

Prediction and Evaluation of Vocal Performance in the Songbird Auditory Cortex

by

Fabiola Duarte Ortiz

Department of Neurobiology  
Duke University

Defense Date: November 22<sup>nd</sup>, 2024

Approved:

Richard Mooney, Advisor

Lindsey Glickfeld, Co-Chair

Marc Sommer, Co-Chair

Steven J. Eliades

Kevin Franks

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor  
of Philosophy in the Department of Neurobiology in The Graduate School of  
Duke University  
2024

ABSTRACT

Prediction and Evaluation of Vocal Performance in the Songbird Auditory Cortex

by

Fabiola Duarte Ortiz

Department of Neurobiology  
Duke University

Defense Date: November 22<sup>nd</sup>, 2024

Approved:

Richard Mooney, Advisor

Lindsey Glickfeld, Co-Chair

Marc Sommer, Co-Chair

Steven J. Eliades

Kevin Franks

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in the Department of Neurobiology in The Graduate School of  
Duke University  
2024

Copyright by  
Fabiola Duarte Ortiz  
2024

## Abstract

Learning and maintaining complex vocalizations requires the brain to generate a vocal motor program that matches an auditory target. One possibility is that this matching process depends on neural circuits that compare vocal auditory feedback to an internal representation or model of the auditory target, and differences in these two representations result in an error signal that is used to modify the vocal motor program. This comparison is thought to be facilitated by vocal-motor or corollary discharge signals that suppress auditory cortical responses to predictable features of vocalization-related auditory feedback. Yet whether predictive suppression characterizes the auditory cortex in animals that produce learned vocalizations remains poorly understood.

I explored this issue in freely behaving adult male zebra finches, which use auditory feedback to learn and maintain their courtship songs. I used miniature microscopes and genetically encoded calcium indicators to measure how vocal-motor signals modulate auditory cortical activity. I found many auditory cortical neurons that were suppressed during singing, even though these same neurons were excited during non-singing epochs by playback of the bird's own song. Additionally, a small proportion of neurons exhibited suppressed activity several hundred milliseconds prior to song onset, consistent with a vocal corollary discharge signal. Finally, perturbing auditory feedback with singing-triggered noise also revealed a subpopulation of neurons that behaved like error-detectors, consistent with a predictive coding process.

To further explore the extent to which the singing-related activity of auditory cortical neurons depend on auditory feedback, I tracked activity before and after deafening, while also using a variational autoencoder to measure deafening-induced changes to song. I

found auditory cortical neurons that maintained similar patterns of singing-related modulation following deafening, but these activity patterns became more variable in parallel with changes in syllable structure. Other neurons that were suppressed or inactive during singing prior to deafening became strongly excited during singing after deafening, supporting the hypothesis of an emergent error-like signal that drives changes in the song. Altogether, these results provide evidence of a vocal corollary discharge signal that functions predictively in the auditory cortex to suppress singing-related feedback and suggest that this signal contributes to the maintenance of stable vocal motor programs.

## **Dedication**

To my Mom and Dad. Thank you for always supporting my path to becoming a scientist.

# Contents

Abstract.....	iv
List of Figures.....	xi
Acknowledgements.....	xiii
1. Introduction.....	1
1.1 Overview .....	1
1.2 Sensorimotor integration .....	2
1.2.1 The refference principle: sensory feedback and motor control .....	2
1.2.2 Corollary discharge signals in the auditory system.....	4
1.2.3 Principles of predictive coding and error detection.....	6
1.3 The zebra finch as a model for feedback evaluation .....	9
1.3.1 Song learning.....	9
1.3.2 The role of auditory feedback in song learning and maintenance .....	11
1.3.3 Circuits for error detection and error correction.....	13
1.3.4 Dopaminergic neurons drive vocal error correction.....	15
1.3.5 Current challenges in studying vocal feedback and evaluation.....	16
1.4 The avian auditory system: anatomy and function.....	17
1.4.1 Organizing principles of the avian pallium and mammalian neocortex .....	17
1.4.2 Early auditory pathway .....	18
1.4.3 The auditory pallium .....	20
2. A vocal corollary discharge modulates the auditory cortex.....	24
2.1 Introduction .....	24

2.2 Results .....	25
2.2.1 The auditory cortex is differentially modulated by singing and playback .....	25
2.2.2 Head and body movements do not modulate activity in the auditory cortex .....	30
2.2.3 Activity modulation across the auditory hierarchy .....	32
2.2.4 Modulation of activity during singing emerges in the auditory cortex .....	36
2.3 Conclusions .....	37
2.4 Methods .....	40
2.4.1 Surgical procedures .....	40
2.4.2 Calcium imaging .....	41
2.4.3 Histology and in situ hybridization .....	41
2.4.4 Playback stimuli .....	42
2.4.5 Fiber photometry .....	43
2.4.6 Movement analysis.....	43
2.4.7 Statistics .....	44
3. The primary auditory cortex evaluates auditory feedback and detects errors.....	45
3.1 Introduction .....	45
3.2 Results .....	46
3.2.1. Error signals are generated in the primary auditory cortex .....	46
3.2.2 Error signals are modulated by different perturbations .....	49
3.2.3 Neurons show tuning to song in a high-dimensional space.....	52
3.3.4 Neurons tune to acoustical features during singing.....	55
3.3 Conclusions .....	57

3.4 Methods .....	60
3.4.1 Distorted auditory feedback .....	60
3.4.2 Variational autoencoder training and alignment .....	60
3.4.3 Acoustic features .....	60
4. Removal of auditory feedback reveals the emergence of an error-like signal .....	62
4.1 Introduction .....	62
4.2 Results .....	63
4.2.1 Early deafening reveals non-auditory modulation in the auditory cortex .....	63
4.2.2 Different stages of deafening produce changes in population dynamics .....	68
4.2.3 Changes in song post-deafening .....	70
4.3 Conclusions .....	72
4.4 Methods .....	73
4.4.1 Deafening procedure .....	73
4.4.2 Cross-registration of neurons .....	73
4.4.3 Statistics and analysis .....	74
5. Conclusions .....	75
5.1 Summary .....	75
5.2 Discussion .....	77
5.2.1 Corollary discharge function and implementation mechanisms .....	77
5.2.3 Vocal motor signals in other animal models .....	80
5.2.4 The meaning of an error signal .....	80
5.2.5 Deafness-related changes .....	81

5.2 Limitations.....	82
5.4 Future directions.....	83
Appendix A.....	85
Developing a naturalistic vocal error manipulation.....	85
A.1 Introduction.....	85
A.2 Results.....	86
A.3 Conclusions.....	90
A.4 Methods.....	90
A.4.1 Syringeal muscle injection.....	90
A.4.3 Electromyography.....	91
A.4.2 Muscle histology.....	91
References.....	92
Biography.....	103

## List of Figures

Figure 1. Schematic of sensorimotor integration in the nervous system. ....	3
Figure 2. Schematic of a forward internal model.....	7
Figure 3. Timeline of song development. ....	10
Figure 4. Schematic of regions involved in song production and perception. ....	14
Figure 5. Connectivity of the auditory pallium.....	21
Figure 6. Experimental conditions.....	25
Figure 7. Single-neuron modulation in CML.....	27
Figure 8. Activity in CML during singing and playback.....	29
Figure 9. Movement measurement.....	31
Figure 10. Activity in Field L during singing and playback.....	33
Figure 11. Activity in NCM during singing and playback.....	34
Figure 12. Activity modulation across the auditory hierarchy.....	35
Figure 13. Activity modulation in thalamic terminals.....	37
Figure 14. Representative responses to white noise perturbation.....	47
Figure 15. Population activity to distorted auditory feedback with white noise.....	49
Figure 16. Representative responses to syllable perturbation.....	50
Figure 17. Population activity to distorted auditory feedback with syllable.....	51
Figure 18. Methods to quantify song.....	53
Figure 19. Analysis of calcium activity and latent features.....	54
Figure 20. Correlation distributions across song features.....	56
Figure 21. Tuning to spectral features during singing.....	57

Figure 22. Identification of neurons pre- and post-deafening.....	64
Figure 23. Activity during playback pre- and post-deafening. ....	65
Figure 24. Activity during singing pre- and post-deafening.....	67
Figure 25. Longitudinal tracking of individual neurons. ....	69
Figure 26. Population trajectories across deafening. ....	70
Figure 27. Song characterization using the VAE across deafening.....	71
Figure 28. ChRmine stimulation approach. ....	87
Figure 29. Validation of ChRmine function. ....	88
Figure 30. Neural responses to syrinx stimulation.....	89

## Acknowledgements

I would like to thank my advisor, Rich, for his guidance throughout all these years. He undoubtedly made me a more critical and overall better scientist and largely contributed to shaping my scientific perspective. I know I often took the path of most resistance, but I am deeply grateful to him for allowing me to work on a project that I was most passionate about. During my time in the lab, I often felt challenged by working on something that was different from the most common themes in the lab. However, this pushed me to develop new analyses and explore creative solutions to the scientific challenges I faced. Rich was always supportive during this time and helped me lose the fear of being the first one to try new things. This is something that I will carry with me for the rest of my career.

I also want to thank Lindsey, Marc, Kevin, and Steve for serving as members of my committee and for simply being additional mentors throughout all these years. I will always remember their support that came in many different forms including one-on-one science talks, making sure I felt ready for my first conference, supporting the next stages of my career, and encouraging me to keep asking questions. Additionally, I would like to thank John Pearson, he always provided enthusiastic support with my project and steered me on the right path when I felt lost in the depths of my analysis.

I would like to thank members of the Mooney and Pearson labs for making the lab a great place to be, especially in the past 4 years. I would like to especially thank Jiaxuan, Scott, Thomas and Tom for always being there when I needed to talk or vent or rant about things. I also often miss Mor and BJ, previous members of the Mooney lab, who taught me how to do many things and offered a lot of guidance through tough times.

While I carried out my Ph.D., it often felt like it was me against the world; thanks to my friends and family, I constantly remembered this was not true, and I would not be where I am today if it weren't for their unwavering support. To those friends who shared the graduate school experience with me, we did it, and we survived. Thank you to Pia, Ale, Liam, Jiaxuan, Arthy, Robin, Janet, Chris, Grecia, Mariana, Marie, Christine, Aryana, Wenxi, and Stuart for all the hikes, museum and gallery visits, fun moments, and overall comic relief from the graduate school journey. Thank you to all my friends back home, Ali, Nisa, Carmen, Rodrigo I, Rodrigo P, Andres, Malinka, Andy, and Eli, who always rooted for me at a distance. I especially want to thank Diana and Ari for the 22 (and counting) years of friendship and for talking me out of my imposter syndrome so many times. All my visits to Mexico, where I got to see all of them and my family, always recharged my battery and allowed me to keep going.

To my partner Evan, thank you for making sure I continued to be a functioning person during some of the most stressful times, for keeping me company during all those late work nights either over Zoom or in the quantum lab, and for always believing in me, especially during all those times that I doubted myself.

Finally, and most importantly, I would like to thank all my family. Even though they were perplexed about why I was still in school after so long, they never stopped being proud of me. I want to especially thank my Mom, Dad, and my brother Hector for all those Sunday night video calls that kept me sane and reminded me of what was truly important during my time in graduate school. I also want to give a special shout-out to my Dad for being the most enthusiastic cheerleader throughout my Ph.D. by so often sending me messages of support. It was not easy to do this, being far away from all of them in a different country, in a different language, and in a different culture, but they always made sure I kept going.

# 1. Introduction

## *1.1 Overview*

Many motor plans require precise sensory feedback to maintain the execution within a desirable range. We need to feel the glass full of water to hold it without dropping it, we need to be able to observe the forest to follow with our gaze the bird that took flight, and we need to listen to the note we just played in the piano to know we are playing the right key. Throughout many regions of the brain, there is a constant interplay between sensory and motor information that allows us to learn, execute, correct, and predict the outcome of our behaviors. Many studies across several species have proposed that specifically motor-to-sensory transformations are neural circuit implementations that function to 1) distinguish sensory feedback produced by our own movements from sensory stimuli from the environment, 2) predict sensory input based on previous experiences and 3) update this prediction.

In the context of vocal communication, one way in which we believe the brain carries out motor-to-sensory transformations is by producing copies of vocal-motor commands that cancel out auditory feedback. Studies in humans have shown that this may occur through suppression of neural activity in the auditory cortex during speech (Hickok et al., 2011; Houde & Chang, 2015). This suppressive effect is conserved across species, even in those that only produce innate vocalizations (Eliades & Wang, 2003; Harmon et al., 2024). While this strategy can be useful in preventing saturation of the auditory cortex due to the loudness of vocalizations, it remains unclear whether it serves as a prediction of the vocal output that contributes to error detection and correction of future vocalizations.

Zebra finches offer the opportunity to study the predictive and corrective nature of vocal-motor commands since they need to constantly monitor the feedback of their vocalizations to

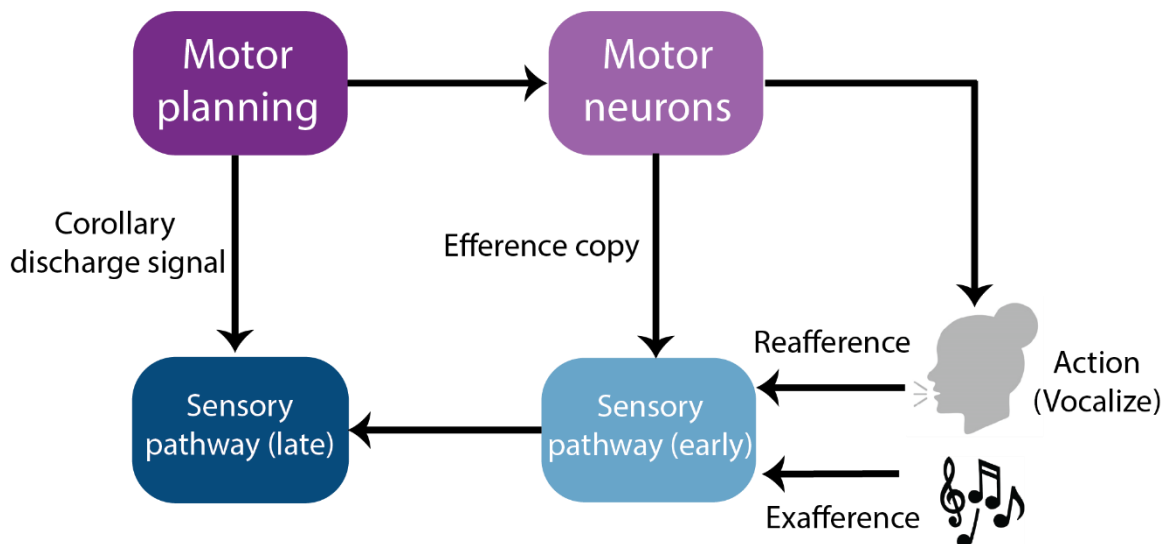
correct and maintain them within a desirable range (Doupe & Kuhl, 1999). In this thesis, I used the zebra finch as a model system to explore the effects of vocal-motor modulation in the songbird equivalent of the auditory cortex. I tested multiple experimental conditions that included playbacks of the bird's own song, distorted auditory feedback, and removal of auditory feedback to test how the auditory cortex perceives and evaluates vocal feedback. In Chapters 2 and 4, I provide evidence that supports the hypothesis of a corollary discharge signal that modulates neurons in the auditory cortex during singing and in the absence of auditory feedback. Additionally, in Chapter 3, I demonstrate how neurons in the auditory cortex monitor and evaluate song feedback through different subpopulations of neurons that detect distorted feedback and closely track the acoustic features of the song. Finally, removing auditory feedback through deafening revealed the emergence of an error signal that parallels with changes in the song. Altogether, these results demonstrate that neurons in the auditory cortex do not simply passively process sounds but rather integrate vocal-motor information, are selective to the auditory signals that they encode during singing, and that their modulation contributes to shaping vocal output.

## ***1.2 Sensorimotor integration***

### **1.2.1 The reafference principle: sensory feedback and motor control**

Any animal that moves needs to distinguish between the sensations produced by their own movements (reafference) and the sensory stimuli that come from the surrounding environment (exafference) (Figure 1). Sensory feedback always lags motor commands, and thus, the brain must have a strategy to bypass these latencies and anticipate the consequences of its movements. One of the first formal descriptions of this phenomenon was made in the 19<sup>th</sup> century by Hermann von Helmholtz in the context of eye movements. He proposed the existence of a signal in the brain that cancels out visual flow to prevent perceiving our eye movements as a shift

in the visual scene (von Helmholtz, 1867). However, it wasn't until almost a century later that some of these ideas were tested experimentally by Erich von Holst and Horst Mittelstaedt (von Holst and Mittelstaedt, 1950) and Roger Sperry (Sperry, 1950). By studying the optokinetic reflex in different species, they found that under normal conditions, there was a signal that canceled out the visual flow and thus prevented perceiving eye movements as external visual input. They coined two different terms to refer to the influence of motor signals at the level at which they are generated: efference copy and corollary discharge (Figure 1).



**Figure 1. Schematic of sensorimotor integration in the nervous system.**

Reafference and exafference are distinguished by motor signals that modulate sensory responses at multiple stages along the processing hierarchy (early vs. late).

Efference copies are generated by the neurons that execute movement, i.e. motor neurons, and they exclusively cancel unwanted reafference. In contrast, corollary discharge signals are not constrained to a specific point along the motor and sensory pathways and can influence sensorimotor integration across the nervous system in many different behaviors (Crapse & Sommer, 2008). Besides canceling out reafference or sensory feedback, corollary discharge signals have been theorized to generate internal forward models. A forward model consists of a

representation of a sensory outcome based on motor commands and can be used to predict sensory feedback (Miall & Wolpert, 1996). While the influence of corollary discharge signals has been recorded across many sensory modalities (Bell, 1982; Conner et al., 2021; Schneider et al., 2014; Sommer & Wurtz, 2002), I will focus on their effects on the auditory system.

### **1.2.2 Corollary discharge signals in the auditory system**

Corollary discharge signals can modify auditory processing by acting at different points along the auditory pathway (Schneider & Mooney, 2018). Generally, they suppress responses to self-generated movements which prevents the auditory system from becoming overwhelmed and helps to maintain sensitivity to sounds in the environment that may be produced by conspecifics or predators.

One of the first systems where corollary discharge signals were shown to dampen auditory signals is the cricket. Despite their small size, male crickets can produce stridulating sounds of over 100 dB by rubbing their forewings together. To prevent adaptation and desensitization to exafferent sounds, an interneuron in the mesothoracic ganglion produces a corollary discharge signal that is synchronized with chirping bouts to suppress both pre and post-synaptic auditory afferent neurons (Poulet & Hedwig, 2002, 2006). This suppression near the periphery prevents them from being deafened by their own chirping.

In mammals, corollary discharge signals modulate activity in the central nervous system across many behaviors, including locomoting, grooming, licking, and vocalizing (Harmon et al., 2024; Schneider et al., 2014; Singla et al., 2017). In the auditory cortex of mice, a projection from the secondary motor cortex has been identified as one source of corollary discharge signals for movement. This projection activates parvalbumin-expressing interneurons in the auditory cortex that, in turn, suppress excitatory pyramidal neurons (Schneider et al., 2014). Importantly, this modulation is predictive: it starts before the onset of locomotion and can be shaped to adapt and

selectively suppress predictable sounds in novel contexts (Schneider et al., 2018). This suppression is also observed when mice vocalize but is different from that seen during locomotion. Specifically, neurons that are excited by playbacks of vocalization are selectively suppressed during vocalization. Additionally, neurons in the auditory cortex can be modulated before vocal onset, and this modulation is scaled by vocal power, suggesting an adaptive and anticipatory function of corollary discharge signals (Harmon et al., 2024). Altogether, these results support the idea that corollary discharge signals can be specific in their function depending on the behavioral context.

In marmoset monkeys, the vocal modulation of the auditory cortex has been studied extensively. During calling, most neurons are suppressed below baseline, and this suppression starts on average 200 milliseconds before vocal onset (Eliades & Wang, 2003). Importantly, this suppression is generalized since it occurs independently of the type of vocalization being produced (Eliades & Wang, 2013).

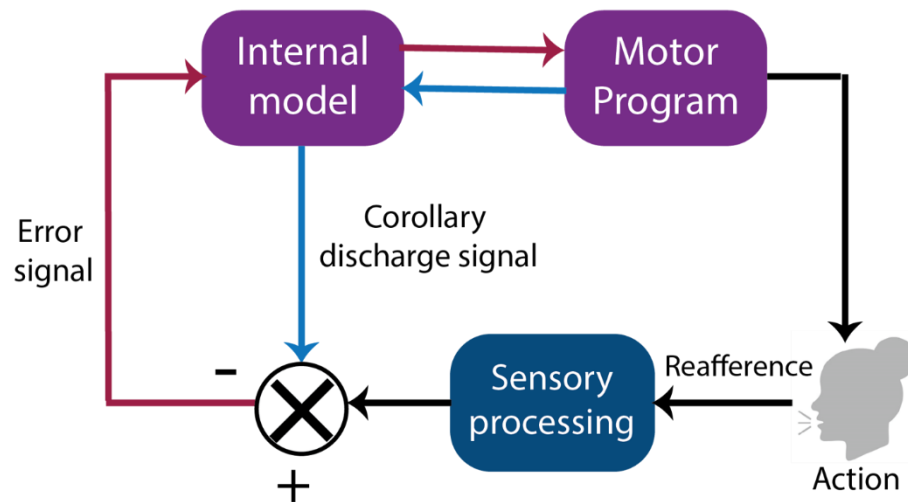
The studies mentioned thus far primarily cover the effects of corollary discharge signals during the production of innate vocalizations. However, corollary discharge signals also figure prominently in models of speech production in humans. Magnetoencephalography scans have shown that responses in the auditory cortex are suppressed relative to passively hearing speech (Curio et al., 1998). Moreover, this suppression is specific to self-produced speech rather than generalized suppression since responses to external sounds that are played during speech production are not dampened (Houde et al., 2002).

These findings demonstrate how corollary discharge signals are a conserved mechanism that cancels auditory feedback from our movements to maintain sensitivity to external stimuli. However, corollary discharge signals have also been proposed to serve a predictive function to monitor, evaluate, and correct motor performance. These processes are particularly important in

the context of learned vocalizations, and work in humans has contributed to the adaption of models of state feedback control to understand how they contribute to producing and maintaining speech (Hickok et al., 2011).

### **1.2.3 Principles of predictive coding and error detection**

The basis of the predictive coding theory is that the brain generates internal models of the world to predict sensory input. Internal models are based on experience and can represent the relationship between movement and its sensory consequences (Keller & Mrsic-Flogel, 2018). There are two types of internal models that solve different aspects of sensorimotor control and integration: inverse internal models estimate the motor command needed for a desired sensory outcome, and forward internal models predict the sensory outcome based on the motor command (Miall & Wolpert, 1996). Corollary discharge signals have been proposed as a mechanism through which forward models are implemented in the brain. By estimating the sensory consequences of movement, they generate a prediction that can be compared with the sensory feedback. These two inputs into the system can then be used to evaluate whether the sensory outcome was correct (match) or not (mismatch) based on the prediction made. Mismatches between expected and received feedback can generate error signals that update the internal model or correct future motor execution (Figure 2). Error signals have been reported in many different species, behavioral contexts, and circuits, including those in the cerebellum, basal ganglia, and cortex (Popa et al., 2017; Schultz et al., 1997). However, the exact mechanisms by which they implement changes in the motor program and update the internal model are not well understood.



**Figure 2. Schematic of a forward internal model.**

A prediction from the internal model is transmitted via a corollary discharge signal. A region downstream integrates information about the prediction and the sensory outcome produced by the action (circle). If the sensory feedback matches the prediction (+) there is no error. If the sensory feedback deviates from the predicted feedback (-) an error signal is used to either update the internal model or modify the motor program. Red arrows represent prediction error signals, while blue arrows represent prediction signals.

Sensory cortical circuits have been shown to be an important locus for integrating predictions of sensory outcomes with sensory feedback to generate error signals (Audette & Schneider, 2023; Jordan & Keller, 2020; Manley, 2017; Rao & Ballard, 1999). In the auditory cortex of marmosets, Eliades and Wang have reported error-like responses to altered auditory feedback. They carried out an experiment where they shifted the pitch of vocalizations through custom-made headphones. With this manipulation, they recorded individual neurons and found a reduced suppression when compared to normal vocalizations (Eliades & Wang, 2008). This finding suggested the presence of an error signal that is generated when the expectation does not match the auditory feedback. A more recent study expanded on this finding and showed that the magnitude of this error signal can be scaled with the magnitude of the mismatch (Eliades & Tsunada, 2024). Despite only producing innate vocalizations, marmoset monkeys may use these

error signals to self-monitor and flexibly modify vocal output in response to external perturbations (Eliades & Wang, 2012; Eliades & Tsunada, 2018; Löschner et al., 2023; Pomerger et al., 2018). While the exact purpose of this flexibility and vocal control is still unclear, it may be useful for adapting to different acoustic environments and identifying conspecifics (Oren et al., 2024).

Error signals have also been incorporated into models of human speech control, where they function to monitor and correct speech to ensure proper communication (Hickok et al., 2011). One proposed model for speech control is the state feedback control model, which encompasses different aspects of speech production. The core of this model is a comparison computation that integrates the prediction of speech through a corollary discharge signal with auditory and somatosensory feedback from the vocal apparatus (Hickok, 2012; Houde & Nagarajan, 2011). This model is supported by results from auditory feedback perturbation experiments (Houde et al., 2002; Tourville et al., 2008), as well as measurements of speech deterioration in hearing-impaired or deaf children and adults (Gold, 1980; Lane & Webster, 1991), which showed that alterations to auditory feedback prevent the maintenance of a stable motor program for spoken language. In contrast, direct evidence for the role of corollary discharge signals in human speech control has been difficult to test. Studies in patients with lesions in regions that have been proposed to either generate or relay vocal corollary signals experience a reduced ability to detect and adapt to speech errors, resulting in disorders such as stuttering and conduction aphasia (Behroozmand et al., 2018; Daliri & Max, 2015).

The consequences of faulty sensorimotor control for speech perception and production highlight the importance of understanding the neural circuits that integrate vocal corollary signals and sensory feedback, especially in the auditory system. Exploring the function of vocal corollary discharge signals and their effects at the single-cell and circuit level is impractical in humans due

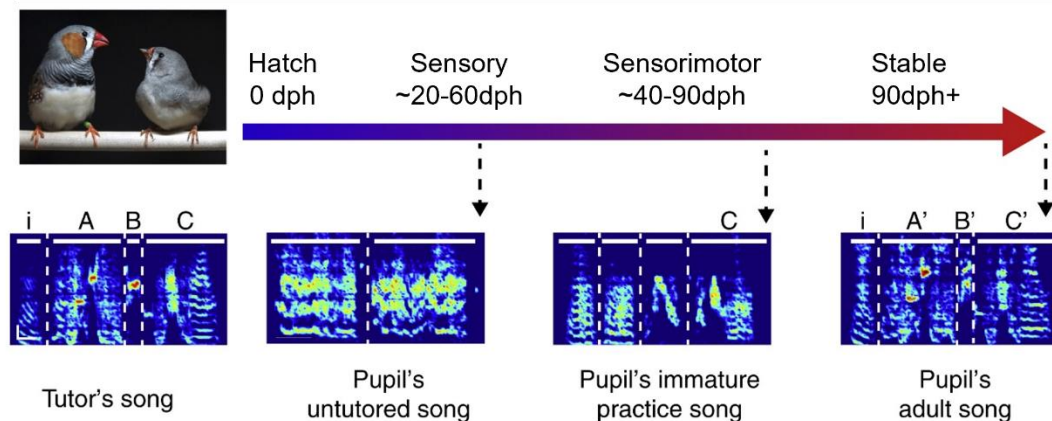
to the low spatial and temporal resolution of non-invasive recording techniques. Moreover, most animal models, such as rodents and non-human primates, produce only innate vocalizations and fail to capture the complex sensorimotor interplay characteristic of speech. One exception is the songbird, they offer a learned naturalistic behavior that requires auditory feedback, making them a great model to study prediction processing and error detection in the context of vocal performance.

### ***1.3 The zebra finch as a model for feedback evaluation***

Vocal communication is essential in many vertebrate species. Vocalizations can transmit information related to resources, danger, conspecific identification, and social affiliation. Most vertebrates only produce innate vocalizations, which do not require auditory feedback and can be produced even if the animal is reared in isolation. However, a handful of vertebrate species are vocal learners: they learn to imitate vocalizations through cultural transmission. Songbirds belong to this category, and studying the natural process through which they learn and maintain their vocalizations can offer many insights into the sensorimotor circuits that coordinate complex and highly skilled behaviors (Doupe & Kuhl, 1999).

#### **1.3.1 Song learning**

Zebra finches are gregarious songbirds native to Australia and Indonesia. They can live in large colonies and are easily adaptable to laboratory conditions. They have a rich vocal repertoire that includes ten different calls and a learned song (Elie & Theunissen, 2016; Zann, 1993). Young chicks use begging calls for feeding, and tonal calls when they are separated from their parents.



**Figure 3. Timeline of song development.**

A mature song consists of intronotes (i) and individual syllables (A-C). A juvenile songbird undergoes a process of learning where he tries to imitate his tutor or father's song. The main stages of this development consist of a sensory phase, where he memorizes a template of his tutor's song, and a sensorimotor phase, where he starts producing sounds to match the template. 90 days post-hatch, juvenile birds can produce a stable song that resembles their tutor's song.

Modified from Mooney 2022.

Adult calls include those for close and distance communication (*tet*, distance), threatening situations (*wsst*, distress, *tuck*), and affiliation or pair bonding between adults or chicks (whine, nest, *thuck*). Both males and females produce all these types of calls, but sexual dimorphism distinguishes between *tet* calls and distance calls. Furthermore, only male zebra finches learn how to sing, and while the song is part of their courtship display to convey fitness, they freely sing in isolation and produce thousands of renditions per day.

The zebra finch song is composed of a string of innate introductory calls followed by 3-6 distinct syllables in a fixed sequence that constitutes a motif (Figure 3). Multiple motifs are concatenated, with the occasional addition of introductory or *tet* calls, to produce a song bout. Adult zebra finch songs are highly stereotyped, but individual syllables have some degree of variability that occupies a distribution within the acoustic space (Goffinet et al., 2021). In the presence of a female, the bird narrows this distribution and performs a less variable song

characterized by higher pitch, faster pace, and longer duration (Kao et al., 2005; Sossinka & Böhner, 1980). This is directed singing, and female birds show a strong preference for this type of song compared to undirected song (produced by the male when he is alone) (S. C. Woolley & Doupe, 2008).

The courtship song is learned over a period of ~90-120 days across two phases: sensory and sensorimotor (Figure 3) (Price, 1979). During the sensory phase, between 20-60 days post-hatch, juvenile birds listen and memorize their tutor's song. This sensory phase coincides with the development of selectivity to conspecific vocalizations in the primary auditory cortex (Amin et al., 2007). The start of the sensorimotor phase is marked by the onset of vocal production and overlaps with the sensory phase. During this time, they produce a subsong that is characterized by broadband sounds and variable duration (Aronov et al., 2011; Tchernichovski et al., 2001). Throughout song development, the juvenile's goal is to match their vocal feedback to the memorized tutor template (Konishi, 1964). After months of vocal practice and correction, they reduce the vocal variability and minimize the difference between their auditory feedback and memorized template to converge into a song that closely resembles that of their tutor (Figure 3).

### **1.3.2 The role of auditory feedback in song learning and maintenance**

The process of song learning heavily relies on auditory feedback. While juvenile birds do not need constant exposure to their tutor song, deafening experiments in different songbird species have demonstrated that they do need to hear themselves sing to develop a normal song (Konishi, 1964). Additionally, the auditory feedback information must be properly conveyed across the auditory pathway to maintain learned vocalizations. This was shown in a study by Lei and Mooney where they used syllable-triggered electrical microstimulation in the auditory thalamus of juvenile birds (70-90dph). This manipulation caused distortion in the spectral

features of the targeted syllable after only two days of stimulation, demonstrating how auditory feedback shapes the vocal motor program (Lei & Mooney, 2010).

Intact auditory feedback is also required in adulthood to maintain stable vocal performance. Many studies have tested this using distorted auditory feedback paradigms, where a closed-loop system monitors the song and targets specific syllables in the song with a playback stimulus. This type of feedback manipulation can be used to bias the bird to produce syllable renditions with either a higher or lower pitch (Tumer & Brainard, 2007). Studies in adult zebra finches and Bengalese finches using this behavioral approach have demonstrated how songbirds change their song when auditory feedback is altered by shifts in pitch (Sober & Brainard, 2009), time (Leonardo & Konishi, 1999; Sakata & Brainard, 2006), or removed entirely (Lombardino & Nottebohm, 2000; Nordeen & Nordeen, 1992).

For auditory feedback to influence vocal performance, there must be strong interactions between auditory and song motor circuits. Two neural pathways are crucial for song learning and production, and are influenced by auditory feedback: the song motor pathway and the anterior forebrain pathway (AFP). The song motor pathway includes HVC, the robust nucleus of the Arcopallium (RA), and the vocal-respiratory targets in the brainstem. It is indispensable for producing songs and provides input to the AFP via HVC<sub>x</sub> projection neurons (Mooney & Prather, 2005). Importantly, the song motor pathway is also influenced by the AFP through the Lateral Magnocellular Nucleus of the Anterior Nidopallium (LMAN) (Figure 4). The AFP is analogous to a cortico-basal ganglia loop in mammals and is primarily involved in feedback-dependent plasticity during song learning and in adults subjected to distorted auditory feedback (Bottjer et al., 1984; Brainard & Doupe, 2000; Kao et al., 2005).

The process of sensorimotor learning can be simplified to an evaluation model where auditory feedback of juvenile vocalizations is compared to the template of their tutor. The goal of

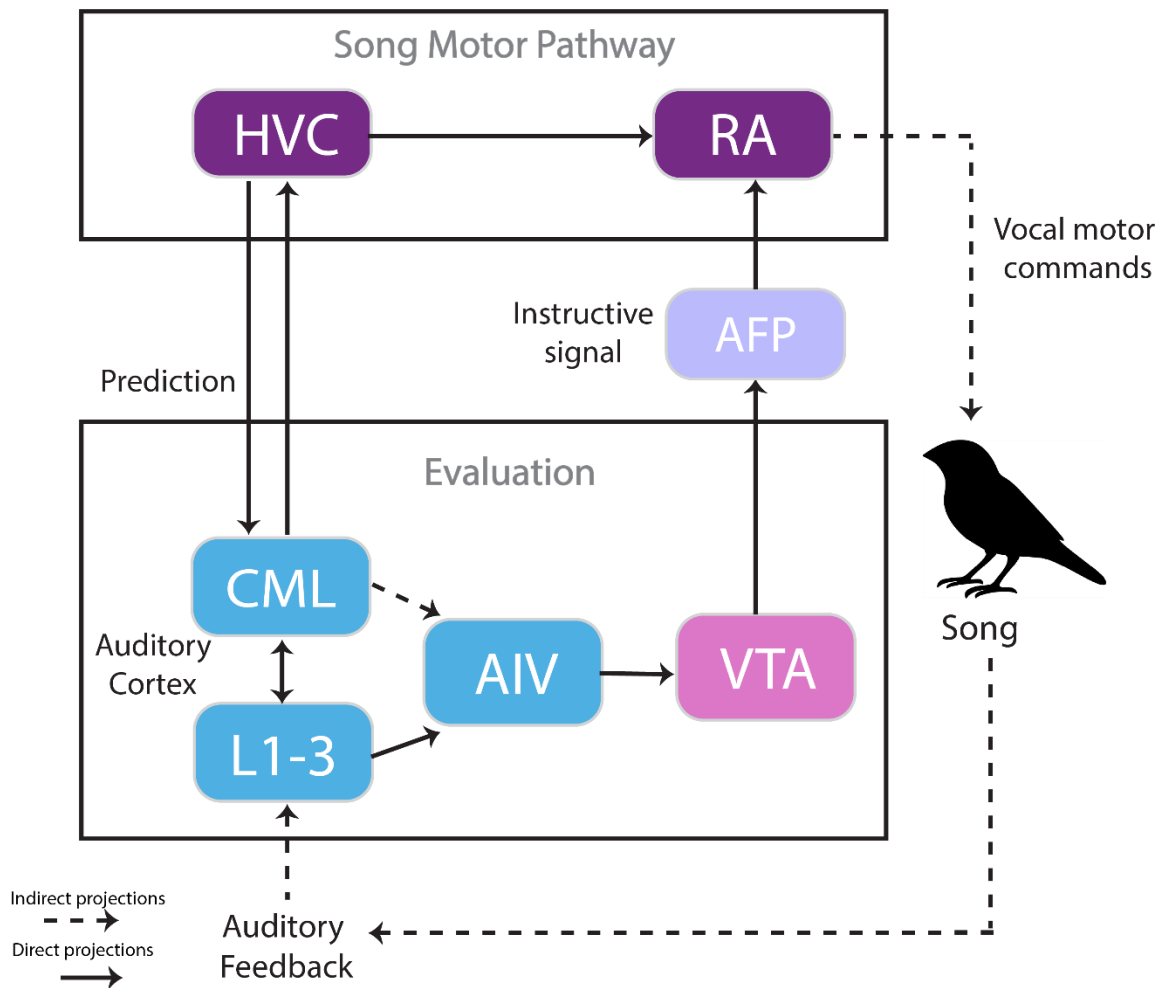
learning is to generate an instructive signal that can drive changes in the song motor pathway to minimize the differences between the auditory feedback and the auditory target. When birds transition into adulthood, this instructive signal is either smaller in magnitude or remains stable, preventing any changes to the song. This model was proposed by Doupe and Brainard based on their experimental observations. When they deafened adult birds, they observed the gradual deterioration of song. However, when they deafened and lesioned the output of the AFP, LMAN, they intercepted the transmission of the instructive signal and prevented the song deterioration normally caused by altered auditory feedback (Brainard & Doupe, 2000). This result demonstrated that an evaluation pathway remains present in the adult bird, but alterations to auditory feedback can drive an active process of song deterioration via an instructive signal that recruits the AFP and drives plasticity in the song motor pathway (Figure 4).

### **1.3.3 Circuits for error detection and error correction**

One possible interpretation of the result described by Doupe and Brainard is that the instructive signal is an error signal. They suggested that LMAN might contain the locus for song evaluation based on sound-evoked responses observed in anesthetized birds (Doupe and Konishi 1991). However, LMAN neurons are not sensitive to distorted auditory feedback during singing, suggesting that the matching computation between auditory feedback and the auditory target must occur upstream (Leonardo, 2004).

Error signals in zebra finches were discovered for the first time in the caudolateral mesopallium (CML) and Field L (Figure 4). These regions make up the auditory pallium and they are analogous to the primary auditory cortex in mammals. Keller and Hahnloser recorded in these regions during auditory feedback perturbations and found that 10% of neurons (n=8/92 neurons) increased their firing rate only when a syllable was perturbed by a short playback stimulus but not during normal singing. These neurons also did not respond to playbacks of songs with

perturbation (Keller & Hahnloser, 2009). This type of response suggested that neurons in the auditory pallium carry out a computation where they compare auditory feedback with an internal model of the song. Whether these signals are important for song learning or maintenance is still unclear. However, they are likely conveyed to other downstream pathways that eventually reach the AFP.



**Figure 4. Schematic of regions involved in song production and perception.**

Auditory information is necessary to maintain a stable song. This diagram summarizes some of the anatomical pathways through which auditory feedback information may influence the song motor pathway. HVC, proper name, RA Robust Nucleus of the Arcopallium, AFP, anterior forebrain pathway, CML, caudolateral mesopallium, AIV, ventral intermediate arcopallium, VTA Ventral Tegmental Area.

Downstream of the auditory regions recorded by Keller and Hahnloser lies the ventral intermediate arcopallium (AIV). AIV has been characterized as part of the auditory pallium with subtelencephalic projections to the VTA that resemble those of deep layers in the mammalian auditory cortex. Electrophysiological recordings in AIV have also revealed error signals in the presence of distorted auditory feedback, demonstrating that these neurons either inherit these responses from upstream auditory regions or that there is more than one site where error signals are generated. Furthermore, another important finding from the same study was that these responses are only present in neurons that project to the ventral tegmental area (VTA) (Mandelblat-Cerf et al., 2014).

#### **1.3.4 Dopaminergic neurons drive vocal error correction**

Dopaminergic neurons in the VTA of mammals are known to encode a reward prediction error. Their basal firing rates allow them to have a bidirectional modulation depending on the outcome of an action. An increase in firing rate represents a better-than-expected outcome (positive prediction error), while a decrease in firing rate or pause represents a worse-than-expected outcome (negative prediction error) (Schultz et al., 1997). These same rules apply to the dopaminergic neurons in the VTA of singing zebra finches, where white noise perturbation of a targeted syllable can produce a pause in VTA firing, while the absence of it produces an increase in firing rate (Gadagkar et al., 2016). Even more surprisingly, these dopaminergic dynamics also guide sensorimotor learning in juvenile birds by tracking song performance (Qi et al. 2024)

Work by Kearney and colleagues have integrated these findings with that of AIV-VTA error-encoding neurons and showed that this population of neurons activates inhibitory interneurons in the VTA, likely driving the pauses seen during negative prediction errors of perturbed song. Additionally, these neurons are important in transmitting an error signal to VTA. During pitch learning in adults, imposing activity patterns by optogenetically stimulating AIV-

VTA terminals, interferes with the accurate transmission of this error signal and prevents birds from shifting their vocalizations in response to white noise (Kearney et al., 2019). The correct transmission of this AIV-VTA signal is also important for song learning in juvenile birds since lesioning AIV prevents young birds from copying their tutor song (Mandelblat-Cerf et al., 2014). These results strongly suggest that VTA dopaminergic neurons facilitate learning by producing an error signal that is partially inherited from upstream auditory regions.

Altogether, the pathway from the auditory pallium to the VTA is one way in which auditory information is conveyed to the AFP to drive changes in vocal performance.

### **1.3.5 Current challenges in studying vocal feedback and evaluation**

Birdsong has been traditionally quantified using a select number of spectrotemporal features hand-picked by experimenters (Tchernichovski et al., 2001). These methods have provided important contributions to understanding some aspects of song learning and production. However, the features that have been previously characterized often show dependencies among them and fail to capture the multi-dimensional variability of acoustic space that exists in natural singing.

Novel methods to characterize acoustic features have allowed us to dive deeper into the neural correlates that produce them. One of these methods was developed by the laboratory of John Pearson, in collaboration with Richard Mooney's group. Autoencoded vocal analysis, or AVA, applies variational autoencoders (VAE) to images of song spectrograms to agnostically characterize acoustic features (Goffinet et al., 2021). VAEs are generative models that learn patterns of the input dataset, extract latent features, and then use these features to reconstruct the input (Kingma and Welling, 2013). For the acoustical analysis of vocalizations, the encoder network uses as inputs high-dimensional spectrograms (128-frequency bins x 128-time bins) of individual syllables and converts them into a low-dimensional or latent representation. AVA

extracts 32 latent features that are then utilized by the decoder network to reconstruct its own input. VAEs are optimized to preserve the most relevant information to make an accurate reconstruction, and for this reason, the low-dimensional latent features that we obtain faithfully preserve the multi-dimensional acoustics of the song.

In songbird research, we can use the VAEs to better understand what aspects of song the birds care about without having to pre-determine them. In Chapters 3 and 4 of this thesis, I used VAEs to better describe the complex tuning of neurons in the auditory cortex.

## ***1.4 The avian auditory system: anatomy and function***

### **1.4.1 Organizing principles of the avian pallium and mammalian neocortex**

Birds have remarkable cognitive abilities that used to be attributed to only primates. Tool development, superb visual and spatial memory, complex social hierarchies, theory of mind, and the rare capability of imitative learning are only some of their outstanding cognitive features (Güntürkün & Bugnyar, 2016; Payne et al., 2021; Watanabe et al., 1995; Weir et al., 2002). For a long period of time, it was believed that only cortical mechanisms would allow for such behaviors. However, this understanding started to change in the second half of the 20<sup>th</sup> century, and decades of research have demonstrated that the avian pallium maintains the same organizational principles as the mammalian neocortex (Colquitt et al., 2021; Jarvis et al., 2005; Karten, 2015).

Gene expression and connectivity from mouse visual and auditory cortex layers IV and V are shared in avian auditory and visual pallial nuclei; similarly, they receive thalamic inputs and have recurrent connections amongst themselves (Dugas-Ford et al., 2012; Vates et al., 1996). Furthermore, the microcircuitry of the auditory pallium maintains a radial columnar organization as that of the mammalian auditory cortex and has very similar cell types, including small

somatodendritic neurons that resemble stellate granule cells from cortical layer IV (Stacho et al., 2020; Wang et al., 2010). Beyond the anatomical and genetic similarities, evidence to support analogy has been shown through functional studies. Calabrese and Woolley carried out electrophysiological recordings in the auditory nuclei of unanesthetized zebra finches and demonstrated a hierarchy of auditory information processing. Furthermore, they discovered that computations carried out by excitatory principal cells and inhibitory neurons of different auditory nuclei closely resembled those in neurons of mammalian layers I-VI (Calabrese & Woolley, 2015).

Recent genetic studies showed that nuclei dedicated to song motor control from the avian arcopallium have distinct core genetic identities to neurons in the mammalian cortex, and they likely arose from the ventral pallium (Colquitt et al., 2021). However, further genetic and developmental studies are still required to truly discard the possibility of homology across the rest of the pallium, where sensory processing regions reside. Independently of the debate over the origin of these structures, the shared anatomy and function between the avian pallium and the mammalian cortex supports the utility of the songbird as a model to understand many of the computation principles of complex learned behaviors that may be shared across species.

### **1.4.2 Early auditory pathway**

Male zebra finches learn to sing and maintain their songs by monitoring auditory feedback and detecting deviations from an expected auditory template. Just like human speech, bird song shares a sequential organization, a similar learning process within a critical period, and a specialization of forebrain areas for vocal production and perception. Additionally, they have similar hearing capabilities for frequencies in the range of 0.3 to 10kHz (Hashino & Okanoya, 1989). Altogether, this has made songbirds a good model for human hearing and speech production.

Despite a 300-million-year common ancestor divergence, the auditory pathway in avian species maintains a similar functional and anatomical organization as that of mammals (Manley, 2017). In the middle ear, sound processing starts with the mechanical transduction of sound energy vibrations to the tympanic membrane. These vibrations are transmitted through the columella, the only middle-ear bone in birds (Peacock et al., 2020), that impinges via a footplate on the oval window, resolving the impedance mismatch of acoustic transmission between air and fluid and mechanically activating hair cells in the basilar papilla. The basilar papilla is tonotopically organized like its mammalian homolog, the organ of Corti; however, these two sensory systems maintain some anatomical differences. These include the homogeneous organization of hair cells along the sensory epithelium as well as the lack of inner and outer hair cells for birds (Gleich et al., 1994). Additionally, avian hair cells have the remarkable capacity to regenerate and recuperate hearing after damage produced by noise or antibiotic exposure (Cotanche, 1999; Ryals et al., 1999).

In addition to the direct propagation of sound waves, sound can also be transmitted by bone conduction, influencing acoustic transmission. Vibrations through the skull bone are transformed in the inner ear to sound by inertia in the inner ear fluid and by compression and expansion of the inner ear space, both of which produce oscillations in the basilar membrane, depolarizing the hair cells (Dauman, 2013).

Overall, the activation of hair cells is similarly transmitted in birds to ganglion neurons that form the cochlear nerve and bilaterally project to the first relay in the brain stem, the cochlear nuclei (Krützfeldt et al., 2010). From there, neurons in the cochlear nuclei project to the dorsal lateral nucleus of the mesencephalon (Mld), which is homologous to the inferior colliculus in mammals. The Mld maintains a tonotopical organization, neurons show little adaptation and have robust responses to a wide variety of auditory stimuli (S. M. N. Woolley & Casseday, 2005).

These diverse responses are reflected in their spectrotemporal tuning; however, population responses can be used to decode song (Schneider & Woolley, 2010).

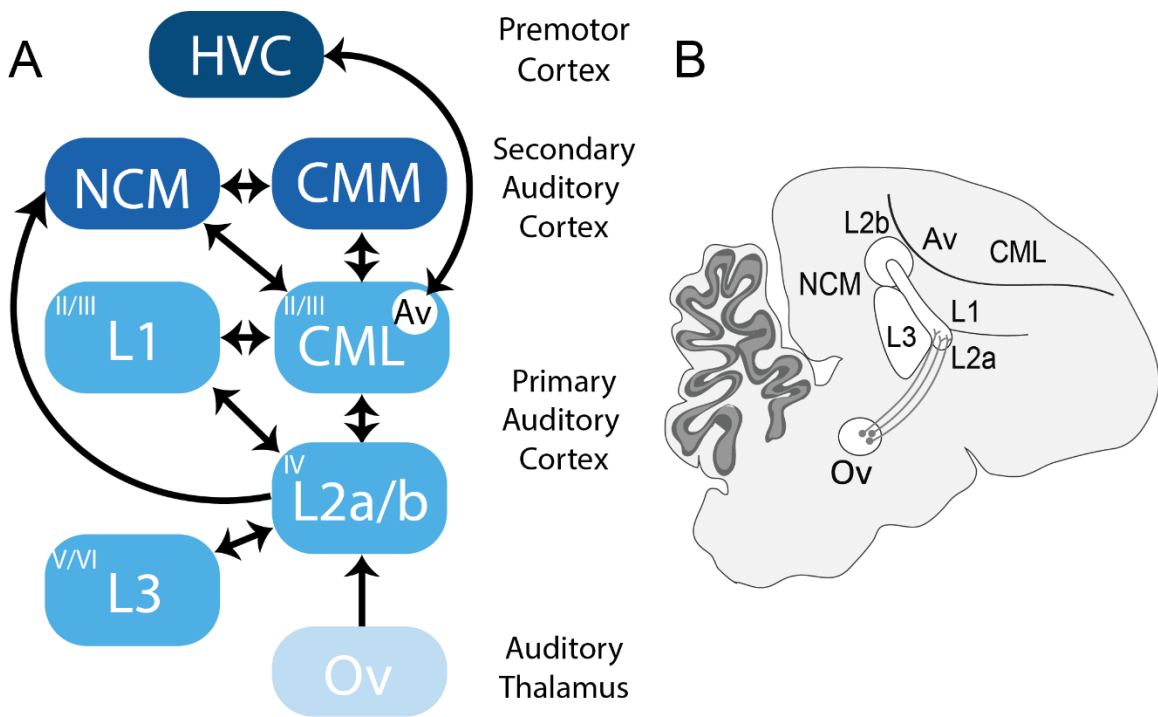
The receptive fields found in the Mld change in the downstream thalamic region, the nucleus Ovoidalis (Ov). Importantly, three different spectrotemporal receptive fields (STRF) in Ov support a better encoding of features of vocalization: broadband, which encodes amplitude envelope; narrow band, which encodes specific frequencies throughout syllables; and frequency sweep selective, which is an acoustic feature commonly observed in natural songs (Amin et al., 2010).

### **1.4.3 The auditory pallium**

Early anatomical and activity-dependent gene expression studies identified Field L, CM (caudal mesopallium), and NCM (caudomedial nidopallium) as the pallial regions involved in processing auditory stimuli (Kelley & Nottebohm, 1979; Mello & Clayton, 1994). Across these regions, individual neurons show different tuning properties, forming an auditory processing hierarchy (Figure 5). In the lower order, neurons have receptive fields that are non-selective, simple, and linear, while receptive fields in higher-order regions are sparse, selective, and non-linear (Calabrese & Woolley, 2015).

The auditory hierarchy in the pallium starts with the transmission of auditory information through a projection from Ov to Field L (Vates et al., 1996). Field L has an internal organization of four regions: L1, L2a, L2b, and L3 (Fortune & Margoliash, 1992). However, only Field L2a and b receive the auditory thalamic input from Ov, and they then project to L1, L3 and CML. While some of the receptive fields in Field L are shared with those in the early auditory pathway, overall, they are more heterogeneous and maintain a hierarchy. Generally, they are narrowly tuned to either spectral dimensions, temporal dimensions, or both (Nagel & Doupe, 2008). In other words, neurons fire at either specific frequencies, at specific time points, or for a mix of

both. This tuning can be modulated by sound intensity; at lower amplitude levels, neurons behave as integrators and accumulate information over time, while at higher intensities, neurons decrease their latency of response and follow changes in amplitude more selectively (Nagel & Doupe, 2006). While the properties above are shared across Field L, the auditory hierarchy becomes evident when looking at their firing rate properties and how they integrate information over time. Neurons in L2 have high driven rates to sound and are tuned primarily to fast temporal frequencies; as opposed to lower driven rates and slower temporal modulation in L1 and L3 (Nagel & Doupe, 2008).



**Figure 5. Connectivity of the auditory pallium.**

Connectivity diagram of the auditory pathway starting at the thalamus. Superscripts in the boxes indicate analogs to the mammalian cortical layers. B Schematic of a sagittal section showing the anatomical organization of the auditory cortex. Ov Ovoidalis, CML Caudolateral Mesopallium, CMM Caudomedial Mesopallium, Av Avalanche, NCM Caudomedial Nidopallium, HVC proper name. The projections between L3 and CML were omitted for simplicity.

Neurons in Field L then project to CM. CM can be subdivided along the mediolateral axis. This distinction not only comes from anatomical differences but from functional responses as well. Neurons on the lateral part (CML) show similar tuning properties as those in L1, and together, they have been classified as being analogous to superficial layers 2/3 of the mammalian auditory cortex (Calabrese & Woolley, 2015). Importantly, CML stands out from the rest of the auditory pallium because it contains a small field named Avalanche (Av). Av can be distinguished by its selective responses to the bird's own song and its anatomical connections that provide privileged access to a recurrent loop between sensorimotor nuclei, including reciprocal connections with HVC (Akutagawa & Konishi, 2010). The medial portion (CMM) has not been widely studied in finches. However, recordings performed in starlings have shown that neurons in this region likely integrate receptive fields from other auditory regions to produce sparse responses to complex elements of song and is analogous to the secondary auditory cortex (Meliza et al., 2010).

NCM receives inputs from most primary auditory nuclei and has also been categorized as a secondary auditory cortex. Neurons in this region show higher selectivity for bird's own song, are sparse-coding, and have invariant representations to vocalizations (Schneider & Woolley, 2013). Importantly, this region is likely involved in different social aspects of auditory-related behaviors, including recognition of conspecifics and memorization of the tutor song (Yanagihara & Yazaki-Sugiyama, Yu et al. 2023).

Despite the vast information regarding passive auditory processing in the auditory cortex, very little is known about the activity of these regions when birds sing and actively listen to their songs. In fact, there are only two studies that have ever recorded from neurons in the auditory cortex while the bird sings (Keller & Hahnloser, 2009; Mandelblat-Cerf et al., 2014).

In the next chapters, I will demonstrate how the auditory cortex is differentially modulated between passive and active listening throughout the auditory hierarchy, how the modulation during singing contributes to monitoring and evaluating song feedback, what non-auditory sources modulate the auditory cortex, and how its activity changes at different stages and alters the vocal output in the absence of auditory feedback.

## **2. A vocal corollary discharge modulates the auditory cortex**

### ***2.1 Introduction***

Songbirds live in noisy environments where they must constantly monitor their vocal feedback either for learning their song or maintaining it in adulthood. Despite the importance of the avian auditory cortex in processing auditory feedback, as well as its potential role in sensorimotor integration, our current knowledge of its function is limited to how neurons encode acoustic features. Most studies have performed electrophysiological recordings and measured playback responses to diverse stimuli in anesthetized or awake head-fixed birds. While these studies have revealed remarkable and unique features of the songbird auditory cortex, we lack an understanding of how it is engaged during singing and whether its features change during this state. To date, only two studies have recorded small populations of neurons in the auditory cortex of singing birds, and these studies did not explore in depth the type of modulation that occurred during singing (Keller & Hahnloser, 2009; Mandelblat-Cerf et al., 2014). Additionally, these studies were limited by the number of neurons that can be recorded with only a few electrodes in unrestrained birds. Calcium imaging enables the recording of larger populations of neurons, up to 80 in a single bird, in contrast to 10-15 neurons per bird reported in previous studies. Additionally, the longitudinal tracking of neurons across days expands the experimental conditions that can be tested and collected from a single bird.

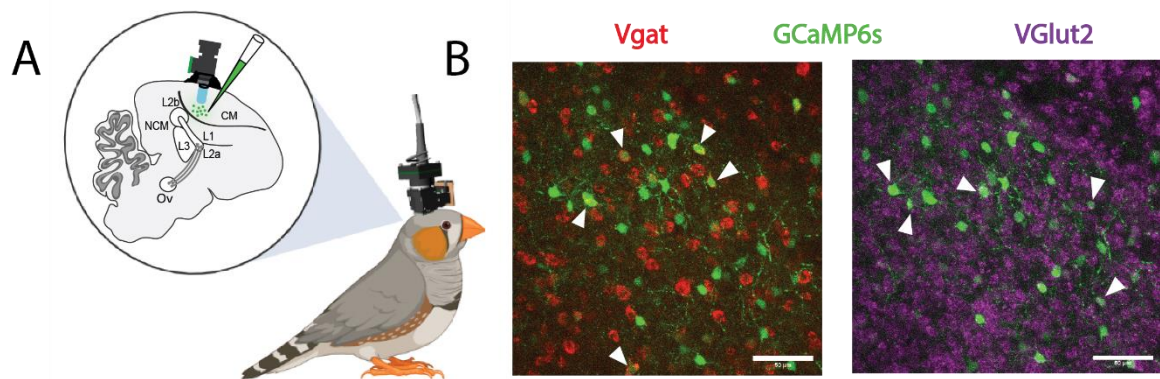
In this chapter, I describe the calcium imaging experiments that I carried out to explore whether and how a corollary discharge signal modulates the auditory cortex. Given the anatomical projections between HVC and the primary auditory cortex CML-Av this region was the focus of my research. However, corollary discharge signals are known to modulate activity at different points along the auditory hierarchy; for this reason, I explored how activity was

modulated in thalamic terminals, Field L (primary auditory cortex without motor input), and NCM (secondary auditory cortex).

## 2.2 Results

### 2.2.1 The auditory cortex is differentially modulated by singing and playback

To test the hypothesis that the primary auditory cortex is modulated by a corollary discharge signal during singing, I used a viral construct to express GCaMP6s in both excitatory and inhibitory neurons and implanted a lens in CML-Av (T. W. Chen et al., 2013). I confirmed the expression of GCaMP6s in these subpopulations of neurons by carrying out *in situ* hybridization with probes for excitatory (VGlut2) and inhibitory (Vgat) neurons (Figure 6A and B).



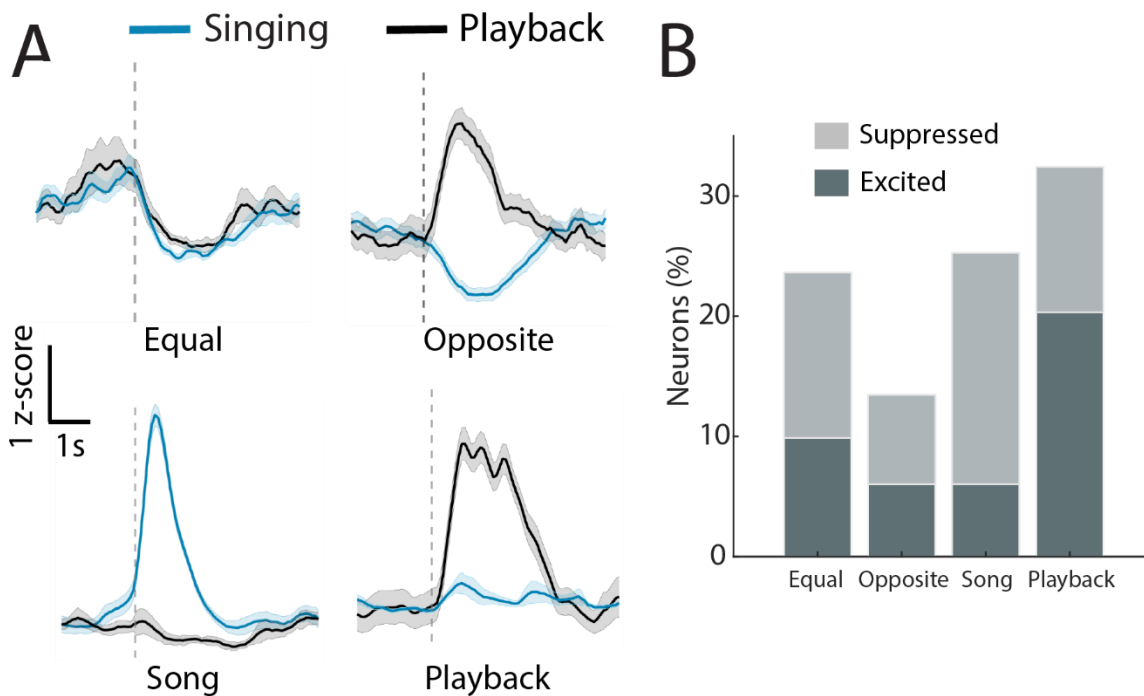
**Figure 6. Experimental conditions.**

**A.** Imaging experimental approach and sagittal section of auditory regions. The lens targeted either CML-Av, Field L or NCM that were previously injected with an AAV2/9 CAG-GCaMP6s virus. **B.** *In situ* hybridization for Vgat (red) and VGlut2 (magenta) neurons expressing GCaMP6s in the auditory cortex.

After at least 3 weeks of expression, I carried out one-photon calcium imaging in freely moving birds to monitor activity in CML-Av. I tested two conditions: playbacks of the bird's own song (BOS) and active singing. On an experimental day, the birds were monitored throughout the day by a system that detected changes in amplitude. If the bird started to sing, this change in

amplitude triggered the imaging system and recorded for periods of one minute with a blocked-out period of 5-10 minutes to prevent photobleaching from consecutive triggering. Audio clips of previously recorded bird's own song were played through a speaker, and imaging was manually triggered during non-singing epochs throughout the day. These conditions allowed me to distinguish between activity modulation caused exclusively by acoustic features of the song and modulation by motor signals.

When comparing the modulation of activity between singing and playback, I observed heterogeneous responses. I classified these responses into four categories: neurons equally modulated by both conditions, opposite modulation, modulated by only song, and only by playback (Figure 7A). Neurons that belonged to the first category and were either equally excited or suppressed by singing and playback were expected in a primary auditory region. In fact, if neurons in this region only passively integrate auditory inputs, this is the only type of response I would anticipate. However, this population only represented 24% of the significantly modulated neurons (Figure 7B). The remaining 76% of neurons displayed striking differences between playback and singing-related activity (Figure 7 A-B). Importantly, suppression dominated the responses during singing, while excitation was predominant in neurons only responding to playback (Figure 7B). These heterogeneous responses highlight the impact of non-auditory sources of modulation in this primary auditory region.



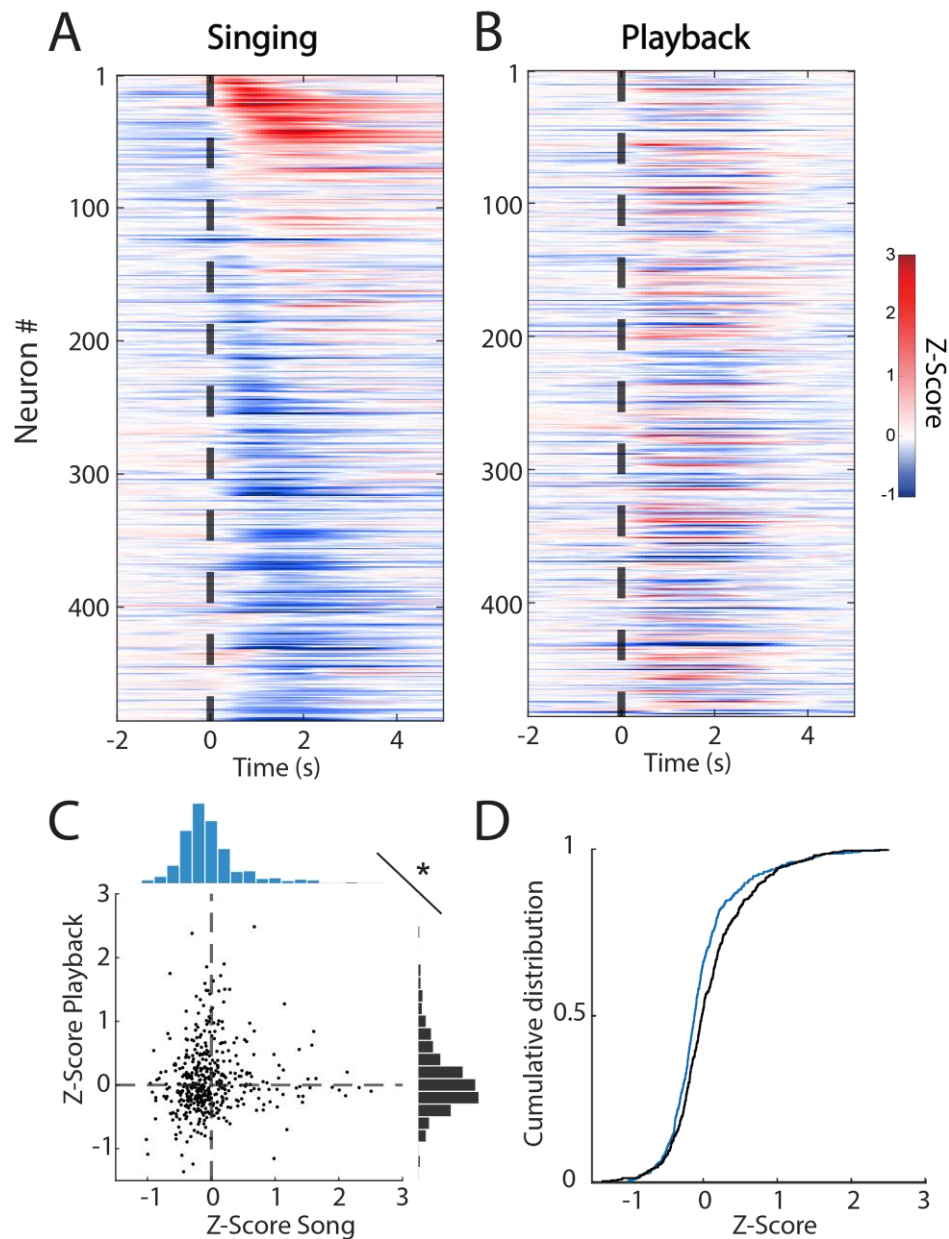
**Figure 7. Single-neuron modulation in CML.**

**A.** Representative neurons were classified into four categories based on their responses to playback and singing. Traces were averaged across trials and aligned to the onset of the first intronote or onset of playback (dotted line, mean $\pm$ s.e.m). **B.** Bar plots illustrating the percentage of modulated neurons categorized by the four groups shown in A, differentiated by the sign of modulation (n=364/485 significantly modulated neurons from 12 birds). For neurons with opposite signs of modulation, dark gray represents excited by song and suppressed by playback while light gray represents suppressed by song and excited by playback.

The suppressive effect across the different categories of modulated neurons was more clearly reflected in the dynamic responses across the entire population (Figure 8 A-B), as well as in the distributions of averaged activity in singing and playback epochs. Here, the distributions shifted towards suppression, as shown in the scatter plot and the cumulative distribution (Figure 8 C-D, Wilcoxon rank sum test,  $p=7.81e-05$ ). This shift towards suppression during singing was attributed to an increase in the number of suppressed neurons, as well as a reduction in the number of excited neurons (Figure 7B, Figure 12A).

To determine whether the modulation was produced by a vocal-motor signal, I analyzed the activity during a premotor window (800 milliseconds before intronote onset) and compared it

to an earlier baseline window during periods of non-singing. The premotor window was selected based on the distribution of vocal premotor modulation in other model systems (Eliades & Wang 2003; Harmon et al., 2024). With this analysis, I found that 11.7% of the population showed a significant modulation before vocal onset (Wilcoxon rank sum test,  $p < 0.01$ ). Among these neurons, the majority was suppressed before the first intronote (8.9%, Figure 12).

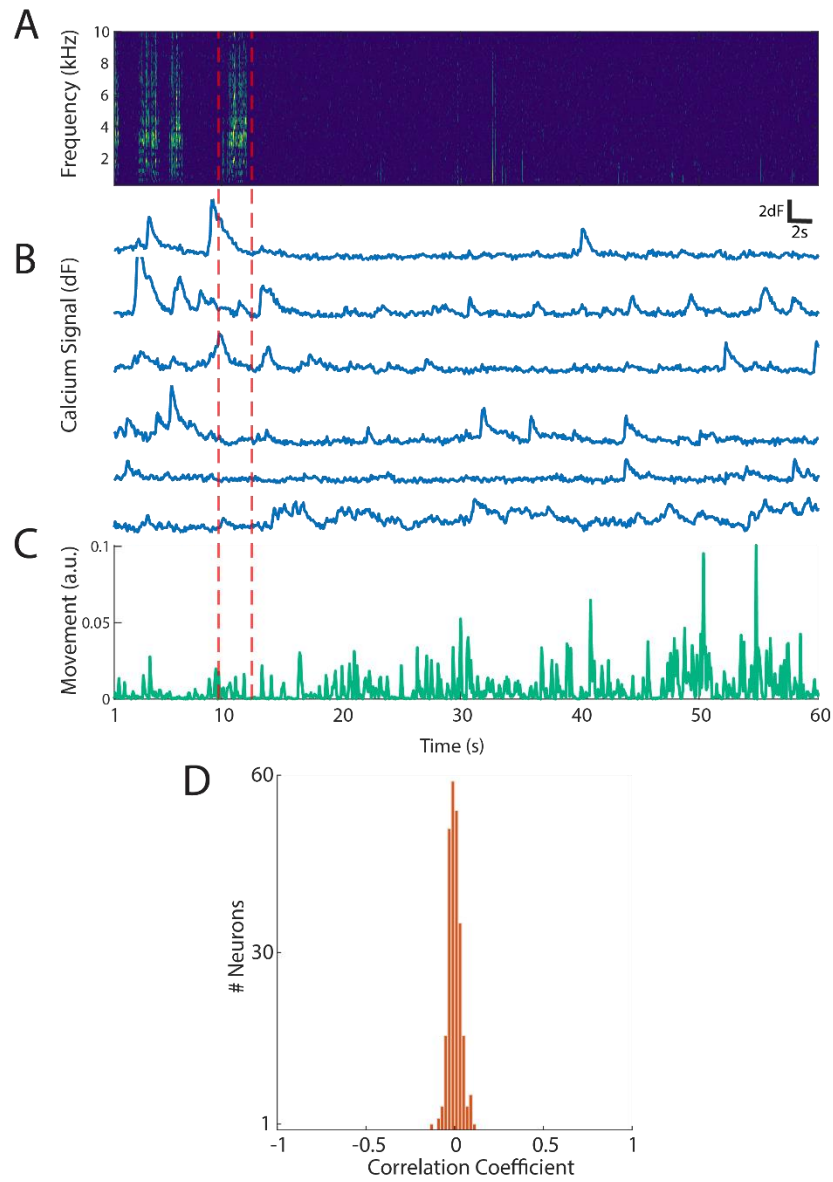


**Figure 8. Activity in CML during singing and playback.**

**A-B.** Heatmaps showing the averaged activity across trials during playback and singing sorted by activity during singing. The time is aligned to the first intronote ( $n = 485$  neurons from 12 birds). **C.** Scatter plot of averaged activity in individual neurons during the first 2 seconds of either song (blue) or playback (black) with a projection of their respective distributions (Wilcoxon rank sum test,  $p=7.81e-05$ ). **D.** The cumulative distribution between conditions shows more clearly the shift towards suppression.

### **2.2.2 Head and body movements do not modulate activity in the auditory cortex**

Movement-related activity is widespread and can modulate the auditory cortex (Schneider et al., 2014). Therefore, the modulation reported in the previous section could simply reflect general movement and not a vocal-motor signal. When singing, birds usually perch, adopt a singing posture, and do not execute many head or body movements. However, to discard the possibility of these movements causing non-specific modulation in the auditory cortex, I took advantage of the accelerometer built-in to the imaging miniscope. I calculated the acceleration across 3 dimensions to obtain a measurement of movement (see Methods, Figure 9 A-C). Then, I calculated the Pearson correlation coefficients between the normalized movement trace and the normalized neural calcium traces of individual neurons. The correlation coefficients were weak across the entire population of neurons and centered around zero (4 birds, n= 249 neurons, Figure 9 D). This distribution of correlation values demonstrated that the modulation of calcium activity in the primary auditory cortex cannot be explained by head and body movements and is specific to a vocal-motor signal.



**Figure 9. Movement measurement.**

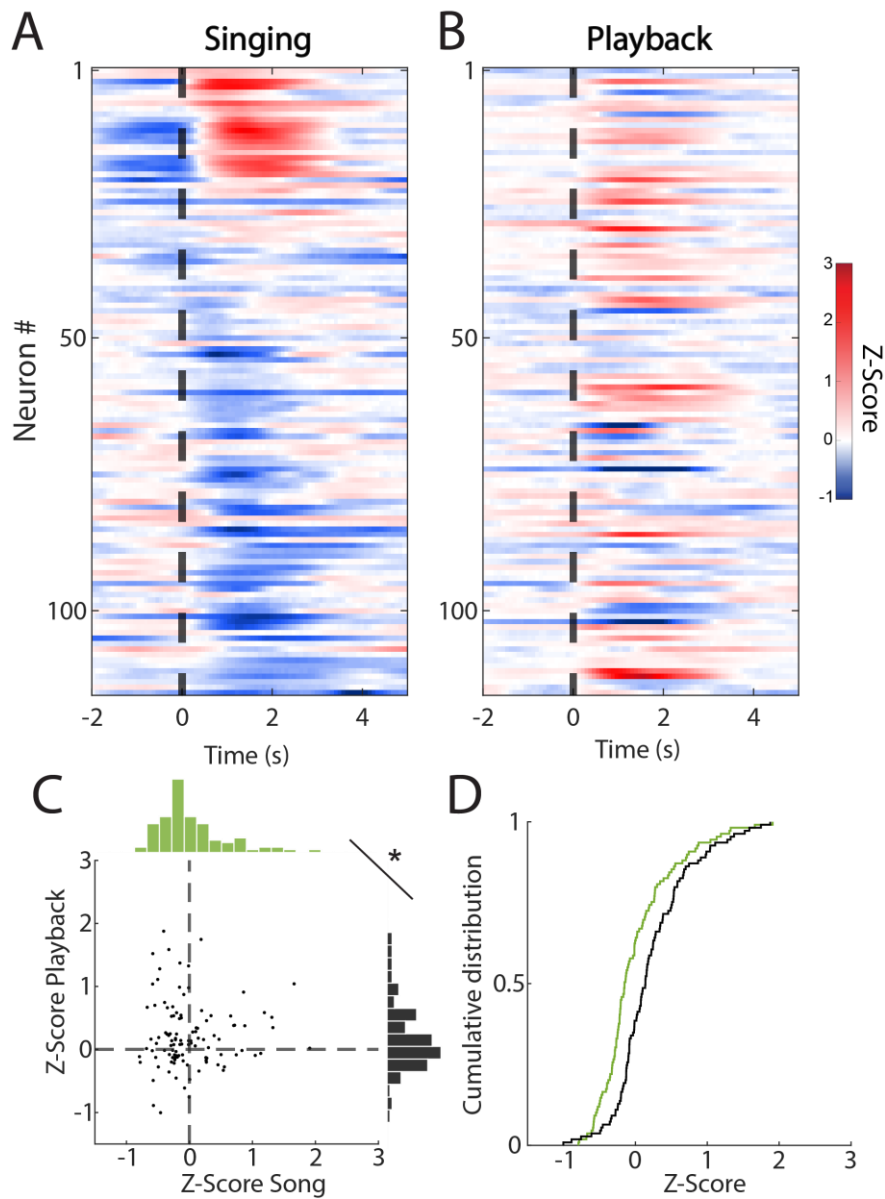
**A.** Representative sound spectrogram of one imaging session, one singing bout is highlighted by red dotted lines that mark onset and offset. **B.** Responses of 6 different representative neurons during that same imaging session. **C.** Movement signal calculated from the accelerometer during the same session. Note the reduced movement in the first 12s when the bird is singing. **D.** Histogram of correlation coefficients calculated for 249 neurons between calcium traces and movement traces for all sessions within a day ( $n=30-35$  concatenated sessions per bird).

### 2.2.3 Activity modulation across the auditory hierarchy

The anatomy positions CML-Av as the region most likely to be modulated by a motor signal. Nevertheless, I was interested in whether this modulation was transmitted to other auditory cortical regions. For this reason, I collected smaller datasets in another analog of the primary auditory cortex, Field L, and in a secondary auditory region, NCM.

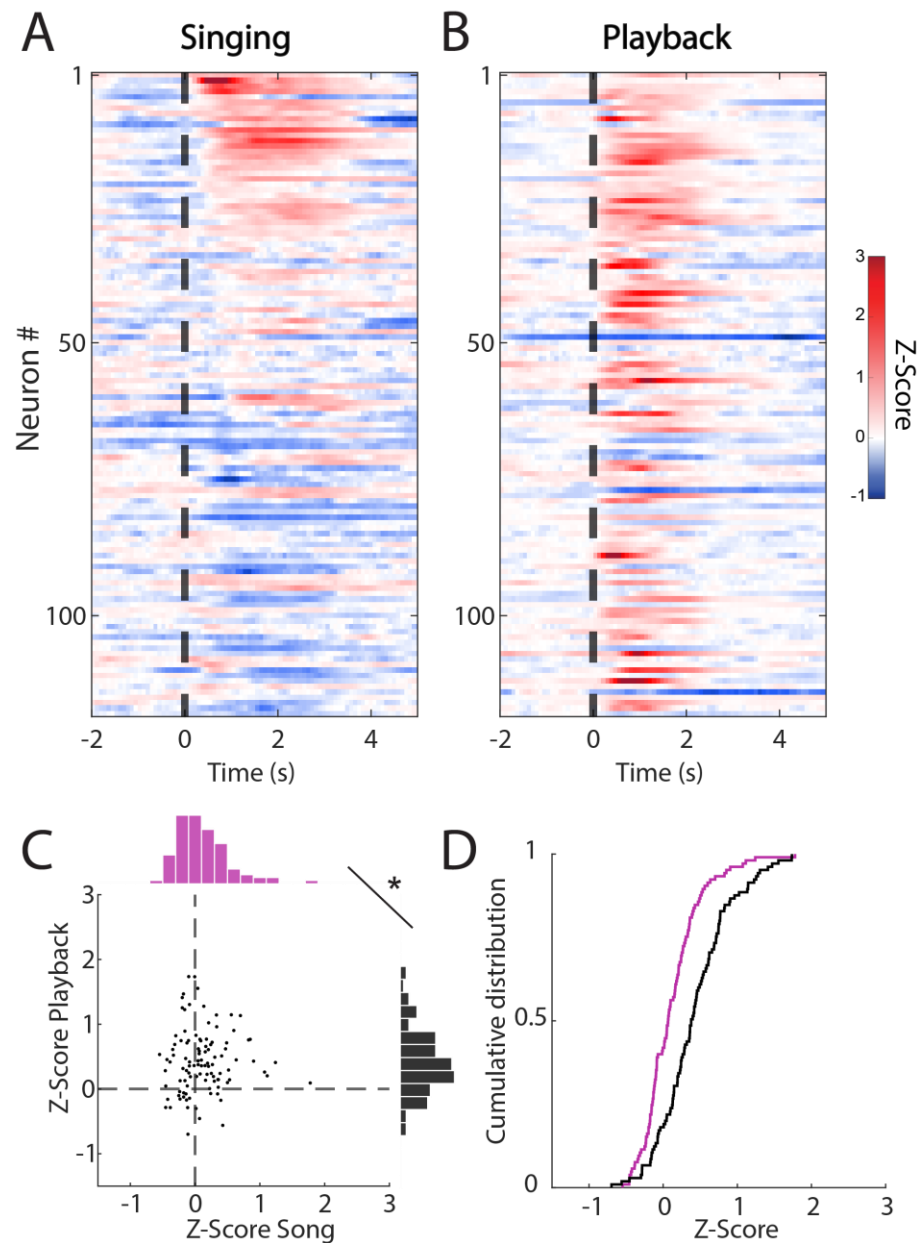
The anatomy of Field L makes it challenging to distinguish among sub-regions when imaging with a large lens. For this reason, I targeted L1, L2, and L3 collectively with viral injections to express GCaMP6s pan-neuronally. Field L and CML are categorized as the primary auditory cortex; for this reason, it was unsurprising to find very similar responses in Field L to song and playback as those reported in CML. Here, most neurons were also suppressed during singing but less so compared to CML-Av ( $n=115$  neurons, 3 birds, Figure 10 A-B, Figure 12 A). This suppression was also shown by a significant shift in the distributions of activity caused by an increase in suppression and a decrease in excitation (Figure 10 C-D, Figure 12 A, Wilcoxon rank sum test,  $p=5.45e-05$ ).

NCM receives direct inputs from CML and Field. However, it showed a different modulation compared to Field L and CML-Av. The most striking difference was the increase in activity to playbacks of song (Figure 11 A-B). This result matches previous findings on higher song selectivity in this region compared to Field L and CML. Activity during singing was also suppressed compared to playback, but this suppression was not as strong as in Field L and CML. However, the differences between singing and playback still separated the distributions of averaged activity (Wilcoxon rank sum test,  $p = 2e-07$ , Figure 11 C,D).



**Figure 10. Activity in Field L during singing and playback.**

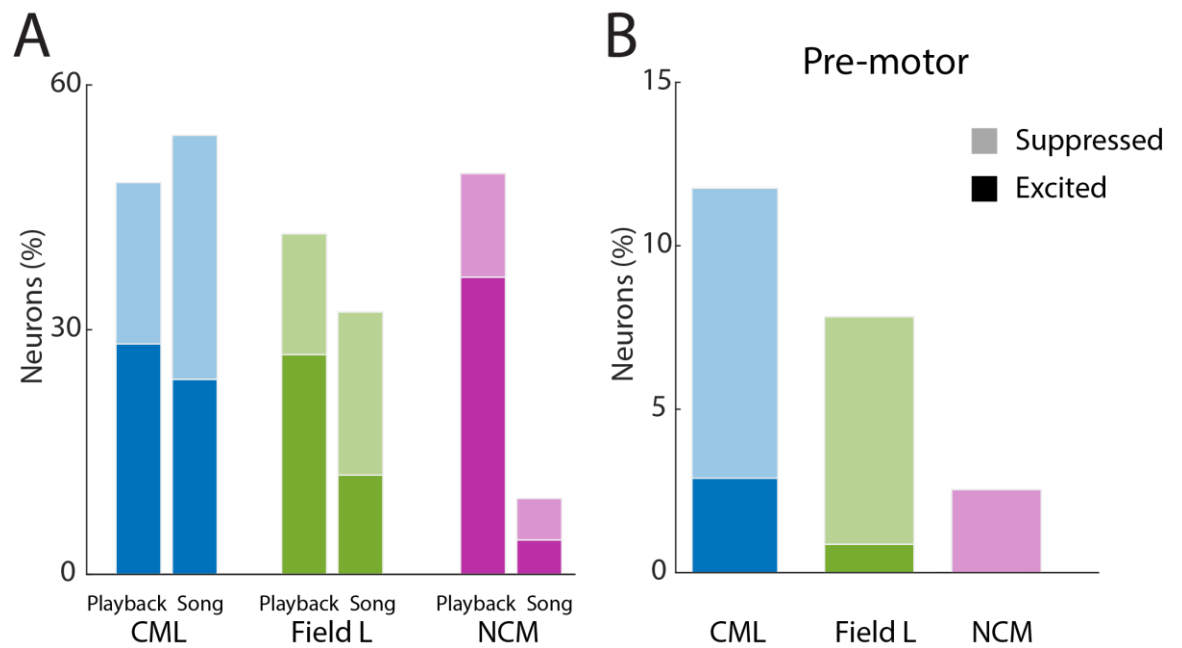
**A-B.** Heatmaps showing the averaged activity across trials during playback and singing sorted by activity during singing ( $n = 115$  neurons from 3 birds). **C.** Scatter plot of averaged activity in individual neurons during the first 2 seconds of either song (green) or playback (black) with a projection of their respective distributions (Wilcoxon rank sum test,  $p = 5.45e-05$ ). **D.** The cumulative distribution between conditions shows more clearly the shift towards suppression.



**Figure 11. Activity in NCM during singing and playback.**

**A-B.** Heatmaps showing the averaged activity across trials during playback and singing sorted by activity during singing ( $n = 118$  neurons from 2 birds). **C.** Scatter plot of averaged activity in individual neurons during the first 2 seconds of either song (magenta) or playback (black) with a projection of their respective distributions (Wilcoxon rank sum test,  $p = 2e-07$ ). **D.** The cumulative distribution between conditions shows more clearly the shift towards excitation for playbacks.

When comparing across regions, I found that the effects of a corollary discharge signal were well represented by the anatomical connectivity. CML had the largest proportion of neurons that were suppressed during singing and the largest proportion of neurons modulated during singing (Figure 12A). Field L showed similar trends to CML but in a decreased proportion. The modulation of activity during singing decreased the most in NCM, supporting previous findings that this region is primarily used to recognize behaviorally relevant stimuli rather than evaluate performance. An important finding was that a higher proportion of neurons were modulated in a premotor window in CML compared to Field L and NCM. This supports the idea that a corollary discharge signal modulates CML, and this signal may be then transmitted to Field L and NCM (Figure 12B).

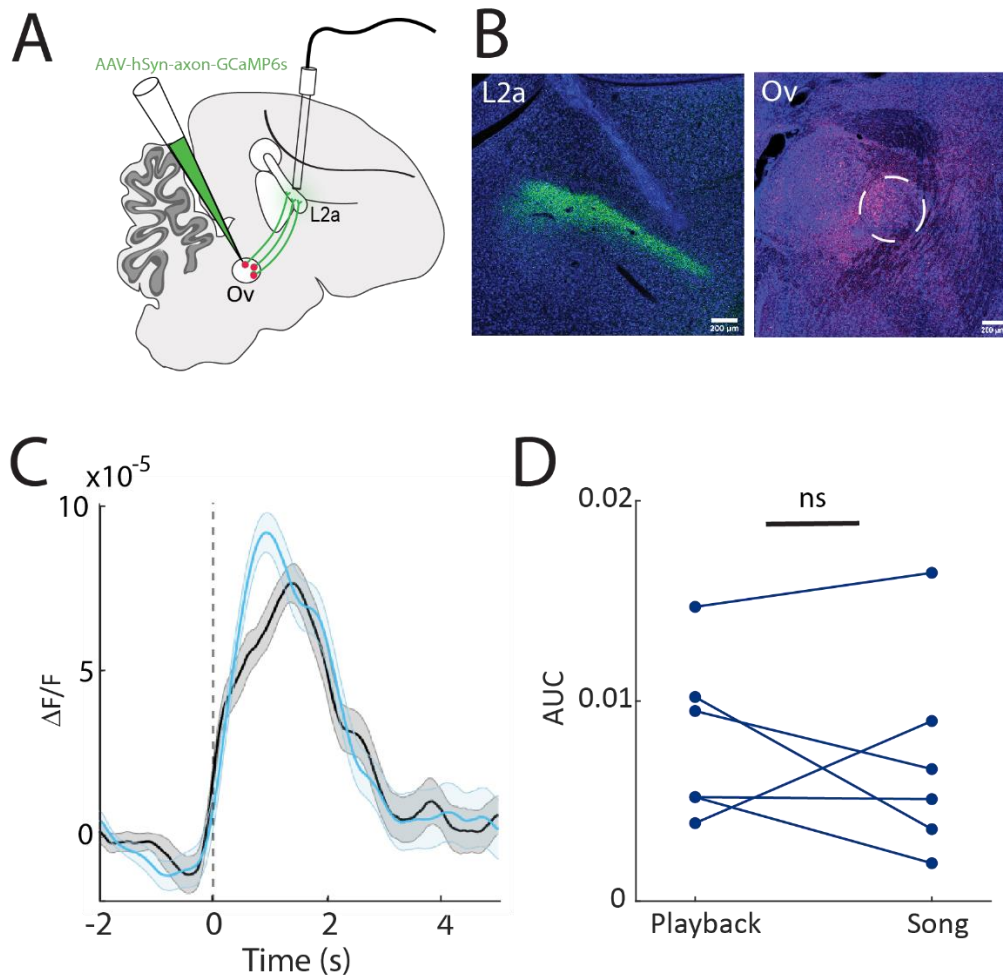


**Figure 12. Activity modulation across the auditory hierarchy.**

**A.** Differentially modulated neurons during song and playbacks relative to baseline for every auditory region (Wilcoxon rank sum test  $p < 0.01$ ). **B.** Differentially modulated neurons during an 800ms window before the onset of the first intronote (Wilcoxon rank sum test  $p < 0.01$ ). Darker colors represent the proportion of neurons that were excited, while lighter ones represent suppressed neurons.

#### **2.2.4 Modulation of activity during singing emerges in the auditory cortex**

Corollary discharge signals can modulate activity at different points along the auditory pathway. One possibility is that this suppression simply arises earlier in this pathway and is just conveyed to the auditory cortex. To determine whether the suppression emerged upstream of the primary auditory cortex, I used a viral construct to express GCaMP6s exclusively in axon terminals. I injected the virus in the thalamic auditory nucleus Ov and imaged these terminals in Field L2a using fiber photometry (Figure 13 A-B). To compare these conditions, I calculated the area under the curve for the first 2 seconds of activity after singing or playback onset. These experiments revealed that there was no significant difference between activity during singing and during playback in these thalamic terminals ( $n = 6$  hemispheres, 4 birds, Wilcoxon tank sum test,  $p = 0.46$ , Figure 13 C-D).



**Figure 13. Activity modulation in thalamic terminals.**

**A.** Schematic of fiber photometry approach to image Ovoidalis (Ov) terminals in FieldL2a. **B.** Expression of axon targeted GCaMP6s in L2a (green) and Ov cell bodies (red). **C.** Representative traces of axonal activity during singing (light blue) and playback (black, mean $\pm$ sem). **D.** Area under the curve calculated for the calcium traces during both conditions ( $n = 6$  hemispheres, 4 birds, Wilcoxon rank sum test  $p = 0.46$ ).

## 2.3 Conclusions

Studies in mammals, including the mouse, the marmoset monkey, and humans, have reported an effect of suppression in the auditory cortex caused by a corollary discharge signal (Curio et al., 1998; Eliades & Wang, 2003; Harmon et al., 2024). Here, I have shown for the first time that the primary auditory cortex of zebra finches is suppressed during singing. This

suppression is predictive as it starts before song onset, providing strong evidence that modulation of activity is influenced by a vocal-corollary discharge signal. This premotor modulation occurred in a larger proportion of neurons in CML-Av. This is the only auditory region that receives direct input from the premotor nucleus HVC. One possibility is that this excitatory projection activates a small population of neurons that then drives a strong suppressive effect that then gets transmitted to other auditory regions. Carrying out latency analysis across these regions would confirm this. However, this type of analysis is limited by the lower yield of neurons that I collected in Field L and NCM (9 and 3 neurons, respectively, that were modulated in a premotor window), as well as the slow kinetics of the calcium sensor that cannot provide an accurate measurement for onset of suppression (T. W. Chen et al., 2013; Forli et al., 2018).

While suppression dominated the modulation, I also found diverse responses across the population that were differentially elicited during singing and playback, demonstrating that activity in the primary auditory cortex cannot be explained by purely acoustic signals. The heterogeneity of responses to vocalizing and playback is an effect that had not been previously reported in mammals. One possible explanation for this is that zebra finches require a larger population of neurons to represent their rich learned vocal repertoire. Additionally, these neurons may also serve different functions when evaluating auditory feedback, which will be the focus of the next chapter.

A caveat from these experiments is that discrepancies may exist between the playback stimulus and what the bird hears while singing. This could be caused by differences in the amplitude of the stimulus or by a different representation of frequencies attributed to bone conduction. The amplitude of the playback stimuli was calibrated to match those of the recorded song. However, an analysis of a dataset that I collected with different decibel levels should clarify this point in the future. Bone conduction is an effect that enhances the representation for low

frequencies and is much harder to control. While the playbacks may have a decreased representation of low frequencies, they still maintain the majority of the spectrotemporal features in the original song, making it unlikely to account for the large differences seen in the modulation of activity between singing and playback.

It is important to note that the results presented in this chapter differ from those reported in the study by Keller and Hahnloser, where they showed no differences in activity during singing compared to playbacks of the bird's song (Keller & Hahnloser, 2009). There are several reasons why these discrepancies may exist. First, differences between recording methods. Electrophysiological recordings have a much better time resolution but may have a bias to sample neurons that have higher firing rates (Wei et al., 2020). Calcium imaging allows me to obtain a higher yield of neurons per individual bird. While this number is still much lower than what can be obtained in optimized models like the rodent, collected datasets of over 100 neurons are extremely rare in the songbird field. Another likely source for the discrepancy between my results and those published before is the age at which the birds were recorded. Keller and Hahnloser recorded in young birds that were still in the sensorimotor phase (60-90 dph). At this age, the birds are still in the process of learning and have not crystallized their song. The modulation of activity that I observe may occur when birds crystallize an updated template of their own song (rather than their tutor song) and learn to suppress this activity predictively. Carrying out longitudinal recordings during the transition period between sensorimotor and crystallization periods would be a way to test this.

My findings on the degree of suppression in Field L and CML compared to that in thalamic terminals suggest that suppression arises via a local mechanism in the auditory cortex. While my fiber photometry experiments showed slightly different trends across birds, other

studies have also provided evidence for equally elevated activity during both singing and playback with multiunit recordings in Ov (Lei & Mooney, 2010).

## ***2.4 Methods***

### **2.4.1 Surgical procedures**

Adult male birds were anesthetized with inhaled 1-2% isoflurane and placed on a stereotaxic apparatus with a temperature-regulated surgical bed. The area of the incision was sterilized with alcohol, and anesthetized locally (bupivacaine 0.25%) before cutting the skin over the skull. Coordinates for the injections and lens implant were measured from the bifurcation of the midsagittal sinus. Birds were injected with 200nL of an AAV2/9 CAG GCaMP6s (Addgene, 100844-AAV9) viral construct on each site at a speed of 9nl/s using a glass needle attached to a Nanoject-II (Drummond Scientific). The coordinates used for targeting various auditory regions were as follow: Field L, head angle of 40, Injection Site (IS) 1: 1.63A, 1.5L, 1.79V, IS 2: 1.6A, 1.6L, 1.59V, IS 3: 2.0A, 1.55L, 1.45V, IS 4: 2.3A, 1.6L, 1.3V, IS 5: 2.25A, 1.5L, 1.2V; CML, head angle of 43, IS 1 and 2: 1.8A, 1.95L, 0.7 and 0.9V, IS 3 and 4: 2.0A, 2.05L 0.65 and 0.95V, IS 5 and 6: 2.2A, 1.95L 0.7 and 0.9V; NCM, head angle of 35, IS 1 and 2: 0.7A, 0.5L 1.6 and 1.8V, IS 3: 0.8A, 0.6L, 1.7V, IS 4 and 5: 0.9A, 0.5L, 1.6 and 1.8V. Immediately after the injections, I implanted a GRIN prism lens (Inscopix 1050-004601) held by a vacuum holder at a low speed (5-10um/s) or an integrated prism lens. After implantation, the lens was cemented to the skull and covered using silicone rubber (Body-Double Fast Set) to prevent damage. In birds that were only implanted with the lens, 3-5 weeks after the lens implant, I examined the field of view and cemented a baseplate (Inscopix 1050-004638) to hold a 1-photon miniscope (nVista 3 system, Inscopix). After 2-3 days of surgery recovery, the bird was attached to the miniscope with a counterweight to allow normal behavior.

### **2.4.2 Calcium imaging**

Imaging data was acquired with an nVista 3.0 data acquisition board and the Inscopix imaging software at a frame rate of 12 Hz, using an LED power of 1-1.4mW/mm<sup>2</sup>. The acquisition board interfaced with custom Matlab code to monitor singing (Mor Ben-Tov). This allowed the computer to detect changes in amplitude (i.e., when the bird started to sing), which triggered the imaging system for sessions of 1 minute and recorded the audio data. A refractory period of 5-30 minutes was set to prevent constant song-triggered imaging from photobleaching of the GCaMP fluorophore. The playbacks of BOS were triggered manually throughout the day.

The imaging files recorded within a day were curated by concatenating all the sessions that contained singing or playbacks together, spatially downsampling the videos, performing motion correction, and extracting the ROIs/neurons using CNMFe (Zhou et al., 2018). The criteria for non-inclusion comprised sessions that had dropped frames or sessions that were accidentally triggered and did not contain song or playback.

The audio recordings containing song or playback were manually labeled to determine onsets and offsets. To avoid contamination of the playback or song signal due to the low calcium sensor decay, only the songs or playbacks that had at least 4 seconds of silence were labeled. Then, the extracted traces were aligned to the onset time stamps, z-scored, and averaged across trials. The data was z-scored with the Matlab zscore function using the CMNFe output signal, which contained concatenated calcium traces from one imaging day.

### **2.4.3 Histology and in situ hybridization**

Birds were perfused transcardially with 4% PFA/PBS and fixed in the same solution overnight and then transferred to 30% sucrose in RNase-free PBS for 2 overnights at 4C. Brains were then sectioned at 75  $\mu$ m and collected into 0.5-1% PFA. Slices were washed twice in PBS

for 3 min, incubated in 5% SDS/PBS for 45 min, rinsed twice with 2x sodium chloride sodium citrate 0.1% Tween 20 (2x SSCT), and put in 2x SSCT for 15 min on a shaker at room temperature. They were pre-incubated in probe hybridization buffer for 30 min at 37C, and later hybridized in 2.5 mL probe set/500 ml probe hybridization buffer overnight at the same temperature. The probes were custom made by Molecular Instruments to detect zebra finch isoforms of VGluT2, and VGAT. The next day, the slices were washed four times for 15 min with 500 uL of probe wash buffer at 37C and twice in 2x SSCT for 5 min. at room temperature on a shaker. Then they were incubated in 500 ul of HCR amplification buffer for 30 min at room temperature on a shaker. Finally, the slices were incubated in a solution containing 300 uL HCR amplification buffer and fluorescent hairpins for the HCR initiator probe for 2 nights in the dark at 25C. On the last day, at room temperature, slices were washed twice with 2x SSCT for 5 min, stained with Neurotrace for 2 hours (1:500, N21479; Invitrogen), rinsed twice with 2x SSCT and mounted on a slide with Fluoromount-G. Hairpins, probe sets and probe hybridization buffer were created by Molecular Instruments.

#### **2.4.4 Playback stimuli**

Song was recorded from experimental birds in an acoustically isolated chamber before and after lens implant surgery using the software Sound Analysis Pro. Song bouts that contained 2-3 motifs were randomly selected and filtered to remove background noise and normalized to -6SPL using Adobe Audition. These 3-5 selected bouts were then concatenated with an inter-stimulus interval of 4-6 seconds between bouts to make up a playback stimulus of ~1 minute long. In a subset of birds, playback stimuli also included calls, tutor song, and conspecific song that were pre-processed in the same way. The auditory stimuli were played through a speaker at a 78 dB peak amplitude level as measured by a sound meter. This amplitude level was selected to be 5-10 dB larger than the amplitude levels produced during singing by the birds as captured by

the recording system. This increase in the amplitude of playbacks aimed to reduce the possible differences that arise between perceived song feedback and heard playback.

#### **2.4.5 Fiber photometry**

The surgical preparations were the same as those described for calcium imaging. A pAAV-hSynapsin1-axon-GCaMP6s-P2A-mRuby3 (axon targeted GCaMP6s) viral construct was injected into the auditory thalamus Ov using the following coordinates: head angle of 55, 3.45A, 0.9L, 4.79 and 4.83V. 200nL of virus was injected at each injection site. In the same surgery, optic fibers were implanted either uni- or bi-laterally into Field L2a using the following coordinates: head angle of 25, 1.4A, 1.2L, 1.7V, or head angle of 50, 1.65A, 1.2L, 1.6V. These coordinates were used on each hemisphere to be able to image both hemispheres simultaneously. The fibers were cemented using Metabond, the skin was pulled over the cement and glued using VetBond to prevent skull exposure.

After 3-4 weeks of viral expression, birds were imaged using a fiber photometry system (FP3001, Neurophotometrics). The calcium sensor was excited with two wavelengths, 470nm for the calcium-dependent signal and 415nm as an isosbestic control of movement artifacts and to generate a control signal. Specifically, the isosbestic signal was smoothed and regressed to the 470nm channel calcium-dependent signal. Then, the model obtained from the regression was used to generate the control predicted signal. Subtracting the recorded raw calcium signal from the predicted signal normalized the calcium signal into a  $dF$  measurement (Mor Ben-Tov).

#### **2.4.6 Movement analysis**

The Inscopix data acquisition system generates files related to events (GPIO) and movement (IMU) that are recorded throughout the entire day. To process IMU files, custom code written in Python was run to calculate Euclidean distances based on the x, y, and z coordinates

measured by the accelerometer. The signal was high-pass filtered and thresholded to obtain movement events. Then, the times when a recording was triggered based on the GPIO files were extracted to obtain the movement during recording times and resampled to match the calcium signal. On the processed movement signal, an absolute value transformation was performed to focus on changes in movement amplitude, and values were normalized in a range from 0 to 1. The calcium traces were also in the same range for individual neurons to then calculate the Pearson correlation coefficient between the normalized calcium signal and normalized movement.

#### **2.4.7 Statistics**

Before carrying out tests to compare the distributions of activity (scatter plots and cumulative distributions), the distributions were tested for normality using a Kolmogorov-Smirnoff test. Based on this result, the appropriate statistical measure was used; in the case of this chapter, all the distributions required non-parametric statistics, I used a Wilcoxon rank sum test to test differences. I used this same statistic when determining the modulation of neurons by comparing activity between the epoch of interest (song vs playback) and baseline. For this comparison, I picked a 2.5-second window from the onset of singing or playback and compared it to a 2.5-second window of silence prior to singing. The baseline window was set to -3.5 to -1 seconds before the onset of singing to avoid including activity that may have been related to a premotor process. For the premotor analysis, I picked an 800-millisecond window before the onset of singing based on distributions of vocal premotor modulation in other model systems (Eliades & Wang 2003; Harmon et al., 2024).

For fiber photometry statistics, I also used the Wilcoxon rank sum test due to the small sample size. Generally, I applied non-parametric statistics when sample sizes were <30.

### **3. The primary auditory cortex evaluates auditory feedback and detects errors**

#### ***3.1 Introduction***

The discovery of error signals in the auditory cortex of zebra finches and their influence on VTA activity proposes a neural pathway through which birds can evaluate and implement corrections of vocal performance. The transmission of this signal is necessary because disrupting the output through AIV hinders song learning and flexible vocal performance (Kearney et al., 2019; Mandelblat-Cerf et al., 2014). However, what this error signal precisely transmits or encodes is still unclear. Error signals in the auditory cortex have only been studied in the context of external auditory feedback distortion, and previous studies in zebra finches only characterized individual neuron responses to the stimulus without providing any context related to the neuron's tuning to determine what acoustic features it is integrating and evaluating.

In juvenile birds, dopaminergic release in the song basal ganglia is directly linked to the learned quality of song performance on a rendition-to-rendition basis and likely drives sensorimotor learning (Qi et al. 2024). These mechanisms are preserved in adults, where cell body spiking of dopaminergic neurons in the VTA that project to the song basal ganglia can also track natural variations of the song (Duffy et al., 2022). These findings place VTA neurons as potential integrators of multi-dimensional acoustical features that are then used to transmit information on song quality. Whether neurons in the primary auditory cortex encode these features during song production and track variability in the acoustical space has not been explored. This could be important to better understand exactly what sensory coordinates the bird is using to guide song learning and maintenance.

In this chapter, I used calcium imaging to describe responses to distorted auditory feedback in the primary auditory cortex (CML-Av) of singing birds. Additionally, I quantified the

acoustic features of the song using a variational autoencoder (VAE) as well as canonical song feature analysis. I used these acoustical analyses to correlate natural variability in song to neural activity to better understand what neurons in the auditory cortex are encoding during singing.

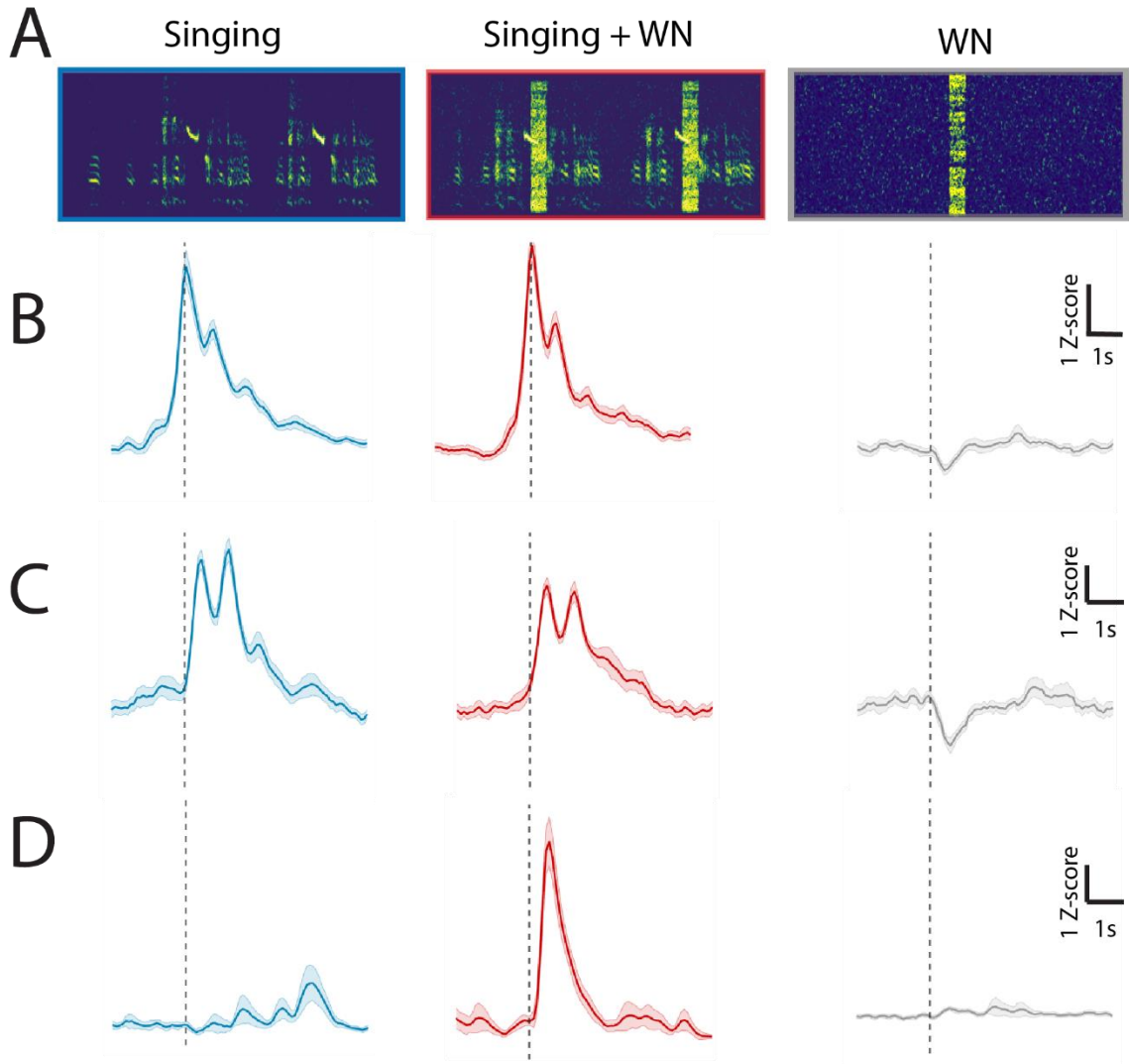
## **3.2 Results**

### **3.21. Error signals are generated in the primary auditory cortex**

To explore whether neurons in the auditory cortex of adult birds can detect mismatches between expected auditory feedback and real-time auditory feedback, I used a closed-loop system that detects song and targets a syllable with a short playback (Tumer & Brainard, 2007). Using this system, I targeted a single syllable of the zebra finch stereotyped song to trigger a short burst of white noise with a 50% probability (Figure 14A). To discard the possibility of neurons being driven by the white noise or by the song's acoustic features, rather than computing a mismatch, I carried out playbacks of isolated white noise.

I identified three distinct types of responses across the population during distorted auditory feedback. First, there were neurons that showed consistent modulation, either excitation or suppression, during both normal singing and when subjected to distorted auditory feedback, independently of their responses to white noise playbacks (Figure 14B). Second, neurons that integrated information linearly, the way they responded to white noise produced an additive or subtractive effect during distorted auditory feedback (Figure 14C). Lastly, I found neurons that behaved as error detectors; these neurons did not respond to white noise but exhibited modulation

during distorted auditory feedback that was different from that elicited by singing (Figure 14D).

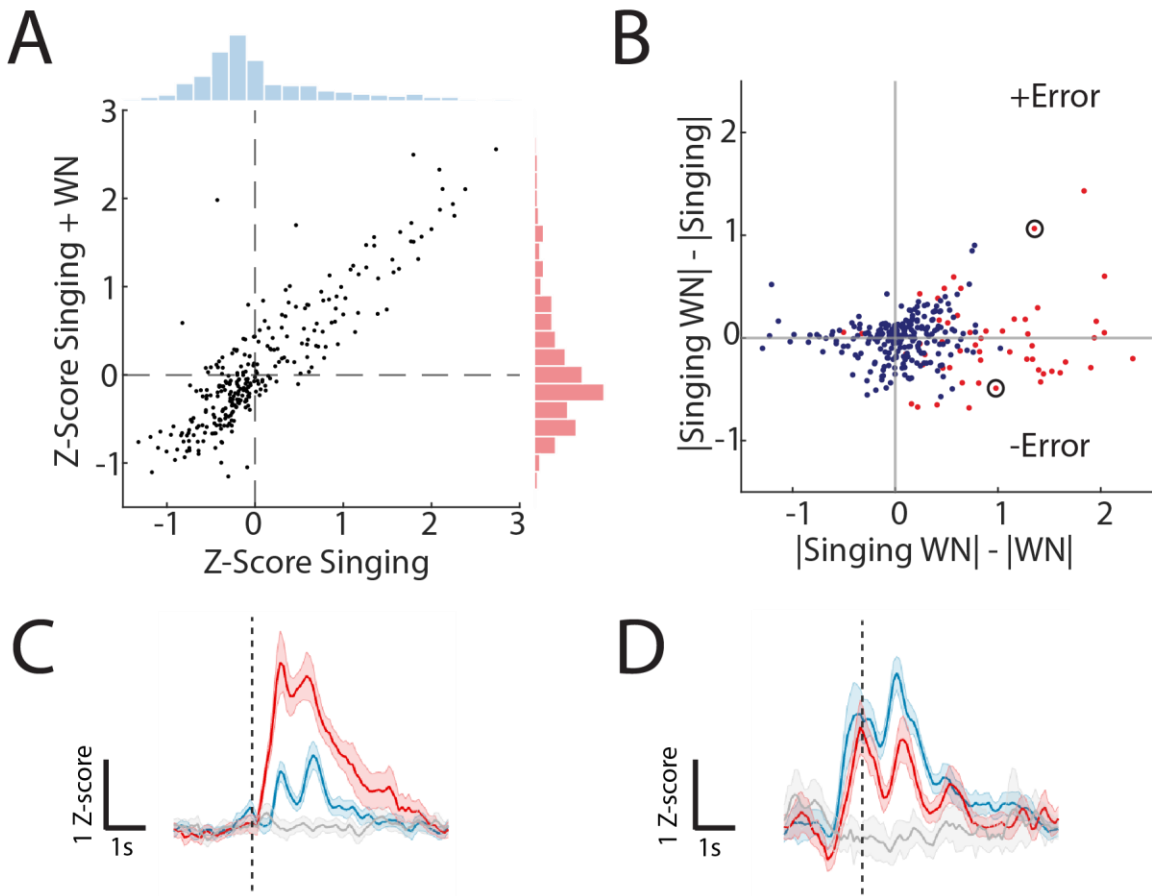


**Figure 14. Representative responses to white noise perturbation.**

**A.** Representation of the experimental conditions, white noise (WN) happened with a 50% probability in randomly selected songs. Spectrograms and neurons are not in the same time scale. **B.** Example of a neuron that is equally modulated during singing and singing with white noise. **C.** Example of a linear neuron that reduces its response when white noise is triggered and is suppressed by white noise alone. **D.** Error neuron that increases its firing rate in the presence of white noise without responding to white noise alone. All neurons are aligned to the onset of the targeted syllable (columns 1-2) or onset of the white noise (column 3).

To highlight these neurons across the population, I plotted averaged responses 1.5s after the onset of the targeted syllable with or without white noise (Figure 15A,  $n = 283$  neurons, 6

birds, Wilcoxon rank sum test  $p=0.39$ ). While there were no differences in the population, this scatter plot showed a small population of neurons that deviated from the diagonal trend of equal modulation. These neurons could potentially act as error detectors. This scatter plot fails to capture whether these neurons may be modulated by white noise to elicit this type of response. To differentiate between neurons that were additive or subtractive from true error detectors, I carried out an absolute value transformation in the responses and subtracted either their response during normal singing or during isolated white noise playback (Figure 15B). This subtraction allowed me to identify neurons that changed their activity exclusively when the bird sang and received white noise. Furthermore, the absolute value transformation allowed me to focus on the magnitude of the modulation rather than the sign. These plots revealed two different types of error responses: those that produced an increase in activity in the presence of white noise (Figure 15C) and those that produced a decrease in activity in the presence of white noise (Figure 15D). Additionally, neurons that either had no modulation or were excited during singing were more likely to produce error-like responses, while suppressed neurons were equally modulated during singing and singing with white noise or were simply linear integrators.



**Figure 15. Population activity to distorted auditory feedback with white noise.**

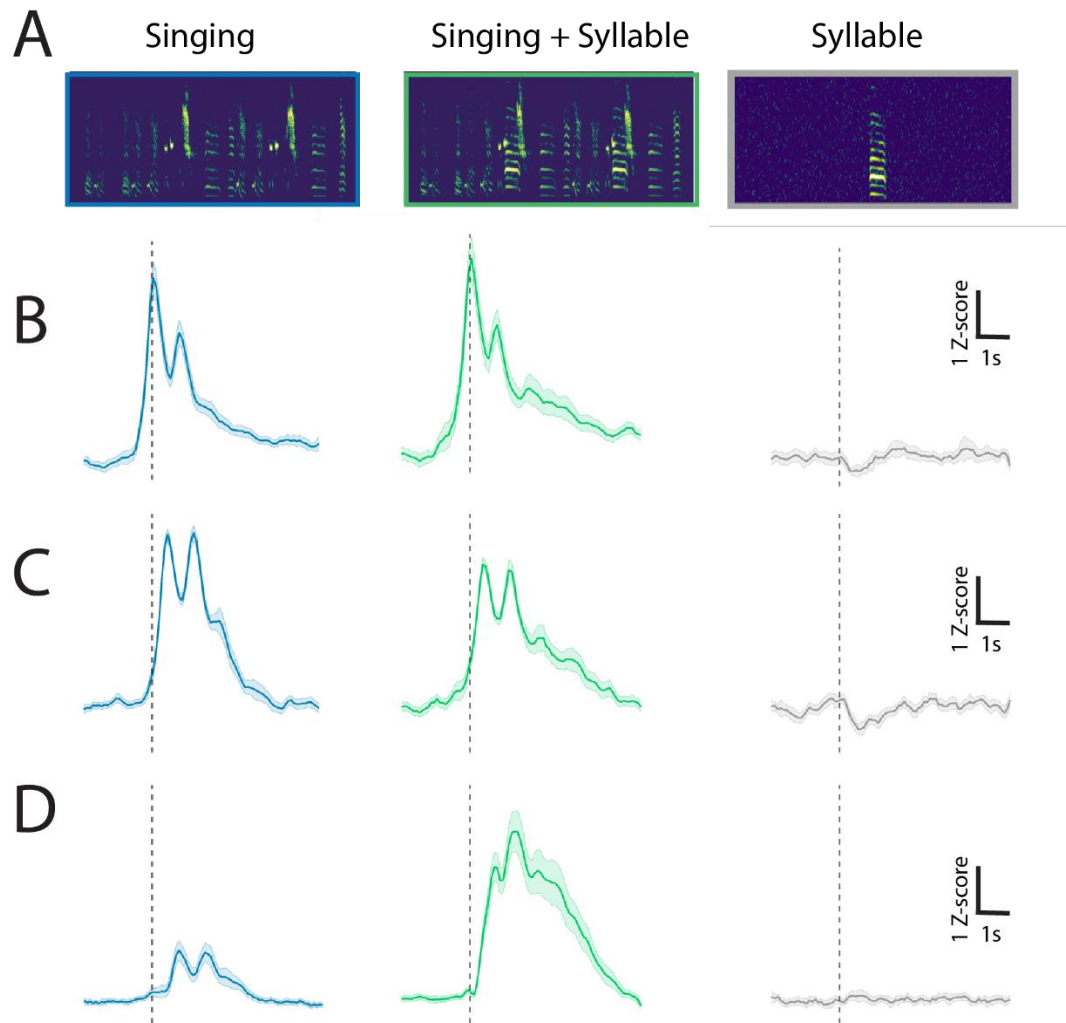
**A.** Scatter plot of averaged responses of the first 1.5s after target syllable onset during singing and singing with white noise ( $n=283$  neurons, 6 birds, Wilcoxon rank sum test  $p=0.39$ ). **B.** Averaged absolute value differences across conditions. Dark blue represents neurons that were suppressed during singing + WN and dark red represents neurons that were excited in the same condition. **C.**

A “positive” error neuron, highlighted in B by a black circle. **D.** A “negative” error neuron, highlighted in B by a black circle.

### 3.2.2 Error signals are modulated by different perturbations

White noise is commonly used to distort auditory feedback. However, the broadband representation of frequencies could make it a salient and even aversive stimulus that may influence activity. To explore this issue and to determine whether neurons generalized their responses to other acoustic contexts, I carried out the same distorted auditory feedback experiment but used a more naturalistic stimulus, a syllable of the bird’s song repertoire (Figure

16A). With this type of manipulation, I found a small subset of neurons with the same modulation in both contexts (Figure 16B-D).



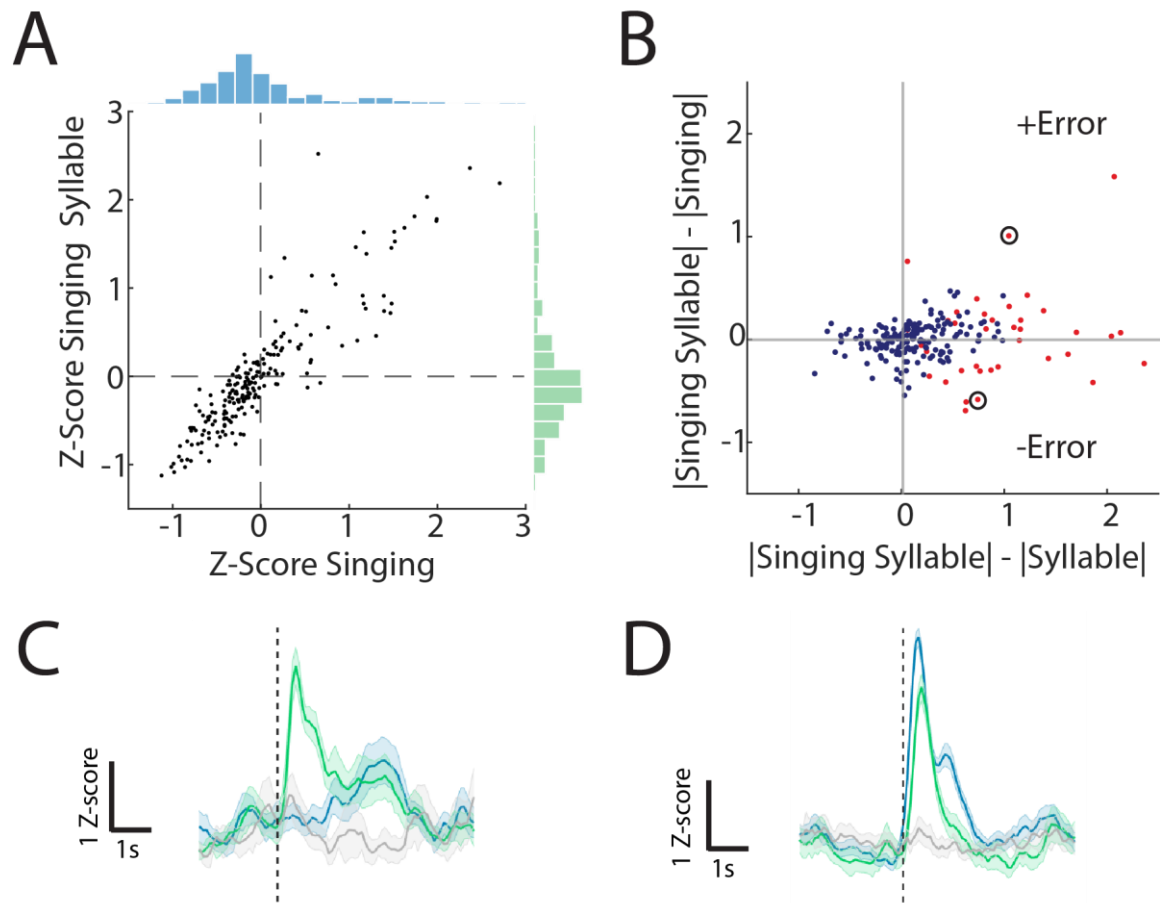
**Figure 16. Representative responses to syllable perturbation.**

**A.** Same as Figure 14. Here, a random syllable from the bird's song was used as a playback stimulus. Spectrograms and neurons are not in the same time scale. **B.** Example of a neuron that is equally modulated during singing and singing with syllable. **C.** Example of a linear neuron that integrates its response to the syllable. **D.** Error neuron that increases its activity in the presence of the syllable without responding to it alone. All neurons are aligned to the onset of the targeted syllable or onset of the playback syllable.

However, when carrying out the same analysis as with white noise perturbation, I found that responses were more similar between singing and singing with a syllable overlaid, as shown

in the narrower distribution in Figure 17A (n=227 neurons, 5 birds). The number of neurons that deviated along the axis of either positive or negative error was also smaller (Figure 17B).

Suggesting that the acoustic features of the perceived auditory feedback matter and can influence activity differentially.

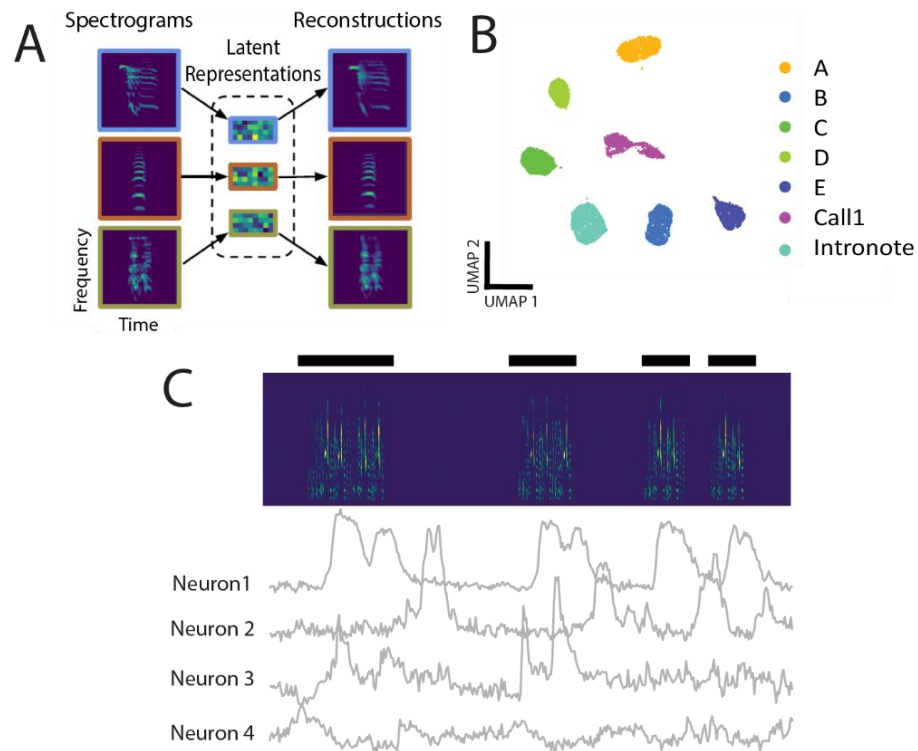


**Figure 17. Population activity to distorted auditory feedback with syllable.**

**A.** Scatter plot of averaged responses of the first 1.5s after target syllable onset during singing and singing with white noise (n=283 neurons, 6 birds, Wilcoxon rank sum test  $p=0.667$ ). **B.** Averaged absolute value differences across conditions. Dark blue represents neurons that were suppressed during distorted feedback, and dark red represents neurons that were excited in the same condition. **C.** A positive error neuron, highlighted in B by a black circle. **D.** A negative error neuron, highlighted in B by a black circle.

### **3.2.3 Neurons show tuning to song in a high-dimensional space**

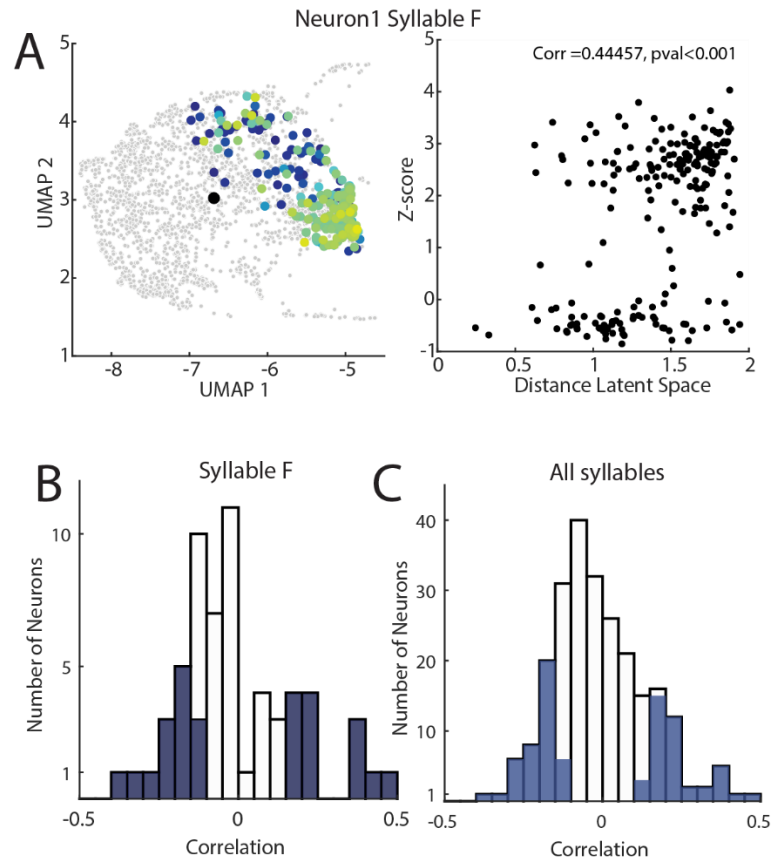
To better understand exactly what type of information the auditory cortical neurons may be encoding during singing, I turned to their natural song. In collaboration with Dr. Scott Smyre, we trained a VAE with data that was collected from all the imaging days of 6 birds (25-35 imaging days per bird). This approach allowed us to have all the vocal renditions in the same high-dimensional space to make comparisons across birds more comparable. After manually selecting the segmentation parameters, the VAE was trained, and we obtained the latent representations or features for all vocal renditions together with their reconstructions (Figure 18A). We then used a uniform manifold approximation and projection (UMAP) to embed the 32 latent features in a two-dimensional space (McInnes et al., 2018). This embedding allowed us to visualize and assign a syllable identity to different clusters (Figure 18B).



**Figure 18. Methods to quantify song.**

**A.** Graphical representation of the VAE model. An encoder network reduces a high-dimensional spectrogram into 32 latent representations. Then, a decoder network reconstructs the spectrogram from the latent representations to determine accuracy, **B.** Two-dimensional embedding of the 32 latent features using a UMAP. **C.** Example spectrogram that contains four song bouts and four representative calcium traces of individual neurons aligned to the audio.

To relate these metrics to neural activity, I analyzed the days that had the most singing renditions. For one bird, I aligned the calcium traces for all neurons to the audio files by picking the frame closest to when the syllable occurred and obtained that normalized calcium value. UMAP embeddings were informative when looking at how calcium activity was modulated in this space (Figure 19A). To relate the activity and the song features, I obtained a measurement of song variability in individual syllables, I calculated the Euclidean distance in the 32-latent space between each syllable rendition and the median of all the renditions. With the Euclidean distance metric and the calcium activity values, I calculated the Pearson correlation coefficient (Figure 19A, right).



**Figure 19. Analysis of calcium activity and latent features.**

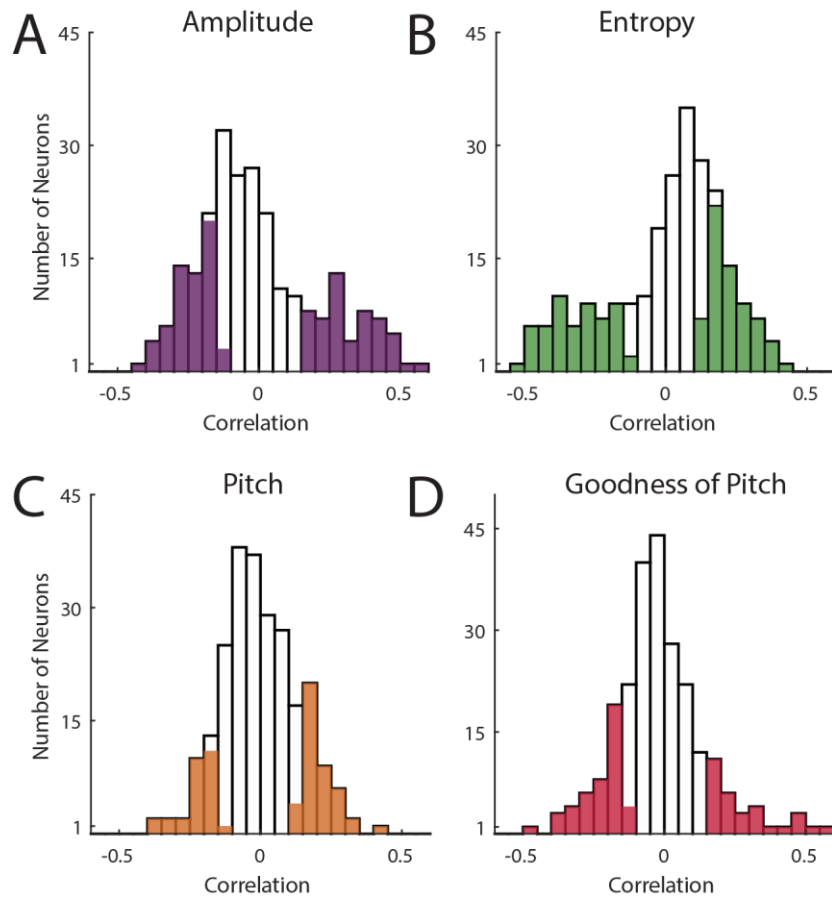
**A.** UMAP embedding of syllable F for one day of recording. Each data point is a single-syllable rendition colored by the calcium activity during that rendition (yellow excitation, blue suppression). Right. Scatter plot showing how the correlation coefficients were calculated between the calcium activity and the distance in latent space. **B.** Distribution of the correlation coefficients for the syllable shown in A. The filled histograms mark the correlation coefficients that showed a  $p\text{-value} < 0.01$ . **C.** Distribution of correlation coefficients extended to all syllables in one bird.

I first assessed this analysis in all the neurons of a single syllable for a single bird (Figure 19B); here, I found significant correlations between the measurements suggesting that the auditory cortical neurons track changes in syllables that largely deviate from median renditions. To assess whether these correlations were meaningful, I carried out the same correlation analysis but shuffled the calcium imaging frames picked. Even with a small number of iterations of the shuffling procedure, the correlation distributions across syllables were narrower,

and increasing the number of iterations narrowed them even more, supporting the finding that these correlations are related to song and do not occur randomly.

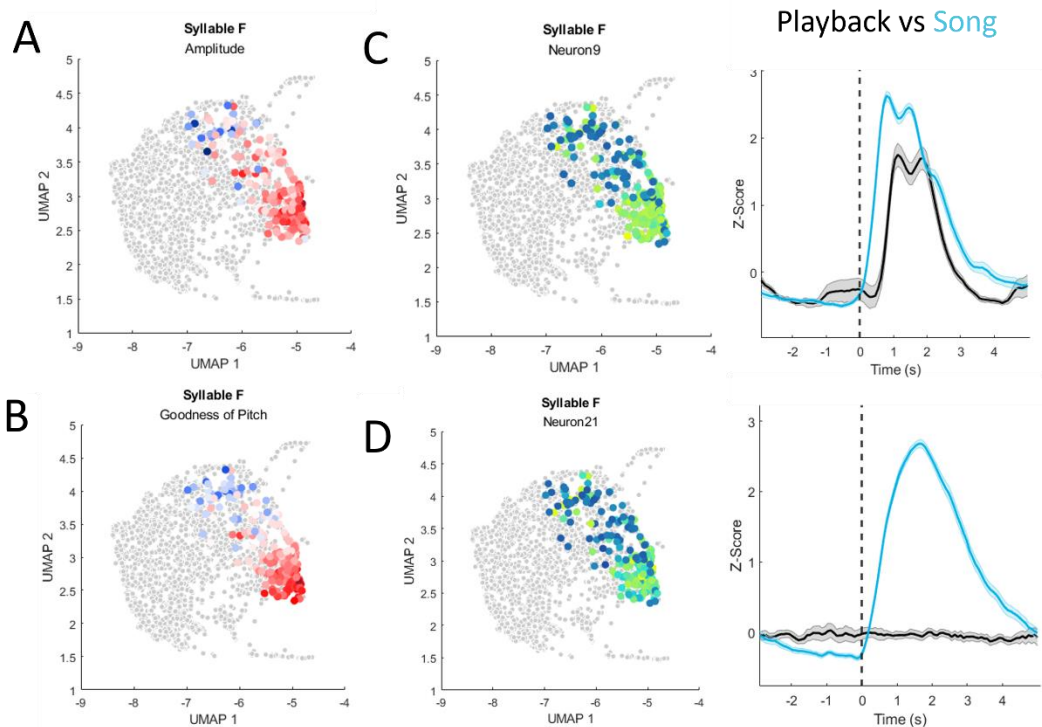
### **3.3.4 Neurons tune to acoustical features during singing**

Auditory neurons show tuning to different spectrotemporal features. To explore whether this was something that neurons tracked during vocal performance. I used four acoustic features that have been previously reported to be encoded by auditory neurons: amplitude, entropy (how broadband or tonal the syllable is), pitch (estimated fundamental frequency), and goodness of pitch (harmonicity). I calculated each one of these features for individual syllables that were recorded in a span of 25-35 days. Similar to the analysis in the previous section I then calculated the Pearson correlation coefficient between the calcium activity for all neurons and the individual song features to determine whether there was any relationship between them (Figure 20A-D). With this analysis, I found that many neurons correlated with at least one of these measures. Using the VAE was useful to determine whether neurons were tuned to more than one of these features, and the UMAP embeddings were informative when visualizing the trends for acoustical features. Specifically, it allowed me to determine that there are neurons that show tuning to different spectral features (Figure 21A-B). Importantly, this tuning can occur during both playback and song, reflecting a purely auditory process (Figure 21C), or it can occur only during singing (Figure 21D), suggesting that these neurons exclusively evaluate feedback. This analysis is limited in power and is preliminary, given that I have only carried it out for one bird. However, it shows a promising way of assessing what these neurons are encoding during singing.



**Figure 20. Correlation distributions across song features.**

**A-D.** Pearson correlation coefficients were calculated between the calcium activity values and the acoustic features indicated for individual syllable renditions. Shaded bars indicate the correlations that had a p-value < 0.05.



**Figure 21. Tuning to spectral features during singing.**

**A.** A UMAP embedding of the syllables extracted by the VAE but colored by amplitude (red higher amplitude blue lower). **B.** Same as A for goodness of pitch. **C.** A neuron that shows mixed tuning to amplitude and goodness of pitch during both playback and singing. **D.** A neuron that is tuned to amplitude and goodness of pitch only during singing.

### 3.3 Conclusions

The results presented in this chapter show that error-like responses that detect mismatches between expected and received auditory feedback are present in the adult zebra finch. However, the population that I found here is slightly smaller compared to the 10% that had been previously reported (albeit this corresponded to 9/92 neurons in Keller & Hahnloser, 2009). This reduced sensitivity in the adult could be due to more subtle vocal adjustments needed to maintain song compared to the more extensive corrections that are needed across sensorimotor learning. The variability in error-like responses suggests that they integrate multiple inputs and may operate in a nonlinear fashion, responding selectively to unexpected auditory feedback rather than

simply canceling out singing responses. I also discovered a population of neurons that maintained stable activity during singing and when exposed to distorted auditory feedback. These neurons likely receive inputs from non-auditory sources. Most of the neurons that maintained the activity during singing and distorted auditory feedback were suppressed, suggesting that this modulation could again be attributed to a corollary discharge signal. One proposed mechanism for error neurons is that a corollary discharge signal generates a “negative image” of the incoming sensory input, thus canceling a response during normal singing. However, this type of mechanism would elicit a strong response to white noise playbacks, which is not the case with prediction error neurons. Thus, it is likely that neurons that respond as error detectors behave non-linearly and integrate multiple inputs.

The comparison between white noise and syllable perturbation showed that the acoustic features of the auditory feedback are something that neurons in the auditory cortex encode during singing. Furthermore, the finding that the auditory stimuli alone (i.e., white noise or syllable) failed to evoke responses when played in isolation could alternatively be explained by the idea that prediction errors heavily rely on the behavioral context and the relevance of the stimulus.

The model by Duffy et al. assessed how VTA neurons track song using an eight-dimensional representation of song that included different acoustical features. Here, I carried out measurements that were agnostic to handpicked features, but I also broke down four different features to determine what was relevant for these neurons beyond high-dimensional representations. This analysis allowed me to directly link the natural variability of the song with the modulation of activity.

Neurons in the auditory cortex of the songbird can be tuned to acoustical features of external sounds (Nagel & Doupe, 2008). However, in my results, I found that this tuning cannot be predicted by only auditory responses since many neurons showed this tuning only during

singing. While these results offer a reasonable explanation of the type of information error signals may transmit, further experiments with an approach that offers better temporal resolution will be extremely valuable.

The results here contrast those found in animals that only produce innate vocalizations. In mice, neurons that are modulated during playback are also active during vocalization and scale with vocal power before the onset of the vocalization. Furthermore, this scaling occurs in deaf mice, suggesting a hardwired mechanism that suppresses vocal feedback (Harmon et al., 2024). While I also observed neurons that scaled their activity during singing and were responsive to playback of the bird's song, the responses to changes in amplitude could not be simply explained by auditory integration since a large proportion of neurons that showed this scaling did not respond to playbacks. Studies in marmoset monkeys have previously shown that neurons that are suppressed during vocalizations are less so when the feedback is shifted in pitch (Eliades & Wang, 2008). In a more recent study, Eliades and Tsunada found that this decrease in suppression during shifted auditory feedback occurred primarily in neurons that were tuned to frequencies within the vocal range (Eliades & Tsunada, 2024). These results match those found through the activity analysis using the VAE, however, the neurons in the songbird auditory cortex seem to have more complex receptive fields that cannot be explained by tuning to a single acoustic dimension like frequency.

Given that the response of a neuron cannot be predicted by its responses to playback, it is likely that these neurons serve a self-monitoring function. This contrasts with the findings in the marmoset monkey, given that neurons that showed sensitivity to perturbed feedback were more likely to be excited by song rather than suppressed. One possible reason for this is that these signals are integrated into downstream regions and need a positive sign.

## ***3.4 Methods***

### **3.4.1 Distorted auditory feedback**

Custom software, EvTAF, was used to deliver either white noise or syllable playback in 50% of the song renditions. To target a syllable, a template based on spectral frequencies was created; the system matches this template in real-time and triggers the system reliably with low jitter. Either white noise or a syllable playback was emitted through a speaker at a ~80dB level.

All the trials included were aligned to the onset of the first instance of white noise perturbation to avoid the influence of slow calcium dynamics. The criteria for exclusion in this experiment included trials where the bird stopped singing immediately after the distorted auditory feedback, this prevented offset responses from being classified as error responses.

### **3.4.2 Variational autoencoder training and alignment**

The VAE was trained with data recorded during calcium imaging sessions. The training data consisted of spectrograms of individual syllables segmented from the recorded audio files. This segmentation was carried out by manually defining amplitude thresholds to determine the onsets and offsets of syllables. Individual syllables were then interpolated to fit a 128x128 frequency-time matrix and fed into the network. After training, the 32 latent features were used for subsequent analysis to identify individual syllables using custom Matlab code (Samuel Brudner). To align the calcium imaging with the syllable timestamps, the closest frame to the onset of the signal was used.

### **3.4.3 Acoustic features**

The Python package VocalPy was used to calculate four different acoustic features: amplitude, entropy, pitch, and goodness of pitch. The amplitude is the absolute deviation from zero and was calculated by the root mean square of the signal. Pitch is the estimated fundamental

frequency of the syllable as determined by the lowest common denominator frequency. Goodness of pitch is an estimate of the harmonic pitch periodicity. Syllables with harmonic stacks have higher values, while tonal elements show lower ones. Entropy measures the spectral uniformity and variance of a signal. Tonal elements have larger negative values while broadband noise approximates zero.

## **4. Removal of auditory feedback reveals the emergence of an error-like signal**

### ***4.1 Introduction***

Auditory feedback is necessary to maintain learned vocalizations. Human infants that are congenitally deaf cannot acquire spoken language, and those that suffer from hearing loss in childhood fail to maintain a stable structure of speech. This strong dependence on auditory feedback changes in adulthood, where hearing loss leads to a much slower and more subtle speech deterioration (Blamey & Sarant, 2013).

This process draws many parallels with songbirds, where changes in learned song post-deafening occur faster in young birds. Specifically, in young adults (120-180dph), these changes start around week 3 post-deafening, while in older birds (180dph+), the time of changes in song structure, measured by manual scoring, is proportional to their age (Lombardino and Nottebohm 2000). In young adult birds, the changes observed do not merely reflect a passive process of song deterioration caused by a drift in the uncorrected output or plasticity changes in dendrites or cell death. While these processes may account for part of the changes, lesion studies have provided evidence to support the idea that song deterioration induced by deafening is an active process that recruits circuits in the AFP to drive changes in the motor pathway (Brainard & Doupe, 2000).

A large gap in our understanding of the effects of deafening in song deterioration is where and how changes at the single-neuron and network level trigger this active process and recruit AFP circuits to produce changes. Tschida and Mooney have provided some insights on this matter by tracking structural and activity changes in the song motor nucleus HVC. They used two-photon calcium imaging to longitudinally track the dendritic spines of HVC neurons before and after deafening. They found that deafening leads to a decrease in the size and stability of dendritic spines in HVC<sub>x</sub> neurons that project to the song basal ganglia and form part of the AFP

(Tschida & Mooney, 2012). Importantly, HVC<sub>x</sub> neurons receive direct input from CML-Av (Trusel et al., 2024). Thus, these changes may be preceded by changes in the primary auditory cortex.

In previous chapters, I provided evidence for a predictive corollary discharge signal that modulates the auditory cortex and demonstrated that the auditory cortex carries out feedback-dependent evaluative processes. Here, I removed auditory feedback by bilaterally extracting the cochlea to test two hypotheses. First, to test whether I can observe the persistence of a predictive corollary discharge signal even in the absence of auditory feedback; and second, whether the absence of auditory feedback generates an error-signal that may precede changes in the vocal-motor output. To determine this, I tracked the activity of the same neurons before and after deafening for up to a month and analyzed the acoustic features of the songs produced using a VAE.

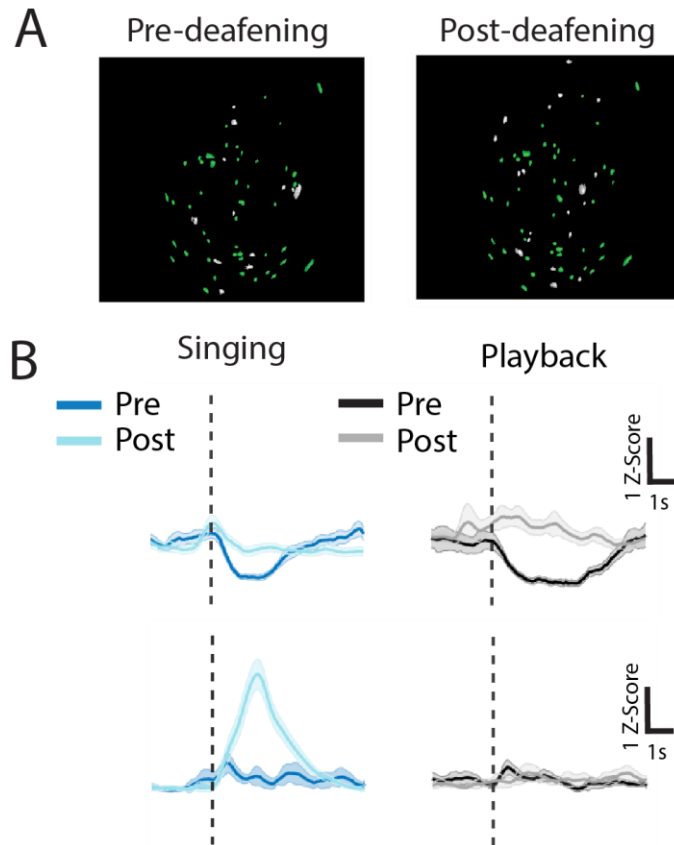
## ***4.2 Results***

### **4.2.1 Early deafening reveals non-auditory modulation in the auditory cortex**

Cochlea removal surgery is a very invasive procedure, and recovery takes between 3-6 days. During this time, the birds were constantly monitored and were only re-attached to the miniscope and imaged until they recovered pre-deafening levels of singing. While this prevented me from obtaining re-wiring changes that may occur early after deafening, it still allowed me to observe changes that were relevant to the behavior.

I used the algorithm CellReg to carry out the longitudinal analysis of neurons throughout different stages of deafening (Sheintuch et al., 2017). Briefly, this algorithm aligns the spatial footprints and generates probabilistic distributions based on centroid distances or spatial correlations across sessions to establish a registration threshold for same or different neurons

(Figure 22A). After this registration procedure, I was able to observe the changes during singing and playback on the first day after singing onset. The changes in activity could be broadly classified by loss of playback activity, maintenance of singing activity (excitatory or suppressive), or changes pre- and post-deafening during singing (Figure 22B).

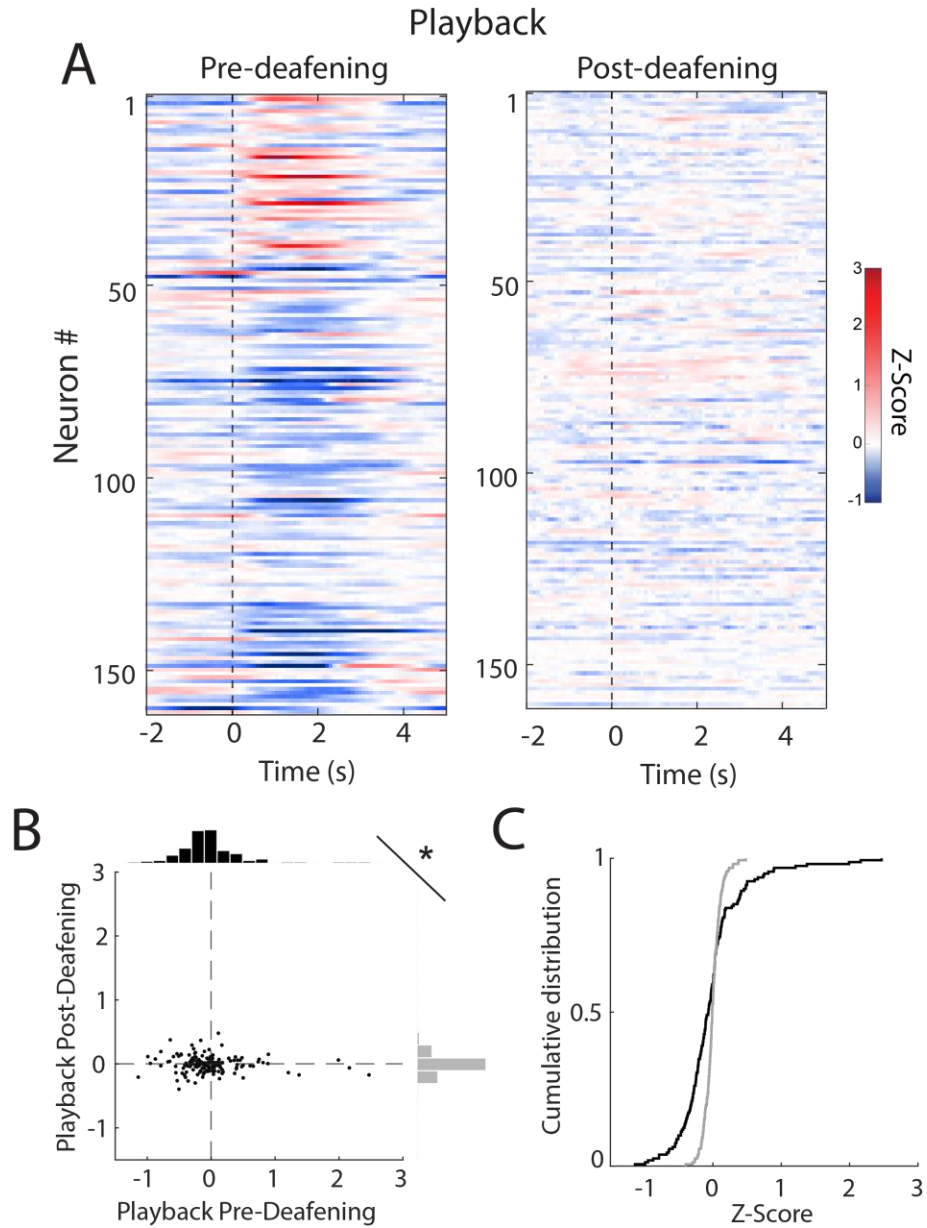


**Figure 22. Identification of neurons pre- and post-deafening.**

**A.** Representative field of view showing in green the neurons that were cross-registered across two sessions before and after deafening. **B.** Example neurons that changed their activity before and after deafening.

By looking at the activity pre- and post-deafening during playback across the entire population (n=161, 6 birds), I was able to confirm the success of the cochlea removal (Figure 23A). All modulation that was either suppressive or excitatory in response to playback was gone

after deafening. This was reflected in the flattening of the scatter plot along the pre-deafening axis and the sharp rise in the cumulative distribution (Figure 23B-C).

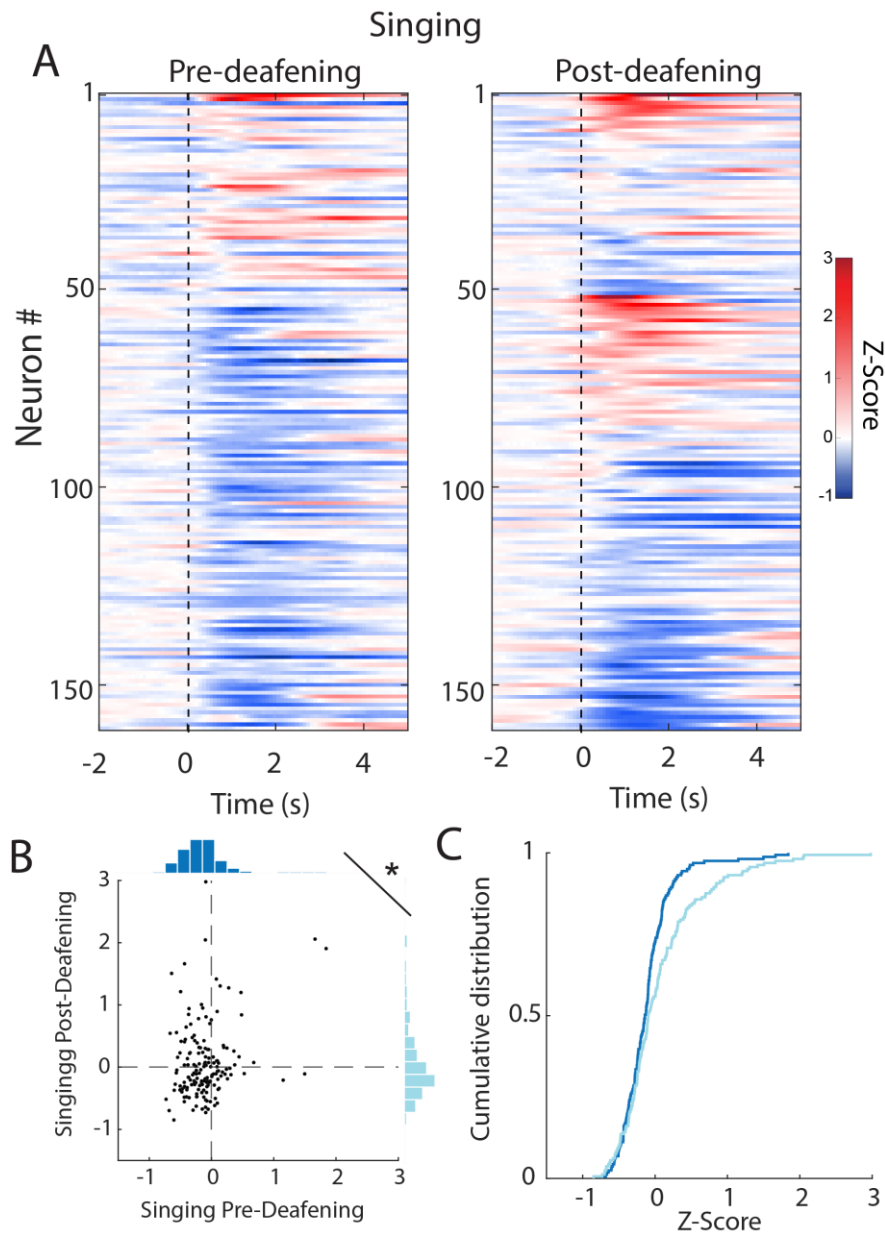


**Figure 23. Activity during playback pre- and post-deafening.**

**A-B.** Heatmaps showing the averaged activity across trials during playback pre- and post-deafening sorted by excitatory and suppressive activity pre-deafening ( $n = 161$  neurons from 6 birds). **C.** Scatter plot of averaged activity in individual neurons during the first 2 seconds before (black) or after deafening (gray) with a histogram projection of their respective distributions

(differences in distributions, Wilcoxon rank sum test,  $p=0.0046$ ). **D.** Cumulative distribution of the data shown in C.

When analyzing changes in activity during singing I found more heterogenous responses across the population (Figure 23A). As mentioned before, there was a small subset of neurons that were only modulated by auditory inputs that lost this activity post-deafening, this population was reflected along the pre-deafening x-axis of the scatter plot and resembled the changes observed during playback (Figure 23B). However, there were two predominant populations of neurons. First, those that maintained their suppressive modulation before and after deafening. These neurons concentrated along the diagonal in the lower left quadrant of the scatter plot and did not shift the distributions on the cumulative distribution (Figure 23B-C). Second, neurons that increased their activity post-deafening. These neurons were distributed along the post-deafening y-axis of the scatter plot and produced a shift in the cumulative distribution at positive z-score values (Figure 23B-C).



**Figure 24. Activity during singing pre- and post-deafening.**

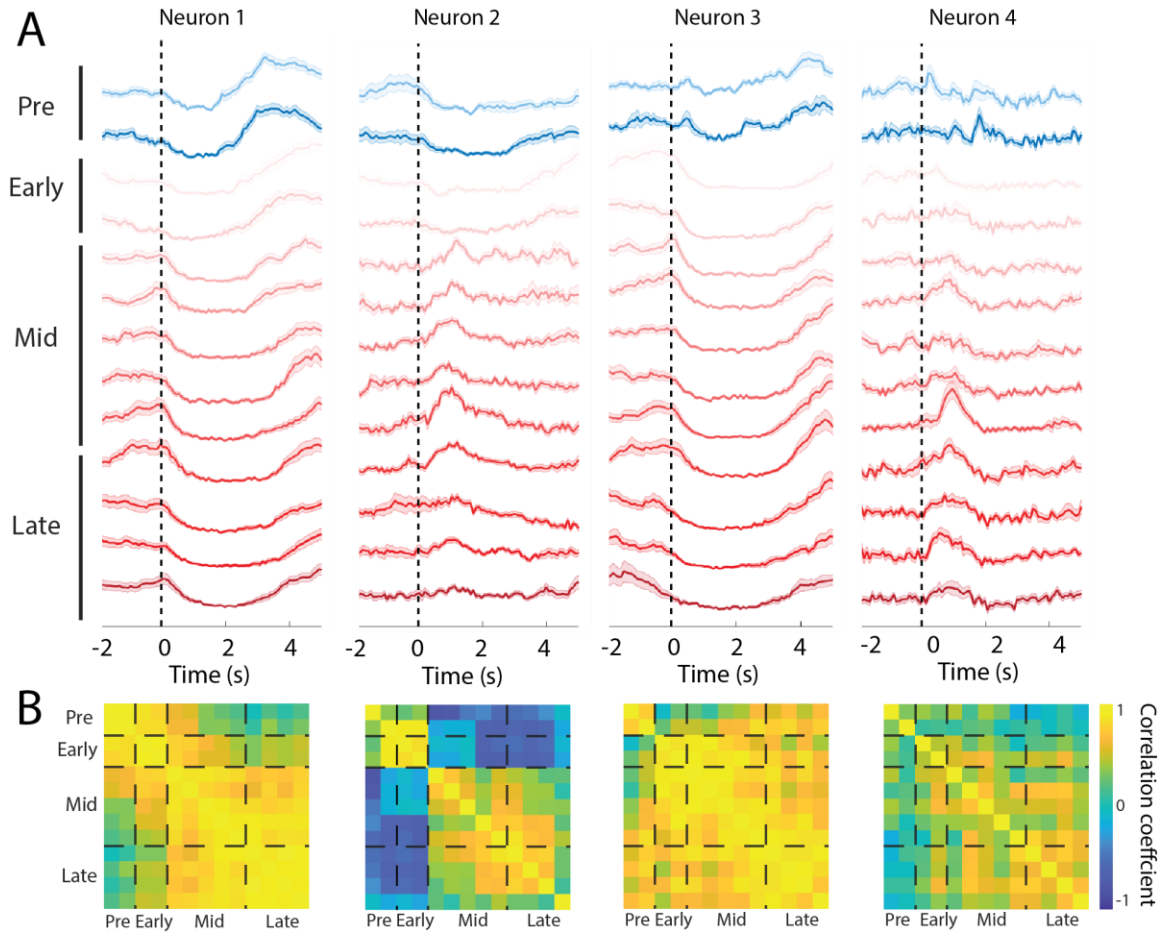
**A-B.** Heatmaps showing the averaged activity across trials during playback pre- and post-deafening sorted by excitatory and suppressive activity pre-deafening and re-sorted for activation and suppression of post-deafening activity to show the changes more clearly ( $n = 161$  neurons from 6 birds). **C.** Scatter plot of averaged activity in individual neurons during the first 2 seconds of singing pre- (dark blue) or post-deafening (light blue) with a histogram projection of their respective distributions (differences in distributions, Wilcoxon rank sum test,  $p=0.013$ ). **D.** Cumulative distribution of data shown in C.

## 4.2.2 Different stages of deafening produce changes in population dynamics

The absence of auditory feedback produces changes in the song. To determine whether changes in the activity in the auditory cortex drove the changes in vocal output I tracked the same neurons up to four weeks post-deafening. For this analysis, I focused on one young adult bird (156dph at the time of deafening) since changes in their song start around three weeks post-deafening. I classified the data into four categories: pre-deafening, early (1-10 days after singing onset), mid (11-20 days), and late deafening (21-30 days). When looking at individual neurons, I found a subpopulation that maintained their suppression across several days. Interestingly, these neurons had a shift in the onset and offsets of their response (neurons 1 and 3, Figure 24A). Additionally, I found another subpopulation of neurons that showed an emergence of activity during mid-deafening (neurons 2 and 4 Figure 25A). To quantify these changes, I calculated cross-correlations of activity across days. These correlation matrices reflected at which point post-deafening the activity was changing relative to other days (Figure 25B). Importantly, these matrices reflect how some neurons did not change their activity immediately after deafening, possibly reflecting a local reorganization event of these circuits

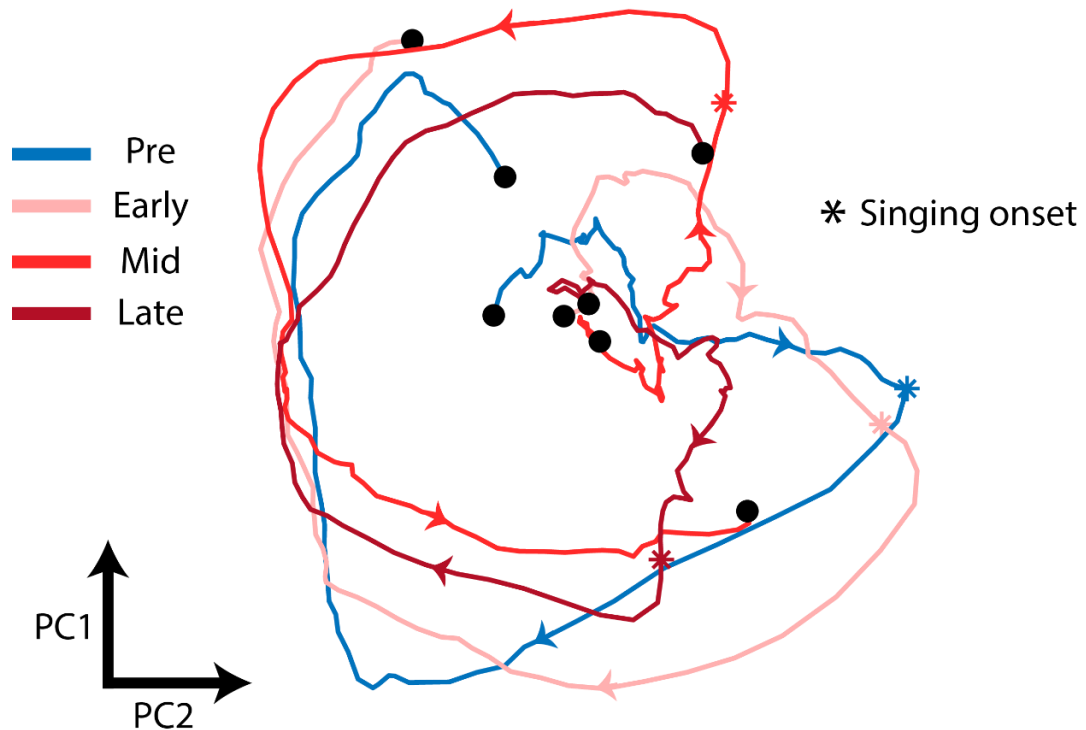
To analyze the changes at the population level across deafening, I carried out a principal component analysis across stages for all neurons (n=19 neurons tracked across 2 days per stage in one bird). I plotted the first two PCs that accounted for 72% of the variance to visualize the population dynamics as neural trajectories (Figure 26). These projections revealed that the trajectories from start to finish between pre- and early-deafening were strikingly similar, demonstrating that even in the absence of hearing, the activity remained stable. However, a drastic change occurred in the trajectory mid-deafening, possibly reflecting the emergence of activity and the increase in variability observed at the single neuron level. Finally, during late

deafening, the trajectory returned to follow the same direction as those during pre- and early deafening, possibly reflecting the stabilization of the network.



**Figure 25. Longitudinal tracking of individual neurons.**

**A.** Four representative neurons tracked pre- and post-deafening (last day of imaging was day 28 post-deafening, 13 days were picked for the analysis). All calcium traces are aligned to the onset of the first intronote. **B.** Correlation matrices across traces shown in A.



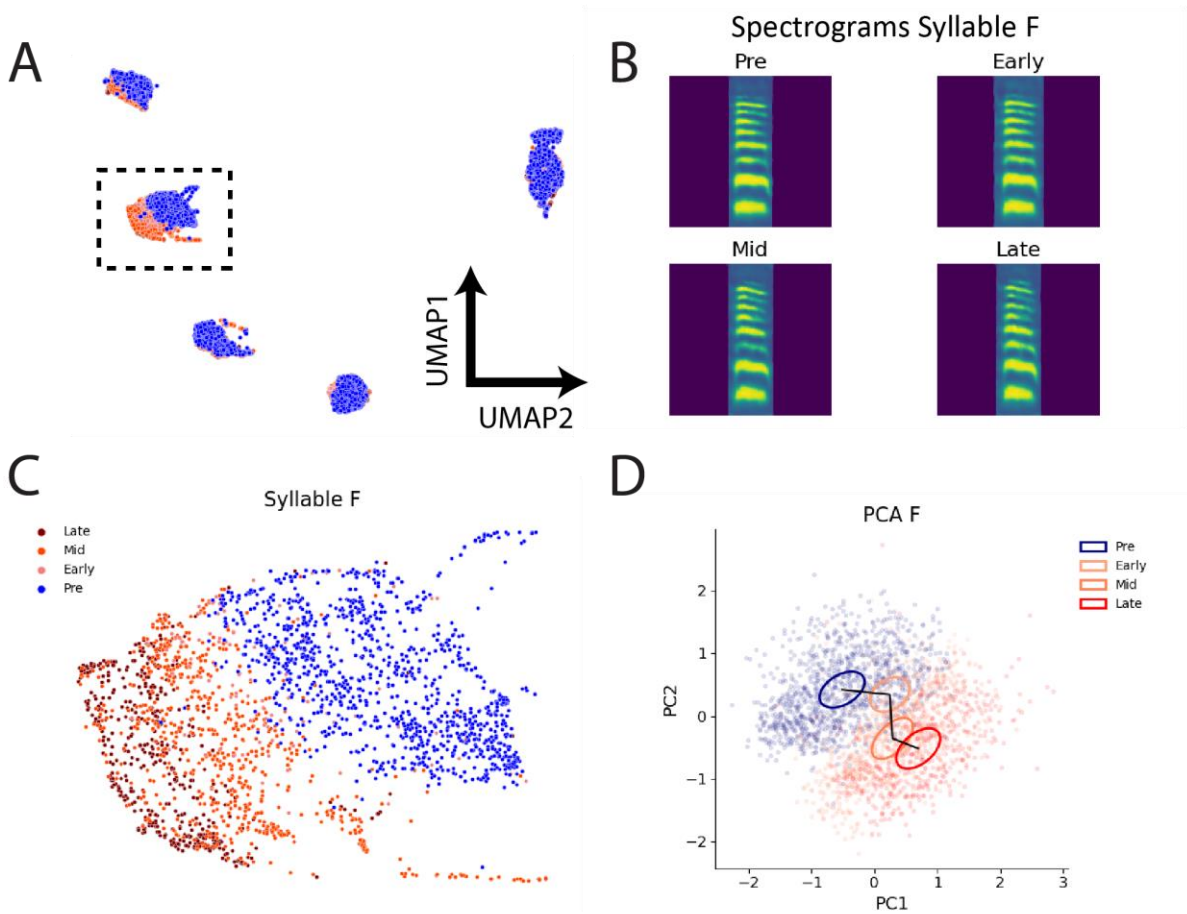
**Figure 26. Population trajectories across deafening.**

Population dynamics are projected into two principal component dimensions (n=19 neurons). The onset of the imaging snippet is at -5 seconds from intronote onset and is shown in black points at the center of the plot; the trajectories ended at +5 seconds post-song onset and terminated in the periphery. Asterisks denote the onset of singing and arrows show the direction of the trajectory forward in time.

#### 4.2.3 Changes in song post-deafening

To determine whether the changes that I saw at the neural level had any relationship to changes in song, I collaborated with Dr. Scott Smyre. We trained a VAE with the data pre- and post-deafening to obtain a low-dimensional representation of 32 latent features of individual syllables (See methods in Chapter 3). We embedded the 32 latent features in a two-dimensional space using a UMAP to see whether there were any changes in the song. Most syllables that belonged to the bird tracked longitudinally did not show large changes in this UMAP space (Figure 27A). However, we focused on one of the syllables that showed a shift pre- and post-deafening and looked at its spectrogram reconstructions. There was a noticeable difference in one

of the harmonic stacks that had a decrease in amplitude during mid-deafening (Figure 26B). To look closer at this syllable, we zoomed into the UMAP space and observed how the renditions shifted throughout deafening. Additionally, to better represent these changes, we projected the latent features on PC space and fitted a Gaussian ellipsoid to show the data distribution. These ellipsoids encompassed one standard deviation of the covariance matrix for both PCs at each deafening stage and showed a larger shift between early and mid-deafening, partially matching what was reported in the neural data (Figure 27D). The results presented in this section are primarily qualitative, and we are continuing to expand on this analysis to determine the best metrics to quantify these shifts across all birds.



**Figure 27. Song characterization using the VAE across deafening.**

**A.** UMAP embedding of the 32 latent features extracted by the VAE. Blue represents syllables pre-deafening and red syllables post-deafening. Each cluster represents an individual syllable. **B.** Reconstructed spectrograms from the 32 latents for syllable F, highlighted by a rectangle on A. **C.** Zoomed image to syllable F in UMAP space. **D.** Shifts in the distribution across deafening stages. The Gaussian ellipsoids represent one standard deviation of the covariance matrix of all syllables and they are connected at the mean of the covariance matrix.

### ***4.3 Conclusions***

The effects of sensory deafference in local cortical circuits have primarily been studied at a short-term timescale. However, how the circuits maintain or change their activity during a behavior that requires sensory input has not been studied. Here, I was able to use the delay of effects post-deafening in the adult zebra finch song to assess neural changes correlated to initial auditory feedback loss, as well as long-term changes in activity linked to changes in the vocal motor output.

During the initial loss of auditory feedback, I found sources of non-auditory modulation. First, neurons that are suppressed before and after deafening support the hypothesis of a stable corollary discharge signal that predicts singing and is not altered by auditory inputs. Second, a population of neurons that, in the absence of auditory feedback, increase their firing rates are consistent with a matching computation that produces changes in the firing rates when the auditory feedback is altered. These two different populations of neurons, together support the hypothesis for a predictive coding process that anticipates auditory feedback and compares it with actual auditory feedback during singing.

When looking at longer periods of time after deafening, I observed an effect across the population between days 10-20 post-deafening. One possibility is that these changes reflect an opening of a plasticity window that produces changes across the network and leads to more variable activity. Interestingly, after this process of re-organization, the patterns of activity return to those that resemble the activity before and early after deafening.

At the behavioral level, we saw changes in some of the syllables pre- and post-deafening that continued to drift through all stages of deafening. When looking at the changes in the covariance distributions in PC space, the largest shifts matched those that occurred in early and mid-deafening at the neural level. However, a deeper analysis between activity and song is still needed to understand whether changes in activity precede or follow changes in song.

## **4.4 Methods**

### **4.4.1 Deafening procedure**

Prior to deafening, I collected calcium imaging data for different conditions reported in the previous chapters for 3-4 weeks. I deafened the birds by bilaterally removing the cochlea. Birds were anesthetized with 20 mg/kg of ketamine and 10mg/kg xylazine and fixed on their side on a movable platform to facilitate access to the ear. The incision ear area was sterilized with alcohol and anesthetized with bupivacaine 0.25%. The ear skin was cut using microscissors, and the tympanic membrane was punctured by a fine scalp. Then, the columella and the footplate were detached with forceps to reveal the oval window. A fine custom tungsten wire hook was introduced into the inner ear canal to remove the cochlea entirely. All cochleas were collected in 4% PFA. After surgery recovery (as measured by normal singing rates), the birds were re-attached to the miniscope to image the neurons after deafening.

### **4.4.2 Cross-registration of neurons**

I used CellReg, an automated algorithm, to cross-register neurons across days (Sheintuch et al., 2017). This algorithm used the spatial footprint matrices of the fields of view generated by the neuron extraction performed by the CNMFe (Zhou et al., 2018), see methods in Chapter 2). Briefly, CellReg translated and rotated the fields of view to match a reference session and yielded a new spatial footprint matrix in the same coordinate system. All the neighboring cell-pairs that

were in close proximity were considered to test if they were the same neuron. For every neighboring cell-pair, the algorithm calculated the distance between the center of mass (centroid distance) and the correlation coefficient between the spatial footprints. These metrics were calculated across sessions between nearest neighbors and far neighbors to generate a probability distribution of same-cell likelihood. These distributions were usually bimodal, and a threshold for distance or correlation to register same-cells could be optimized based on their separability. The centroid and correlation distributions are then used to produce a joint model of the data. This model uses a Bayesian framework to calculate the probability of the neurons being the same or different. I set this threshold to be  $p \geq 0.95$  to determine same-cell pairs and register the neurons across sessions. The output of CellReg produced a matrix of indices that corresponded to the same neurons across the sessions.

#### **4.4.3 Statistics and analysis**

To compare distributions pre- and post- deafening I averaged the activity of a 2-second window after singing onset and used a Wilcoxon rank sum test to compare the distributions of this averaged activity before and after deafening.

To generate the matrix of Pearson correlation coefficients, I used 120 data points that spanned from -5 seconds prior to song onset to +5 seconds post-singing onset. I correlated these data vectors for every neuron across all days that were included in the analysis.

To generate the neural trajectories, I selected an equal number of days for each stage. Pre-deafening included days 4 and 5 before surgery, early days 9 and 10 post-deafening, mid-days 16 and 19, and late days 24 and 28. I carried out a PC analysis concatenating the data for these days and used the first two PCs that explained 72% of the variance to generate the trajectories.

## 5. Conclusions

### 5.1 Summary

Most studies on the zebra finch have focused on the motor circuits required to learn and maintain a highly stereotyped song. A large body of this work tried to understand how birds use auditory feedback to produce a learned song. These sensory and motor aspects of song learning and maintenance have primarily been integrated in a feed-forward manner, where auditory information enters the song motor pathway to influence activity. However, how top-down motor signals modulate activity in auditory regions and what this modulation may be important for has not been studied extensively.

Previous studies from the Mooney lab have established the identity of circuits that transmit a motor signal to the zebra finch's primary auditory cortex via the song motor nucleus HVC and those that transmit evaluation-related information downstream of it via AIV (Kearney et al., 2019; T. F. Roberts et al., 2017). Here, I explored how the activity in the primary auditory cortex is influenced by a corollary discharge signal and how neurons here monitor and evaluate auditory feedback.

I carried out one-photon calcium imaging to gain access to neural activity in freely moving birds and record large populations of neurons in the auditory cortex. First, I monitored the activity during singing and during passive listening to playbacks of the bird's song. I found that the majority of neurons were suppressed compared to activity evoked during non-singing epochs by playback. In 13% of the neurons, this suppression started up to 800ms before singing onset. This singing-related suppression and premotor modulation supported the hypothesis of a vocal corollary discharge signal that operates in the zebra finch auditory cortex.

I extended my analysis to other auditory regions to determine whether the modulation by a corollary discharge in CML-Av was transmitted to other regions to exert a similar effect. I

performed the same set of experiments in another primary auditory region, Field L, and in a secondary auditory region, NCM. In both areas, I saw a decrease in the number of neurons that were excited during singing compared to song playbacks. Additionally, the proportion of neurons that were suppressed was larger in CML-Av than in Field L and NCM, demonstrating that suppression is more dominant where the motor nucleus HVC projects. For this reason, I focused the rest of my research on CML-Av neurons.

The modulation of activity before singing onset supports the idea of a corollary discharge signal executing a predictive process. To expand on whether the auditory cortex may use a corollary discharge signal as a prediction to evaluate feedback in the auditory cortex, I carried out feedback manipulations by introducing a burst of white noise or the playback of a syllable out of place in the motif. These manipulations allowed me to test whether auditory cortical neurons carry out a predictive coding computation by comparing an internal model, or corollary discharge signal, with the incoming auditory feedback (Figure 2). With these experiments, I found a small subset of neurons that modulated their activity in response to changes in the expected auditory feedback. I was able to identify neurons that encoded either a positive or negative prediction error. Positive prediction error neurons increased their activity during the perturbation of the song but were not modulated by either white noise or the syllable playback alone. In contrast, there was a smaller number of negative prediction error neurons that decreased their activity in response to the perturbation of the song. Error signals were more likely to occur in neurons that either did not have any modulation during normal singing or were excited.

Finally, to determine whether changes in activity in the auditory cortex precede changes in vocal motor output I removed auditory feedback and tracked activity throughout several weeks. These experiments brought to light two different subpopulations of neurons, those that maintained their activity even several days post-deafening and those that increased their activity

transiently. The neurons that maintained a stable modulation throughout deafening were more likely to be neurons that were suppressed during singing. Demonstrating that even if local reorganization occurs due to the absence of excitatory auditory input, motor signals remain the same. The stability of the motor vocal plan was translated at the behavioral level, where we did not find drastic changes in song during a span of 4-6 weeks. The reorganization observed mid-deafening likely corresponded to the trigger of homeostatic plasticity following the removal of excitatory drive from bottom-up auditory information.

Altogether, these results provide evidence of a learned vocal-motor signal that functions predictively to suppress singing-related feedback and suggests that this signal contributes to the maintenance and execution of stable motor programs.

## ***5.2 Discussion***

### **5.2.1 Corollary discharge function and implementation mechanisms**

Corollary discharge signals in the auditory cortex are important for several reasons. They can prevent the auditory cortex from becoming saturated by the proximity to the source of auditory feedback (i.e., the syrinx in songbirds). A decrease in the firing rate across the population caused by corollary discharge signals may allow it to maintain a dynamic range of activity instead of being saturated by the loudness of vocalizations. In the population of neurons that I imaged from, I did not see a large activation response to playbacks of the bird's song; in fact, 37% of the significantly modulated neurons had an excitatory response to playback, while 32% of them were suppressed. While playbacks may have a lower amplitude than active singing, this nearly equal modulation suggests that activity in primary regions is unlikely to be saturated by excitation in response to sound. Nevertheless, there is a population of neurons that was suppressed during singing but excited by playback, this is likely a population of neurons that

contributes to maintaining sensitivity to external sounds when the bird is singing, and a corollary discharge signal facilitates this process.

Another plausible function of corollary discharge signals in the auditory cortex is to contribute to the monitoring and evaluation of own vocalizations. These processes are important for birds to distinguish their own vocalizations in noisy environments, and to maintain a stable vocal motor program during adulthood. The neurons that are modulated exclusively during singing provide evidence for a monitoring function. Furthermore, those that are excited and tuned to acoustic features only during singing, together with neurons that can generate positive or negative prediction errors, support the idea that a population of neurons tracks auditory input exclusively to evaluate it. Overall, these findings support the hypothesis that corollary discharge signals exert effects in two different subpopulations of neurons through either excitation or suppression that together support the monitoring and evaluation of own vocalizations.

One possible source for a corollary discharge signal that influences activity in the auditory cortex is HVC. Work from the Mooney lab has demonstrated that the HVC-Av projection is important for vocal flexibility and learning. These neurons are active before song onset and maintain their activity even in deafened birds. Furthermore, ablating these neurons prevents a juvenile bird from copying their tutor's song, attenuates song timing defects post-deafening, and hinders the ability of adults to modify song timing in response to song-contingent reinforcement with noise (T. F. Roberts et al., 2017). Altogether, these results demonstrate that this projection has an important role in evaluating auditory feedback and is likely producing the effects that I observe in the auditory cortex.

One of the key questions in the field of corollary discharge signals is how exactly motor commands are transformed into sensory coordinates. One possibility is that reciprocal connections between sensory and motor regions facilitate this during song learning in the zebra

finch. HVC is the only song system nucleus known to receive direct input from auditory regions, and as mentioned before, it forms connections with CML-Av. During song learning, one can imagine that auditory information enters HVC, which then projects back to CML-Av and learns a rule to match the auditory consequences of the vocal motor commands. Some of these ideas are adapted from one of the best-studied corollary discharge signal models, the electric fish (P. D. Roberts & Bell, 2000). In the zebra finch, some of these ideas are currently being tested by a collaborator, Ziyi Gong, from the Pearson lab. Ziyi Gong is building theoretical models on different excitation-inhibition networks that receive auditory and premotor inputs to establish how the cancelation of auditory feedback and the emergence of error signals may arise via local recurrent anti-Hebbian plasticity mechanisms during sensory and sensorimotor learning.

Finally, another aspect to consider in singing-related activity is the modulation that may be caused by either proprioceptive feedback or changes in the internal state. Proprioceptive feedback from the air sacs, lungs, and syrinx can be conveyed to CML-Av via a projection from the thalamic nucleus uvulaeformis (Uva) (Akutagawa & Konishi, 2010; Burke et al., 2024; Schmidt et al., 2012). This non-auditory feedback is important to learn and maintain vocal patterning (Veit et al., 2011), and birds are sensitive to changes in this type of sensory feedback. Suthers et al. demonstrated that songbirds can quickly detect changes in air sac pressure and compensate for it in the range of milliseconds to maintain stable vocal output (Suthers et al., 2002). While it is experimentally challenging to access the airsacs, I explored whether proprioceptive feedback from the syringeal muscle influences activity in the auditory cortex. I developed a new method using an ultra-sensitive, red-shifted, pump-like cation channel-rhodopsin to stimulate the contraction of the syrinx. These findings are summarized in Appendix A.

### **5.2.3 Vocal motor signals in other animal models**

Vocal communication is important in many animal species. Sensorimotor integration in the auditory cortex has been studied in other model systems (Curio et al., 1998; Eliades & Wang, 2003; Harmon et al., 2024). My findings match previous work that reported how suppression of activity in the auditory cortex has the largest effect at the population level. However, when looking at individual neurons, I see a larger variety of responses during singing compared to those during playback. For example, Harmon et al. reported that in mice, pyramidal neurons and interneurons that were excited during playbacks of ultra-sonic vocalizations showed a decrease in this excitation during vocalization (Harmon et al., 2024). In contrast, in the songbird auditory cortex, I found that approximately one-third of the neurons were gated off during singing but modulated by playback, and a smaller percentage of neurons had opposite signs of modulation during singing and playback.

Another type of response that differed between animals that produce innate vocalizations and songbirds was in response to distorted auditory feedback. These two differences in responses during vocalizations in animals that produce innate vocalizations and songbirds may reflect differences in circuit motifs that are unique to animals that learn their vocalizations and have to actively maintain a stable vocal output that is dependent on auditory feedback.

### **5.2.4 The meaning of an error signal**

The simplest model for how corollary discharge signals may contribute to generating error signals is through the generation of a negative mirror image that predicts and cancels out the excitatory auditory inputs. If the auditory input deviates from the expectation, then the activity increases. However, this model leaves out exactly what these neurons are encoding. This is particularly challenging to determine because neurons in the auditory cortex have a wide range of

tuning, and responses to the stimulus that distorts the auditory feedback could be driving the receptive field of the neuron instead of encoding a true error signal.

By looking at the natural distribution of the song, my work provides some insights into how neurons may be encoding this error. I found neurons that are tuned to different acoustic features, including amplitude, pitch, harmonicity, and entropy. These neurons showed a correlation between their activity and the acoustic features. Importantly, some of these neurons seem to be only active while the bird is singing, demonstrating that there are mechanisms that grant access to auditory responses only in the context of singing, perhaps based entirely on a prediction of the auditory feedback. Another possible mechanism for this could be through neuromodulatory systems that act in different brain regions to coordinate such effect (Ben-Tov et al., 2023; Jaffe & Brainard, 2020; Singh Alvarado et al., 2021).

### **5.2.5 Deafness-related changes**

Changes in circuit connectivity occur naturally to modify behavior and adapt to new environments or sensory experiences. During instances where there are large changes in sensory input caused by experimental manipulations, injury, or sensory deafference, plasticity mechanisms trigger reorganization in many circuits across the brain (Hübener & Bonhoeffer, 2014). In mammalian auditory processing, lesions to the cochlea produce changes in primary auditory cortex connectivity at different stages. First, the decrease in excitatory drive leads to a decrease in GABAergic transmission that is followed by hyperexcitability of pyramidal cells. After 5-8 days these changes in activity are rebalanced and the network returns to a pre-lesion working range and can sustain auditory function even with a large magnitude of hearing loss (Resnik & Polley, 2017). While these changes largely depend on the magnitude of hearing loss and many studies have but are likely to be guided by behavioral use.

Here, I showed what happens to a circuit that undergoes sensory loss but maintains a stable behavior even in its absence. One of the more striking results is that I observed neurons that increased their activity early during deafening but others that underwent this change in reorganization until days 10-20 after the bird started singing. These slower processes in the songbird likely reflect a reorganization and stabilization of the network that contributes to maintaining a stable behavior for a longer period of time.

## ***5.2 Limitations***

The zebra finch comes with a unique behavior that is difficult to find in other models, but they have the disadvantage of having fewer tools available to manipulate and record neural activity. During the early stages of this project, I tried a series of viruses and promoters to obtain cell-type specificity of GCaMP expression. However, I was unsuccessful in obtaining expression that was equivalent to or surpassed that of the pan-neuronal construct under the CAG promoter. Being able to distinguish between excitatory and inhibitory neurons in these circuits will help us make better models and predictions on how these circuits integrate information across different conditions.

Additionally, while calcium imaging is a great tool to obtain a higher yield of neurons and track the same neurons through long periods of time, it lacks temporal resolution. Electrophysiological recordings in this area during singing could provide important insights into when and how neurons are modulated and could help characterize the circuit by defining cell types based on waveform or opto-tagging. Having a higher temporal resolution would also improve tuning measurements, that I was able to initially characterize, and determine whether this tuning is syllable-specific or agnostic to both the type of syllable and its place in the motif sequence.

The experiments that I carried out using distorted auditory feedback are well-established in the field. However, they have the disadvantage of superimposing a sound over ongoing singing and may fail to generate a true perceived error. To mitigate this limitation, I tried developing a new method where I could manipulate the contraction of the syrinx to generate a more naturalistic error. My progress on this approach is discussed in Appendix A.

#### ***5.4 Future directions***

I believe that an important step in understanding the neural correlates of song evaluation will be to determine how bottom-up information coming from auditory feedback is integrated with top-down corollary discharge signals. To test this, reversible manipulation, such as expressing an inhibitory opsin or DREADDs in HVC-Av neurons and auditory thalamic terminals, could help contribute to dissecting these circuits and understanding the mechanisms through which the integration of auditory, motor, and non-auditory feedback, producing the large repertoire of responses that I observed.

Additionally, reversible manipulations of auditory feedback that distort song but do not trigger drastic network changes as those caused by deafening, could be useful in understanding how the system transiently adapts to vocal errors. Some of these manipulations could include pitch learning, injecting botulin toxin in the syringeal muscle, or introducing birds to heliox chambers. It would be interesting to explore whether the plasticity mechanisms that are triggered by these manipulations are similar to those observed early in deafening.

Predictive coding is a booming field in modern neuroscience. A lot of research has focused on understanding how internal models are generated and updated. One of the main challenges is to find a behavior where this is tractable. The zebra finch offers a unique system to

understand how internal models are learned, and the work I carried out in this thesis lays the groundwork to go to a deeper level and study how internal models are generated and updated.

## **Appendix A**

### **Developing a naturalistic vocal error manipulation**

#### ***A.1 Introduction***

Most studies investigating auditory feedback distortion have used superimposed sounds while the bird sings. While this approach can produce drastic and adaptive changes in the song, it does not recapitulate naturalistic errors (Leonardo & Konishi, 1999; Tumer & Brainard, 2007). Other researchers have used creative approaches to manipulate vocal output by introducing the birds to a heliox chamber, decompressing the air sacs, or injecting botulinum toxin in the syringeal muscles to transiently and partially paralyze them (Amador & Margoliash, 2013; Nowicki, 1987; Pytte et al., 2011).

One way in which a naturalistic error could be induced is by reversibly manipulating the contraction of the vocal organ, the syrinx. The syrinx is composed of four major muscle groups that contribute in different ways to the vibration of the labia (similar to vocal folds in humans) that are located on each side of the syrinx and produce the sound (Schmidt & Wild, 2014).

Additionally, developing an approach that alters the contraction of the syrinx would allow us to study more in-depth the contributions of proprioceptive feedback to song monitoring and evaluation. While the syringeal muscles do not have muscle spindles, proprioceptive feedback is likely transmitted by an ascending projection of the tracheo-syringeal nerve (Bottjer & Arnold, 1982). These projections, together with stretch receptors from the airsacs, can make their way into the song motor pathway and the auditory cortex via projections from the thalamic nucleus Uva (Schmidt et al., 2012).

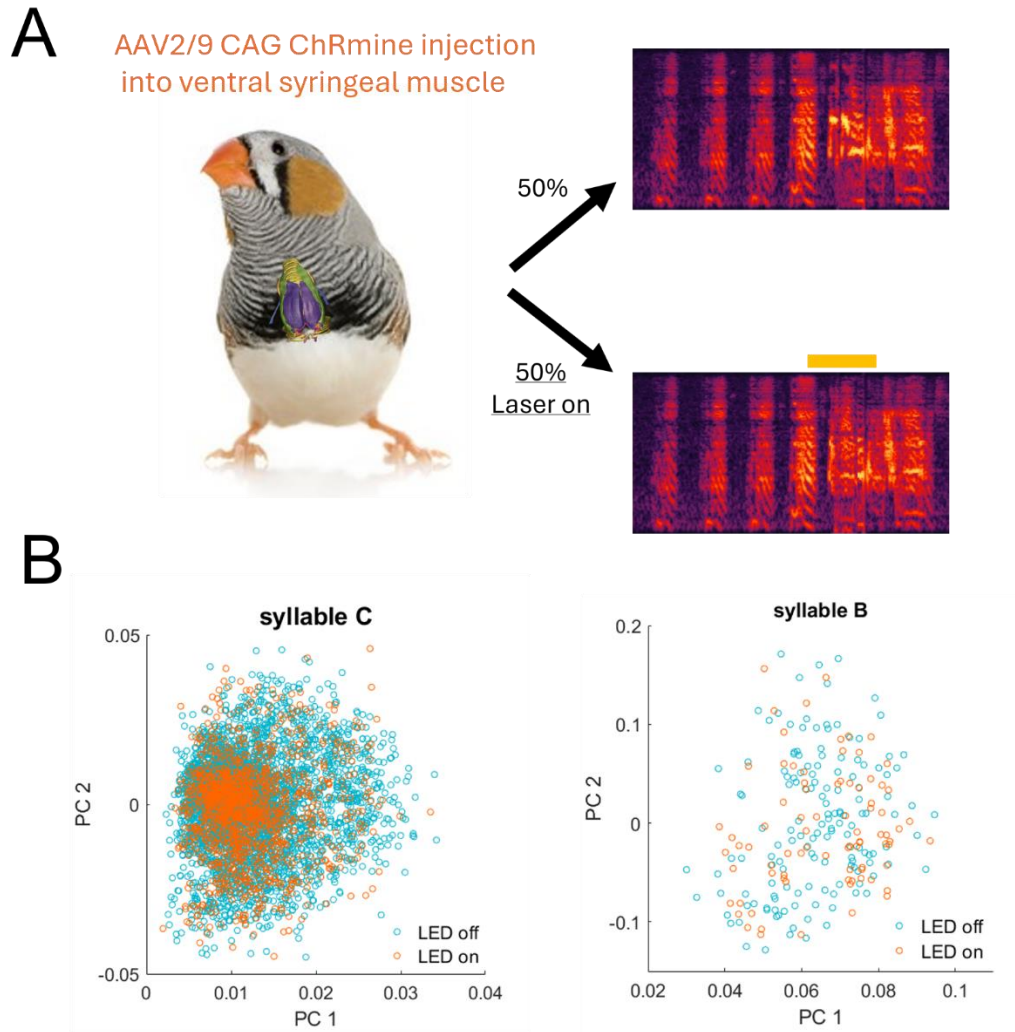
To produce a naturalistic error that allowed me to study error signals and proprioceptive feedback, I developed a new method that aimed to stimulate the syrinx vocal muscles to distort

the song. I targeted the ventral syringeal muscle because it has been directly implicated in controlling the fundamental frequency of syllables (Adam et al., 2021; Düring et al., 2013). I injected the ventral syringeal muscle with a viral construct to express ChRmine, a recently developed highly sensitive red-shifted pump-like channelrhodopsin (Kishi et al., 2022). This channelrhodopsin has been previously shown to non-invasively activate neurons in the midbrain and produce contraction in the heart (R. Chen et al., 2021; Hsueh et al., 2023). The amber 591nm wavelength that is used to activate the channels can penetrate deep into the tissues making it a great tool to target muscles in the periphery.

## ***A.2 Results***

To produce a ChRmine-induced contraction in the syrinx, I developed a custom-made vest that had two serial LEDs that were positioned in the interclavicular space and shone light towards the ventral syringeal muscle (Figure 28A).

I used the same system as in Chapter 3 to track the song in real-time and deliver a TTL pulse of light in 50% of the renditions to stimulate ChRmine and produce contraction in the muscle to change the vocal output. When carrying out this manipulation, I was not able to see any changes in song production. However, I carried out this analysis using the VAE tools described in previous chapters and only performed a qualitative evaluation of the syllables in a PC space (Figure 28B, variance explained = 64% for syllable C, 88% for syllable B). Further acoustic analysis is needed to truly discard the possibility that this manipulation cannot alter the song.

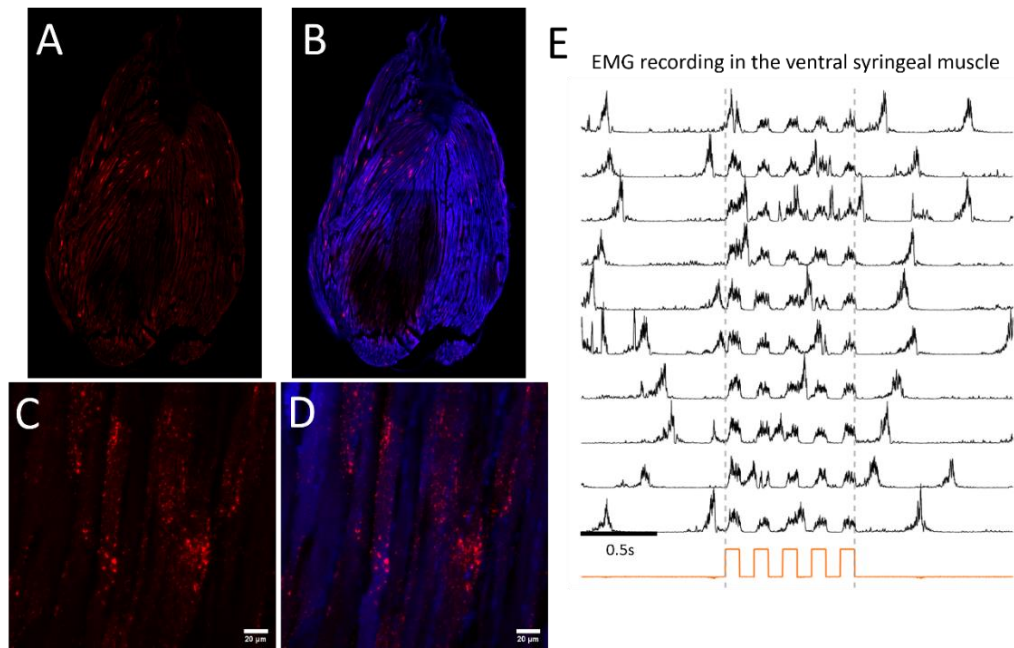


**Figure 28. ChRmine stimulation approach.**

**A.** The viral construct was injected into the ventral syringeal muscle, shown in purple (syrinx model adapted from Düring et al., 2013). To alter the vocal output, I targeted 50% of the renditions with an LED pulse. **B.** A principal component analysis on the 32 latent features of two syllables that were targeted by the LED stimulation.

To validate the method and determine why I could not see any effects in my preliminary analysis, I first tested the opsin functionally by stimulating with an LED while I recorded an electromyogram in the ventral syringeal muscle. I was able to successfully evoke contractile movement when I emitted light pulses through the LED (Figure 29 E). Then, I carried out

histology on the muscle and I was able to obtain good expression of the reporter protein mScarlet in several fibers of the ventral syringeal muscle (Figure 29A-D). Importantly, this expression was strong even without any amplification of the signal through immunohistochemistry.

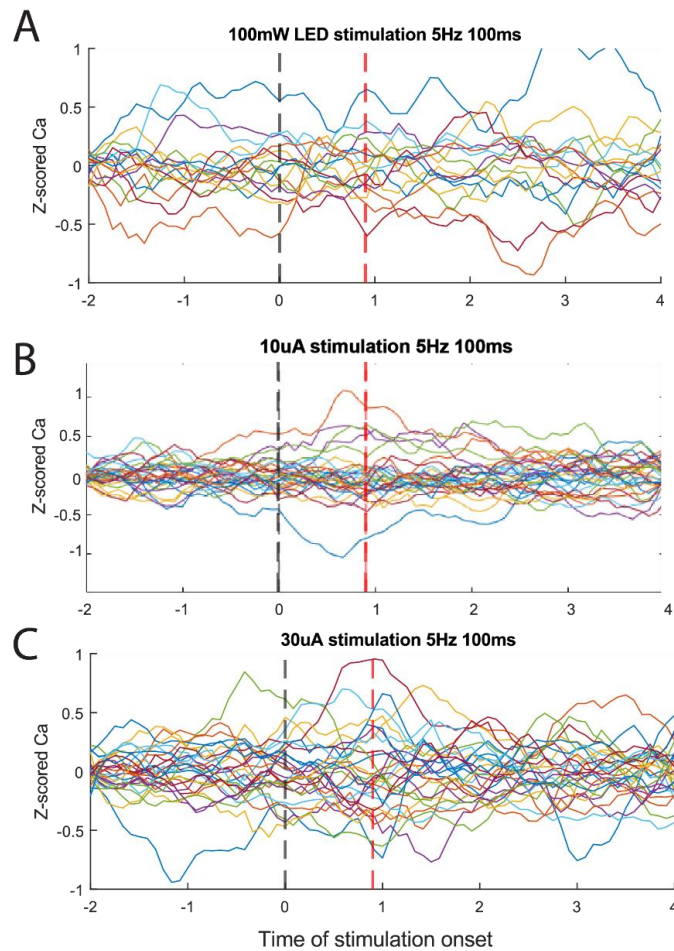


**Figure 29. Validation of ChRmine function.**

**A-D.** Coronal section of the ventral syringeal muscle showing the expression of the viral construct ChRmine-mScarlet in red and a DAPI staining of the polynucleated muscle fibers in blue **E.** Electromyography in the ventral syringeal muscle while stimulating with an LED (orange pulses) from the same bird. The period of stimulation occurred between the dotted lines, and every trace represents an individual trial. The large transients outside of the stimulation window are respiration-related contractions that occur naturally in the syringeal muscles.

Finally, I wanted to determine whether I could evoke activity in the auditory cortex when I stimulated the ventral syringeal muscle. I imaged neural activity in a bird that was previously injected to express GCaMP6s in CML-Av in the brain, as well as with the viral construct to express ChRmine in the syringeal muscle. Under anesthesia, I stimulated the muscle with the LED and recorded the responses. Prior to this stimulation, I had confirmed the function of ChRmine by listening to the EMG responses while I emitted light pulses. My analysis revealed that the light stimulation did not strongly drive activity in most of the neurons that I was able to

image from (n=15 trials, 19 neurons, Figure 30A). To discard the possibility of having uneven expression in the muscle and ensure I was evoking a muscle contractile response, I also stimulated the muscle with electrical stimulation at different magnitudes (n=15 trials, 19 neurons, Figure 30B-C). This stimulation revealed a few neurons that had some small degree of modulation during stimulation. However, this modulation was not as strong as that evoked by auditory feedback. Further statistical analyses of these responses are necessary to determine whether these responses can be attributed to the stimulation.



**Figure 30. Neural responses to syrx stimulation.**

**A.** Averaged traces of individual neurons for 15 trials of stimulation with an LED power of 100mW measured from the light source. The onset and offset of the stimulation pulse are marked by dotted lines. One neuron (orange) shows a potential excitatory response to this stimulation. **B-**

C. Same as A but with electrical stimulation to 10 and 30 micro-amperes targeted towards the lower right ventral syringeal muscle, the lens was implanted on the left hemisphere.

### ***A.3 Conclusions***

The experimental approach I pioneered in this chapter has the potential to become a great tool for studying the proprioceptive contributions to vocal performance. While my preliminary analyses using the 32 latent features from the VAE did not reveal an online change in acoustical aspects of song, a deeper dive into specific acoustic features like pitch and amplitude may provide a better quantification of possible effects. However, there are a few explanations as to why there

I was not able to find strong neural responses to syrinx stimulation in the anesthetized bird. One possible explanation for this is that the syrinx contraction is closely synchronized to breathing cycles, and producing a contraction out of cycle may not evoke the same response that would occur naturally during singing.

### ***A.4 Methods***

#### **A.4.1 Syringeal muscle injection**

Birds were anesthetized with an intramuscular chest injection of ketamine (20mg/kg) and xylazine (10mg/kg). They were placed on a rotating platform to facilitate access to the intraclavicular cavity. The interclavicular area was sterilized with alcohol and betadine, and locally anesthetized with bupivacaine 0.25%. I followed a previously published protocol to access the syrinx (Kim et al., 2020). Briefly, the skin was cut open along the midline with microscissors, and the adipose tissue was partially cut and pulled upwards to reveal the interclavicular airsac. The airsac was punctured with a scalpel to access the syrinx. A Hamilton syringe was loaded with 1-1.5 uL of the viral construct AAV9-CAG- ChRmine-mScarlet-WPRE, and the muscle was injected in 4 different sites targeting the upper and lower left and right ventral syringeal muscle. After the

injection, the adipose tissue was placed over the interclavicular space and sealed using Vetbond, and the skin was sutured closed.

### **A.4.3 Electromyography**

The birds were anesthetized and prepped as previously described. Two tungsten electrodes of low impedance (0.5 ohms) were bent at the tip and introduced into the ventral syringeal muscle to perform differential recordings. Two amber LEDs (591nm) were placed over the interclavicular area angled at the syrinx; they were lit at a range of 40-100mW by TTL pulses triggered by custom LabView programs. The EMG recording signals were rectified and filtered to observe the contractile events for individual trials.

### **A.4.2 Muscle histology**

Birds were transcardially perfused and the syrinx was extracted afterwards in 4% PFA. After 1 day in PFA it was transferred to a 30% sucrose -PFA solution for 2 days. The syrinx was flash-frozen in a block of O.C.T. compound. To preserve the anatomy of the syrinx, the muscle was sliced in a cryostat and immediately collected into mounting slides. A DAPI mounting media was applied, and the samples were cover-slipped and imaged with a confocal microscope.

## References

- Akutagawa, E., & Konishi, M. (2010). New brain pathways found in the vocal control system of a songbird. *Journal of Comparative Neurology*, *518*(15), 3086–3100. <https://doi.org/10.1002/cne.22383>
- Amador, A., & Margoliash, D. (2013). A mechanism for frequency modulation in songbirds shared with humans. *Journal of Neuroscience*, *33*(27), 11136–11144. <https://doi.org/10.1523/JNEUROSCI.5906-12.2013>
- Amin, N., Doupe, A., & Theunissen, F. E. (2007). Development of selectivity for natural sounds in the songbird auditory forebrain. *Journal of Neurophysiology*, *97*(5), 3517–3531. <https://doi.org/10.1152/jn.01066.2006>
- Amin, N., Gill, P., & Theunissen, F. E. (2010). Role of the zebra finch auditory thalamus in generating complex representations for natural sounds. *Journal of Neurophysiology*, *104*(2), 784–798. <https://doi.org/10.1152/jn.00128.2010>
- Aronov, D., Veit, L., Goldberg, J. H., & Fee, M. S. (2011). Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. *Journal of Neuroscience*, *31*(45), 16353–16368. <https://doi.org/10.1523/JNEUROSCI.3009-11.2011>
- Audette, N. J., & Schneider, D. M. (2023). Stimulus-specific prediction error neurons in mouse auditory cortex. *Journal of Neuroscience*, *43*(43), 7119–7129. <https://doi.org/10.1523/JNEUROSCI.0512-23.2023>
- Behroozmand, R., Phillip, L., Johari, K., Bonilha, L., Rorden, C., Hickok, G., & Fridriksson, J. (2018). Sensorimotor impairment of speech auditory feedback processing in aphasia. *NeuroImage*, *165*, 102–111. <https://doi.org/10.1016/j.neuroimage.2017.10.014>
- Bell, C. C. (1982). Properties of a modifiable efference copy in an electric fish. *Journal of Neurophysiology*, *47*(6), 1043–1056. <https://doi.org/10.1152/jn.1982.47.6.1043>
- Ben-Tov, M., Duarte, F., & Mooney, R. (2023). A neural hub for holistic courtship displays. *Current Biology*, *33*(9), 1640–1653.e5. <https://doi.org/10.1016/j.cub.2023.02.072>
- Blamey, P. J., & Sarant, J. Z. (2013). The Consequences of Deafness for Spoken Language Development. In A. Kral, A. N. Popper, & R. R. Fay (Eds.), *Deafness* (pp. 265–299). Springer New York. [https://doi.org/10.1007/2506\\_2013\\_10](https://doi.org/10.1007/2506_2013_10)
- Bottjer, S. W., & Arnold, A. P. (1982). Afferent neurons in the hypoglossal nerve of the zebra finch *Poephila guttata*: Localization with horseradish peroxidase. *Journal of Comparative Neurology*, *210*(2), 190–197. <https://doi.org/10.1002/cne.902100209>
- Bottjer, S. W., Miesner, E. A., & Arnold, A. P. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*, *224*(4648), 901–903.

<https://doi.org/10.1126/science.6719123>

- Brainard, M. S., & Doupe, A. J. (2000). *Interruption of a basal ganglia±forebrain circuit prevents plasticity of learned vocalizations*. [www.nature.com](http://www.nature.com)
- Burke, J. E., Perkes, A. D., Perlegos, A. E., & Schmidt, M. F. (2024). *A neural circuit for vocal production responds to viscerosensory input in the songbird*. *October 2023*, 304–310. <https://doi.org/10.1152/jn.00400.2023>
- Calabrese, A., & Woolley, S. M. N. (2015). Coding principles of the canonical cortical microcircuit in the avian brain. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3517–3522. <https://doi.org/10.1073/pnas.1408545112>
- Chen, R., Gore, F., Nguyen, Q. A., Ramakrishnan, C., Patel, S., Kim, S. H., Raffiee, M., Kim, Y. S., Hsueh, B., Krook-Magnusson, E., Soltesz, I., & Deisseroth, K. (2021). Deep brain optogenetics without intracranial surgery. *Nature Biotechnology*, *39*(2), 161–164. <https://doi.org/10.1038/s41587-020-0679-9>
- Chen, T. W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., Schreiter, E. R., Kerr, R. A., Orger, M. B., Jayaraman, V., Looger, L. L., Svoboda, K., & Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, *499*(7458), 295–300. <https://doi.org/10.1038/nature12354>
- Colquitt, B. M., Merullo, D. P., Konopka, G., Roberts, T. F., & Brainard, M. S. (2021). Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. *Science*, *371*(6530). <https://doi.org/10.1126/science.abd9704>
- Conner, J. M., Bohannon, A., Igarashi, M., Taniguchi, J., Baltar, N., & Azim, E. (2021). Modulation of tactile feedback for the execution of dexterous movement. *Science*, *374*(6565), 316–323. <https://doi.org/10.1126/science.abh1123>
- Cotanche, D. A. (1999). Structural recovery from sound and aminoglycoside damage in the avian cochlea. *Audiology and Neuro-Otology*, *4*(6), 271–285. <https://doi.org/10.1159/000013852>
- Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience*, *9*(8), 587–600. <https://doi.org/10.1038/nrn2457>
- Curio, G., Numminen, J., Neuloh, G., Jousmäki, V., & Hari, R. (1998). Speaking modifies utterance-related activity of the human auditory cortex. *NeuroImage*, *7*(4 PART II), 183–191. [https://doi.org/10.1016/s1053-8119\(18\)30881-4](https://doi.org/10.1016/s1053-8119(18)30881-4)
- Daliri, A., & Max, L. (2015). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language*, *143*, 59–68. <https://doi.org/10.1016/j.bandl.2015.03.002>
- Dauman, R. (2013). Bone conduction: An explanation for this phenomenon comprising complex mechanisms. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *130*(4), 209–213. <https://doi.org/10.1016/j.anorl.2012.11.002>

- Doupe, A. J., & Kuhl, P. K. (1999). *BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms*. 567–631.
- Duffy, A., Latimer, K. W., Goldberg, J. H., Fairhall, A. L., & Gadagkar, V. (2022). Dopamine neurons evaluate natural fluctuations in performance quality. *Cell Reports*, 38(13), 110574. <https://doi.org/10.1016/j.celrep.2022.110574>
- Dugas-Ford, J., Rowell, J. J., & Ragsdale, C. W. (2012). Cell-type homologies and the origins of the neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42), 16974–16979. <https://doi.org/10.1073/pnas.1204773109>
- Düring, D. N., Ziegler, A., Thompson, C. K., Ziegler, A., Faber, C., Müller, J., Scharff, C., & Elemans, C. P. H. (2013). The songbird syrinx morphome: A three-dimensional, high-resolution, interactive morphological map of the zebra finch vocal organ. *BMC Biology*, 11, 1. <https://doi.org/10.1186/1741-7007-11-1>
- Eliades, S.J., & Tsunada, J. (2018). Auditory cortical activity drives feedback-dependent vocal control in marmosets. *Nature Communications*, 9. <https://doi.org/10.1038/s41467-018-04961-8>
- Eliades, S. J., & Tsunada, J. (2024). Vocal error monitoring in the primate auditory cortex. *Bioarxiv*, 29(2), 1–24. <https://doi.org/10.37202/kmmr.2024.29.2.1>
- Eliades, S. J., & Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*, 89(4), 2194–2207. <https://doi.org/10.1152/jn.00627.2002>
- Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, 453(7198), 1102–1106. <https://doi.org/10.1038/nature06910>
- Eliades, S. J., & Wang, X. (2012). Neural correlates of the lombard effect in primate auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(31), 10737–10748. <https://doi.org/10.1523/JNEUROSCI.3448-11.2012>
- Eliades, S. J., & Wang, X. (2013). Comparison of auditory-vocal interactions across multiple types of vocalizations in marmoset auditory cortex. *Journal of Neurophysiology*, 109(6), 1638–1657. <https://doi.org/10.1152/jn.00698.2012>
- Eliades, S. J., & Wang, X. (2017). Contributions of sensory tuning to auditory-vocal interactions in marmoset auditory cortex. *Hearing Research*, 348, 98–111. <https://doi.org/10.1016/j.heares.2017.03.001>
- Elie, J. E., & Theunissen, F. E. (2016). The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal Cognition*, 19(2), 285–315. <https://doi.org/10.1007/s10071-015-0933-6>

- Forli, A., Vecchia, D., Binini, N., Succol, F., Bovetti, S., Moretti, C., Nespoli, F., Mahn, M., Baker, C. A., Bolton, M. M., Yizhar, O., & Fellin, T. (2018). Two-Photon Bidirectional Control and Imaging of Neuronal Excitability with High Spatial Resolution In Vivo. *Cell Reports*, 22(11), 3087–3098. <https://doi.org/10.1016/j.celrep.2018.02.063>
- Fortune, E. S., & Margoliash, D. (1992). Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taenopygia guttata*). *The Journal of Comparative Neurology*, 325(3), 388–404. <https://doi.org/10.1002/cne.903250306>
- Gadagkar, V., Puzerey, P. A., Chen, R., Baird-Daniel, E., Farhang, A. R., & Goldberg, J. H. (2016). Dopamine neurons encode performance error in singing birds. *Science*, 354(6317), 1278–1282. <https://doi.org/10.1126/science.aah6837>
- Gleich, O., Manley, G. A., Mandl, A., & Dooling, R. J. (1994). Basilar papilla of the canary and zebra finch: A quantitative scanning electron microscopical description. *Journal of Morphology*, 221(1), 1–24. <https://doi.org/10.1002/jmor.1052210102>
- Goffinet, J., Brudner, S., Mooney, R., & Pearson, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *ELife*, 10, 1–23. <https://doi.org/10.7554/eLife.67855>
- Gold, T. (1980). Speech production in hearing-impaired children. *Journal of Communication Disorders*, 13(6), 397–418. [https://doi.org/10.1016/0021-9924\(80\)90042-8](https://doi.org/10.1016/0021-9924(80)90042-8)
- Güntürkün, O., & Bugnyar, T. (2016). Cognition without Cortex. *Trends in Cognitive Sciences*, 20(4), 291–303. <https://doi.org/10.1016/j.tics.2016.02.001>
- Harmon, T. C., Madlon-Kay, S., Pearson, J., & Mooney, R. (2024). Vocalization modulates the mouse auditory cortex even in the absence of hearing. *Cell Reports*, 43(8), 114611. <https://doi.org/10.1016/j.celrep.2024.114611>
- Hashino, E., & Okanoya, K. (1989). Auditory sensitivity in the zebra finch (*Poephila guttata castanotis*) PACS number: 43.80.Lb, 43.66.Gf. *Journal of the Acoustical Society of Japan (E) (English Translation of Nippon Onkyo Gakkaishi)*, 10(1), 51–52. <https://doi.org/10.1250/ast.10.51>
- Hickok, G. (2012). Computational neuroanatomy of speech production. In *Nature Reviews Neuroscience* (Vol. 13, Issue 2, pp. 135–145). <https://doi.org/10.1038/nrn3158>
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. In *Neuron* (Vol. 69, Issue 3, pp. 407–422). <https://doi.org/10.1016/j.neuron.2011.01.019>
- Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. In *Current Opinion in Neurobiology* (Vol. 33, pp. 174–181). Elsevier Ltd. <https://doi.org/10.1016/j.conb.2015.04.006>
- Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. In *Frontiers*

in *Human Neuroscience* (Issue OCTOBER). Frontiers Media S. A.  
<https://doi.org/10.3389/fnhum.2011.00082>

- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, *14*(8), 1125–1138. <https://doi.org/10.1162/089892902760807140>
- Hsueh, B., Chen, R., Jo, Y. J., Tang, D., Raffiee, M., Kim, Y. S., Inoue, M., Randles, S., Ramakrishnan, C., Patel, S., Kim, D. K., Liu, T. X., Kim, S. H., Tan, L., Mortazavi, L., Cordero, A., Shi, J., Zhao, M., Ho, T. T., ... Deisseroth, K. (2023). Cardiogenic control of affective behavioural state. *Nature*, *615*(7951), 292–299. <https://doi.org/10.1038/s41586-023-05748-8>
- Hübener, M., & Bonhoeffer, T. (2014). *Neuronal Plasticity : Beyond the Critical Period*. 1989, 727–737. <https://doi.org/10.1016/j.cell.2014.10.035>
- Jaffe, P. I., & Brainard, M. S. (2020). Acetylcholine acts on songbird premotor circuitry to invigorate vocal output. *ELife*, *9*, e53288. <https://doi.org/10.7554/eLife.53288>
- Jarvis, E., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D. J., Shimizu, T., Striedter, G., Martin Wild, J., Ball, G. F., Dugas-Ford, J., Durand, S. E., Hough, G. E., Husband, S., Kubikova, L., Lee, D. W., ... Butler, A. B. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*(2), 151–159. <https://doi.org/10.1038/nrn1606>
- Jordan, R., & Keller, G. B. (2020). Opposing Influence of Top-down and Bottom-up Input on Excitatory Layer 2/3 Neurons in Mouse Primary Visual Cortex. *Neuron*, *108*(6), 1194–1206.e5. <https://doi.org/10.1016/j.neuron.2020.09.024>
- Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contribution of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature*, *433*(7026), 638–643. <https://doi.org/10.1038/nature03127>
- Karten, H. J. (2015). Vertebrate brains and evolutionary connectomics: On the origins of the mammalian ‘neocortex.’ *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1684). <https://doi.org/10.1098/rstb.2015.0060>
- Kearney, M. G., Warren, T. L., Hisey, E., Qi, J., & Mooney, R. (2019). Discrete Evaluative and Premotor Circuits Enable Vocal Learning in Songbirds. *Neuron*, *104*(3), 559–575.e6. <https://doi.org/10.1016/j.neuron.2019.07.025>
- Keller, G. B., & Hahnloser, R. H. R. (2009). Neural processing of auditory feedback during vocal practice in a songbird. *Nature*, *457*(7226), 187–190. <https://doi.org/10.1038/nature07467>
- Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron*, *100*(2), 424–435. <https://doi.org/10.1016/j.neuron.2018.10.003>
- Kelley, D. B., & Nottebohm, F. (1979). Projections of a telencephalic auditory nucleus-field L-in

- the canary. *The Journal of Comparative Neurology*, 183(3), 455–469.  
<https://doi.org/10.1002/cne.901830302>
- Kim, Y., Mizuguchi, D., & Kojima, S. (2020). Long-term Devocalization of Zebra Finches. *Bio-Protocol*, 10(18), 1–14. <https://doi.org/10.21769/BioProtoc.3752>
- Konishi, M. (1964). *EFFECTS OF DEAFENING ON SONG DEVELOPMENT IN TWO SPECIES OF JUNCOS*.
- Krützfeldt, N. O. E., Logerot, P., Kubke, M. F., & Wild, J. M. (2010). Connections of the auditory brainstem in a songbird, *Taeniopygia guttata*. I. projections of nucleus angularis and nucleus laminaris to the auditory torus. *Journal of Comparative Neurology*, 518(11), 2109–2134. <https://doi.org/10.1002/cne.22334>
- Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America*, 89(2), 859–866.  
<https://doi.org/10.1121/1.1894647>
- Lei, H., & Mooney, R. (2010). Manipulation of a Central Auditory Representation Shapes Learned Vocal Output. *Neuron*, 65(1), 122–134.  
<https://doi.org/10.1016/j.neuron.2009.12.008>
- Leonardo, A. (2004). Experimental test of the birdsong error-correction model. *Proceedings of the National Academy of Sciences of the United States of America*, 101(48), 16935–16940.  
<https://doi.org/10.1073/pnas.0407870101>
- Leonardo, A., & Konishi, M. (1999). *Decrystallization of adult birdsong by perturbation of auditory feedback*. [www.nature.com](http://www.nature.com)
- Lombardino, A. J., & Nottebohm, F. (2000). *Age at Deafening Affects the Stability of Learned Song in Adult Male Zebra Finches*.
- Löschner, J., Pomberger, T., & Hage, S. R. (2023). Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior. *iScience*, 26(3).  
<https://doi.org/10.1016/j.isci.2023.106219>
- Mandelblat-Cerf, Y., Las, L., Denissenko, N., & Fee, M. (2014). A role for descending auditory cortical projections in songbird vocal learning. *ELife*, 2014(3), 1–23.  
<https://doi.org/10.7554/elife.02152>
- Manley, G. A. (2017). Comparative Auditory Neuroscience: Understanding the Evolution and Function of Ears. *JARO - Journal of the Association for Research in Otolaryngology*, 18(1), 1–24. <https://doi.org/10.1007/s10162-016-0579-3>
- McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. <http://arxiv.org/abs/1802.03426>
- Meliza, C. D., Chi, Z., & Margoliash, D. (2010). Representations of conspecific song by starling

- secondary forebrain auditory neurons: toward a hierarchical framework. *Journal of Neurophysiology*, 103(3), 1195–1208. <https://doi.org/10.1152/jn.00464.2009>
- Mello', C. V., & Clayton<sup>2</sup>, D. F. (1994). Song-induced ZENK Gene Expression in Auditory Pathways of Songbird Brain and Its Relation to the Song Control System. In *The Journal of Neuroscience* (Vol. 14, Issue 11).
- Miall, C., & Wolpert, D. M. (1996). *Forward Models for Physiological Motor Control* (Vol. 9, Issue 8).
- Mooney, R., & Prather, J. F. (2005). *The HVC Microcircuit : The Synaptic Basis for Interactions between Song Motor and Vocal Plasticity Pathways*. 25(8), 1952–1964. <https://doi.org/10.1523/JNEUROSCI.3726-04.2005>
- Nagel, K. I., & Doupe, A. J. (2006). Temporal Processing and Adaptation in the Songbird Auditory Forebrain. *Neuron*, 51(6), 845–859. <https://doi.org/10.1016/j.neuron.2006.08.030>
- Nagel, K. I., & Doupe, A. J. (2008). Organizing Principles of Spectro-Temporal Encoding in the Avian Primary Auditory Area Field L. *Neuron*, 58(6), 938–955. <https://doi.org/10.1016/j.neuron.2008.04.028>
- Nordeen, K. W., & Nordeen, E. J. (1992). Auditory Feedback Is Necessary for the Maintenance of Stereotyped Song in Adult Zebra Finches. In *BEHAVIORAL AND NEURAL BIOLOGY* (Vol. 57).
- Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: Evidence from birdsongs in a helium atmosphere. In *Nature* (Vol. 325, Issue 6099, pp. 53–55). <https://doi.org/10.1038/325053a0>
- Oren, G., Shapira, A., Lifshitz, R., Vinepinsky, E., Cohen, R., Fried, T., Hadad, G. P., & Omer, D. (2024). Vocal labeling of others by nonhuman primates. *Science (New York, N.Y.)*, 385(6712), 996–1003. <https://doi.org/10.1126/science.adp3757>
- Payne, H. L., Lynch, G. F., & Aronov, D. (2021). Neural representations of space in the hippocampus of a food-caching bird. *Science*, 373(6552), 343–348. <https://doi.org/10.1126/science.abg2009>
- Peacock, J., Spellman, G. M., Greene, N. T., & Tollin, D. J. (2020). Scaling of the avian middle ear. *Hearing Research*, 395, 108017. <https://doi.org/10.1016/j.heares.2020.108017>
- Pomberger, T., Risueno-Segovia, C., Löschner, J., & Hage, S. R. (2018). Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of Marmoset Monkeys. *Current Biology : CB*, 28(5), 788-794.e3. <https://doi.org/10.1016/j.cub.2018.01.070>
- Popa, L. S., Streng, M. L., & Ebner, T. J. (2017). Long-Term Predictive and Feedback Encoding of Motor Signals in the Simple Spike Discharge of Purkinje Cells. *ENeuro*, 4(2). <https://doi.org/10.1523/ENEURO.0036-17.2017>

- Poulet, J. F. A., & Hedwig, B. (2002). A corollary discharge maintains auditory sensitivity during sound production. *Nature*, *418*(6900), 872–876. <https://doi.org/10.1038/nature00919>
- Poulet, J. F. A., & Hedwig, B. (2006). The Cellular Basis of a Corollary Discharge. *Science*, *311*(5760), 515–518. <https://doi.org/10.1126/science.1120937>
- Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, *451*(7176), 305–310. <https://doi.org/10.1038/nature06492>
- Price, P. H. (1979). Developmental Determinants of Structure in Zebra Finch Song. In *Journal of Comparative and Physiological Psychology* (Vol. 93, Issue 2). Waser & Marler.
- Pytte, C., Yu, Y. Lo, Wildstein, S., George, S., & Kirn, J. R. (2011). Adult neuron addition to the zebra finch song motor pathway correlates with the rate and extent of recovery from botox-induced paralysis of the vocal muscles. *Journal of Neuroscience*, *31*(47), 16958–16968. <https://doi.org/10.1523/JNEUROSCI.2971-11.2011>
- Qi, J., Schreiner, D., Martinez, M., Pearson, J., Mooney, R. (2024). Dual Neuromodulatory Dynamics Underline Birdsong Learning. In press.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. <https://doi.org/10.1038/4580>
- Resnik, J., & Polley, D. B. (2017). Fast-spiking GABA circuit dynamics in the auditory cortex predict recovery of sensory processing following peripheral nerve damage. *ELife*. <https://doi.org/10.7554/eLife.21452.001>
- Roberts, P. D., & Bell, C. C. (2000). Computational Consequences of Temporally Asymmetric Learning Rules: II. Sensory Image Cancellation. In *Journal of Computational Neuroscience* (Vol. 9).
- Roberts, T. F., Hisey, E., Tanaka, M., Kearney, M. G., Chattree, G., Yang, C. F., Shah, N. M., & Mooney, R. (2017). Identification of a motor-to-auditory pathway important for vocal learning. *Nature Neuroscience*, *20*(7), 978–986. <https://doi.org/10.1038/nn.4563>
- Roy, A., & Mooney, R. (2009). *Song Decrystallization in Adult Zebra Finches Does Not Require the Song Nucleus Nif*. 979–991. <https://doi.org/10.1152/jn.00293.2009>.
- Ryals, B. M., Dooling, R. J., Westbrook, E., Dent, M. L., MacKenzie, A., & Larsen, O. N. (1999). Avian species differences in susceptibility to noise exposure. *Hearing Research*, *131*(1–2), 71–88. [https://doi.org/10.1016/S0378-5955\(99\)00022-2](https://doi.org/10.1016/S0378-5955(99)00022-2)
- Sakata, J. T., & Brainard, M. S. (2006). Real-time contributions of auditory feedback to avian vocal motor control. *Journal of Neuroscience*, *26*(38), 9619–9628. <https://doi.org/10.1523/JNEUROSCI.2027-06.2006>

- Schmidt, M. F., Mclean, J., & Goller, F. (2012). *Breathing and vocal control : the respiratory system as both a driver and a target of telencephalic vocal motor circuits in songbirds*. 4, 455–461. <https://doi.org/10.1113/expphysiol.2011.058669>
- Schmidt, M. F., & Wild, J. M. (2014). The respiratory-vocal system of songbirds: Anatomy, physiology, and neural control. In *Zeitschrift für Krebsforschung* (Vol. 54, Issue 212). <https://doi.org/10.1007/BF01621461>
- Schneider, D. M., & Mooney, R. (2018). How movement modulates hearing. *Annual Review of Neuroscience*, 41, 553–572. <https://doi.org/10.1146/annurev-neuro-072116-031215>
- Schneider, D. M., Nelson, A., & Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517), 189–194. <https://doi.org/10.1038/nature13724>
- Schneider, D. M., Sundararajan, J., & Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. *Nature*, 561(7723), 391–395. <https://doi.org/10.1038/s41586-018-0520-5>
- Schneider, D. M., & Woolley, S. M. N. (2010). Discrimination of communication vocalizations by single neurons and groups of neurons in the auditory midbrain. *Journal of Neurophysiology*, 103(6), 3248–3265. <https://doi.org/10.1152/jn.01131.2009>
- Schneider, D. M., & Woolley, S. M. N. (2013). Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes. *Neuron*, 79(1), 141–152. <https://doi.org/10.1016/j.neuron.2013.04.038>
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)*, 275(5306), 1593–1599. <https://doi.org/10.1126/science.275.5306.1593>
- Sheintuch, L., Rubin, A., Brande-Eilat, N., Geva, N., Sadeh, N., Pinchasof, O., & Ziv, Y. (2017). Tracking the Same Neurons across Multiple Days in Ca<sup>2+</sup> Imaging Data. *Cell Reports*, 21(4), 1102–1115. <https://doi.org/10.1016/j.celrep.2017.10.013>
- Singh Alvarado, J., Goffinet, J., Michael, V., Liberti, W., Hatfield, J., Gardner, T., Pearson, J., & Mooney, R. (2021). Neural dynamics underlying birdsong practice and performance. *Nature*, 599(7886), 635–639. <https://doi.org/10.1038/s41586-021-04004-1>
- Singla, S., Dempsey, C., Warren, R., Enikolopov, A. G., & Sawtell, N. B. (2017). A cerebellum-like circuit in the auditory system cancels responses to self-generated sounds. *Nature Neuroscience*, 20(7), 943–950. <https://doi.org/10.1038/nn.4567>
- Sober, S. J., & Brainard, M. S. (2009). Adult birdsong is actively maintained by error correction. *Nature Neuroscience*, 12(7), 927–931. <https://doi.org/10.1038/nn.2336>
- Sommer, M. A., & Wurtz, R. H. (2002). *A Pathway in Primate Brain for Internal Monitoring of Movements*. <http://science.sciencemag.org/>

- Sossinka, R., & Böhner, J. (1980). Song Types in the Zebra Finch *Poephila guttata castanotis*. *Zeitschrift Für Tierpsychologie*, 53(2), 123–132. <https://doi.org/10.1111/j.1439-0310.1980.tb01044.x>
- Sperry, R. W. (1950). *NEURAL BASIS OF THE SPONTANEOUS OPTOKINETIC RESPONSE PRODUCED BY VISUAL INVERSION*.
- Stacho, M., Herold, C., Rook, N., Wagner, H., Axer, M., Amunts, K., & Güntürkün, O. (2020). A cortex-like canonical circuit in the avian forebrain. *Science*, 369(6511). <https://doi.org/10.1126/science.abc5534>
- Suthers, R. A., Goller, F., & Martin Wild, J. (2002). Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), 5680–5685. <https://doi.org/10.1073/pnas.042103199>
- Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science*, 291(5513), 2564–2569. <https://doi.org/10.1126/science.1058522>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- Trusel, M., Zhao, Z., Alam, D., Marks, E., Ikeda, M., & Roberts, T. F. (2024). *Synaptic Connectivity of Sensorimotor Circuits for Vocal Imitation in the Songbird*.
- Tschida, K. A., & Mooney, R. (2012). Deafening Drives Cell-Type-Specific Changes to Dendritic Spines in a Sensorimotor Nucleus Important to Learned Vocalizations. *Neuron*. <https://doi.org/10.1016/j.neuron.2011.12.038>
- Tumer, E. C., & Brainard, M. S. (2007). Performance variability enables adaptive plasticity of “crystallized” adult birdsong. *Nature*, 450(7173), 1240–1244. <https://doi.org/10.1038/nature06390>
- Vates, G. E., Broome, B. M., Mello, C. V., & Nottebohm, F. (1996). Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (*Taenopygia guttata*). *Journal of Comparative Neurology*, 366(4), 613–642. [https://doi.org/10.1002/\(SICI\)1096-9861\(19960318\)366:4<613::AID-CNE5>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1096-9861(19960318)366:4<613::AID-CNE5>3.0.CO;2-7)
- Veit, L., Aronov, D., & Fee, M. S. (2011). Learning to breathe and sing: Development of respiratory-vocal coordination in young songbirds. *Journal of Neurophysiology*, 106(4), 1747–1765. <https://doi.org/10.1152/jn.00247.2011>
- Wang, Y., Brzozowska-Precht, A., & Karten, H. J. (2010). Laminar and columnar auditory cortex in avian brain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(28), 12676–12681. <https://doi.org/10.1073/pnas.1006645107>

- Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' Discrimination of Paintings By Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, 63(2), 165–174. <https://doi.org/10.1901/jeab.1995.63-165>
- Wei, Z., Lin, B. J., Chen, T. W., Daie, K., Svoboda, K., & Druckmann, S. (2020). A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. *PLoS Computational Biology*, 16(9), 1–29. <https://doi.org/10.1371/journal.pcbi.1008198>
- Weir, A. A. S., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in new caledonian crows. *Science*, 297(5583), 981. <https://doi.org/10.1126/science.1073433>
- Woolley, S. C., & Doupe, A. J. (2008). Social context-induced song variation affects female behavior and gene expression. *PLoS Biology*, 6(3), 0525–0537. <https://doi.org/10.1371/journal.pbio.0060062>
- Woolley, S. M. N., & Casseday, J. H. (2005). Processing of modulated sounds in the zebra finch auditory midbrain: Responses to noise, frequency sweeps, and sinusoidal amplitude modulations. *Journal of Neurophysiology*, 94(2), 1143–1157. <https://doi.org/10.1152/jn.01064.2004>
- Yanagihara, S., & Yazaki-Sugiyama, Y. (2016). Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms11946>
- Yu, K., Wood, W. E., Johnston, L. G., & Theunissen, F. E. (2023). Lesions to Caudomedial Nidopallium Impair Individual Vocal Recognition in the Zebra Finch. *Journal of Neuroscience*, 43(14), 2579–2596. <https://doi.org/10.1523/JNEUROSCI.0643-22.2023>
- Zann, R. (1993). *Structure, Sequence and Evolution of Song Elements in Wild Australian Zebra Finches* Author (s): Richard Zann Published by : Oxford University Press Stable URL : <https://www.jstor.org/stable/4088626>. 110(4), 702–715.
- Zhou, P., Resendez, S. L., Rodriguez-Romaguera, J., Jimenez, J. C., Neufeld, S. Q., Giovannucci, A., Friedrich, J., Pnevmatikakis, E. A., Stuber, G. D., Hen, R., Kheirbek, M. A., Sabatini, B. L., Kass, R. E., & Paninski, L. (2018). Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *ELife*, 7, 1–37. <https://doi.org/10.7554/eLife.28728>

## **Biography**

Fabiola Duarte Ortiz grew up in Mexico City, Mexico. She received a BS in Basic Biomedical Research from the National Autonomous University of Mexico in 2017. At Duke University, she was the co-first author of the following paper: “A Neural Hub for Holistic Courtship Displays”, published in *Current Biology* in 2023. This work was not included in this dissertation. She is currently preparing a first-author manuscript based on the findings of Chapters 2-4 from this dissertation.

While at Duke, she received support from the Chancellor’s International Fellowship, the Ruth K. Broad Award for Graduate Students, and the Society of Neuroscience through a Trainee Professional Development Award. Additionally, she was awarded the inaugural Michael Mutersbaugh Community Award by the Department of Neurobiology at Duke University.