animal model for proof-of-concept system development in which closed-loop interaction with prosthesis designs can be rigorously studied. Systems derived and refined in this model system could then be translated to the human clinical setting, in which such rigor and repeatability are more challenging to achieve. These features and this approach provide a starting point for further analyses that look to zebra finch and other songbirds as more than a model for vocal learning and sequence generation, but also as a model of vocal prediction and neural prosthesis development.

3.6.2 Comparing and Contrasting Symbolic Decoding and Acoustic Synthesis and their implications for Human Speech BCIs

In this chapter we present methods and results for both Symbolic Decoding and Acoustic Synthesis of bird song. Every one of the methods described use neural features that precede the production of the vocalization and can be implemented in real-time systems. Although their end goal is the same, to decode the intended vocalization from neural activity, the features they use and the route they take to get there vary. The methods for Symbolic Decoding use LFP which are more readily accessible using invasive recording techniques, while methods for Acoustic Synthesis use neuronal spiking activity which are not as readily accessible but if recorded from many neurons simultaneously can provide higher fidelity readout of brain state than LFP.

Both methods have their advantages and disadvantages. Symbolic Decoding can reconstruct a vocal communication signal with high fidelity using acoustic templates of the symbolic units used. However it is limited by the size of the bank of symbolic units used to train it. Acoustic Synthesis has much higher flexibility and generalizability and could potentially encapsure prosody, emphasis, and tone. These components of the vocal signal are generally lost when utilizing Symbolic Decoding approaches. However Acoustic Synthesis systems require larger datasets and more complex systems architecture to be implemented. In addition, evaluation of their performance is significantly more difficult and many of the currently used metrics may not be functionally relevant.

Another point of departure from the two methods are their evaluation metrics. *Symbolic Decoding* is straightforward: (1) classifier accuracy, (2) Onset prediction error. Classifier accuracy has several well defined metrics that are easily adapted to benchmark Symbolic Decoding work. Onset timing is basically how close to the time of vocalization initialization the decoder predicts the vocalization. However it is important to keep in mind that decoder accuracy is limited by what

percentage of the entire vocal repertoire of the subject is being covered by the classifier. For example for the work presented in this chapter all of the syllables of the song in addition to the introductory note are being classified. However this doesn't represent the full vocal repertoire of the zebra finch. The majority of their vocalizations are loosely grouped together as 'Calls'. So the accuracy of our system is representative of its performance only when the bird is singing and does not describe how accurately the system would perform if run over all periods of time, whether singing, vocalizing calls, or silent. *Acoustic Synthesis* is a bit more complex. The work presented utilized similarity metrics such as Correlation, Root Mean Square Error (RMSE), and Earth mover's distance. Although these metrics have succinct mathematical definitions their relation to the perception of the communication signal produced, the most important performance benchmark, is unclear.

Overall, these methods showcase the versatility of the songbird model for testing proposed methods for human speech BCI research regardless of output strategy. They open new doors for testing and developing proof-of-concept systems architecture before testing them in humans. In addition, given the extensive history of using songbirds for auditory perception research they stand to be an excellent model for developing novel perceptually derived evaluation metrics for Acoustic Synthesis systems. Collectively these works exhibit the utility of the songbird animal model for testing decoding strategies that can be feasibly translated to human speech BCI research.

3.7 Acknowledgment

Chapter 3, in part, has been adapted from material as it appears in PLOS Computational Biology, 2021, “Local Field Potentials in a Pre-motor Region Predict Learned Vocal Sequences,” D. E. Brown II, J. I. Chavez, D. H. Nguyen, A. Kadwory, B. Voytek, E. Arneodo, T. Q. Gentner, and V. Gilja. The dissertation author was the primary investigator and author of this paper. Chapter 3, in part, also contains material as it appears in Current Biology, 2021, E. M. Arneodo, S. Chen, D. E. Brown II, V. Gilja. The dissertation author was the primary author of this chapter.

3.7 References

1. Mugler EM, Tate MC, Livescu K, Templer JW, Goldrick MA, Slutzky MW. Differential Representation of Articulatory Gestures and Phonemes in Precentral and Inferior Frontal Gyri. J Neurosci. 2018;38: 9803–9813. doi:10.1523/JNEUROSCI.1206-18.2018
2. Ramsey NF, Salari E, Aarnoutse EJ, Vansteensel MJ, Bleichner MG, Freudenburg ZV. Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. NeuroImage. 2018;180: 301–311. doi:10.1016/j.neuroimage.2017.10.011
3. Mugler EM, Patton JL, Flint RD, Wright ZA, Schuele SU, Rosenow J, et al. Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng. 2014;11: 035015. doi:10.1088/1741-2560/11/3/035015
4. Lotte F, Brumberg JS, Brunner P, Gunduz A, Ritaccio AL, Guan C, et al. Electrocorticographic representations of segmental features in continuous speech. Front Hum Neurosci. 2015;09. doi:10.3389/fnhum.2015.00097
5. Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic Feature Encoding in Human Superior Temporal Gyrus. Science. 2014;343: 1006–1010. doi:10.1126/science.1245994
6. Kellis S, Miller K, Thomson K, Brown R, House P, Greger B. Decoding spoken words using local field potentials recorded from the cortical surface. J Neural Eng. 2010;7: 056007. doi:10.1088/1741-2560/7/5/056007
7. Moses DA, Leonard MK, Makin JG, Chang EF. Real-time decoding of question-and-answer speech dialogue using human cortical activity. Nat Commun. 2019;10: 3096. doi:10.1038/s41467-019-10994-4
8. Moses DA, Metzger SL, Liu JR, Anumanchipalli GK, Makin JG, Sun PF, et al. Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. N Engl J Med. 2021;385: 217–227. doi:10.1056/NEJMoa2027540
9. Herff C, Heger D, de Pesters A, Telaar D, Brunner P, Schalk G, et al. Brain-to-text: decoding spoken phrases from phone representations in the brain. Front Neurosci. 2015;9. doi:10.3389/fnins.2015.00217
10. Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, Panko M, et al. A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. Ben-Jacob E, editor. PLoS ONE. 2009;4: e8218. doi:10.1371/journal.pone.0008218
11. Angrick M, Herff C, Mugler E, Tate MC, Slutzky MW, Krusienski DJ, et al. Speech synthesis from ECoG using densely connected 3D convolutional neural networks. J Neural Eng. 2019;16: 036019. doi:10.1088/1741-2552/ab0c59
12. Herff C, Diener L, Angrick M, Mugler E, Tate MC, Goldrick MA, et al. Generating Natural, Intelligible Speech From Brain Activity in Motor, Premotor, and Inferior Frontal Cortices. Front Neurosci. 2019;13: 1267. doi:10.3389/fnins.2019.01267
13. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. Nature. 2019;568: 493–498. doi:10.1038/s41586-019-1119-1
14. Angrick M, Ottenhoff MC, Diener L, Ivucic D, Ivucic G, Goulis S, et al. Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. Commun Biol. 2021;4: 1055. doi:10.1038/s42003-021-02578-0
15. Brown DE, Chavez JI, Nguyen DH, Kadwory A, Voytek B, Arneodo EM, et al. Local field potentials in a pre-motor region predict learned vocal sequences. Theunissen FE, editor. PLOS Comput Biol. 2021;17: e1008100. doi:10.1371/journal.pcbi.1008100

16. Ludwig KA, Miriani RM, Langhals NB, Joseph MD, Anderson DJ, Kipke DR. Using a Common Average Reference to Improve Cortical Neuron Recordings From Microelectrode Arrays. J Neurophysiol. 2009;101: 1679–1689. doi:10.1152/jn.90989.2008
17. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Mach Learn PYTHON. 2011; 6.
18. Arneodo EM, Chen S, Brown DE, Gilja V, Gentner TQ. Neurally driven synthesis of learned, complex vocalizations. Curr Biol. 2021; S0960982221007338. doi:10.1016/j.cub.2021.05.035
19. Perl YS, Arneodo EM, Amador A, Goller F, Mindlin GB. Reconstruction of physiological instructions from Zebra finch song. Phys Rev E. 2011;84: 051909. doi:10.1103/PhysRevE.84.051909
20. Pachitariu M, Steinmetz NA, Kadir SN, Carandini M, Harris KD. Fast and accurate spike sorting of high-channel count probes with KiloSort. Adv Neural Inf Process Syst. 2016;29: 4448–4456.
21. Wild JM. Neural pathways for the control of birdsong production. J Neurobiol. 1997;33: 653–670. doi:10.1002/(SICI)1097-4695(19971105)33:5<653::AID-NEU11>3.0.CO;2-A
22. Goller F, Larsen ON. A new mechanism of sound generation in songbirds. Proc Natl Acad Sci. 1997;94: 14787–14791. doi:10.1073/pnas.94.26.14787
23. Titze IR. The physics of small-amplitude oscillation of the vocal folds. J Acoust Soc Am. 1988;83: 1536–1552. doi:10.1121/1.395910
24. Perl YS, Arneodo EM, Amador A, Mindlin GB. NONLINEAR DYNAMICS AND THE SYNTHESIS OF ZEBRA FINCH SONG. Int J Bifurc Chaos. 2012;22: 1250235. doi:10.1142/S0218127412502355
25. Kozhevnikov AA, Fee MS. Singing-Related Activity of Identified HVC Neurons in the Zebra Finch. J Neurophysiol. 2007;97: 4271–4283. doi:10.1152/jn.00952.2006
26. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17: 261–272. doi:10.1038/s41592-019-0686-2
27. Arneodo EM, Perl YS, Goller F, Mindlin GB. Prosthetic Avian Vocal Organ Controlled by a Freely Behaving Bird Based on a Low Dimensional Model of the Biomechanical Periphery. Ayers J, editor. PLoS Comput Biol. 2012;8: e1002546. doi:10.1371/journal.pcbi.1002546
28. Sitt JD, Amador A, Goller F, Mindlin GB. Dynamical origin of spectrally rich vocalizations in birdsong. Phys Rev E. 2008;78: 011905. doi:10.1103/PhysRevE.78.011905
29. Riede T, Suthers RA, Fletcher NH, Blevins WE. Songbirds tune their vocal tract to the fundamental frequency of their song. Proc Natl Acad Sci. 2006;103: 5543–5548. doi:10.1073/pnas.0601262103
30. Riede T, Schilling N, Goller F. The acoustic effect of vocal tract adjustments in zebra finches. J Comp Physiol A. 2013;199: 57–69. doi:10.1007/s00359-012-0768-4
31. Amador A, Perl YS, Mindlin GB, Margoliash D. Elemental gesture dynamics are encoded by song premotor cortical neurons. Nature. 2013;495: 59–64. doi:10.1038/nature11967
32. Boari S, Perl YS, Amador A, Margoliash D, Mindlin GB. Automatic reconstruction of physiological gestures used in a model of birdsong production. J Neurophysiol. 2015;114: 2912–2922. doi:10.1152/jn.00385.2015

33. Lohr B, Dooling RJ. Detection of changes in timbre and harmonicity in complex sounds by zebra finches (Taeniopygia guttata) and budgerigars (Melopsittacus undulatus). J Comp Psychol. 1998;112: 36–47. doi:10.1037/0735-7036.112.1.36
34. Bouchard KE, Chang EF. Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Chicago, IL: IEEE; 2014. pp. 6782–6785. doi:10.1109/EMBC.2014.6945185
35. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. Nature. 2019;568: 493–498. doi:10.1038/s41586-019-1119-1
36. G. Makin J, Edward F. Chang, A. Moses D. Machine translation of cortical activity to text with an encoder-decoder framework. : 22.

Chapter 4 : Dynamic Counterbalance to Enable Chronic Invasive Electrophysiology Studies in Small Animals

4.1 Abstract

Both basic science and translational medical research are highly dependent on small animal models. In fact 95% of all lab animals are mice and rats, according to the Foundation for Biomedical Research (FBR)[5]. This is especially true for neurotechnology research. However, working with small animals has several technological hurdles. One of the most difficult engineering hurdles is the small size and low weight tolerance of small animal models, which not only impedes the rate of scientific progress but also reduces the types of experiments that are possible to conduct in both an experimental and clinical setting. At present the weight of recording instruments used in small animal research is addressed by using a simple counterweight system. While great in theory this approach both doubles the inertia on the animal making it harder for them to move and creates a force that can stress the subject, thus, introducing behavioral confounds into the data being collected. We have created a dynamic counterbalance system that not only reduces the inertia of instrumented subjects but also reduces the forces experienced by the animals. Not only will this potentially reduce the stress on the animals used in research but also increase the total weight of instruments that can be used safely in these types of research paradigms. This will not only improve data acquisition of existing experiment paradigms by removing the stress due to instrument weight, but also enable novel experimental protocols that were not feasible due to current weight constraints.

[5] https://fbresearch.org/biomedical-research/

4.2 Problem Statement

Utilizing small animal models for long term chronic experiments involves an inherent hurdle of accommodating for their limited weight bearing abilities. Current solutions either require employing simple mechanical solutions — i.e. simple counterweight systems, commutators, and rubber bands—or utilizing cutting edge lightweight wireless technology. The current gold standard counterweight approach creates new problems, namely doubling the effective inertia experienced by the test subject and creating corrective forces that try to move the animal back to the center of their enclosure. These forces and reduced mobility are thought to stress the animals, thus negatively impacting data collection and reducing the time that experiments can be safely conducted. The lightweight technology approach is still not ideal due to both limits in both existing fabrication techniques and current wireless technology. For high throughput recording wired recording systems are still better suited for many chronic recording instruments. There exists a need for a dynamic counterbalance system that can counterbalance the weight of recording instruments without creating restorative forces that will unnecessarily stress the recording subject.

4.3 Background

Small animal models (e.g. mice, rats, and songbirds) are heavily employed throughout medical research. They are employed to study behaviors and biological phenomena that are difficult to study in humans or larger animal models (e.g. non-human primates, pigs, etc). Some of the major strengths of small animal models are their size and availability, as well as the accessibility of reliable scientific tools to study them at multiple scales. However, for those who study complex behaviors this small size creates limits on what instruments can be used to record

vital data from their subjects. The problem is simple: small animals have a limit to how much additional weight they can bear. Scientists and engineers have largely addressed this problem using three methods: (1) Head fixing the animal, (2) Miniaturizing instruments so that they weigh less, (3) Counterweighting the recording instruments.

(1) Head Fixing overcomes the weight issue by having recording instruments self-supported and attaching the animal to the recording rig. The animal is held in place and any movement it makes does not cause it to physically move and instead moves a part of the recording setup. Although the animal is freely moving this behavior does not fully encapsulate all of the behavior researchers wish to understand and the animal cannot safely be continuously recorded for days without long resting periods.

(2) Miniaturizing instruments directly addresses the weight problem by reducing the size and weight of the recording instruments. However, there is a fundamental limit to how small and how light recording instruments can be made even with continuous improvements in fabrication techniques.

(3) Counterweighting reduces the stress the subject experiences from the weight the recording instruments by tying a weight of equal or slightly lesser weight. This is often the method of choice for researchers who conduct chronic recordings of free behavior in small animals.

Both (2) and (3) are the current best methods for conducting long chronic scientific research with small animal models. However they both have major weaknesses. Only (3) reduces the weight the animal has to bear from the recording instruments, and it does so at the cost of doubling the experienced inertia the animal has to contend with. This means that the animal has to try twice as hard to move with the instrument and the counterweight than if it had to move with the weight of the recording instruments alone. What's more, counterweights often only balance properly at a

single point in the cage and produces a slight pull towards that equilibrium point at any other point in the cage the animal tries to move. The field needs a method that will address the recording instrument weight issue without causing additional stress to the animal.

To illustrate how significant the weight issue is for the field I will use our own experimental setup as an example. Our recordings use Neuropixels probes to record neural activity from awake free-behaving zebra finches. Neuropixels, which are the current state of the art for neural recordings, weigh about 1.8 grams and are impressive in terms of signal count and signal quality for their small size. They are the result of years of improvements in fabrication techniques for miniaturizing electronics. However a large male zebra finch may weigh at most 16 grams, which means the neural pixel alone weighs about 11.25% of the subject's weight. This is the equivalent of attaching a 20.25 lbs weight to the top of a 180 lb man. This estimate doesn't include the weight of the protective headcap that is installed to protect the probe and the bird, the dental cement used to secure the headcap, or the electrical tether that connects the probe to the recording system.

4.3 Solution

Here we have developed a dynamic counterbalance system using a Variable Radius pulley [1], a miniature compound pulley system, a dynamic adjustment arm, and a dynamic pulley. The designed parts are capable of being 3D printed and are significantly cheaper to make than existing technologies often deployed in science and industry research. (Note: currently most teams only deploy commutators that only address concerns regarding the electrical tether connected to the recording instruments and Do Not address the weight of the system).

Chapter 4: Dynamic Counterbalance

Variable radius pulleys (VRP) are mechanical systems that convert the linearly increasing force of a spring into a constant force. This is done by adjusting the radius of one side of the pulley to change such that the torque on one end stays constant as the other linearly increases with the spring. These mechanical systems have found use in medical robotics and are significantly smaller than constant force pulleys. The Variable radius Pulley developed for this invention can be fine-tuned at the design stage, using inhouse written python scripts, to counterbalance a specified weight. The first prototype was designed to counterbalance a weight of 1.8 grams.

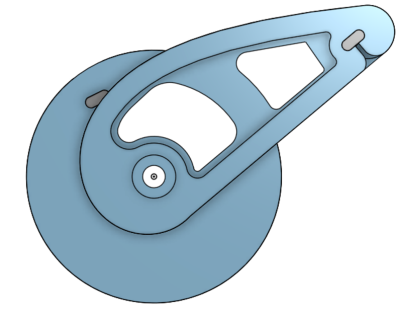


Figure 4.1: CAD Rendering of Variable Radius Pulley

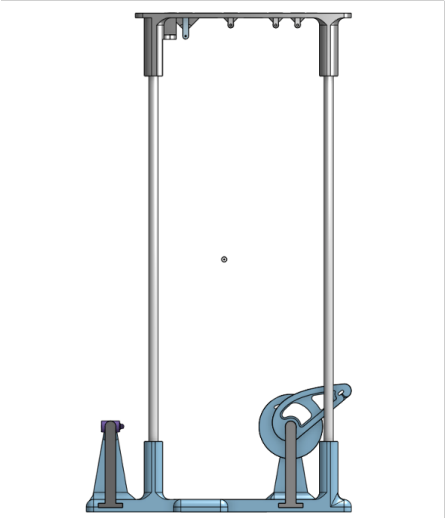


Figure 4.2: CAD Rendering of counterbalance system (Compound Pulley System and Spring Excluded)

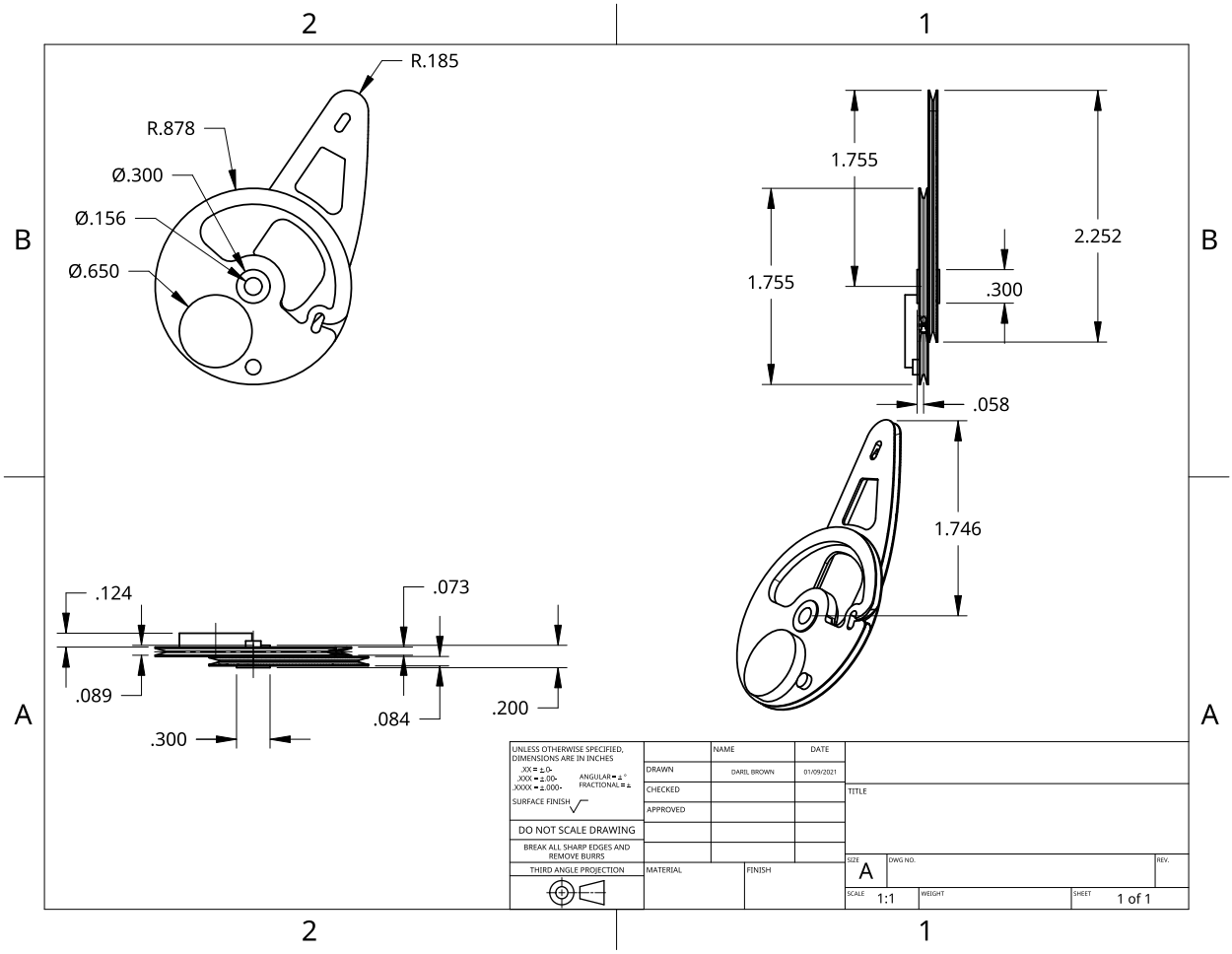


Figure 4.3: Engineering Drawing of VRP

As most commercial springs are too strong to use at the regime that small animal model recording instruments weigh, we developed custom light weight pulley units to create a compound pulley system. This compound pulley system is used connected in reverse such that the system is doing work on the VRP & Spring instead of the animal subject. Pulley systems allow one to do more work by reducing the force necessary to move an object by increasing the distance you must pulley to move it.

This approach gives our system two benefits: (1) it allows us to use commercially available springs for our system (2) it increases the distance our system is able to counterbalance a weight. The supportive structure used to support both the VRP and the compound pulley system is compact and designed to fit within more recording rigs employed in both academia and industry.

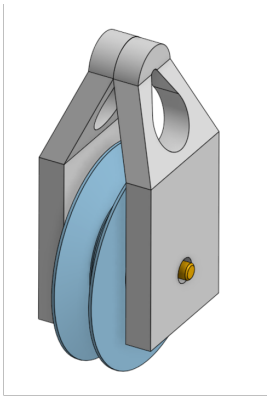


Figure 4.4: CAD Rendering of an individual Pulley Unit

To address the problem of only having only one point in the animal enclosure being balanced we developed a dynamic adjustment arm that inverts the problem. Typically, when using a counterweight system the equilibrium point is the center of the animal enclosure, however for most small prey animals the center of the enclosure is the least desirable location for them. The dynamic adjustment arm makes it so that most of the enclosure will have the system balanced and only at the center or the extreme corners of the enclosure will have a slight off angle imbalance. The adjustment arm has a bore hole through it which allows any commercially available electrical tether to pass through it. This means that the system is compatible with all active and passive commutator systems available on the market. In addition, the adjustment arm can work with either the counterbalance or with a traditional counterweight.

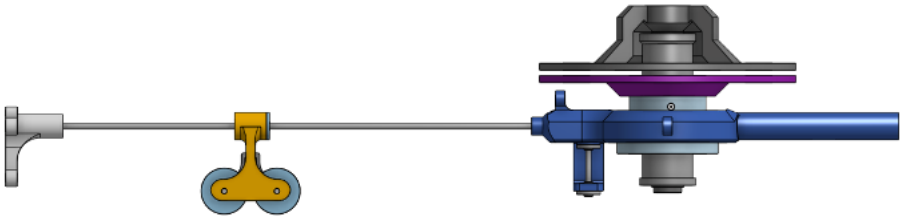


Figure 4.5: CAD Rendering of Adjustment Arm

For highly mobile small animals such as songbirds it is necessary to prevent the tethers — the electrical wire which connects to the recording system and the mechanical wire which connects to the VRP — from wrapping around each other in a way that limits the animals mobility. Having two wires that move independently of each other means that there is no way the tethers will not wrap around each other, so we designed a lightweight dynamic pulley system that will allow the two tethers to wrap each other without compromising either tether's functionality or limiting the mobility of the animal subject. This dynamic pulley system can be 3D printed and weighs under one gram.

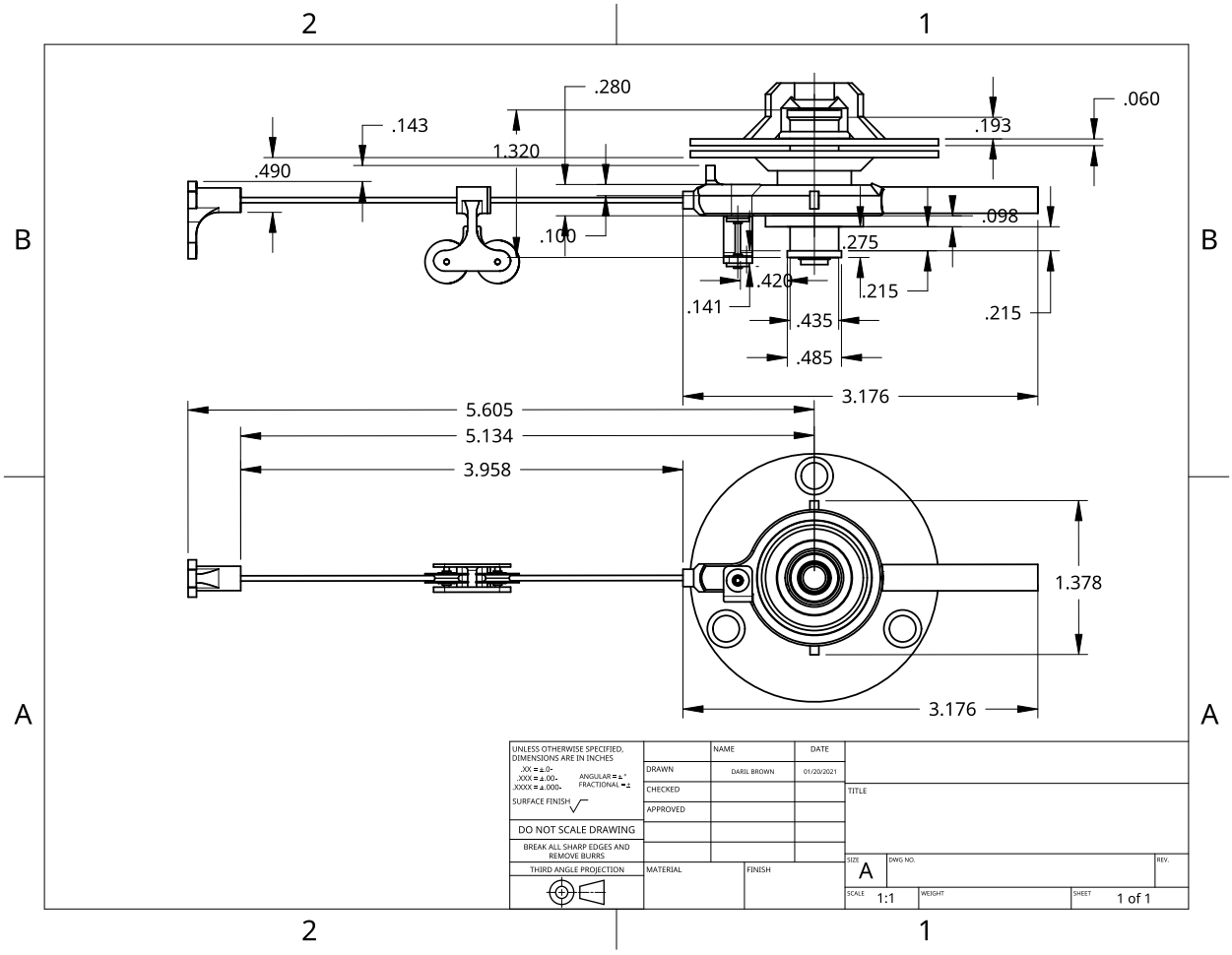


Figure 4.6: Engineering Drawing of Adjustment Arm

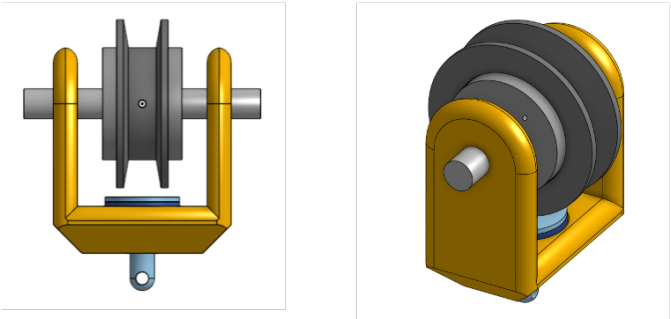


Figure 4.7: CAD Rendering of Dynamic Pulley Unit

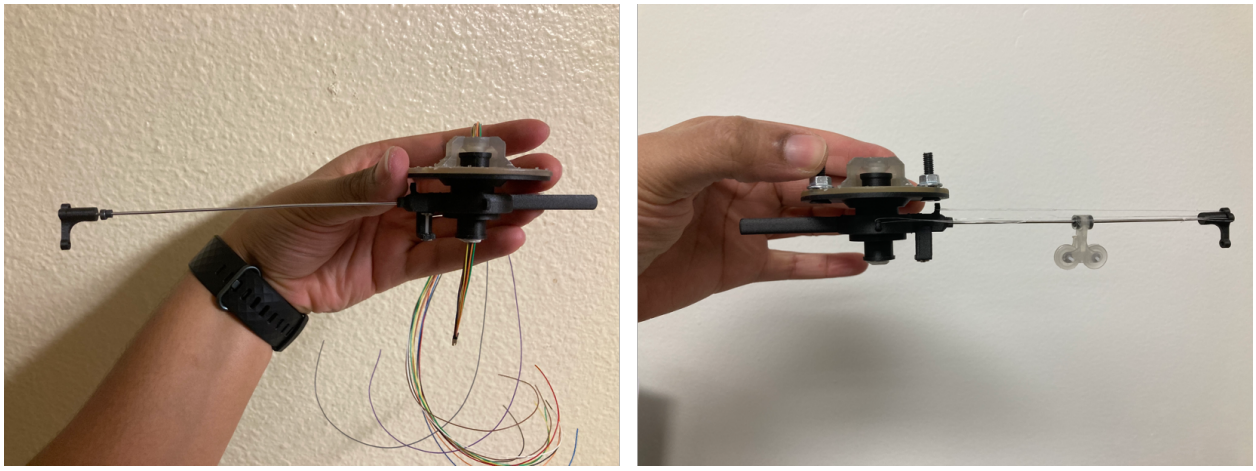


Figure 4.8: Prototyped of Adjustment arm.
(Left) Adjustment arm prototype with example electrophysiology cable passing through the center of rotation. (Right) Adjustment arm with its' traveling pulley resting in the center of the arm.

Figure 4.9: Prototype of full counterbalance system.
The full prototype of the dynamic counterbalance system. Shown is the counterbalance casing (Black), the compound pulley system which comprises of 5 individual pulley units (Center and clear in color), the VRP (lower right and clear in color), and the spring (Lower left and silver in color). The weight suspended to the left weighs 1.8 grams.

4.4 Conclusion

Together the dynamic counterbalance system, dynamic adjustment arm, and dynamic pulley provide a low cost system that addresses several of the pressing weight constraints of chronic small animal experiments. Together they will enable researchers to conduct experiments that would not have previously been possible due to weight restrictions of existing recording instruments.

4.5 Discussion

Initially there will be a big barrier to entry for neuroengineering researchers who wish to begin working with songbirds. These barriers will not only prevent interested researchers from incorporating songbirds into their work, but also slow the wider adoption of the animal model by the field. This delayed adoption translates to slower progress in research advances and impedes the translation of such findings into clinically viable devices that would eventually help patients who greatly need them now. By leveraging my mechanical engineering skills to solve these major pain points, documenting the solutions, and making them readily available for others I aim to help lower these barriers to entry and help speed up the adoption of the songbird animal model. This also makes research findings easier to reproduce by using a common framework for conducting free behavior experiments in small animals. These actions can potentially set a standard for the field that will undoubtedly pay dividends for the acceleration of speech prosthesis development.

4.6 Acknowledgement

The variable radius pulley was developed in collaboration with Eugene Duval who has previously patented methods for designing Variable radius pulleys [2], however the inventions presented in this white paper work at a weight regime that is currently not addressed in Industry.

Chapter 4, in full, has been submitted in a provisionary patent, 2021, “Dynamic Counterbalance to Enable Chronic Free-behaving Research with Small Animals”, Daril EVan Brown II, Eugene Duval, Derek Hung Nguyen, Vikash Gilja, Timothy Q. Gentner. The dissertation author was the primary author of the white paper that informed the provisionary patent applicant and the primary author of this chapter.

4.7 References

1. Yigit CB, Bayraktar E, Boyraz P. Low-cost variable stiffness joint design using translational variable radius pulleys. Mech Mach Theory. 2018;130: 203–219. doi:10.1016/j.mechmachtheory.2018.08.006
2. Eugene F. Duval. Dual Pulley Constant Force Mechanism. 7,677,540B1, 2010.

Chapter 5 : Conclusion

5.1 Summary of Contributions

This thesis develops songbirds as an animal model to supplement and help advance human speech BCI research. Chapter 1 sets the context for this work by describing the current state of the field and the hurdles impeding its continued progress. I describe the two main target outputs modalities for speech BCIs, namely *Acoustic Synthesis* and *Symbolic Decoding*. I also give a brief overview of the historic benefit of having a translatable animal model, such as the rhesus macaque, which accelerates progress in the motor limb BCI field. That chapter presents our initial motivations for targeting the songbird as a potential animal model to fill the needs of the speech BCI field and describes what characteristics must be verified to justify this approach.

Chapter 2 helps validate the songbird animal model by characterizing a nucleus crucial for song production and elucidating several LFP features that are qualitatively similar to neural features known to be useful for speech decoding in human and non-human mammalian motor cortex. These features were found to be correlated with vocal production and are modulated prior to the onset of vocal units. Volume conductance signals, such as LFP, are more readily accessible than action potentials and thus the bulk of human speech BCI research uses volume conductance signal features as inputs for BCIs. Prior to this work there was limited knowledge about the characteristics of LFP activity in zebra finch HVC. This work fills the gaps in the literature and helps strengthen the argument for using songbirds to supplement human vocal production BCI research.

Chapter 3 demonstrates several methods for decoding songbird vocal activity from neural activity that precedes vocal onset. The first two methods, which are components of a Symbolic

Decoding approaches, utilizes the LFP features first described in chapter 2. The first method decodes the syllable identity and the second method predicts vocal onset. In addition, to further validate the zebra finch model we leveraged the nondeterministic intra-motif note to test the decoder's generalizability. The final method described comes from work done in collaboration with lab mates and developed Acoustic Synthesis systems for bird song. Together these methods demonstrate the versatility of the songbird animal model for testing and prototyping approaches that can feasibly be translated to human clinical speech BCI research.

Finally, in Chapter 4 I document mechanical designs for an integrated counterbalance and tether management system that lowers the stress on the subject caused by the weight of the recording equipment. This is a major pain point for conducting long duration chronic free behaving experiments in songbirds. Although designed with songbird in mind, the system is species agnostic and can easily be adapted to most small animal research such as mice and rats. Contributions such as these are crucial for long term invasive experiments with small animals where wireless recordings can be unfeasible.

Collectively, these neurophysiology and neural decoding contributions enrich the literature connecting human and avian vocal-motor production. They demonstrate proof-of-concept systems that can be translated to humans. Opening new opportunities to ask basic science questions regarding vocal motor control as well as prototyping novel proof-of-concept vocal motor BCI systems, and thus, bringing the lofty goal of this thesis closer to reality.

5.2 Contextualizing the Songbird Animal Model in the Broader Speech BCI Research Field

The development of high performance speech BCI will require invasive studies in humans. The songbird model is not intended to replace such human work, but supplement it. Despite anatomical differences in the musculature that produce their vocalizations, songbirds and humans have a few key similarities that strengthen the argument for this approach. Notably songbirds have similar auditory requirements to Humans during vocal production [1–3]. Just like humans, when songbirds sing they are constantly listening to themselves to evaluate their vocal performance. Implementing any of the song decoding methods documented in this thesis in real-time would allow researchers to ask basic science questions regarding the design and performance of closed loop BCI systems. A range of questions can be explored with this model system, such as, (1) what are the salient aspects of intended behavior that the BCI system must produce for the subject to be satisfied, and (2) what are the basic system requirements to implement a basic BCI for real-time online applications. The results of the first questions could be used to develop physiologically relevant evaluation metrics for Acoustic Synthesis systems or to help define an adequate set of symbols for effective Symbolic Decoding Systems. The second question can inform system architecture for human speech BCI systems. This approach of asking basic science questions in the songbird animal model then leveraging newly gained knowledge to inform human experiments could help accelerate the rate of progress for the field.

5.3 Future Work and Potential Directions

Most of the work discussed in this thesis focused on the learned vocalizations of the zebra finch, namely song syllables and introductory notes. However, this is just a small portion of the zebra finch's vocal repertoire. The majority of the zebra finch's vocal behavior consists of vocalizations collectively referred to as 'calls'. Some calls are learned and others are not [4–6]. While previous work has found that neural activity in pre-motor nuclei (i.e HVC and RA) encodes information regarding which syllable the bird will sing [7–9], few studies focus on elucidating what activity, if any, relates to call production [4–6]. Future research that would elucidate the contribution, if any, of HVC and RA to producing Calls and leveraging them for vocal decoding would significantly strengthen the songbird animal model. Our lab has found some preliminary success investigating this question, however, this preliminary work while initially promising is not mature enough to be included in this thesis.

Finally, future efforts should broaden songbird BCI research by incorporating songbirds with greater song structure complexity; namely Starlings and Canaries. This thesis focuses on Zebra Finch for good reason, their songs have much simpler sequential structure than Starlings which are lifelong learners. This simpler, though still relevant, song structure of the zebra finch allowed us to verify the feasibility of the proposed approach. Now equipped with the preliminary work described in this thesis, future work should reproduce these findings in other songbird species. In doing so it will ensure that the songbird animal model fulfills its full potential for supplementing human speech BCI research. By this I mean having a suite of songbird animal models with varying degrees of complexity of song structure to choose from when designing

research experiments. This will equip researchers with the means of studying vocal motor production with a previously unimaginable degree of control.

It is my hope in writing and completing this thesis that a flag has been set and a call to arms is heard by the field. There is a lot of uncharted intellectual territory to be explored with this approach, and it could potentially help a lot of people.

5.4 Reference

1. Brainard MS, Doupe AJ. Auditory feedback in learning and maintenance of vocal behaviour. Nat Rev Neurosci. 2000;1: 31–40. doi:10.1038/35036205
2. Keller GB, Hahnloser RHR. Neural processing of auditory feedback during vocal practice in a songbird. Nature. 2009;457: 187–190. doi:10.1038/nature07467
3. Sakata JT, Brainard MS. Real-Time Contributions of Auditory Feedback to Avian Vocal Motor Control. J Neurosci. 2006;26: 9619–9628. doi:10.1523/JNEUROSCI.2027-06.2006
4. Elie JE, Theunissen FE. The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. Anim Cogn. 2016;19: 285–315. doi:10.1007/s10071-015-0933-6
5. Elie JE, Soula HA, Mathevon N, Vignal C. Dynamics of communal vocalizations in a social songbird, the zebra finch (Taeniopygia guttata). J Acoust Soc Am. 2011;129: 4037–4046. doi:10.1121/1.3570959
6. Elie JE, Mariette MM, Soula HA, Griffith SC, Mathevon N, Vignal C. Vocal communication at the nest between mates in wild zebra finches: a private vocal duet? Anim Behav. 2010;80: 597–605. doi:10.1016/j.anbehav.2010.06.003
7. Brown DE, Chavez JI, Nguyen DH, Kadwory A, Voytek B, Arneodo E, et al. Local Field Potentials in a Pre-motor Region Predict Learned Vocal Sequences. Neuroscience; 2020 Jun. doi:10.1101/2020.06.30.179861
8. Arneodo EM, Chen S, Brown DE, Gilja V, Gentner TQ. Neurally driven synthesis of learned, complex vocalizations. Curr Biol. 2021; S0960982221007338. doi:10.1016/j.cub.2021.05.035
9. Liberti WA, Markowitz JE, Perkins LN, Liberti DC, Leman DP, Guitchounts G, et al. Unstable neurons underlie a stable learned behavior. Nat Neurosci. 2016;19: 1665–1671. doi:10.1038/nn.4405