DISS ETH NO. 30144

Song-related neuron populations in the zebra finch forebrain chronically recorded with a custom-designed Neuropixels device

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by

CORINNA LORENZ

M.Sc. of Cognitive Science, University of Tübingen

born 15.09.1989

accepted on the recommendation of

Prof. Dr. Richard H. R. Hahnloser

Dr. Nicolas Giret

Dr. Daniela Vallentin

Prof. Dr. Manfred Gahr

2024

On either side of the Manger that night it was quiet at one A.M. and it was quiet at two A.M. and it was quiet at three A.M. and it was such a *loud* quietness at four A.M. that everyone blinked, sat up in bed, and *listened*.

Ray Bradbury, “We’ll Always Have Paris” (2009)

Abstract

Songbirds (*Passeri*) have evolved a network of brain areas dedicated to one specific task – learning to produce complex song. This close link between structure and behavior is a rare trait in the animal kingdom, presenting them as an excellent model for the study of neural motor learning and control. Research in this field, however, still lacks behind technical advances of high-yield electrophysiology because songbirds are particularly small and sensitive to the weight of a chronic implant.

To address these challenges, we developed a lightweight chronic Neuropixels implant assembly (NIA). Enabling independent placement of the headstage and the NIA, it surpasses previous designs in its adaptability and flexibility to target brain areas at more extreme angles while maintaining a minimal weight. Combined with refined surgical and behavioral protocols, this enabled a fast recovery time and reduced burden when tested in freely moving zebra finches (*Taeniopygia guttata*). Consequently, birds resumed singing within days after implantation, facilitating the collection of an extensive dataset (> 350 hours) consisting of spiking activity from both the cortex-like premotor region LMAN (lateral magnocellular nucleus of the anterior nidopallium) and the basal-ganglia-like area including Area X.

We tested the neural recording quality over several days and evaluated spike sorting results for sessions with significant amount of singing. The results showcase the effectiveness of the custom-designed NIA in facilitating high-yield recordings, capturing up to 100 units simultaneously. Despite a slow but gradual decline in spiking signals, we collected comprehensive datasets with substantial singing and robust neural counts (> 140 h). In part of these datasets, we effectively classified neuronal cell types in Area X, affirming the validity of our approach by both replicating and building upon previous research. Crucially, these recordings allowed the first-ever assessment of neuronal interactions between LMAN and Area X during singing, offering unprecedented insights into the coupling between these two areas and unveiling a complex interplay within the neural network.

To gain further insight into the relationship between LMAN and song behavior, we analyzed pitch variability from one rendition to the next using a latent dynamic model combined with song recordings from birds both before and after LMAN lesioning. Our analysis revealed a stochastic pattern in pitch variability which is closely linked to LMAN, supported by a robust linear relationship between lesion volume and the reduction in estimated variability. Extending this approach to our Neuropixels recordings, we find significant neuro-behavioral correlations that align with the idea of LMAN's involvement in pitch variability. However, comparing neuro-behavioral correlates between estimated and observed pitch variability yielded inconclusive results and hints towards a more intricate neural control mechanism responsible for regulating vocal variability.

Lastly, we developed a semi-supervised method for efficiently segmenting and clustering vocalizations of adult zebra finches, a previously tedious task often impeded by background noises. We use a dimensionality reduction technique to embed short, overlapping segments of

the sound spectrogram in a 2D plane. In a human-in-the-loop approach, we allow users to explore and define a given vocalization type by specifying pairs of high-density regions in the 2D embedding, one region associated with vocalization onsets and the other with offsets. We demonstrate our two-neighborhood (2N) extraction method by identification of various vocalization types from continuous data streams. Our approach yields lower false positive rates than comparable approaches and, owing to our purpose-built graphical user interface (GUI) for visualizing and annotation data, can be performed in a fast and user-friendly manner.

Taken together, we present a novel implant design facilitating the recording of high-yield electrophysiological data from freely moving and singing birds. Notably, the chronic nature of the implant allowed the acquisition of an extensive dataset of neural activity during intrinsically motivated singing, female elicited singing, playback of the birds own song, and sleep, and will hopefully help to further our understanding of the neuronal code in the LMAN/Area X network with respect to the questions addressed in this thesis and beyond.

Zusammenfassung

Singvögel (*Passeri*) besitzen ein dezidiertes Netzwerk an Hirnstrukturen, das sich einer spezifischen Aufgabe widmet – dem Erlernen und Produzieren von komplexen Gesängen. Diese enge Verbindung zwischen Struktur und Verhalten ist einzigartig im Tierreich und postuliert Singvögel dadurch als exzellentes Forschungsmodell für die Untersuchung von neuronalen Mechanismen der motorischen Kontrolle und des motorischen Lernens. Aktuell hält die Forschung in diesem Bereich jedoch nicht mit den technischen Fortschritten hochskalierter elektrophysiologischer Messungen mit, da Singvögel besonders klein sind und empfindlich auf das zusätzliche Gewicht neuartiger Messgeräte in einem chronischen Implantat reagieren.

Um diese Herausforderungen anzugehen, haben wir eine leichtes, chronisches Hirnimplantat auf Basis von Neuropixels Elektroden (NIA) entwickelt. Durch die unabhängige Platzierung der Hilfselektronik und dem NIA übertrifft es frühere Entwürfe in Bezug auf die experimentelle Anpassungsfähigkeit und Flexibilität. Es kann dadurch Hirnareale unter extremen Winkeln ansteuern, während es ein minimales Gewicht beibehält. In Kombination mit verfeinerten chirurgischen und verhaltensbiologischen Protokollen ermöglicht unser Ansatz eine schnelle Erholungszeit nach der Implantierung und reduziert die generelle Belastung, wie Tests in frei beweglichen Zebrafinken (*Taeniopygia guttata*) zeigen. Die getesteten Vögel fingen innerhalb der ersten Tage nach der Operation wieder an zu singen. Dies ermöglichte die Aufnahme eines umfangreichen Datensatzes (> 350 Stunden), der erstmalig simultan die neuronale Aktivität in dem kortikalen prämotorischen Areal LMAN (lateraler magnozellulärer Kern des vorderen Nidopalliums) und der gesangsrelevanten Region in den Basal Ganglien Area X erfasst.

Wir testeten die Qualität der neuralen Aufzeichnungen über mehrere Tage und bewerteten die Ergebnisse nach Zuordnung der Aktionspotentiale für Aufnahmen mit einer hohen Anzahl an Gesangswiederholungen. Die Ergebnisse zeigen die Wirksamkeit des speziell entwickelten NIAs für die skalierte Messung intrakortikaler Aktionspotentiale, bei denen bis zu 100 individuelle Neurone bzw. Gruppen davon gleichzeitig erfasst werden konnten. Trotz eines langsamen, aber stetigen Rückgangs der neuronalen Signale sammelten wir umfassende Datensätze mit erheblichem Gesang und einer robusten Anzahl neuronalen Einheiten (> 140 Stunden). In einem Teil dieser Datensätze klassifizierten wir neuronale Zelltypen in Area X, was die Gültigkeit unseres Ansatzes sowohl durch die Replikation als auch durch den Ausbau vorheriger Forschung untermauert. Ein wesentlicher Beitrag unserer Arbeit ist, dass diese Aufzeichnungen die allererste Untersuchung neuronaler Interaktionen zwischen LMAN und Area X während des Singens ermöglichen. Dies bietet ein beispielloser Einblick in die flächenübergreifende Kopplung und enthüllten ein komplexes Zusammenspiel innerhalb der verschiedenen Hirnregionen.

Um weitere Einblicke in die Beziehung zwischen LMAN und dem Gesangsverhalten zu gewinnen, analysierten wir die Variabilität in der Tonhöhe einer Gesangssilbe die über mehrere Wiederholungen zu beobachten ist. LMAN ist eine zentrale Quelle dieser Variabilität, doch ist bisher unklar, wie gross der Anteil an Variabilität bezogen auf LMAN ist und welche Merkmale ihn beschreiben. Anhand eines latenten Variablenmodellen analysierten wir die Gesangsaufnahmen von Vögeln vor und nach bilateraler Läsionierung von LMAN. Unsere Analyse deutet auf ein

stochastisches Muster in der Variabilität der Tonhöhe hin, das mit LMAN assoziiert ist. Die Korrelation zwischen der Läsionsgrösse und der Veränderung dieser Art latenter Motorvariabiltät war nahezu perfekt, und sie war stärker als die Korrelation mit der Veränderung der Tonhöhenvariabilität. Wir erweiterten diesen Ansatz auf unsere Aufnahmen mit Neuropixels Elektroden erweiterten und fanden signifikante Korrelationen zwischen prämotorischer neuronaler Aktivität in LMAN und Tonhöhe, in Übereinstimmung mit der Idee das LMAN zu Variationen in der Tonhöhe beiträgt. Der Vergleich von Korrelationen zwischen latenter und beobachteter Tonhöhenvariabilität lieferte jedoch inkonsistente Ergebnisse und deutet auf einen komplexeren neuralen Kontrollmechanismus hin, der für die Regulierung der vokalen Variabilität verantwortlich ist.

Zuletzt entwickelten wir eine semi-überwachte Methode zur effizienten Segmentierung und Annotieren von Gesängen erwachsener Zebrafinken, eine zuvor mühsame Aufgabe, die oft durch Hintergrundgeräusche erschwert wird. Unser Ansatz basiert auf Methoden, die die Dimensionalität grosser Datenmengen reduziert und kurze, überlappende Segmente des Audiospektrogramms in einer 2D-Ebene einbettet. Im Gegensatz zu anderen Methoden, bezieht unser Ansatz Benutzer aktive in die Entscheidungsfindung mit ein. Wir ermöglichen es den Benutzern einen bestimmten Vokalisationtyp zu erkunden und zu definieren, indem sie Regionen hoher Dichte im zweidimensionalen Embedding angeben: eine Region, die mit dem Anfang assoziiert ist, und eine mit dem Ende. Wir veranschaulichen unsere Extraktionsmethode durch die Identifikation verschiedener Arten von Vokalisation in Gesangsaufnahmen, die sich über mehrere Tage erstreckten. Unser Ansatz führt zu niedrigeren falsch-positiv Raten als vergleichbare Ansätze und kann dank unserer speziell entwickelten grafischen Benutzeroberfläche (GUI) für die Visualisierung und Annotieren von Daten auf schnelle und benutzerfreundliche Weise durchgeführt werden.

Zusammenfassend präsentieren wir ein neuartiges Design für Hirnimplantate, die die Aufzeichnung von skalierten neuronalen Messungen in sich frei bewegenden und singenden Vögeln erleichtert. Insbesondere die chronische Natur der Implantate erlaubte es uns, einen reichhaltigen Datensatz von intrakortikaler neuronaler Aktivität zu sammeln – und zwar während des intrinsisch motivierten Singens, des Balzgesangs, beim Abspielen des eigenen Gesangs sowie während des natürlichen Schlafzustandes. Diese Erkenntnisse erhoffen wir, werden einen wertvollen Beitrag leisten, um das Verständnis der neuronalen Kodierung im Vorderhirn der Vögel zu vertiefen, sowohl in Bezug auf die in dieser Arbeit adressierten Fragestellungen als auch darüber hinaus.

Acknowledgements

I have come to realize that pursuing a PhD shares many similarities with running a marathon. Both journeys are long, demanding, and filled with ups and downs—moments when progress feels effortless and exhilarating, contrasted by times when the finish line seems impossibly distant.

In both situations, the path often feels lonely, and having people cheering and supporting from the sideline makes all the difference. In this aspect, I consider myself exceptionally fortunate. I am deeply grateful for the encouragement I've received throughout the years, and I hope my words can adequately convey my appreciation.

First, I want to thank my advisor Prof. Dr. Richard Hahnloser for his continuous support during the past years. Without his trust in our exploration to find new methods, this thesis would not have been possible. Furthermore, I would like to thank Dr. Nicolas Giret, who kindly agreed to serve as my second supervisor early on. His guidance and encouragement have been invaluable in the progress and successful completion of this thesis. I am particularly grateful to him for welcoming me into his group, providing support and helpful advice at all steps of my experiments, and always maintaining an open-door policy (even virtually).

I am also deeply grateful to Dr. Ezequiel M. Arneodo and Dr. Anja T. Zai, who have significantly influenced this work, contributing directly to many aspects of the projects and also in more subtle, yet equally impactful ways. Ezequiel is the truest experimentalist I know. It was thanks to his visionary perspective that we pursued the idea to adapt Neuropixels for use in songbirds, and he taught me many of the essential skills that enabled me to independently conduct the experiments. Importantly, working with him, while ambitious and challenging, was always fueled by joyful curiosity and an experience I will always hold dear. Anja's impact on this thesis has been immense. Right from the start, she has been both a wellspring of inspiration and a pillar of support. Our discussions about the intricacies of songbirds and the AFP have provided me valuable insight and reflection that enriched my understanding and approach to the research. Learning from her and working alongside her has been a remarkably fortunate and stimulating experience I am forever grateful for.

I also want to thank Linus Rüttimann, Tomas Tomka, Rik Ubaghs, Diana Rodriguez, and Sophie Cave-Lopez, for being my peers and sharing many of the moments on this journey together, as well as all members of the current and previous members of the songbird and NLP group at INI and the Neural Code and Auditory Perception group at Paris-Saclay Institute of Neuroscience for fruitful scientific discussions, fun lab disassembling and assembling events, and joyful social events.

Furthermore, I want to express my gratitude to Simone Rickauer, Nathalie Desvignes and Bernadette Wiszniowksi for help with histological preparations, and Heiko Hörster for excellent animal caretaking. Special thanks are also due to Diana, Ezequel, Sophie, Anja, and Nicolas for patiently teaching me surgical procedures and sharing invaluable insights from their own experiences.

Lastly, I want to thank my friends and family. Thank you for showing up and cheering for me every step of the way.

Content

List of abbreviations

Neuroanatomy and neurochemistry

Ach	Cholinergic neuron
AFP	Anterior forebrain pathway
Area X	Area in the AFP homologue to the mammalian basal ganglia
BMI	Bicuculline methiodide
DA	Dopamine
DLM	Dorsolateral thalamus
FS	Fast-spiking neuron
GABA	Gamma-aminobutyric acid
GPe	Neuron of the external Globus palliudus
GPi	Neuron of the internal Globus palliudus
HVC	Proper name
LC	Locus coeruleus
LMAN	Nucleus of the anterior nidopallium
LTS	Low-threshold spiking neuron
MSN	Medium spiny neuron
NA	Noradrenaline (sometimes also called Norepinephrine)
RA	Robust nucleus of the archistriatum
VTA	Ventral tegmental area

others

dph	Days post hatch
FD	Female directed singing
NIA	Neuropixels implant assembly
MEA	Multielectrode array
UD	Undirected singing (in isolation)
WN	Brief, loud burst sound played through the loudspeaker

Introduction

Functional neurobiology with extracellular electrophysiology

From intrinsic electricity to fine neural potentials

The constant drive to understand the forces at play in the coding of our nervous system that ultimately underly the complex behavior observed in humans and animals has gone hand in hand with advances in neural recording techniques.

The first landmark in this journey was the recognition that electricity is the fundamental mean of neural communication in the body of all animals. This was showcased by Luigi Aloisio Galvani in 1791 when he published a series of carefully conducted experiments demonstrating that electrical stimulation elicit muscle twitches (Galvani, 1791). Although there had been similar observations before, it was the thoroughness of the experiments and conclusions drawn from the results that convinced scientists at the time that electricity must play a fundamental role in neural signaling. The pivotal shift in the comprehension of the nervous system did not convince everyone. Opponents, such as Alessandro Volta, continued pointing out that the experiments did not conclusively establish the intrinsic nature of the observed electricity (reviewed in Finger, 2005). It was clear that a device capable of detecting electrical voltage in an unstimulated animal was necessary to fully substantiate Galvani's hypothesis.

Almost 40 years after Galvani's publication, Emil Du Bois-Reymond mastered such recordings. With the aid of an electromagnetic galvanometer[1], he was able to measure the current flowing through the body of various animals (and himself). Technically skilled, he also managed to measure the current during muscle contraction. He overcame the galvanometer's insensitivity to transient deflections with the brilliant idea to apply high-frequency stimulation to a dissected muscle. Upon tetanic contraction, he was able to observe negative deflections (Du Bois-Reymond, 1843; Finkelstein, 2013) which he named 'action potential' to contrast them from the positive deflections observed in passive preparations termed 'resting potential'.

Following Santiago Ramon y Cajal's histological study on the avian cerebellum (Ramón y Cajal, 1888) and the subsequent shift toward the "neuron doctrine", which posits that the brain consists of individual neurons connected by synapses (Sherrington, 1906), there emerged an increased interest in understanding the functional properties of these units. However, the recording techniques employed such as Du Bois-Reymond's galvanometer proved too coarse to detect the small potentials generated by individual nerves.

[1] A galvanometer operates on the laws of electromagnetism and measures electric current using a wire coil placed inside a permanent magnetic field. Passing current through the coil generates a magnetic field that aims to align with the outer magnetic field, subsequently rotating the coil and pivoting a coil-attached needle. In an astatic setting, the deflection of the coil is proportional to the intensity of the current passing through it.

The experimental limitation prompted another technological evolution, marked by the development of amplification systems capable of multiplying signals up to 7000 times with minimal distortion (summarized in Finger, 2005b). This methodological revolution enabled the detection of subtle electrical fluctuations, first in peripheral nerve bundles (Gasser & Erlanger, 1922) and later in single nerves, revealing fundamental features of neural coding such as the all-or-nothing principle, rate coding and adaptation in both the sensory and motor domain (Adrian, 1926; Adrian & Bronk, 1928; Adrian & Zotterman, 1926b, 1926a). A decade later then, Renshaw, Forbes, and Morrison successfully adopted electrophysiology to perform the very first intracortical recordings of action potentials (Renshaw et al., 1940). They actuated a thin electrode in juxtaposition to hippocampal neurons to measure their fine neuronal potentials in anesthetized cats. Notably, accomplishing such experiments required a high technical understanding of signal processing in combination with surgical talent and most importantly perseverance. For instance, David H. Hubel and Torsten N. Wiesel, known for their groundbreaking work on neural processing in the visual cortex, used 40 cats and yielded an average of 7.5 neurons per animal in their acute experiments (Hubel & Wiesel, 1962). In the following years, experimental recordings of neurons remained an immense effort.

Scaling intracortical recording techniques

Today, electrophysiological recording methods remain the gold standard in neuroscience because of their ability to measure neural activity with high temporal resolution in almost any brain location and any animal model. Intracellular recording techniques, albeit more insightful with respect to changes in the resting potential, are experimentally more challenging. Especially in freely moving animals, it is difficult to keep the recording micropipette in place for long enough and even in head-fixed settings, it can happen that the pipettes break. Extracellular recording methods offer more reliability, are commonly employed in freely moving settings and have the advantage of capturing multiple signals at once.

Owing to the evolution in electronic manufacturing, characterized by miniaturization, increased automation, the emergence of advanced semiconductor technologies, and the shift towards more sustainable and efficient manufacturing practices, electrophysiology has evolved into an almost plug-and-play discipline. Furthermore, the technical advancements of the past decades have given rise to new electrode types, specifically designed for extracellular electrophysiology aiming to scale neural yield. These various types are illustrated in Figure 1.3 and are succinctly reviewed in chronological order.

Microwire electrodes (Figure 1A) consist of thin and stiff wires, usually made from metal such as tungsten, metal alloys such as nichrome, or carbon fibers that can be inserted into the brain to record the activity of individual neurons. They are covered by a biocompatible, insulating material, e.g. glass or Parylene-C so that only the tip is exposed and in contact with the tissue. The goal of microelectrode recordings is to capture action potentials of single neurons with excellent signal to noise ratio. This is usually accomplished by placing the electrode on a drive so that it can be

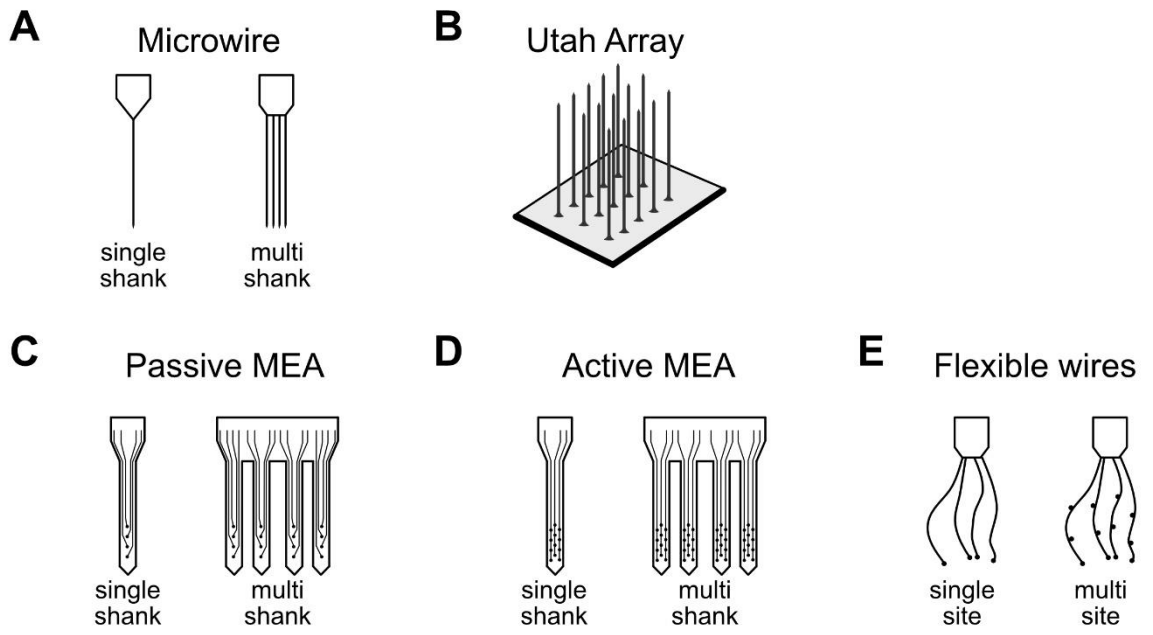


Figure 1.1 Different electrode types used for in vivo extracellular electrophysiology. *Points indicate electrode sites. MEA: Multi-electrode array. See section Scaling intracortical recording techniques for details.*

actively positioned in close proximity to a neuron. Their main shortcoming is the limited scalability and neural recording yield. With one contact site per wire, the width of a multi-wire shank grows linearly with the number of wires, which increases undesirable brain damage. Furthermore, controlling the position of individual wires can be challenging because the necessary mechanical or electrical components add weight that scales proportionally with the number of individually actuated units (e.g. Voigts et al., 2013).

An alternative approach is to configure a combination of multiple shanks of wires or wire bundles to increase the recording yield. Such an approach has been realized with Utah arrays (Figure 1.1B). Developed in the 1990s, they consist of a silicon-based 2D or 3D array of closely spaced microelectrodes (Rousche & Normann, 1998). With each shank carrying one electrode contact site at the tip, this configuration allows for high spatial resolution and increases the likelihood of recording many neurons simultaneously. Compared to actuated microwires, Utah arrays are implanted and placed permanently. They are particularly useful for brain-machine interfaces and neural prosthetics as they provide a particular good balance between independent placement, neural readout, and long-term stability in primates (X. Chen et al., 2023; Sponheim et al., 2021). However, because of the configuration of one electrode contact per shank, Utah arrays keep an undesirable low electrode sites/penetration ratio.

Advances in technology led to the development of more sophisticated multielectrode arrays (MEA), including passive MEAs, named so because they simply transmit signals from one contact point in the brain to another at the connector without passing the signal through active components. Using semiconductor lithography technology, passive MEA such as Michigan-style

arrays, can consist of multiple electrodes on a single substrate either in a single or multi-shank configuration (Figure 1.1C). Passive MEAs thus provide a compromise between the spatial resolution of microwires and the scalability of Utah arrays. They allow for large-scale recordings from multiple neurons and are often used in research involving chronic implants. The channel count is limited by the number of electrode lines fitting into a shank and similar to microwire or Utah arrays, by the interface cable connecting the MEA with substrate that carries electrical processing components.

Active MEAs (Figure 1.1D) are an evolution of passive arrays incorporating the necessary signal processing circuitry on the same substrate. Compared to passive MEAs, they must be powered which is used to the advantage of multiplexing the signals into one data stream. They also allow for on-site amplification of neural signals, reducing noise and improving signal quality. Because of the actively powered circuitry and small components, active MEAs cannot be used for electrical stimulation experiments yet. However, because they can incorporate modern complementary metal-oxide-semiconductor (CMOS) technology, it is possible to place many electrode contacts in line which yields an unprecedented balance between shank size and electrode number.

Interestingly, another branch of microwire electrodes has recently emerged. Combining modern microfabrication techniques and material science, it became possible to manufacture flexible electrical wires that can carry multiple contact sites (Figure 1.1F). Because the wire can move with the tissue, it promises long-term recording stability with minimal tissue irritation(Lycke et al., 2023; Zhao et al., 2023). A major challenge, however, is the initial placement of the wires. Strategies involving a needle-guided insertion have been proposed but the width of these needles is to-date larger than those of commercially available active MEAs (Hong & Lieber, 2019). It is very likely that these will advance in time and such designs will become commercially available.

Promises and challenges of large-scale neural recordings with multielectrode arrays

The field of neuroscience has been fundamentally transformed by the advent of large-scale electrophysiological recording techniques, particularly through the use of passive and active MEAs. A key promise of these techniques lies in their ability to simultaneously record the activity of multiple neurons (Steinmetz et al., 2018). Additionally, the non-specific nature of electrophysiology allows for the recording of a diverse range of cell types, each characterized by unique electrophysiological properties (Goldberg et al., 2010a; Goldberg & Fee, 2010a; Mitchell et al., 2007). This capability provides a rich dataset crucial in unraveling the neural codes and networks that underlie various cognitive processes and behaviors, marking a significant step forward in our understanding of neural dynamics.

Another major stride in electrophysiology is its ability to record neural populations within and across distributed brain areas (Girardeau et al., 2017; Lemke et al., 2019; Semedo et al., 2022; Steinmetz et al., 2019) coupled with the capacity to access deep brain structures (Gadagkar et al., 2016; Girardeau et al., 2017; Hesse & Tsao, 2020; Ruder et al., 2021), a feat challenging for other techniques such as calcium imaging. Recording across various regions offers a comprehensive view of brain functionality, crucial for understanding the coordination and integration of information across the brain, essential in studying complex cognitive functions.

Furthermore, the depth penetration achieved by electrophysiology is invaluable for investigating subcortical regions involved in fundamental neural processes and implicated in numerous neurological disorders. This dual capability of encompassing both superficial and deep brain areas provide a more complete picture of neural activity, which is crucial for fully understanding the brain's intricate architecture and functioning.

However, these advancements come with significant challenges. Spike sorting, the process of categorizing the action potentials of different neurons from recorded signals, remains a complex and time-consuming task. As the number of simultaneously recorded neurons increases, the complexity of spike sorting escalates exponentially, requiring sophisticated algorithms and substantial computational resources (reviewed in Carlson & Carin, 2019). This challenge is compounded by the massive volume of data generated by large-scale recordings. Managing, storing, and analyzing this data demand advanced data processing techniques and significant computational power. On the analytical side, statistical tools have been proposed that leverage the high dimensionality of data (Cunningham & Yu, 2014; Gokcen et al., 2022; Semedo et al., 2020) which ask for a new understanding of brain representations yet to be reconciled with the classical single-neuron doctrine (Stevenson & Kording, 2011; Yuste, 2015). These challenges represent substantial hurdles in fully harnessing the potential of large-scale neural recording techniques in electrophysiology. As the field progresses, developing more efficient spike sorting methods and data management strategies will be crucial for advancing our understanding of the brain.

A remaining challenge is the adoptability of MEA designs, especially in freely moving animals. All recent developments in large-scale neural recordings heavily focus on just a few animal models studied under very restricted experimental conditions (Keifer & Summers, 2016). Instead of solely pursuing new developments, there is growing interest in transferring existing tools to non-classical animal models and studying the neural correlates of their natural behavior.

Functional neurobiology requires behavior

The relevance of neuroethological studies

Animals exhibit a rich diversity of behaviors that are highly dynamic and rooted in the interaction with their ecosystem, their ontogenesis, and evolution (Tinbergen, 1963, 1989). The study of such natural behavior has led to key insights into the neural basis of toad’s pray detection (Ewert, 1974), bat’s echolocation (Yartsev et al., 2011; Yartsev & Ulanovsky, 2013), or vocal learning (Konishi, 1965; Nottebohm et al., 1986) just to name few. Neuroethology in general, and computational neuroethology in particular, are growing fields of research that accelerated in recent years thanks to a scaling of technical and analytical tools. Computational neuroethology aims to gain insight into the neural principles that enable and control natural behaviors by means of big data collections and simulations (Beer, 1990; Chiel & Beer, 1997). Practically, advanced recording systems that sample information simultaneously from multiple sensors allow to track individual animals spontaneously behaving in natural conditions. Analytical tools, on the other hand, are beginning to become available to leverage these extensive datasets. Based on modern

machine learning algorithms, analytical tools, on the other hand, are available to describe the vast amount of collected data by discretizing and labelling actions and motifs either in supervised or un-supervised fashion (Datta et al., 2019; Sainburg & Gentner, 2021).

Songbirds for neuroethological research

Songbirds are exemplary animal models in (computational) neuroethology. They demonstrate a rich array of behaviors, particularly in their vocalizations, which are easily observable and can be efficiently recorded, even for large-scale studies. The most notable characteristic is that they learn a vocal song repertoire through imitation of conspecifics (Marler, 1970; Marler & Peters, 1977; Marler & Tamura, 1964; Thorpe, 1958). This is a rare trait in the animal kingdom, and compared to mammals, much more common in birds as it includes songbirds (>5000 species), parrots and hummingbirds (Hyland Bruno et al., 2021). Furthermore, the ability to learn vocalizations in songbirds shares genetic, developmental, and neural commonalities with human speech acquisition (Haesler et al., 2007; Lipkind et al., 2019; Prather et al., 2017; Teramitsu et al., 2004; reviewed in Doupe & Kuhl 1999).

Zebra finches (*Taeniopygia guttata*) are the most commonly studied songbird in neuroscience and animal vocal learning research (Lattenkamp & Vernes, 2018). These small and hardy birds easily adapt to captivity, and their capability for year-round breeding, coupled with a rapid reproductive cycle, makes them ideally suited for research in a laboratory setting.

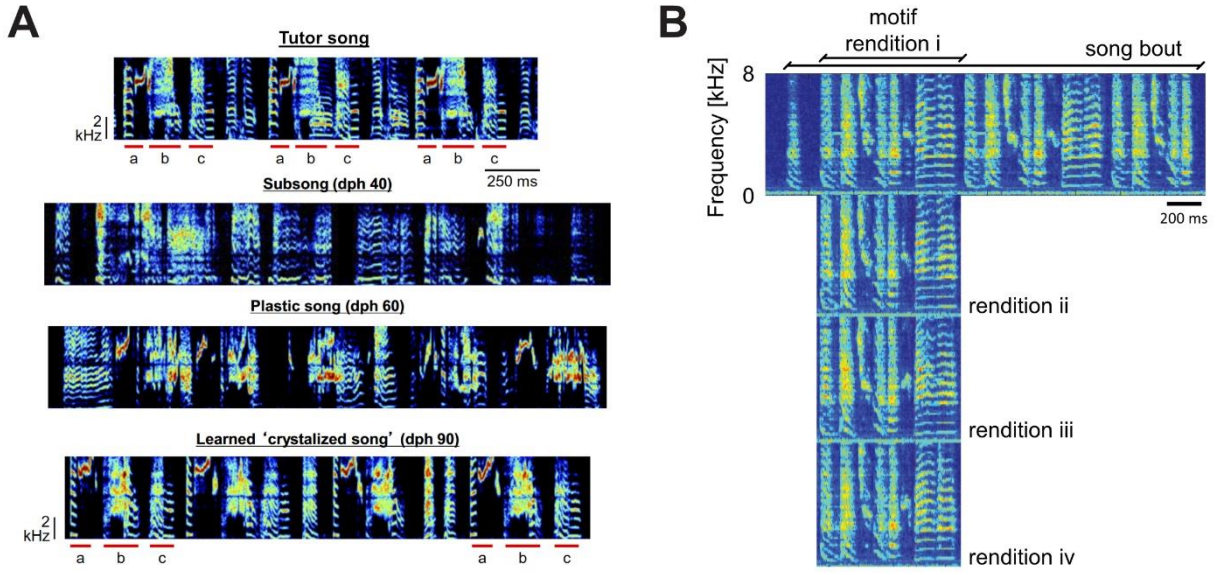


Figure 1.2 Zebra finch vocal ontogeny and adult song. A *Spectrogram of the tutor song and the juvenile's evolving copy. he tutor's song motif, characterized by the 'a b c' syllable sequence, is illustrated at the top. Progressively, as shown in the lower spectrograms, the juvenile gradually adopts this sequence. The transformation begins with the initial babbling in the subsong, advances to a more mature temporal structure in the plastic song phase and culminates in a well-defined imitation of the tutor's song. Graphic adapted from Goldberg and Fee (2011).* ***B*** *Example spectrogram of an adult zebra finch song. The song bout begins with an introductory note and is followed by a set of syllables grouped into reoccurring motif. The last motif in this example was aborted prematurely. The lower spectrograms are randomly chosen motif renditions from other song bouts. Note the subtle variations in spectral features such as the frequency contour in the harmonic syllable from one rendition to the other in the overall highly stereotyped behavior.*

In zebra finches, only males sing. They learn their song from a tutor, usually their father, during a critical developmental period (reviewed in Bolhuis & Gahr, 2006; Doupe & Kuhl, 1999; Kuhl, 2004). After approximately three months after hatching, they crystalized their initial babbling into their final acoustically complex structured song (Funabiki & Konishi, 2003, Figure 1.2A). The adult song structure typically begins with introductory notes and proceeds to one to ten repetitions of a fixed sequence of syllables (vocalizations separated by brief moments of silence), known as the motif (Sossinka & Böhner, 1980). Each male has a unique motif. Figure 1.2B shows an example song together with a selection of randomly chosen motif renditions demonstrating that while the motif exhibits overall stereotypy, each motif rendition displays subtle spectral and temporal differences.

Interesting for neuroscience studies, songbirds evolved with a special set of brain regions dedicated to song learning and control (Nottebohm & Arnold, 1976; Ölveczky et al., 2005a; Roberts et al., 2012). It consists of two distinct premotor pathways (Figure 1.3). Both converge on the downstream analogue of the mammalian motor cortex (the robust nucleus of the arcopallium, RA) but each of them appears to provide different functionality. The 'motor' pathway directly connects the premotor region HVC (former acronym now used as a proper name) to RA and is required to produce stable adult song (Nottebohm et al., 1976; Simpson & Vicario, 1990). The anterior forebrain pathway (AFP) indirectly connects HVC to RA, *via* a loop that includes the cortical-like nucleus LMAN (lateral magnocellular nucleus of the anterior nidopallium), the basal ganglia Area X, and a thalamic nucleus (medial portion of the dorsolateral thalamic nucleus, DLM). The AFP is a homologue of the mammalian cortico-basal ganglia-thalamo-cortical loop (Gale & Perkel, 2010; Reiner, 2002; Reiner et al., 1998, 2004) and is relevant for song plasticity, e.g.,

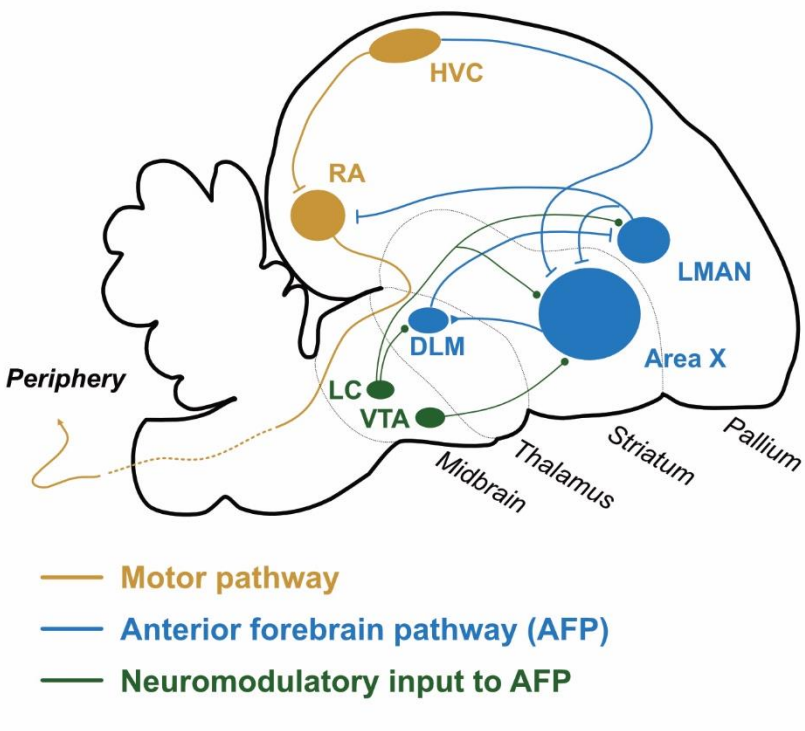


Figure 1.3 *Schematic of a songbird brain in sagittal section highlighting song-related neural structures. The motor pathway is required for the faithful production of the learned song, the anterior forebrain pathway (AFP) is responsible for initial song learning and adult plasticity. Neuromodulatory input from midbrain regions to the AFP has recently been demonstrated vital to the AFP's functionality. (⊢ excitatory, ►inhibitory, ● modulatory input; LC locus coeruleus, VTA ventral tegmental area, LMAN lateral magnocellular nucleus, DLM medial portion of the dorsolateral thalamic nucleus, RA the robust nucleus of the arcopallium, HVC and Area X proper names).*

during initial song acquisition in juveniles (S. Bottjer et al., 1984; Ölveczky et al., 2005a) or later to adapt features of the song to environmental changes in adults (Ali et al., 2013a; Andalman & Fee, 2009a). Research in recent years has further elucidated the critical roles of neuromodulatory input to the AFP from midbrain regions. Dopaminergic innervation from the ventral tegmental area (VTA) and noradrenergic input from the locus coeruleus (LC) were demonstrated to significantly influence vocal plasticity and variability in songbirds (Alvarado et al., 2021a, 2023; Hisey et al., 2018; Hoffmann et al., 2016; Leblois & Perkel, 2012; Xiao et al., 2018).

This close link between structure and behavior presents songbirds as an excellent model for the study of neural motor learning and control in general, and in the context of neuroethology in particular since songbirds sing and learn to sing naturally without external experimental incentive.

Neural control of vocal variability: current state of research

Despite the availability of large-scale neural recording methods in neuroscience, they been applied only sparingly in songbird research. While single-cell recordings in the motor pathway have shown structured neural correlates of song (Hahnloser et al., 2002a; Leonardo & Fee, 2005; Lynch et al., 2016; Sober et al., 2008; Vallentin et al., 2016), which led to a progressive understanding of the neural computations on a systems level, research on song variability and adaptability of the AFP on the systems level is comparably limited. We will briefly review the current understanding of the AFP and highlight open questions that we believe could be addressed through neuron population recording studies.

The causal role of the AFP

The AFP is the main driver of spectral song variability and plays a crucial role in vocal learning, especially during the initial song ontogeny (S. W. Bottjer et al., 1984; Ölveczky et al., 2005b; Scharff & Nottebohm, 1991). During development, the neural functionality of the AFP appears to shift from LMAN initially driving RA to later modulating it (Aronov et al., 2008). A residual of this capacity remains after song crystallization and the AFP's role in the subtle acoustic variability of adult song has been conclusively demonstrated by studies involving lesions, temporary inactivation, or disruption of synaptic transmission (Hampton et al., 2009; Kao & Brainard, 2006; Ölveczky et al., 2005b; Stepanek & Doupe, 2010; Warren et al., 2011), especially on the level of the AFP's output, LMAN.

In LMAN, these interventions consistently lead to reduced spectral variability mostly tested on fundamental frequency of a part of a syllable with stable harmonic stack (pitch). Additionally, targeted electrical microstimulation of LMAN subregions disrupt and alter the vocal output transiently after a short latency (around 50 ms) indicative of a direct contribution of LMAN to the ongoing vocal performance (Giret et al., 2014a; Kao et al., 2005a; Kojima et al., 2018a).

Causal experiments in Area X and DLM in adults are comparably sparse. An earlier study examined their impact performing lesions in birds with disrupted motor pathway. After lesioning HVC, Area X ablation caused a slowing and simplification of the acoustic structure of song. This effect was also observed after DLM lesion but manifested itself even stronger suggesting that

along the AFP pathway complexity increases (J. R. Chen et al., 2014). A more fine-grained acoustic analysis of pitch after Area X lesion was reported by Kojima et al. (2018a). The authors found Area X lesion to decrease pitch fluctuations within a harmonic stack without affecting the variability across renditions (that might stem from LMAN).

Experimental paradigms of behavioral relevant motor variability

In the study of AFP-induced variability, three experimental paradigms have led to key insights into its behavioral relevance:

Firstly, inducing **deafness** in birds, such as through cochlear removal, leads to a progressive degradation of their song, a phenomenon not observed in birds with normal hearing (Lombardino & Nottebohm, 2000; Nordeen & Nordeen, 1992). However, this degenerative effect is prevented when lesions are made to LMAN, highlighting the AFP's essential role in preserving the integrity of the learned song (Brainard & Doupe, 2000).

Secondly, **reinforcement learning paradigms** have been employed to study the intricacies of the AFP's role in spectral song adaptation. In these experiments, birds receive feedback, usually a loud white noise burst, if they sing a particular syllable rendition higher (or lower) than they usually do. The birds adapt their pitch and systematically shift in the direction that escapes the aversive stimulus (Tumer & Brainard, 2007). This behavior, often referred to as '*pitch-shifting*', depends on the AFP (Ali et al., 2013b; Andalman & Fee, 2009b).

Thirdly, the **social context** significantly influences song variability. Songs directed at females exhibit less spectral variability compared to those sung in isolation, indicating a more consistent and refined vocal output during courtship or social interactions (Kao et al., 2005a; Kao & Brainard, 2006; Stepanek & Doupe, 2010). This distinction suggests that females are attuned to and show a preference for less variable songs, a behavior confirmed through phonotaxic tasks demonstrating females' discernment of these nuanced differences (Woolley & Doupe, 2008). Importantly, the reduction in song variability due to social context is negated when LMAN is bilaterally lesioned or inactivated (Kao et al., 2005a; Kao & Brainard, 2006; Stepanek & Doupe, 2010).

These findings collectively underscore that the variability induced by the AFP is not just a byproduct but is actively employed in song modification. The distinct differences between directed and undirected singing lead to the hypothesis that directed song represents a 'performance' state, whereas undirected singing embodies a 'vocal exploration' state, essential for ongoing song maintenance and adaptability.

Neural correlates of vocal variability

A fundamental question remains as to the neural correlate of vocal variability and how it is actively controlled, e.g., during female directed singing. LMAN as the output of the AFP has been the target of several studies investigating this question.

LMAN

One line of research has highlighted the role of spiking regularity and bursting. When singing towards a female, which exhibits less spectral variability, LMAN neurons spike regularly. In contrast, during undirected singing, spiking is more irregular from one rendition to the next with an increase in bursting (Hessler & Doupe, 1999c; Kao et al., 2008a). The causal role of bursting in conveying deviant behavior and inducing plasticity has been recently shown by chronic augmentation with bicuculline methiodide (BMI) (Moorman et al., 2021a). BMI increases bursting in LMAN by blocking of $GABA_A$ receptors and potassium channels. The sustained bilateral administration of BMI had widespread effects on the birds' adult temporal and acoustic song structure, including production of longer, distorted syllables, stuttering and variable sequencing. Interestingly, after BMI cessation, the birds recovered within several days to the original song (Moorman et al., 2021a).

On a network level, Darshan and colleagues (2017a) proposed that LMAN neurons are uncorrelated. In theory, such input would average out in a random feedforward network and have no effect on downstream motor units. However, assuming recurrent amplification and topographically organized excitatory projections, their model simulations show how LMAN could contribute to behavioral variability through irregular and uncorrelated neural activity. They support their model with simultaneous recordings of neurons within LMAN that show no correlated activity.

These indirect and qualitative measures have been highly valuable, however, a direct account on the relationship between spiking in LMAN and vocal output is still missing. Considering that previous studies typically involved sequential single-cell recordings, it's plausible that vocal variability is more closely tied to the collective activity of neuron populations. In other words, neuron population activity could reveal features of neural coding not apparent on the level of a single neuron. This hypothesis aligns with findings in rodent and non-human primate studies, where analysis of motor variability from trial to trial (or rendition to rendition) helped to understand the relationship between variable neural spiking and motor control (reviewed in Gallego et al., 2017).

Furthermore, it's important to acknowledge that vocal motor variability is likely to arise from multiple processes. For example, even after bilateral lesion of LMAN or during singing directed at females, a degree of spectral variability persists. To gain a deeper understanding of how the brain governs movement, behavioral modeling, complemented by extensive, long-term recordings, appears promising. This approach could unravel the hidden contribution of LMAN and reveal its effects of neural processes contributing to motor variability in vocalization.

Thus, research on LMAN's role in vocal motor variability could benefit from two promising areas: behavioral modeling of longitudinal recordings and neuron population recordings to better understand LMAN's role in vocal motor variability. Current research indicates a complex relationship between LMAN spiking patterns and vocal output, yet a direct link remains unestablished. Investigating this at a population level, alongside comprehensive behavioral modeling and extensive neural recordings, could unveil the intricate contributions of LMAN to vocal variability, enhancing our understanding of neural mechanisms in vocal motor control.

Area X and DLM

In Area X and DLM, neurons are highly song selective (Goldberg et al., 2010a; Goldberg & Fee, 2010a, 2012; Hessler & Doupe, 1999a; Woolley et al., 2014a). Interestingly, Area X is a heterogeneous cell hub containing both striatal and pallidal cell types (Person et al., 2008) representing it as an amalgamation of cells that are anatomically distributed in the mammalian basal ganglia (Gerfen & Bolam, 2010; Watkins & Jenkinson, 2016). All these neuron types exhibit song-related activity which is modulated by social context (Alvarado et al., 2021a; Goldberg et al., 2010a; Goldberg & Fee, 2010a; Woolley et al., 2014a). Notably, the social modulation of cholinergic neurons has not been tested yet and more recently, Budzillo and collegues (2017) traced a new neuron type with *in-vitro* similarities to subthalamic neurons that drive pallidal output neurons in the mammalian brain (Budzillo et al., 2017). Whether and how this subthalamic-like neuron is song-and socially modulated, and how these neurons interact is an unanswered question.

Correlates on the network level

Due to a lack of data, there remains an unresolved question about the communication and functional connectivity within the AFP and how this results in vocal variability. LMAN projections to Area X are mostly GABAergic and DLM projections to LMAN are mostly glutamatergic (S. W. Bottjer et al., 1989; Farries & Perkel, 2002; Goldberg et al., 2010b; Livingston & Mooney, 1997a; M. Luo & Perkel, 1999a, 1999b). Given the diversity of cell types in Area X, the mechanisms of information integration and transmission remain unclear.

Evidence suggests the existence of both direct and indirect pathways between striatal and pallidal neurons (Gale & Perkel, 2010). The direct pathway posits that striatal neurons transmit GABAergic projections to pallidal neurons, leading to an inhibition of these pallidal neurons. This inhibition, in turn, results in the disinhibition of DLM and subsequently LMAN neurons. Conversely, the indirect pathway suggests that activation of striatal neurons indirectly leads to the disinhibition of pallidal neurons, which would then increase the activity of both DLM and LMAN neurons (Gale & Perkel, 2010). Additionally, considering the recurrent projections from LMAN neurons back to Area X, it is plausible that LMAN also exerts a neural influence on Area X. This complex interplay of pathways and projections raises intriguing questions about the regulatory mechanisms within this circuit and the consequences for communication schemes within the AFP.

Cross-areal recordings could be instrumental in uncovering the mechanisms of communication and signal transmission within the AFP. Such methodologies have already provided insights into distributed coding schemes and the development of inter-area brain coupling during sensory processing and learning in rodents and non-human primates (Girardeau et al., 2017; Kondapavulur et al., 2022; Lemke et al., 2019; Santos et al., 2015; Semedo et al., 2022; Steinmetz et al., 2019).

Applying simultaneous recordings in both LMAN and Area X could reveal crucial insights. Specifically, such an approach could help to determine whether Area X exerts an excitatory or inhibitory influence on LMAN. Moreover, simultaneous recordings could elucidate how the interactions between LMAN and Area X vary between female-directed singing and undirected

singing. Understanding these dynamics is essential for a comprehensive picture of how the AFP controls spectral song variability.

Objectives and outline of this dissertation

This dissertation focuses on advancing songbird research in neuroscience by integrating large-scale neural recording techniques. Our primary aim is to enhance the understanding of vocal control and learning in songbirds through high-yield electrophysiology. Specifically, we investigate the functional connectivity between LMAN and Area X, aiming to uncover the dynamics of recurrent signal transmission. Additionally, we explore the neural-behavioral relationship between LMAN spiking and vocal variability, such as the rendition-to-rendition pitch variability.

In the first chapter, we present the cornerstone of this thesis: the design, testing, and validation of a novel Neuropixels implant for neural recordings. Neuropixels represent an advanced multi-electrode array technology, capable of recording from 384 sites along a 1 cm shank. Our study targets both LMAN and Area X, utilizing the chronic capabilities of the device to capture neural activity across various behavioral states including singing, silent wakefulness, sleep, playback, and pitch-shifting. This has resulted in a dataset of unparalleled depth, allowing us to both replicate and expand upon previous single-cell recording studies, as well as analyze network interactions between LMAN and Area X that have previously not been possible. The chapter is an expanded version of a manuscript prepared for submission to a peer-reviewed journal, adapted to the thesis format and offering additional insights into the developmental process and the extensive dataset collected.

In chapter 2, I contributed to a project spearheaded by Dr. Anja Zai investigating motor control theories revealed by different sources of motor variability. She developed a stochastic dynamical systems model to fit different sources of pitch variability and validated the estimated impact of LMAN in a lesion study (Zai, 2019). My contribution consisted of additional data, conducting similar lesion experiments but employing a pitch-shifting paradigm before the lesion surgery. I also contributed data from chapter 1 and performed the analysis on the correlation between LMAN activity and observed pitch variation versus an estimate of the hidden LMAN's contribution. The study is in preparation to be submitted to a peer-reviewed journal. I wrote the entire chapter in the scope of this thesis. Ideas, methods and figures taken from co-authors are clearly labeled.

Chapter 3 is an excursion into vocal data analysis and describes a novel approach of segmenting and clustering vocalizations of microphone recordings that contain vocalizations of single and multiple zebra finches amid background noises. In contrast to classical machine learning methods, a user-guided strategy was developed that allows both exploration of big data sets as well as immediate segmentation and classification. My contribution was in refining the graphical user interface and the preparation of the manuscript, as well as accompanying the revision and publishing process. The study was published in Frontiers in Bioinformatics (Lorenz et al., 2021).

Chapter 1
Custom-designed Neuropixels device for chronic recordings in freely moving zebra finches enable high-yield recordings in the cortico-basal ganglia pathway

The last decades have seen the emergence of neural recordings techniques in freely moving and singing birds. While performing such experiments is already difficult in itself, another dimension of this difficulty in applying these techniques to freely moving and singing birds is the weight on the head that might be supported by an animal. Zebra finches, one of the most common songbird species used, usually weight ~15g. Chronic implants should thus be less than 2g, since adding weight is a burden and causes a decrease in singing rate (Yamahachi et al., 2020). Previous chronic recording devices developed in songbirds are mostly based on single electrodes or tetrodes, sometimes mounted on a microdrive that can be motorized or not (Fee & Leonardo, 2001; Jovalekic et al., 2017; Schregardus et al., 2006). These devices can be wired or wireless, the latter bringing one more difficulty related to the weight of the battery required for powering the recording system. These various techniques contributed to elucidating the neuronal coding associated with song production (Giret et al., 2014b; Hahnloser et al., 2002b), learning (Crandall et al., 2007; Ölveczky et al., 2011) or perception (Menardy et al., 2014; Prather et al., 2008). Despite these advances, neuronal data acquired are usually restricted to one or a few single neurons recorded simultaneously in a single brain area.

Indeed, multielectrode arrays have been adopted in prior songbird studies focusing on auditory perception, decision making or sleep (e.g., Beckers and Gahr, 2010, Beckers et al., 2014, Lim et al., 2016, Cazala et al., 2019, Theilman et al., 2021). The experimental settings benefit from the fact that perception and sleep can be studied in head-fixed settings and decision-making tasks can be studied with external incentives (e.g., food reward). However, songbirds don't sing naturally in head-fixed settings and external incentives contrast the self-motivated nature of the behavior.

More recently, several studies managed to translate existing technology to songbirds and perform neural recordings during singing through means of imaging (Moll et al., 2023, Cohen et al., 2020, Alvarado et al., 2021) or multichannel extracellular recordings (Das and Goldberg, 2022, Arneodo et al., 2021, Egger et al., 2020, Elmaleh et al. 2021). While imaging is limited to a surface and can only allow for limited depth access, electrophysiological recordings were restricted to one brain area in the mentioned studies (Das and Goldberg, 2022, Arneodo et al., 2021, Elmaleh et al. 2021). Yet, the song system is a network of interconnected brain nuclei that are densely populated with neurons. To better understand the neuronal dynamics at play within the song system, there

is a clear need for designing new recording techniques that must allow the simultaneous recordings of several neurons within and between different brain areas.

Here we designed a new implant assembly to perform large-scale chronic electrophysiological recordings in freely moving adult male zebra finches using Neuropixels probes. The assembly allows simultaneous recordings from neurons in the premotor region LMAN and the song related basal ganglia Area X. We report the recording quality in terms of number of units per electrode channel in each brain area and number of song modulated units that are socially modulated. Further, we provide the fraction of cell types that could be identified in the heterogeneous cell hub of Area X. Comparing these numbers to anatomical estimates is a quality assessment that, to our knowledge, has not been used before. Lastly, we leverage the acquired data to investigate the communication between LMAN and Area X.

Part I – Mechanical design and data acquisition

In this part, we introduce the Neuropixels probe and present the mechanical design and methods of acquiring neurobehavioral data chronically in small birds.

The Neuropixels probe and headstage

Neuropixels represent a cutting-edge advancement in neuroscientific instrumentation, combining miniature multielectrode arrays (MEA) that seamlessly integrate contemporary complementary metal-oxide semiconductor (CMOS) technology with on-chip amplification, filtering, and digitization. Developed by an international collaboration of neuroscientists and the *Interuniversity Microelectronics Center* (imec), the Neuropixels probe is designed to record neurophysiological activity at high yield. The following will describe the probe and acquisition components briefly.

The first released Neuropixels 1.0 probe comprises an array of 960 switchable electrodes, or pixels, each measuring only 12 x 12 um. These electrodes are tightly distributed (16 um column pitch, 20 um row pitch, Figure 2.1) along the shank that has a mere cross-section of 70 x 24 um.

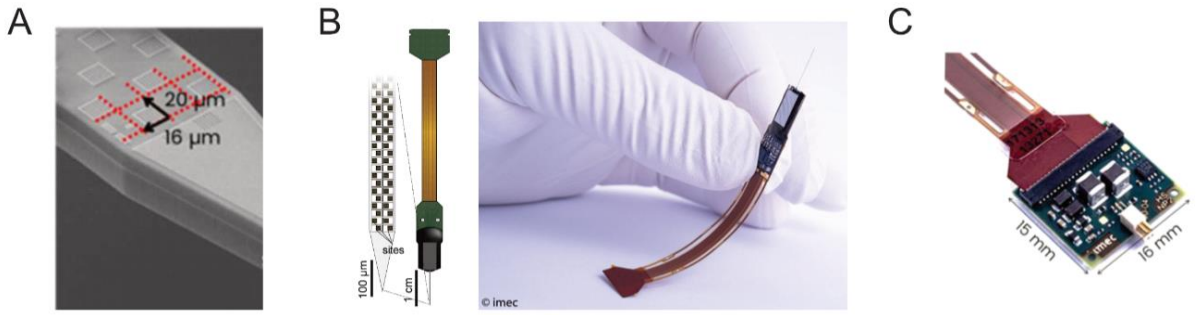


Figure 2.1 The Neuropixels probe 1.0 and headstage. A *The Neuropixels probe has 960 low-impedance TiN contact sites scattered along the shank arranged in a chessboard configuration.* ***B*** *The electrode sites are distributed along the 1 cm long shank. Their signals are processed (amplified, filtered, digitized) on chip, allowing the direct transfer of digital data from the Neuropixels probe. A flex cable connects the probe with the headstage via a ZIF (zero insertion force) connector.* ***C*** *The headstage organizes the data transmission between the Neuropixels and the National Instrument acquisition system and contains additional control and debugging functionally. Pictures taken from www.neuropixels.org.*

Surfaced with porous titanium nitride (TiN), the contact sites feature a low impedance of ~150 kOhms (at 1 kHz in PBS) and promise to capture neural activity from multiple neurons simultaneously at low noise (5.5 - 8 μV_{rms} (Lopez et al., 2017)). Because of the spatial organization along the shank, spike signals are likely captured by multiple neighboring electrodes. This is a deliberate property as differences in spatial co-occurrence can allow differentiating spikes stemming from different sources.

Of the 960 electrode sites, 384 can be chosen to stream input to the integrated circuit on the upper end of the shank. The signal is amplified and split into a high-frequency component accommodating spiking activity, and the low-frequency LFP signal. In a final step, the multiplexed signal is digitized and streamed to the headstage via the 4 mm long flex cable.

The headstage is a 15 x 16 mm printed circuit board (PCB) that organizes the power supply of the MEA implanted in the brain and transmits the neural signals recorded by the MEA to the remote acquisition system. It connects to the MEA via a ZIF connector and to the acquisition cable via a reliable Omnetics connector. Besides interfacing, it provides some additional functionality for testing and debugging.

One Neuropixels 1.0 probe weighs 0.4 g and the headstage weighs 0.9 g. Here, we only tested Neuropixels 1.0. A newer version (Steinmetz et al., 2021) became available recently in 2023.

Existing chronic Neuropixels implant designs

Neuropixels probes were not designed to be used for chronic recordings yet the trade-off between the number of electrode contacts and its weight and size make it an appealing choice for this purpose. To our knowledge, at least four designs have been reported that enable the Neuropixels system to be used in a chronic fashion in freely moving animals, all of them adopting the system for use in rodents (Figure 2.2).

The first design for chronic use published by Juavinett et al. (2018) features an all-in-one implant accommodating both the headstage and the probe altogether weighing only 2.0 g. The design was successfully tested in freely moving mice showing that the animals with implant behave comparable to naïve mice when placed in an experimental arena (Juavinett et al., 2019). A slightly heavier but more protective version was proposed by Van Daal and colleagues in 2021, encompassing the electronic components with added material to safeguard against potential damage in experiments involving stronger rodents such as rats. Both previous designs underwent testing over several weeks. Luo and colleagues (2020) further extended the evaluation period to months-long recordings using rats. Their design, the heaviest at 3.5 g, situates the headstage in elongation to the probe, thus, contrasting the adjacent placement in previous designs. Given that adult laboratory rats (commonly weighing around 300 g) can better accommodate heavier implants, the need for increased protection can be met by stronger and heavier enclosing material.

More recently, a fourth design was introduced (Steinmetz et al., 2021), in line with the encapsulation approach of the initial two designs but allowing for the placement of two Neuropixels 2.0 probes with the updated headstage. This newer design, totaling 2.76 g, maintains an unparalleled balance between weight and functionality. It was successfully tested in mice and rodents (Steinmetz et al., 2021). However, like the previous designs, it shares a fixed alignment of the headstage with the probe, foregoing the potential flexibility of independent placement. This choice likely stems from the complexity and increased surgical time associated with a more intricate scaffold or multiple components. However, based on our previous observations, deviations and misalignments of implants negatively impact a bird's well-being. Extreme angles or positions of an implants can significantly alter the center of mass, introducing torque in the neck that animals must cope with. Thus, to harness the advantages of Neuropixels probes for chronic, freely moving recordings in zebra finches, we developed a novel implant. This design, while comparable in weight to the lightest existing option (~2 g), introduces increased flexibility by allowing independent positioning of the headstage and probe.

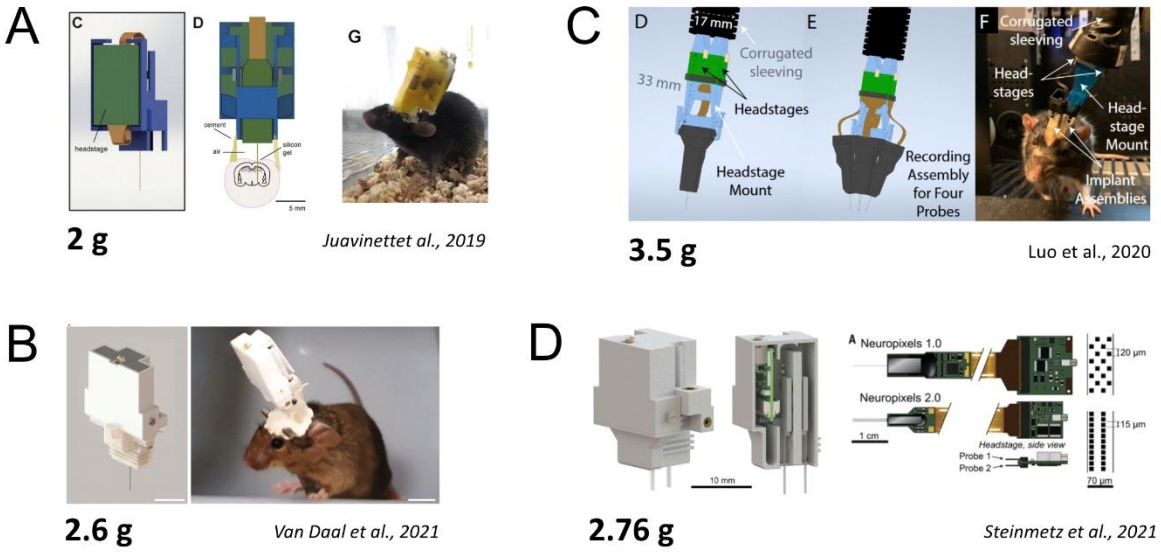


Figure 2.2 Existing chronic Neuropixels implant designs. ***A*** *The first and lightest, all-in-one mechanical implant design by Juavinett et al. (2019). It holds the probe and headstage in a scaffold that only weighs around 2 g including all components.* ***B*** *Van Daal and colleagues (2021) proposed a similar but slightly heavier design that additionally covers the electronical parts.* ***C*** *Luo et al. (2020) chose to place the headstage in elongation to the Neuropixels probe. The single probe design adds up to 3.5 g, the dual probe option doubles to 6.8 g.* ***D*** *The most recent publication of a chronic Neuropixels design allows for recordings with the new Neuropixels 2.0 probe and the corresponding headstage that can connect two probes (Steinmetz et al., 2021). The design is very similar to A and B, combining all electronical components in one 3D printed mechanical part. The simple design with one probe as shown weighs 2.67 g. All designs unites that the headstage is aligned to the probe and therefore the penetration angle which is usually defined by the target location. Photos taken from their original publication.*

Design of the novel implant assembly

Mechanical components

Our assembly consists of a Neuropixels 1.0 probe and two custom-designed parts: the holder and the casing (Figure 2.3A). The holder is manufactured from Torlon(R) with a computerized milling machine. This material comprises two key aspects for the design: it is very hard but at the same time extremely light. Together with the screws, this component only weighs 0.4g. It securely frames and carries the Neuropixels probe. The two protruding M1 screws act as fixation points on the skull. Their length can be chosen according to the implantation requirements. The casing is 3D printed and designed such that it can slide along the holder and protect the delicate shank from the side and front. The backside of the cavity is covered by the tightly bent flex cable (Figure 2.3B) and the headstage (Figure 2.3C). Together with all parts, the final implant adds up to 1.9g (with cement ~2g, Table 1.1.).

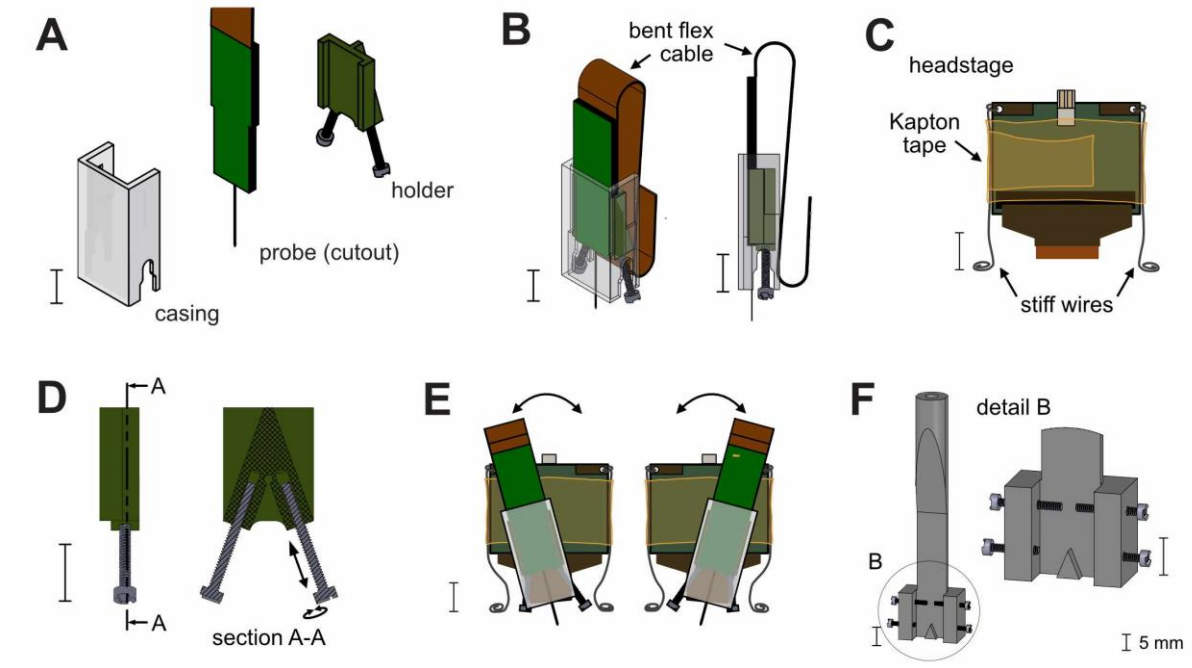


Figure 2.3 Neuropixels 1.0 implant design for chronic use. *(A) The Neuropixels implant assembly (NIA) consists of a lightweight holder that carries the Neuropixels 1.0 probe permanently and a 3-D printed casing that protects the protruding part of the probe from the front and side. (B) The bent flex cable protects the probe from the back. (C) The headstage is fixated to the skull independently using only two stiff wires loosely tied to it. Kapton tape protects the electric parts and enforces the alignment of the wires. (D) The NIA allows for flexibility and reusability. The threads in the back of the holder allow the M1 screws to be replaced e.g. after explantation and their length can be chosen to accommodate different penetration angles (E). The independent placement of the NIA allows to keep the headstage in a position orthogonal to the animal's natural head position. (E). A custom-designed stereotactic holder facilitates the surgical implantation and explantation steps. The NIA is safely secured in the holder with lateral screws. All custom-designed elements except for the casing were manufactured with a computerized milling machine. Threads were created manually.*

Functionality and flexibility

Besides the weight, orientation and placement are two additional factors that determine the success of a chronic recording experiment. Especially in freely moving animals, the external components of an implant should always be placed as close as possible to the head (reducing torque) and at an angle that is perpendicular and in elongation to the natural head position. Oftentimes, however, a more extreme orientation or placement of the external components is required to penetrate the multiple brain areas of interest. Our NIA design allows for flexibility in this scenario as it can accommodate a wider range of penetration angles relative to the animal's head posture. First, the screws of the holder can be adjusted in length to accommodate the placement on the skull (Figure 2D). Second, the headstage and NIA are placed independently. This independence ensures that the headstage can always be placed at a 90 degrees angle, in elongation to the animal's natural position (Figure 2.3E). Together, the implant can be placed more optimally in the weight bearing axis of the body (Figure 2.4).

Component	***Weight [g]***
Neuropixels probe 1.0	0.40
Holder	0.27
Screws (x2)	0.08
Casing	0.26
Headstage 1.0	0.90
2 cm long stiff wire	0.05
∑	1.96

Table 1.1 Weight of the individual components and net-weight of the full NIA with headstage. *Components with less than 0.01 g are neglected (silver wire for grounding the animal, adhesives used during assembling). Together with adhesives for fixating the implant on the skull, the weight adds up to ~ 2 g.*

Reusability

The NIA is fully reusable also for experiments with different requirements. The screws and wires can be removed and replaced. The wires fixating the headstage can simply be cut and the casing can slide off the skull at the end of an experiment. Explanting the Neuropixels and its holder (Figure 2.2B) requires more manual work specially to free the screws from cement. We designed a custom-designed surgical holder (Figure 1F), to secure the implant during the implantation surgery and, more importantly, during the explantation step when the probe must be pulled up once it is no longer fixated on the skull.

We regularly re-used probes. After successful extraction, we could consistently see that the shank was almost clean with only minimal residual tissue covering it, owing to the smooth extraction procedure. We cleaned the probe as suggested by the manufacturer with enzyme detergent and found that in each cycle, the number of usable channels decreased by a maximum of one channel.

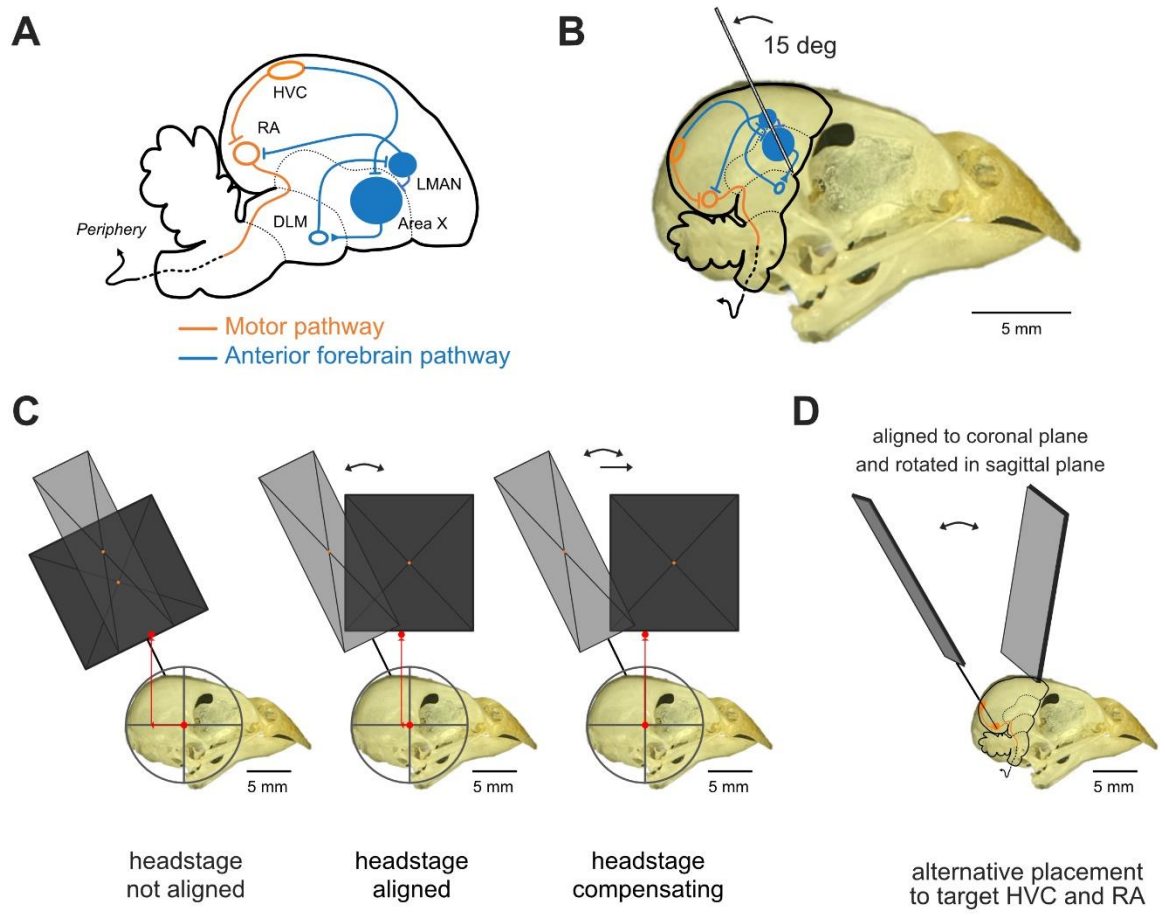


Figure 2.4 NIA configurations. A *Sagittal schematic of a zebra finch brain with annotated brain regions of interest.* ***B*** *Approximate placement of the brain in the bird skull and ideal penetration angle. Shown is the schematic in a photo of a zebra finch skull (© Lorenz). The skull is rotated to depict the birds natural head orientation. Targeting LMAN and Area X in one line is most successful at 15 degrees off.* ***C*** *Estimation of the center of mass (red points) without and with alignment of the headstage. From left to right: Without any alignment, the center of mass moves 17.42 mm up and 3.89 mm posterior. Aligning the headstage with the weight bearing axes reduces the posterior displacement by 73 % to 1.04 mm. A more extreme placement in the anterior direction can compensate for the additional weight distribution. The calculations were taken with following parameters: Natural center of mass (red dot): 2 g head weight around center of most weight bearing organs (eyes, brain, tongue). Neuropixels: 0.04 g with 0.5 mm shank protruding and 22.9 x 7.2 mm size. Headstage: 0.09 g with 15 x 16 mm size, lower end aligned to Neuropixels PCB. Note that the estimations neglect the additional effect of the tether that impacts the displacement additionally if not aligned.* ***D*** *Schematic of an alternative implant configuration. When targeting HVC and RA, which require a more lateral penetration, the configuration can be adapted. The NIA and the headstage can be placed parallel to the coronal plane and individually rotated to ensure for the headstage, connector and tether to align with the weight bearing axis.*

The fact that the probe is permanently attached to the holder did not prevent us from re-using it in different configurations. We usually tested the implant design by targeting the anterior forebrain in zebra finches. In this case, the optimal configuration is to place all components close to the

midline in parallel to the sagittal plane (Figure 2.4A and B). We could successfully replicate this configuration in several birds even with re-used probes. More importantly though, we also tested the design implanting a reused probe in HVC and RA in a canary (*Serinus canaria forma domestica*). This required a 90-degree rotation of the entire implant and a rotation of the probe relative to the headstage (Figure 2.4D). The surgery was successful, and the animal appeared recovering in the first hours after the surgery. However, the bird died in the night after the implantation likely due to respiratory problems unrelated to the implant or surgery. The alternative placement including the rotation of the implant (facing dorsal instead of lateral) confirms the flexibility of the design.

Protocols

Detailed protocols on the assembling of the NIA (P1), the steps of surgical implantation (P2) and explantation (P3) are attached in the appendix (Appendix I).

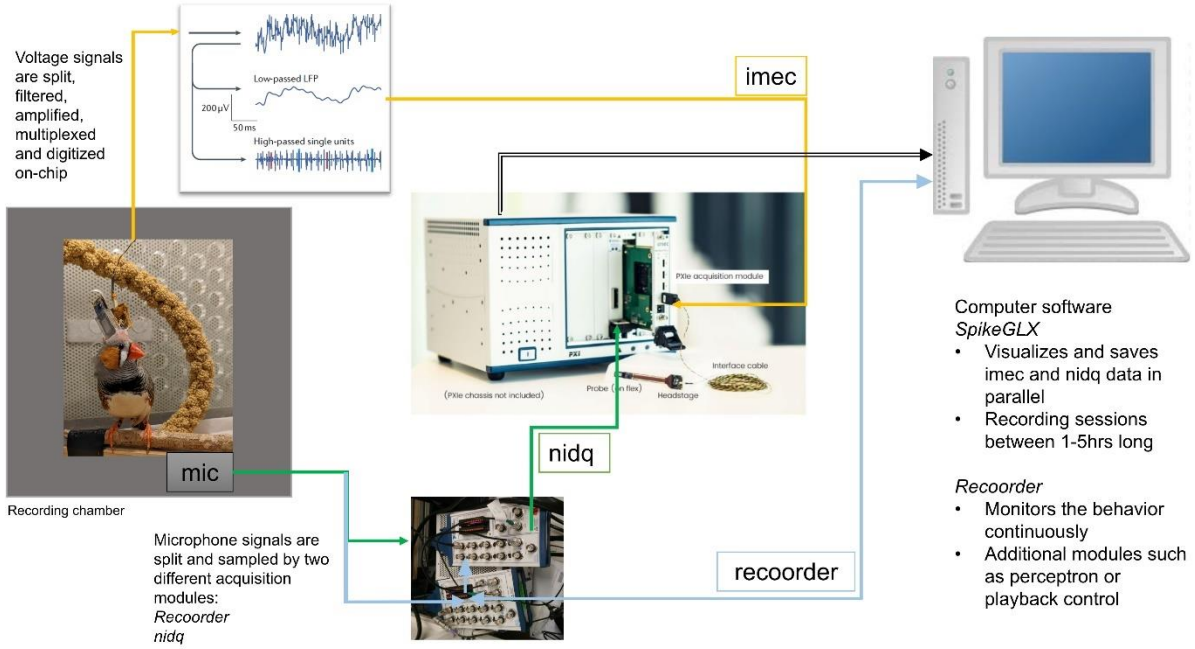


Figure 2.5 Acquisition systems and interfaces. *Neuropixels data (imec, highlighted in yellow) undergoes division into low-pass filtered signals (LF) and high-pass filtered signals (AP), both of which are concurrently recorded with auxiliary input (e.g., audio signals, nidq, depicted in green) through a shared PXI module. Simultaneously, a custom-written LabVIEW program (RecOOrder) samples the microphone signal (depicted in blue) and captures vocal behavior outside the neural recording sessions. This system offers additional functionality, including online syllable detection (perceptron) and playback. On the software side, SpikeGLX is employed to manage imec data acquisition and to visualize incoming data streams online during neural recording sessions. The RecOOrder interface serves to visualize the audio signal and continuously monitor behavior.*

Data acquisition

Hard- and software configuration

The Neuropixels acquisition system comprises two modules, one dedicated to organizing neural data transfer (imec) and one managing auxiliary input (nidq). We configured settings and established connections to capture microphone audio signals as auxiliary analog input, and in subsequent stages, we expanded the input to incorporate a digital switch indicating experimental conditions (female absent or present). Due to the rapid accumulation of acquired neural data with the system (~60 GB/h), recordings were performed in sessions that lasted between one and six hours. Throughout these sessions, data was acquired continuously to monitor potential drift in the neural spiking data, with intermittent breaks for interventions such as placing the female in the box or unwinding the tether. The modules were visualized and recorded with SpikeGLX (https://billkarsh.github.io/SpikeGLX).

In addition to the Neuropixels acquisition system, a custom-written LabVIEW software, the *RecOOrder*, (Herbst et al., 2023) facilitated continuous monitoring of audio events, including vocalizations and songs even outside the neural recording sessions, and offered real-time control of syllable detection and feedback. Selected signals from this process were looped back into the nidq system, enhancing data post-processing and analysis capabilities within one domain.

Figure 2.5 illustrates the interconnected hardware and signal routing. A stepwise protocol is given in the appendix (Appendix II).

Recording routine and file structure

Birds were routinely connected to the interface cable in the morning. Due to the lack of a commutator[2] it was not possible to keep them connected continuously. However, thanks to the 3 m long and flexible interface cable, it was possible to unwind the cable, when necessary, on the bird's end and move the entanglement to the acquisition end where it would not cause any experimental disturbance. We conducted between one and three sessions per day.

To ensure smooth data management, raw files were saved in folders that adhered a strict convention of folder naming:

```
/animal_ID/Ephys/raw/session_ID/                    % raw neural and audio data
/animal_ID/Ephys/spksort/session_ID/                % pre-processed neural data
/animal_ID/ArchiveCollection/…/session_ID/          % pre-processed audio data
/animal_ID/Analysis/…/session_ID/                   % analyzed data
/animal_ID/Figures/…/session_ID/                    % figures and image data
```

[2] We tested a mercury commutator (Dragonfly®) and a custom-made slip ring. Both were able to provide the power supply to the Neuropixels chip but unable to support the data transfer that runs on an ultra-high data rate. A recent publication mentioned success using the Adafruit 1196 (De La Crompe et al., 2023, personal correspondence).

The animal ID was the unique identifier taken from the animal database. The session ID was a composition of the date of recording and a running number of recording sessions within a day, i.e. YYYYMMDD_nn.

In case of short interruptions, e.g. when the recording chamber was opened to disentangle the animal, SpikeGLX recordings were paused and subsequently resumed after closing the chamber. This resulted in several “epochs” per recording session that reflect as such in the file structure:

`/animal_ID/Ephys/raw/session_ID/epoch01/` % first epoch

`…/epoch02/` % second epoch

In the beginning, we used epochs to indicate different experimental conditions (e.g. directed vs undirected singing) and later incorporated a digital switch indicating the social context.

Post-acquisition pre-processing

Temporal alignment

The nidq and imec system each have their own clock and run asynchronously. Additionally, sample clock rates are temperature sensitive. This introduces drift and ultimately temporal misalignment between data streams. SpikeGLX offers an option to correct this misalignment by calibrating the drift based on a digital synchronization pulse. We found that even after calibration, especially in long recordings, the drift was not accounted for completely. Thus, in a preprocessing

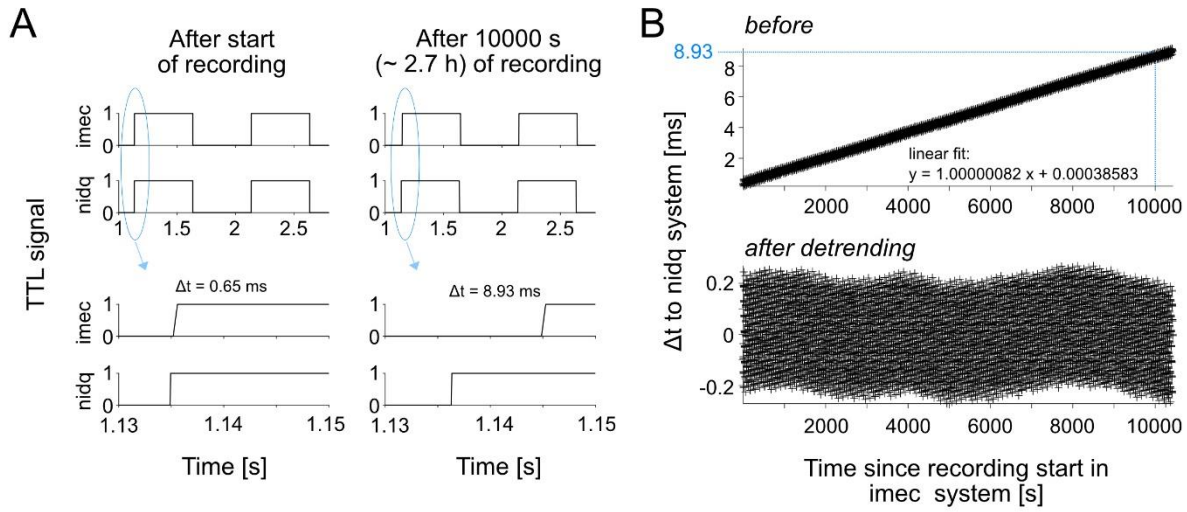


Figure 2.6 Temporal alignment of nidq and imec data streams. ***A*** *Example of the aligned synchronization pulse in the nidq and imec system. Shown is the TTL signal at the beginning (left) and end (right) of a recording session. The bottom panel provides a close-up view of the moment when the TTL signal transitions from 0 to 1, with annotation indicating the time difference between the detection of the switch in both systems. After 2h 46 min, the difference accumulated to almost 9 ms.* ***B*** *Analysis of time discrepancy in detecting TTL switches between the imec and nidq systems (top). After correcting for a linear drift with offset, the signal is detrended, revealing transient fluctuations likely attributable to temperature variations that affect the clock. The oscillating values within the range of -0.2 and 0.2 ms result from the different sampling rates (imec lf 2.5 kHz, nidq fs = 20 kHz).*

step, we estimated the additional drift using the changepoint times in the TTL synchronization pulse in the imec and nidq acquisition. Calculating the linear relationship between the two times series, it is possible to derive a post-hoc correction factor for the timing of spikes relative to the behavioral data streams (Figure 2.6). The parameters are saved together with the figures in the specified figure folder (see File structure).

File concatenation

The spike sorting algorithm, Kilosort (Pachitariu et al., 2023) is designed to process a single file as input. In sessions that involve multiple epochs, it becomes necessary to concatenate the high-pass filtered *.ap files for spike sorting. To streamline this procedure, I developed a user-friendly Graphical User Interface (GUI), details of which are provided in the appendix (Appendix III). Given the nested structure, the GUI facilitates browsing through different sessions in a convenient and fast manner or to run and assess the temporal alignment.

Part II – Behavioral testing and data evaluation

In this second part, we assess the behavior of animals carrying the novel NIA and evaluate the quality of recorded neural signals. Methods are attached in Appendix IV.

Zebra finches return to a stable singing rate during neural recordings

We tested the NIA in eight adult's male zebra finches, placing the Neuropixels probe 15 degrees relative to the natural head orientation in order to penetrate two nuclei of the AFP: the cortex-like premotor nuclei LMAN and the song-related basal ganglia region Area X (Figure 2.7A, see Methods, Appendix IV). The shank was inserted up to 5 mm leaving the upper 5 mm standing out.

First, we were interested in how the animals can cope with the added weight of the implant. Zebra finch birds have an intrinsic motivation to sing but experimental procedures and interventions such as social isolation, the additional weight of an implant, and tethering impair the behaviour and usually cause a decrease in singing (Yamahachi et al., 2020). To counteract this implication, a weight-reliever in the form of a counterweight was attached to the headstage. The weight of the counterweight was adjusted for each individual bird daily throughout the experiment based on visual assessment. We observed that birds adapted to the weight of the implant within the days after the implantation and could cope without the weight reliever.

To habituate the birds to the acquisition cable and to maintain a predictable routine, birds were tethered daily after the surgery. It happened routinely that the tether cable entangled due to the lack of a commutator still within the recording chamber which required a short intervention. To reduce the stress of catching the animal, we decided to take advantage of the 3 m length of the cable and moved the twist to a remote part. With these measures, birds recovered to a stable singing rate comparable to birds with other types of implants (e.g. manual/motorized microdrive of 1.5 g, Yamahachi et al., (2020)) within a few days after the surgery (Figure 2.7B) which shows that zebra finches can cope with the challenge of a chronic Neuropixels implant with our NIA.

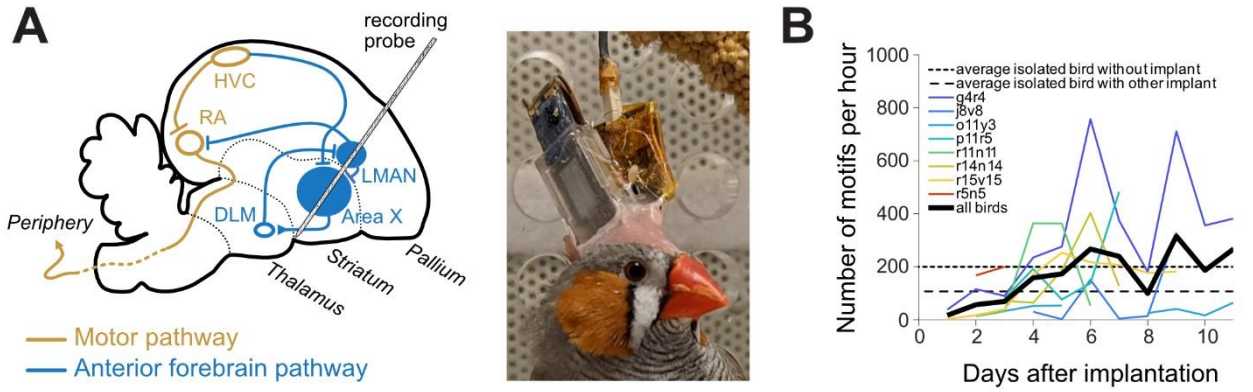


Figure 2.7 Chronic implant of a Neuropixels probe in freely moving birds. *(A) Sagittal schematical view of the motor and anterior forebrain pathways. In order to properly hit LMAN and Area X, the probe must be implanted with a specific angle. (B) Picture of an adult male zebra finches with the headstage holding the Neuropixels probe. (C) Chronically implanted birds resume singing in a few days* post-surgery. *Singing rates under different conditions taken from* (Yamahachi et al., 2020).

Bird	Surgery date	Age at surgery	Npxls number	# useable channels	Implant coord. (AP, ML, DV) [mm] and head angle	Experimental conditions	# good sessions	Npxls recovered
g4r4	11.07.2019	112	18005112662	384	5.1, 1.75, 4.8 51 deg	UD, FD, PB, PCAF, SLEEP	21	Yes
o11y3	21.10.2019	133	18005112742	384	4.85, 1.7, 4.3 51 deg	UD, FD, PB, PCAF, SLEEP	5	No**
j8v8*	19.02.2020	75	19051017451	384	5.0, 1.75, 4.5 50 deg	UD, FD, PB, PCAF, SLEEP	21	Yes
r5n5	19.10.2020	81	19051008742 (2nd use)	383	5.5, 1.75, 4.5 50 deg	UD, SLEEP	2	Yes
r15v15	30.10.2020	112	19051017451 (2nd use)	383	5.2, 1.7, 4.3 50 deg	UD, FD, PB, PCAF, SLEEP	28	Yes
r11n11	14.11.2020	73	19051008742 (3rd use)	383	5.3, 1.7, 4.5 52 deg	UD, SLEEP	2	Yes
r14n14	04.12.2020	93	18005102621	384	5.3, 1.7, 4.75 50 deg	UD, FD, PB, PCAF, SLEEP	12	Yes
p11r5	06.12.2021	134	19051017152	384	5.0, 1.6, 4.5 51 deg	UD, FD	NA*	Yes

Table 1.2 Data summary. *List of birds detailing surgical parameters, experimental conditions, and outcome evaluations. Birds r5n5, r15v15, and r11n11 underwent implantation with previously recovered Neuropixels probes which exhibited one defective channel. The implant coordinates are given relative to the lower boundary of lambda, representing the confluence of sinuses. Experimental conditions included singing in isolation (UD), presence of a female (FD), playback of the bird's own song (PB), pitch-contingent auditory feedback using white noise (PCAF), and recordings during dark phases (SLEEP). Sessions were considered 'good' if over 100 motif renditions were detected, excluding conditions of SLEEP and PB. * This bird had a two-month recording break due to a COVID-19 related lockdown, after which spiking signals were no longer detectable. Thus, despite the bird maintaining good health with the implant intact, data from the subsequent acquisition period was excluded. ** Neural data could not be processed because the files were corrupted likely arising during data transfer. *** The shank of the Neuropixels broke while removing cement during the explantation.*

Neural recording quality

Because of the rather quick recovery of the birds after the implantation, we were able to perform several consecutive recording sessions in awake, freely moving and singing birds. Spike sorting was carried out offline with the Kilosort (v2.0) algorithm (Pachitariu et al., 2023), followed by a manual curation using phy (https://phy.readthedocs.io/). Several units located along the entire shank of the electrode were simultaneously recorded (Figure 2.8A). We reconstructed the channels covering the two areas of interest, LMAN and Area X, from the trace of the probe and the depth of its tip that were both easily visible on a few consecutive histological slices thanks to the applied red dye on the probe before the implantation (Figure 2.8B). Units were thus detected in the two areas of interest and also above (pallium), in between and below (striatum) these areas.

A fundamental aspect of the usability of our NIA is the signal quality not only during a single recording session, or day, but also over several consecutive days. We evaluated signal quality as the number of neural units after spike sorting and manual curation in recording sessions with at

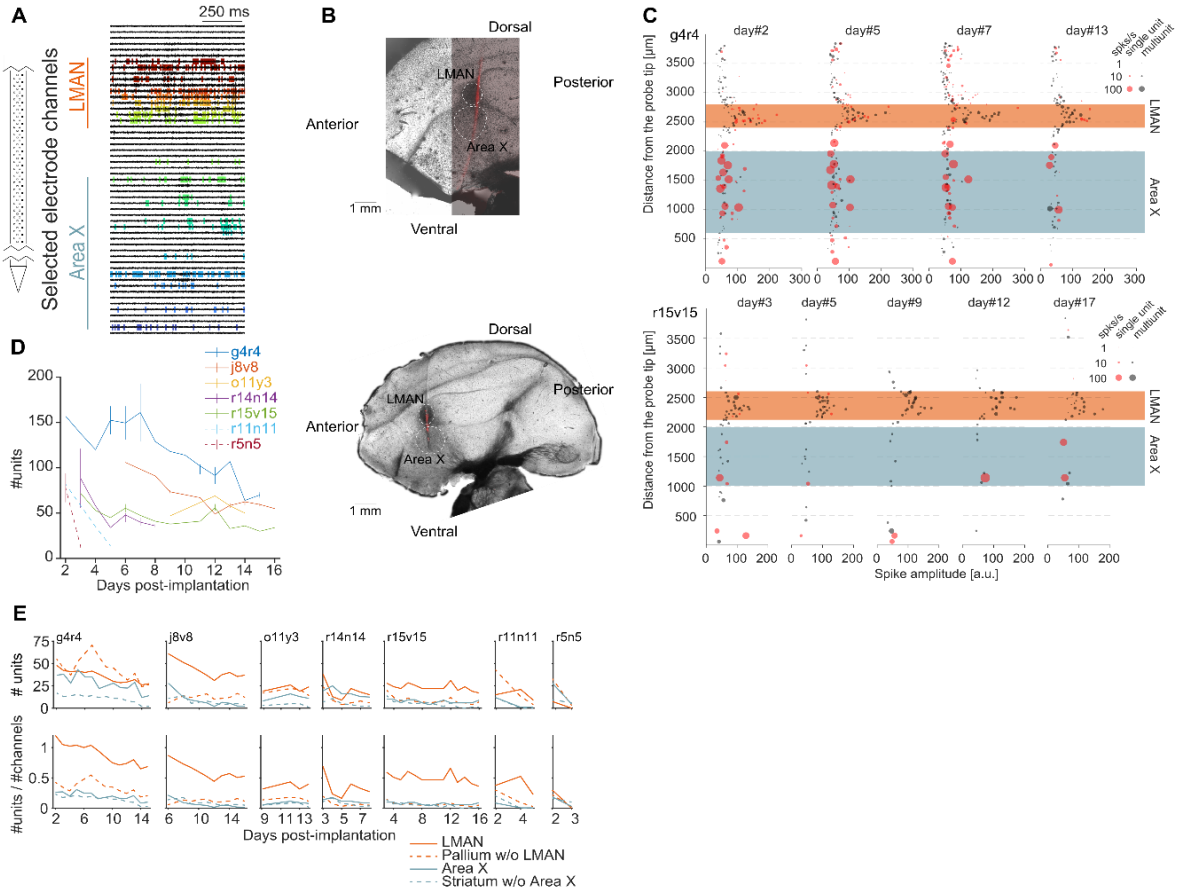


Figure 2.8: Number of detected units remains stable for a few days post-implantation. *(A) Raw traces of a subset of simultaneously recorded channels along the shank of the Neuropixels probe (schematic on the left) in an awake non-singing bird. (B) Sagittal slices of two example birds showing in red the trace of the implanted probe previously surrounded with a red dye. Limits of LMAN and Area X are highlighted by the dashed white lines. (C) Distribution of single and MUA along the shank of the probe according to the spike amplitude in two example birds (same birds than in B) over several days. Marker size reflects spike rate. Channels covering LMAN and Area X are highlighted by the red and blue boxes. (D) Number of units detected following days post-implantation in all the birds. The two birds with dashed lines (r11n11 and r5n5) had loose implant that fell off a few days post-implantation. (E top) Absolute number of units detected on several days post-implantation in LMAN, the pallium without LMAN, Area X and the striatum without Area X in all the birds. (E bottom) Relative number of units detected according to the number of channels covering the areas of interest.*

least 100 renditions of a bird's motif. Both single and multi-units were detected in our recordings up to two weeks after the implantation (Figure 2.8C). Different units with low or high firing rates, especially in the Area X (Figure 2.8C), were still been detected after days of recordings, yet with a large decrement compared to the first recording sessions post-implantation. We also observed that spike clustering was more difficult in LMAN and subsequently resulted in larger fractions of multiunit clusters as compared to Area X, for instance (Figure 2.8C).

Overall, the number of detected units progressively decreased over days, with ~100 units on the first few days post-implantation to ~50 units after two weeks of recordings. There were also

interindividual differences in the number of detected units and how this number change over days (Figure 2.8D,E). For two birds (r11n11 and r5n5), the number of units almost dropped to zero in less than a week after the implantation (Figure 2.8D, dashed lines). We later found that these birds were actually having a loose NIA due to a lack of fixation at the anterior end of the skull where the bone structure is generally thinner. To circumvent this problem, we recommend placing insect pins in between the bone layers. In other birds (e.g. g4r4, j8v8), the signal remains of very good quality for many days, allowing the detection of tens of units two weeks after the implantation. We also wondered whether the overall decrement of detected units was similarly affecting the probe, independently of the location of the channels. We thus determined the number of units in various areas of interest: LMAN, the pallium without LMAN, Area X and the striatum without Area X. First, we noticed that the number of units detected in LMAN is usually higher for all the birds, which reflect the high neuronal density within that nucleus paired with the larger cell bodies (Figure 2.8E top). In comparison, the number of units detected in the Area X is quite low, in spite of its relatively large size. This is particularly clear after normalizing the number of units by the number of channels covering the area of interest (Figure 2.8E bottom). While units were detected on more than half of the channels covering LMAN for most of the birds, and for several days of recordings, units were usually detected on less than 10% of the channels covering Area X, consistent with a systematic dependency between neural yield and anatomical position reported by (Luo et al., 2020).

In all but one bird, thanks to our customized NIA holder (Figure 2.3F), we managed to recover the probe at the end of the experiment (Table 1.2). After proper cleaning and checking, the probe was reusable for another implant. In three cases, we used a recovered probe. We did not specifically test for the loss of SNR but could obtain signal quality as good as for the first use.

Singing-related neuronal activity

Recordings sessions were carried out on a daily basis and lasted at least 1 and up to 6 hours. During a session, the bird was either left alone or presented with a female partner in order to elicit female directed singing. We compared the spike waveform of each individual unit at the beginning and at the end of a recording session to ensure the stability of the signal processing (Figure 2.9 Suppl Fig 1). In spite of the fixed position of the implanted probe, single and multi-units detected on two consecutive days were never considered as being the same.

We conducted a more in-depth analysis of singing-related neural activity in one recording session of each of the five birds with good and stable neural readout (solid lines in Figure 2.8D). Figure 2.9A shows an excerpt of the raw neural recording traces covering LMAN and Area X from the example bird shown on the bottom of Figure 2.8B. Spike times and assigned clusters are annotated and condensed in the multisite raster in Figure 2.9B, which reflects the strong song-modulation of recorded units, especially pronounced around the song offset.

In the pooled subset, we computed song-related activity related to the stereotyped and compared it to spontaneous activity during silent periods motif (see Methods, Appendix IV). Most of the units recorded in LMAN (118/139) and Area X (69/75) exhibited song modulated activity (Figure 2.9A), with a higher firing rate when the bird is singing (LMAN: 20.73 ± 14.10 Hz, Area X: 62.19 ± 60.32

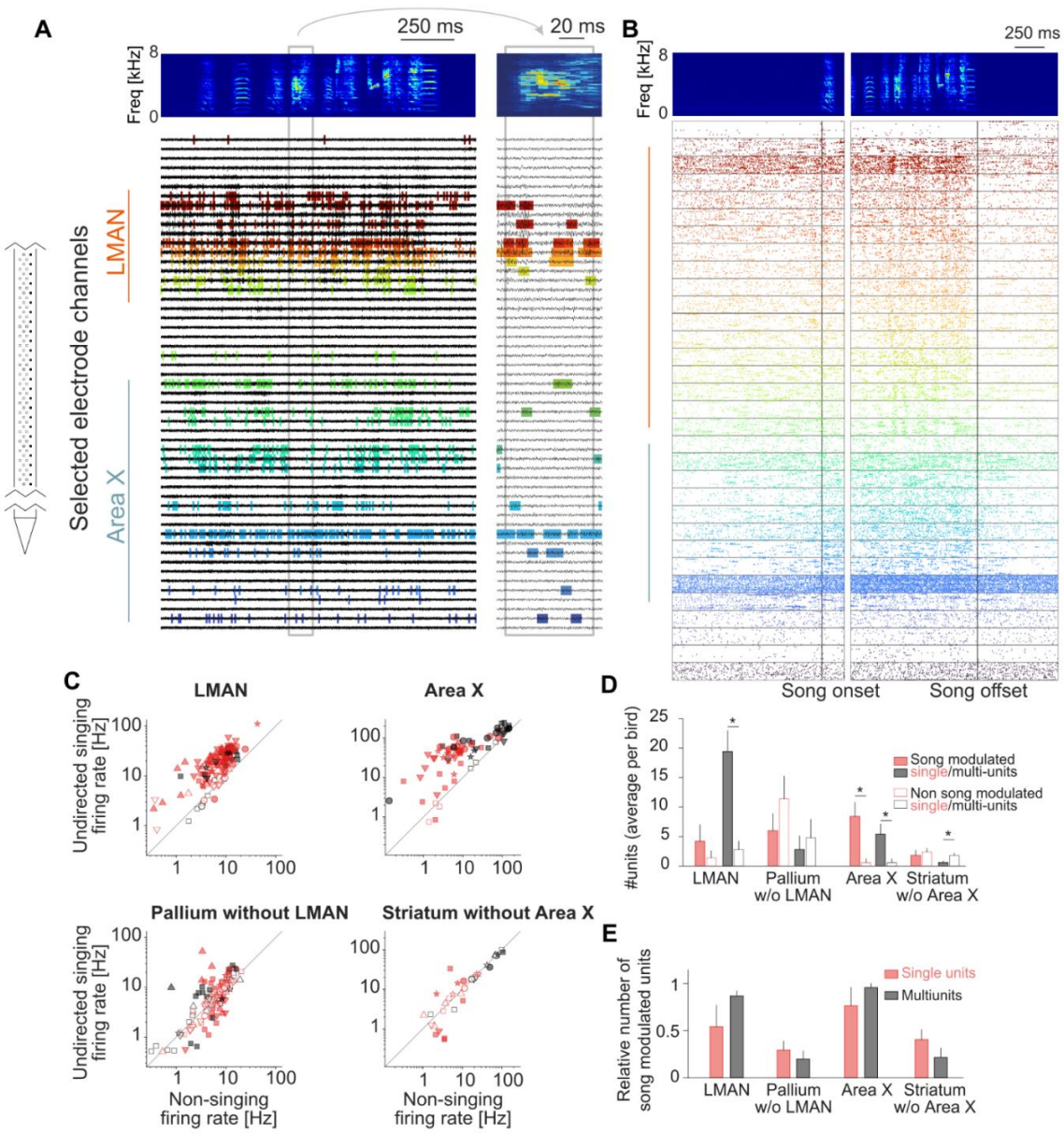


Figure 2.9 Chronic Neuropixels probe implants allow the recording of several song-modulated units. ***A*** *Raw traces of a subset of simultaneously recorded channels along the shank of the Neuropixels probe (schematic on the left) in an awake singing bird (log-power spectrogram on top). Spike-sorted units are color coded and boundaries of LMAN (orange) are Area X (blue) are shown on the left.* ***B*** *Spike raster plot of several LMAN (orange line) and Area X (blue line) units simultaneously recorded during ten song motor renditions.* ***C-D*** *Comparison of the average firing rate of each individual unit recorded in all the birds (one marker type per bird, n = 5 birds; marker O corresponds to the example bird in A and B) in LMAN (n= 139 units), Area X (n= 75 units), the pallium without LMAN (n= 125 units) and the striatum without Area X (n= 33 units) when the bird is singing alone vs not singing (see Methods). Single and multi-units are in red and grey, respectively, with filled markers for significantly song-modulated units.* ***D*** *Absolute number of units detected in LMAN, Area X, the pallium without LMAN and the striatum without Area X per bird.* ***E*** *Same than in D but with the relative number of units to the total number of channels covering the area of interest. *, p<0.05.*

Hz, mean ± SD during UD) compared to silence (LMAN: 8.14 ± 5.21 Hz, Area X: 28.27 ± 43.10 Hz, mean ± SD during silent periods). Much less fluctuations of the firing rate according to song production were measured in the units located in the pallium but outside of LMAN and in the striatum without Area X (Figure 2.9C,D).

Types of Area X neurons

The avian BG Area X is an amalgamation of cell types that exhibit specific electrophysiological properties in relation to the singing behavior. In our subdataset of 75 Area X neurons collected in 5 adult birds, most of the units fired more during singing than during silent period (69/75=92%; Figure 2.10A). We identified high-frequency (HF) and low-frequency (LF) neurons, with spontaneous non-singing firing rates >60 Hz (n=16, mean=107.5 ± 28.6) and <50 Hz (n=75, mean=8.1 ± 10.4), respectively. Following previous reports (Farries & Perkel, 2002; Goldberg & Fee, 2010b; Leblois et al., 2009; Woolley et al., 2014b), HF neurons are putative pallidal neurons while LF neurons are putative striatal neurons and can be further classified in various neuronal categories.

In the mammalian BG, previous reports identified two types of pallidal cells that are both located in the Globus Pallidus (GP). The first population of neurons makes external projections from the BG to the thalamus (GPe) while the other population of neurons locally projects to the BG (GPi). In the avian BG Area X, GPe and GPi-like neurons that share anatomical and firing properties with the mammalian GPe and GPi were also described. Both GPe and GPi-like neurons exhibit high-firing (HF) rates. GPe-like neurons exhibit bursts with high peak firing rates and pauses in firing, especially during singing, while GPi-like neurons rarely burst. GPi-like neurons also increase their firing during singing and display regular fast spiking. To determine whether the set of putative pallidal units include both types of neurons, we computed the Fano Factor against the spontaneous firing rate (Figure 2.10B). We set a threshold of the Fano factor at 0.8 to distinguish units on their burstiness. Units with a Fano factor higher than the threshold fell into the GPe-like category.

Four types of striatal units were previously described: medium spiny neurons (MSN), fast-spiking interneurons (FS), low-threshold spiking neurons (LTS) and cholinergic (or tonically activated) units (ACH). MSNs make up the largest number of striatal cells in Area X (Gale & Perkel, 2010). They exhibit sparse firing often locked to behavioral events, such as a birdsong syllable, and very low spontaneous firing rate (Alvarado et al., 2021b; Goldberg & Fee, 2010b). To characterize whether MSN were present in our dataset, we first plotted the firing rate during undirected singing against a sparseness index (see Methods, Appendix IV). This allowed us to identify a subset of 8 units with sparse firing (sparseness index >0.15) and low firing rate (<15Hz). We then plotted the spontaneous firing rate of this subset of units as a function of the median interspike interval to extract bursty neurons and so the putative MSNs neurons (Figure 2.10C). We were thus able to isolate 5 putative MSNs. A representative example is shown on Figure 2.10F (other examples are shown on Figure 2.10 Suppl 1).

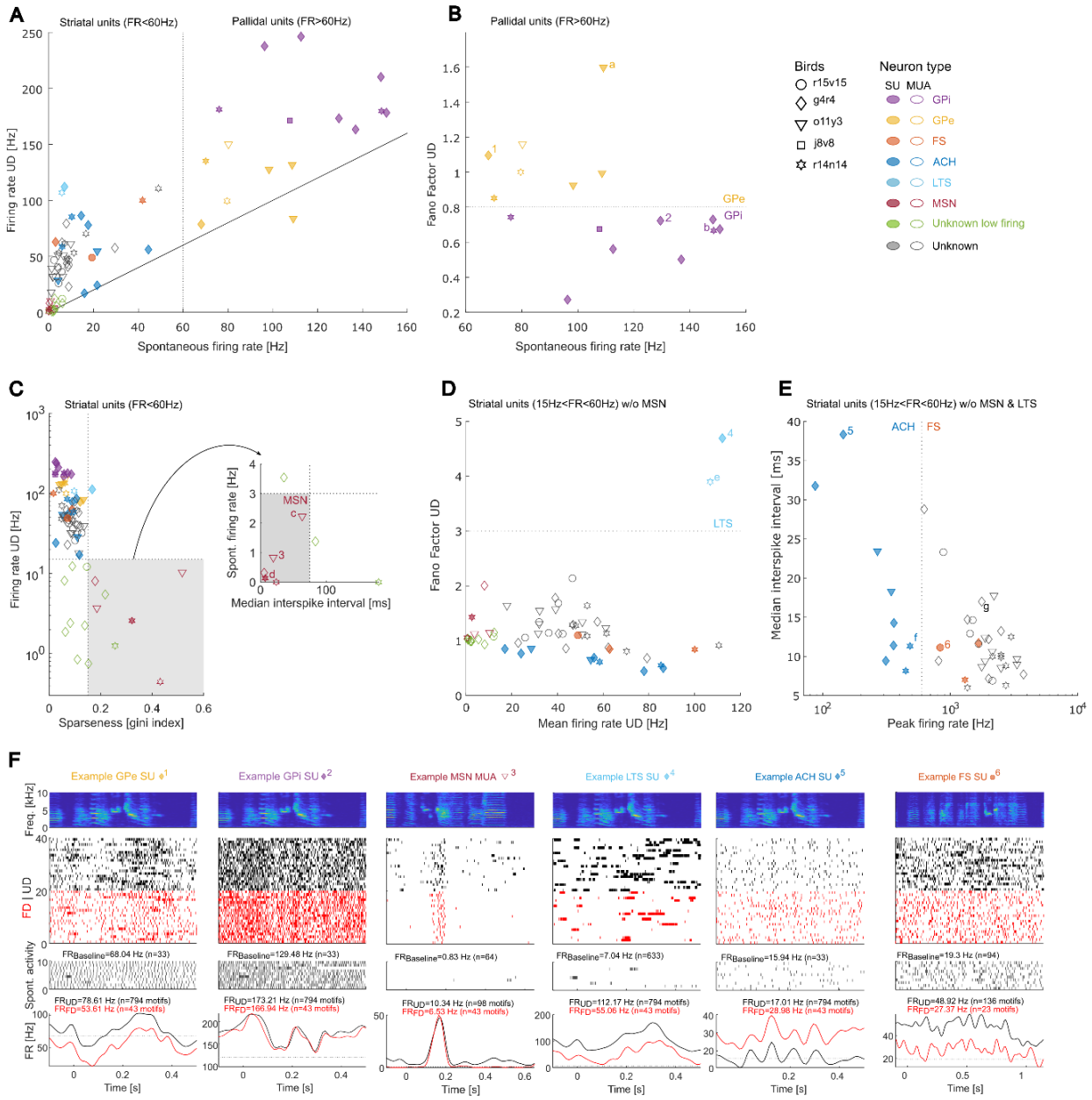


Figure 2.10 Identification of Area X neuron types. ***A*** *Striatal and pallidal units distinguished by spontaneous firing rate, threshold at 60Hz, with higher rates in undirected songs (UD); black line shows average rate.* ***B*** *GPe and GPi cells classified by Fano factor on UD song motifs; threshold at Fano Factor = 0.8.* ***C*** *Striatal units' firing rate during UD singing versus sparseness index (SI); subset under 15 Hz rate and over 0.15 SI plotted against median interspike interval, identifying putative MSN.* ***D*** *Striatal units (excluding MSN) Fano Factor versus UD singing rate to identify putative LTS neurons based on their high variability, Fano Factors > 3.* ***E*** *Median interspike interval versus peak firing rate for striatal units (excluding MSN & LTS) to differentiate putative ACH from FS units at 500 Hz peak rate.* ***F*** *Examples of putative GPe, GPi, MSN, LTS, Ach, and FS units, including spectrograms, spike raster plots for UD and FD songs, spontaneous activity, and average firing rates. Additional markers in subpanels relate to Fig. 5 Suppl. Fig. 1. UD: Undirected singing, FD: Female-directed singing, GPe: External Pallidal Cells, GPi: Internal Pallidal Cells, MSN: Medium spiny neurons, LTS: Low threshold spiking neurons, ACH: Cholinergic neurons, FS: Fast spiking neurons.*

Low threshold spiking (LTS) neurons display long, high-frequency bursts during singing, and much less during silent periods. LTS neurons discharge mostly irregularly throughout the song. Here, we computed the Fano factor to account for this irregularity from rendition to rendition against the firing rate during undirected singing. We were able to identify two putative LTS neurons in our dataset (Fano Factor > 3, Figure 2.10D). These two putative LTS neurons were highly bursty during singing. Fast spiking (FS) interneurons were previously reported as exhibiting steady low-firing rate during silent period. During song production, they display high-firing rate, brief high-frequency bursts, and, contrary to MSN, are active during the entire song (Goldberg & Fee, 2010b). These properties are similar to those of ACH units that are tonically activated during non-singing epochs and that display a higher firing rate during singing. A key difference between FS and Ach is supposed to be the peak of firing, with ACH units expressing lower peaks than FS. To identify these two Area X neuron types in our dataset, we plotted the median interspike interval of the putative striatal units (after removing putative MSN and LTS units) against the peak firing rate (Figure 2.10E). It allowed us to identify two well separated clusters according to the peak of firing rate, with a threshold set at 500 Hz (Goldberg & Fee, 2010b).

Social modulation of singing related neuronal activity

Since we managed to perform recordings when the birds were either singing alone in their cage (undirected singing, UD) or for a female (female directed singing, FD), our data provide insights on whether and how social context impacts both LMAN and Area X singing related neuronal properties, and whether it differs according to the putative cell types (Figure 2.10 Suppl Fig 2A). Several earlier studies highlighted the social modulation of LMAN neuronal activity (Hessler & Doupe, 1999d; Kao et al., 2005b, 2008b). Our data confirms that singing related activity of most of LMAN neurons was lower (lower firing rate and peak of firing rate), less bursty (higher median interspike interval) and more reliable (lower Fano factor and higher cross-correlation of the instantaneous firing rate) during FD than UD.

In Area X (Figure 2.10 Suppl. Fig 2B), putative pallidal cells exhibit lower firing rate and peaks of firing rate, and higher Fano factor, interspike intervals and cross-correlation of the instantaneous firing rate during FD than UD, thus illustrating a decrement of burstiness and cross-rendition irregularity when the bird is producing FD. Putative striatal units do not exhibit major social modulation, with the notable exception of the putative MSN. As shown by the examples in Figure 2.10F and Figure 2.10 Suppl Fig 2B, we noticed that MSN tend to exhibit more sparse firing for FD than for UD. Overall, MSN tend to fire less during FD than UD. While most of the putative ACH units (7/10) were upregulated by the social condition with a higher mean firing rate for FD than for UD, the overall difference did not reach significance (-6.9 ± 12.17 Hz for FD – UD FR). This is the first report on the modulation of ACH firing properties and the first-time upregulated spiking activity has been observed in Area X from UD to FD to our knowledge.

Interestingly, a recent *in vitro* study by Budzillo and colleagues (2017), described non-cholinergic neurons within Area X that propagate an increased excitatory drive in pallidal neurons when exposed to dopamine, a neuromodulator whose concentration in Area X is higher during FD than UD (Sasaki et al., 2006). These anatomically identified cells bear resemblance to subthalamic cells identified in mammals, exhibiting spontaneous activity at approximately 20 Hz in slice

preparations (Budzillo et al., 2017). It remains a possibility that some of the putative acetylcholine (ACH) units could be of this novel subthalamic-like cell type, but it remains challenging to definitively distinguish between the two given the unknown spiking properties of subthalamic cells *in vivo*. Further research and functional studies are yet to be conducted to clearly unravel the specific role of these cell types in the complex neural network during singing behavior.

Since we were able to identify only two putative LTS and three putative FS neurons, we cannot provide reliable assessment on their property, although it is interesting to note that both LTS and 2/3 FS exhibit an important decrease of singing-related activity during FD singing (LTS unit 1 shown in Figure 2.10 from 112.15 to 55.06 Hz, LTS unit 2 shown in Figure 2.10 Suppl Fig 1 from 106.85 to 16.37 Hz). Finally, many units from the Area X that we were not able to assign to a specific class do exhibit significant social modulation consistent with previous reports (Hessler & Doupe, 1999c).

Cross-areal interaction between LMAN and Area X during singing

One notable advantage of Neuropixels probes is their capability to facilitate simultaneous recordings from distributed brain regions, providing a unique opportunity to comprehensively characterize neuronal dynamics across different areas of the brain. In our study, we focused on recording the concurrent activities of neuronal populations in LMAN and Area X. To investigate their functional connectivity and the distinct influences they may exert on each other, we employed two analytical methods on the beforementioned data from our five birds.

Initially, we applied Canonical Correlation Analysis (CCA; see Methods, Appendix IV) to the neural data from Area X and LMAN. CCA, a multivariate linear regression approach, seeks to maximize the correlation between two neural populations, thereby estimating the upper limit of potential interactions or coupling between brain regions (Semedo et al., 2020). In the case of one bird, g4r4, the population coefficient distribution shows a pronounced peak, indicative of a functional coupling between LMAN activity around 150 ms after motif onset and activity in Area X preceding by 50 ms (Figure 2.11A). This peak exceeds the shuffle predictor benchmark, suggesting a possible real-time influence of Area X on LMAN.

While a 50 ms lag might seem substantial, a similar analysis in the macaque visual system conducted by Semedo et al. (2022) identified feedback-driven periods during early cortical processing of visual stimuli with similar time lags (around 50 ms). While feedforward-driven interactions in this study were temporally sharper, peaking at a delay of 2 – 3 ms, feedback interactions displayed less temporal precision. This could be attributed to factors like polysynaptic connectivity or a modulatory influence from an alternative source (Semedo et al., 2022). Similarly, the impact of Area X on LMAN might also exhibit a delay and less precision, potentially due to a polysynaptic route that includes transmission DLM. This hypothesis is supported by from (Kojima et al., 2013) which found that lesions in Area X led to reduced bursting patterns in LMAN. However, it's also possible that an alternate source could simultaneously affect both LMAN and Area X, resulting in slightly varied delays due to distinct influences within their neural networks.

Contrastingly, in the other four birds studied, CCA failed to identify moments of correlated neural activity surpassing the shuffle predictor threshold (Figure 2.11B).

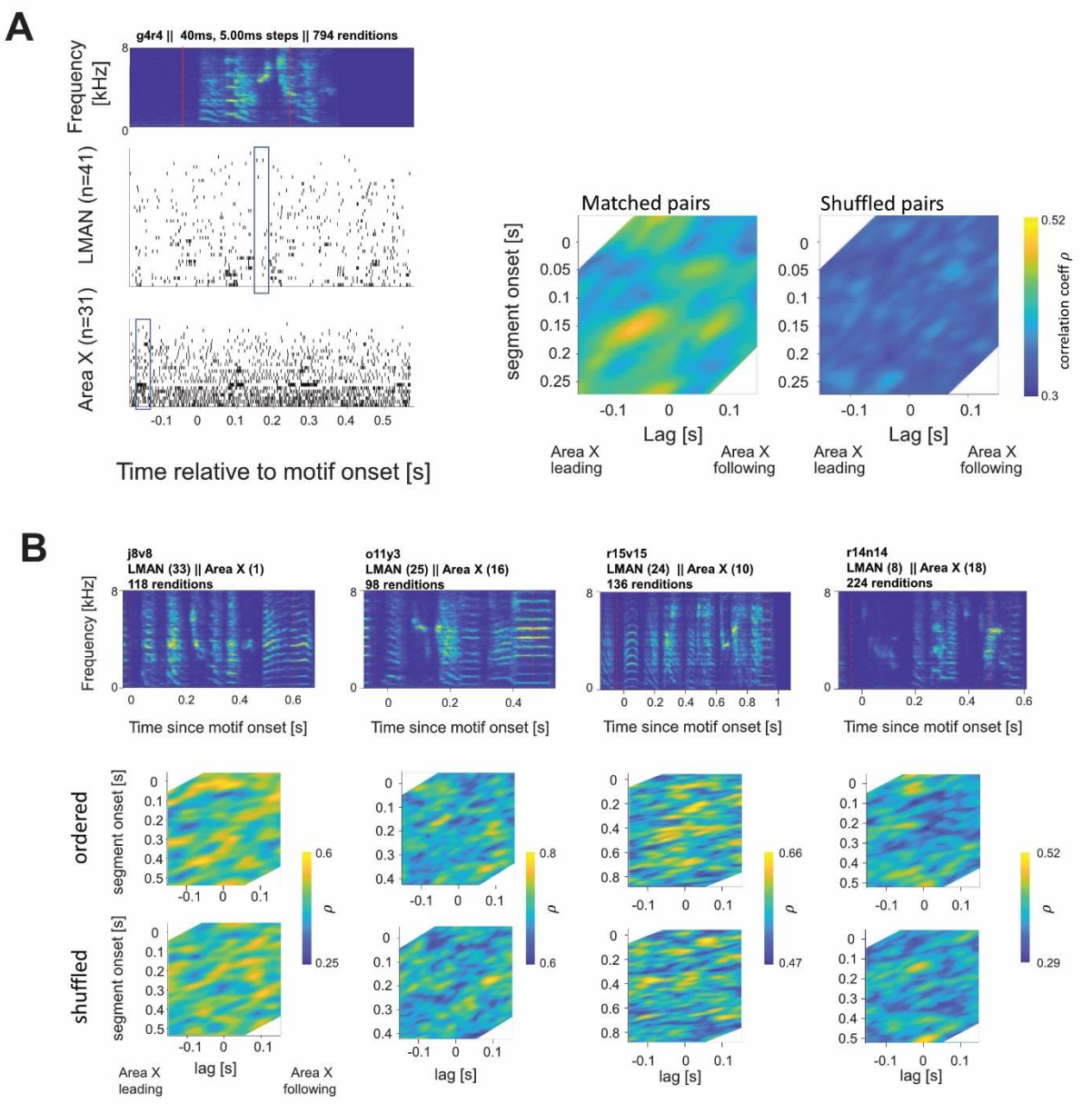


Figure 2.11 Bird-specific effects of Area X and LMAN coupling revealed by Canonical Correlation Analysis. A *Left: An example of a raster plot displaying LMAN and Area X units sorted by firing rate and aligned to the motif onset that is shown as a spectrogram on the top. Canonical Correlation Analysis (CCA) was conducted on paired segments of each 40 ms sliding in 5 ms steps of the analysis window -50 ms before motif onset to 100 ms before motif offset. Right: Leveraging the data from 794 motif renditions, correlation coefficients around 150 ms after the motif onset reveal a robust correlation between the neuron populations, with Area X leading by around 50 ms. This effect is not observed in shuffled controls.* ***B*** *The same CCA analysis conducted in four other birds did not yield conclusive results. Correlation coefficients fall within the range of shuffled control data. Note the color coding of correlation coefficients is scaled for each bird individually but the same across conditions.*

As LMAN and Area X share direct and indirect connections in the recurrent anterior forebrain pathway (Figure 2.12A), we next asked whether it is possible to identify the timing and valence of propagated activity from Area X to LMAN and vice versa between individual pairs of neurons. We used the cross-correlation between spike trains as a function of temporal lag and the cross-covariance as a readout of the momentary physiological transmission (see Methods, Appendix IV). We hypothesized that the direct, glutamatergic projection from LMAN to Area X causes an increase in the cross-covariance function at a short time lag when LMAN leads. Vice versa, increased activity in Area X can transmit negative feedback by inhibiting the excitatory drive of LMAN from DLM (S. W. Bottjer et al., 1989; Farries & Perkel, 2002; Goldberg et al., 2010b; Livingston & Mooney, 1997a; M. Luo & Perkel, 1999a, 1999b), which we expected to cause a negative deflection when Area X leads. Accounting for the multisynaptic propagation and many indirect pathways, we expected this effect to occur at a larger time lag (8 - 12 ms) with a potentially broader peak.

Figure 2.12B presents an example of a pair consisting of an LMAN neuron and a classified MSN, both of which display a notable increase in activity during the motif. The cross-correlation analysis of this pair, as depicted in Figure 2.12C, reveals a peak at approximately 100 ms, mirroring the peak in the shuffle predictor's cross-correlation function (refer to Methods section for details). To better understand the immediate impact of the pair-wise interactions and reduce the bias of average firing patterns, we utilized cross-covariance analysis. This method calculates the difference between the cross-correlation of matched pairs of renditions and their shuffled counterparts.

The average cross-covariance function for all neuron pairs showed no significant deviations within the +/- 50 ms lag, suggesting that diverging effects might be averaging out. Therefore, we conducted a detailed analysis of individual neuron pairs to identify those with significant deviations in their cross-covariance functions. We evaluated the magnitude and timing of these deviations under various conditions: undirected singing, female-directed singing, and baseline (no singing). This analysis involved determining whether and when the values surpassed a significance threshold, set at three standard deviations from a control based on Jackknife resampling, across lags from 2 to 50 ms.

Figure 2.12D presents the cross-covariance functions for two representative neuron pairs, illustrating the variable effects across conditions. The first pair exhibits increased functional connectivity with the LMAN neuron leading during undirected singing; this connectivity decreases and shifts, with Area X taking a leading role during female-directed singing. In contrast, the second neuron pair shows no significant functional connectivity, as indicated by a flat cross-covariance function during undirected singing. However, a negative deflection occurs in the LMAN leading component when the bird sings to a female, highlighting the diverse neuronal interactions observed.

At the population level, our analysis shows that approximately 25% recorded Area X units influence simultaneously recorded LMAN units. The proportion of functionally connected units per Area X unit demonstrates variability across different conditions: undirected singing (UD) at an

average of 25±1.2%, female-directed singing (FD) at 26±1.3%, and non-singing baseline (BL) at 31±1.3%. Statistical analysis using a one-sided ANOVA revealed significant differences (p = 0.014), with a higher fraction of units functionally connected during baseline than during undirected singing, as confirmed by post-hoc comparisons (p=0.014). Time lags were consistent across all conditions, centered around 18±10 ms (one-sided ANOVA p = 0.78). Interestingly,

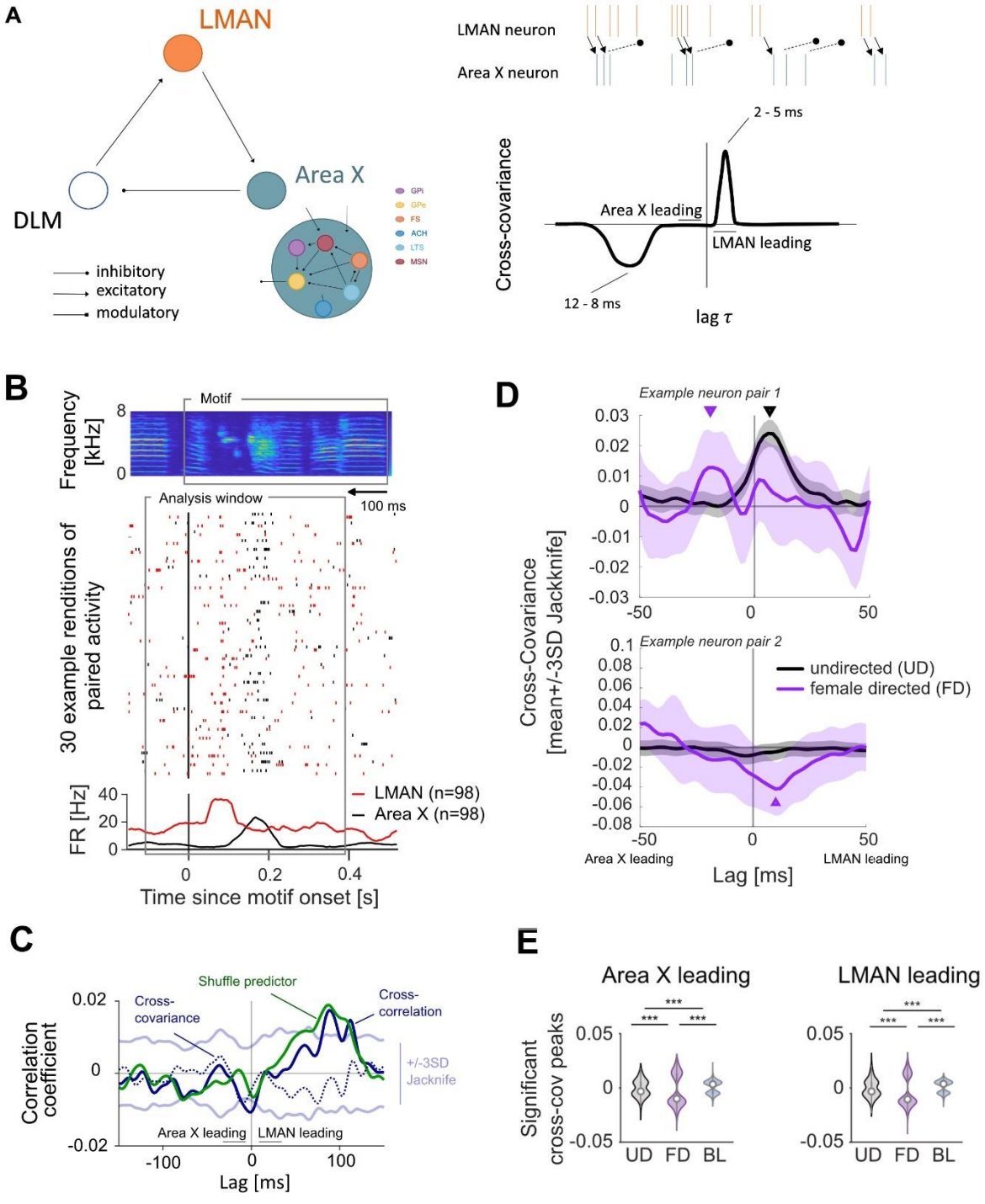


Figure 2.12 Cross-covariance analysis in simultaneously recorded Area X and LMAN units during various behavioral states. A *Illustration of the neural circuitry between LMAN and Area X, highlighting the direct and indirect pathways involved on the left. On the right, a hypothesis of the interaction between pairs of LMAN and Area X units is depicted. Because of the anatomical connectivity, we hypothesized that LMAN exerts a direct excitatory effect on Area X neurons whereas the effect of Area X has a larger lag accounting for the multisynaptic connection and exerts and inhibitory effect.* ***B*** *A representative pair of an LMAN neuron and a classified MSN, which both exhibit pronounced activity during the motif, is given. The panel depicts an example spectrogram of the motif and aligned spike raster, with average firing rates plotted below.* ***C*** *Cross-correlation analysis of the example pair from B, showing a peak at around 100 ms, which coincides with the shuffle predictor's peak, suggesting the need for further analysis to discern genuine connectivity. This led to the use of cross-covariance analysis.* ***D*** *Two example neuron pairs and their cross-covariance function during undirected and directed singing are shown.* ***E*** *Displayed is the distribution of covariance coefficients at the first significant deviation for various conditions: undirected singing (UD), female-directed singing (FD), and non-singing (BL). The data exhibits a bimodal distribution in all conditions, with more pronounced extremes during singing compared to baseline. Notably, the size of the correlation coefficients is significantly different across conditions. Female-directed singing shows lower coefficients than undirected singing, and both singing conditions display lower coefficients than the non-singing baseline.*

covariance coefficients displayed a bimodal distribution, indicating both positive and negative influences from Area X to LMAN (Figure 2.12E, left). We observed significant variability in these coefficients across all conditions (one-sided ANOVA $p < 10^{-6}$), with the smallest values recorded during undirected singing (UD: $-6.16*10^{-04}$), followed by silent periods (SP: $7.86*10^{-04}$), and the strongest during female-directed singing (FD: -0.0026). Post-hoc analysis revealed significantly stronger coefficients during female-directed singing compared to both undirected singing ($p = 0.008$) and silent periods ($p < 10^{-4}$).

Similarly, when examining the influence of LMAN leading times, we noted a small yet significant difference in the proportion of functionally connected cells across conditions: UD at 23±1.2%, FD at 24±1.3%, and SP at 30±1.4% (one-sided ANOVA $p = 0.015$). There was a higher fraction of connected units during baseline than during singing conditions, as evidenced by post-hoc comparisons (BL vs UD: $p = 0.001$; BL vs FD: $p = 0.008$, UD vs FD: $p = 0.83$). Time lags during spontaneous periods were consistent, averaging around 18 ms in all conditions (one-sided ANOVA $p = 0.14$). Additionally, as in the Area X leading analysis, cross-covariance coefficients showed significant variability across all conditions, with the most extreme values recorded during female-directed singing (UD: $-5.45*10^{-04}$, FD: 0.004, SP: 0.0013; one-sided ANOVA $p < 10^{-13}$). Post-hoc comparisons further highlighted marked differences between all conditions (UD vs BL: $p = 0.01$; other comparisons $p < 0.0001$).

These results highlight the intricate dynamics of neural interactions within this songbird circuit, with bimodal distributions (Figure 2.12E) underscores differences in connectivity patterns across neuron pairs, and particularly during female-directed singing, where correlation coefficients are significantly lower compared to other conditions.

Discussion

Neuropixels probes are a promising tool for large scale electrophysiological recordings in freely moving animals. Systems were previously designed for allowing recordings in freely moving rats. Here, we aimed at going one step further in the usability of these probes by adapting it to the constraint of a small and light animal model, the songbird zebra finch. We were facing several unknowns. First, we were unsure of the capacity of the birds to support the implant and to sing after recovery. Second, previous studies have tested Neuropixels only in the mammalian brain and the possibility to use the probe to perform recordings in two brain areas simultaneously in an avian brain remains unclear. Third, we wanted to ensure the capacity of the probe to provide sufficient signal-to-noise ratio allowing a proper discrimination not only of single units but also of the unit types. Lastly, we wondered whether the recording quality will remain at a sufficient level over multiple days to cover a minimum period allowing birdsong learning. We were able to design, adjust and validate a new headstage that is light enough to allow small birds (~15g) expressing their natural singing behavior. We successfully recorded single units simultaneously in two brain areas critical for the birdsong behavior, thus opening a great opportunity to better understand neuronal dynamics at the network level during the expression of a naturally occurring behavior.

Probe casing and protection

Since the development and commercial release of the Neuropixels, several groups successfully designed their own probe holder system to perform chronic recordings in rodents, in relationship with their own experimental goals. Yet, none of these designs are adapted for the constraints of light-weight animals standing on two legs such as small songbirds. Carrying chronic electrophysiology with zebra finches implies designing an implant weighting 2 g maximum including cement (Fee & Leonardo, 2001; Yamahachi et al., 2020). In comparison, rats and mice can carry up to 25 g and 5 g on the cranium, respectively. Previous Neuropixels holders designed for rodents weighted between 2g and 2.6g (Juavinett et al., 2019; T. Z. Luo et al., 2020a). While the model designed for mice is light enough to theoretically be transferable to birds, implantation strategy is not compatible because having the center of mass far away from the skull with only a narrow elongation is difficult to balance for birds. We thus decided to separate the probe from the headstage. It has the double advantage of offering flexibility in the angle of implantation of the probe, thus allowing to target areas in various locations, and to ensure that the headstage remains at 90 degrees angle in elongation to the animal's natural position whatever the probe implanting site, which is an optimal position for body balance. The probe is protected with a 3D printed casing securely fixed to the skull thanks to the two light screws which length can be adjusted to experimental needs (according to the brain areas of interest). We also designed a specific stereotaxic arm to maintain the probe holder and that can be used both for implantation and explantation surgeries. Successful explantation critically depends on the amount of cement used for implantation. After proper cleaning of the explanted probes, and as reported before (T. Z. Luo et al., 2020a), we were able to perform recordings of comparable quality than first-time implanted probes.

Extension to Neuropixels 2.0 edition

A newer version of the probe, the Neuropixels 2.0, accompanied by a new headstage was reported recently with improved specifications (Steinmetz et al., 2021). A single probe weighs only ~0.16 g, i.e. less than half of the initial version. The new headstage also decreased to merely 0.6 g and a smaller size of 10 x 14 mm. Both the probe and headstage also advanced in terms of functionality. In the first edition, only one probe could be connected and sampled from per headstage whereas now, two probes can be handled by one headstage. Additionally, the probes can now be purchased in a version with one shank carrying 1280 electrode sites, or a multi-shank version with four shanks overall comprising 5120 electrodes. Although one can still 'only' choose a subset of 384 electrodes to stream and record from, the increased spatial spread allows capturing more independent brain tissue in the coronal or sagittal plane in one animal.

It was not possible to include and test the new Neuropixels 2.0 probe in this study because it only became available in the second quarter of 2023 and the ordering process has a lead time of six months. Nevertheless, we believe that the results with respect to the neural quality and the behavioral recovery will be comparable if not better. The holder and casing need slight modification as the attached part of the probe became slimmer. We believe that these modifications will be minor and straight forward to implement, even holding the potential to reduce the weight. Moreover, replacing the electrical component by the new and lighter versions in a single probe configuration theoretically results in an implant weight of less than 1.42 g. We can only speculate at this point, but it is reasonable based on our experience that the lighter weight speeds up the recovery process after the implantation surgery, therefore increasing the chance of sessions early on with good neural readout and much singing. The overall lighter implant further creates the possibility of applying the design in even smaller birds. For instance, juvenile birds that are still in the critical sensory and sensorimotor song learning phase are smaller and thus more affected by an implant's weight. Collaborators successfully implanted the here described design in juvenile zebra finches of around 60 dph. At least two juveniles started singing in the days after the implantation surgery. A lighter design will likely accommodate their needs further and increase the success rate of these experiments, and potentially allowing experiments even in younger zebra finches.

Alternatively, one could also consider extending the configuration to leverage two probes, for instance with one probe per hemisphere. Such an implant including two Neuropixels 2.0 probes and one single headstage of the newer edition would add up to approximately 2.03 g. Although this weight compares to the weight of the here presented NIA, it is unclear how the animals would cope with it. It is possible that the added size that comes with an additional probe, holder, and casing will impact them in their natural behavior, e.g., during pecking, but it remains to be tested at this time. If the added dimensions are manageable, we would expect a similar recovery and habituation process as seen in the here presented results. A challenge will remain to find a configuration to place and attach all components on the skull. Additionally, the implantation surgery will require more time. The mapping of an additional brain area and the cementing of an additional probe will take approximately an hour that needs to be considered. If these challenges

are met, however, such a configuration would scale up the number of recorded units and give exciting insight, e.g., into the cross-hemispheric interplay during singing.

Limits & consequences

During the course of the experiments, we encountered several issues that we report below, together with some suggestions for changes and/or improvements. First, after a few hours of recordings, in absence of a commutator between the cable plugged on the bird's head and the acquisition system, the cable plugged on the bird's head was unsurprisingly getting untangled. Cable loops in the vicinity of the bird's head were manually moved away without unplugging the cable, allowing the bird to keep freely moving in his home cage and allowing us to pursue with the recordings. Yet, this is not a perfect solution and there is a clear need for a reliable commutator, especially for ones aiming at performing long-term and undisturbed recording sessions.

Second, an unexpected issue that we faced considering claims in previous reports (T. Z. Luo et al., 2020a) was the decrement in signal quality in terms of the number of (single or multi) units isolated. There was overall a continuous loss of the number of units recorded from one day to another and after about one-week post-implantation, almost no units were detected anymore. Luo and colleagues (2020) reported that the number of units was rather stable for more than two months on channels deeper than 2mm. Here, the decrease was affecting the entire shank, independently of the depth of the channels. Since reimplanted probes exhibited performance in a similar range that first-implanted probes, it argues in favor of changes in the neural tissue around the probe rather than in technical issues of the implanted probe itself. Neuropixels probes are made of silicon, a material that induces higher activation of glial cells, which will eventually constitute a glial scar around the probe, and thus reduce the signal quality (Moshayedi et al., 2014). Recently developed biocompatible polymer probe-based systems may allow solutions to circumvent these limitations of the silicon probes (Chung et al., 2019; Salatino et al., 2017). Another way to extend the usability of the probe would be to connect it to a microdrive which would offer the possibility to compensate micromovements of the probe and to move the probe post-implantation. This would require adjustments of our design since it would add more weight on the bird's head. A simpler possibility would be to attach the NIA to a small screw. In both cases, one remaining issue is that about half of the probe stands out of the bird brain. Future designs of the probe itself could include a length option so it could be more appropriate to small brains or to target areas located not deeper than 5 mm. Instead of the loss of recorded units in a few days, we were able to follow the same units for several hours on a single recording session. This is promising because it allows following putative changes in neuronal tuning over song practice or conditions, as we report between undirected *vs* female directed singing.

Large-scale electrophysiological recordings in freely moving birds

In the past twenty years, methods for performing chronic electrophysiological recordings in birds have moved from single to multi-channel systems, increasing progressively the number of simultaneously recorded channels: 1, 8, 16, 32, 256 channels using either tungsten wires, carbon fibers or glass pipettes (Arneodo et al., 2021; Cazala et al., 2019; Das & Goldberg, 2021; Elmaleh

et al., 2021; Giret et al., 2014b; Guitchounts et al., 2013; Hahnloser et al., 2002b; Jovalekic et al., 2017). Some methods were done only in head-fixed or restrained animals while others aimed at performing recordings in freely moving, and singing, birds. Most of these previous methods only allowed recordings in a single brain area. However, the singing behavior is under the control of a set of interconnected brain nuclei and a proper understanding of the dynamics at play requires multi-areal recordings. To our knowledge, only one study intended to simultaneously record in freely moving and singing birds more than one brain area using electrodes (single tungsten wires) implanted into the RA and LMAN nuclei (Darshan et al., 2017b).

Given the nuclei distribution in the songbird brain, the length of the Neuropixels probe and the angle flexibility for implantation provided by our designed headstage, many combinations of multi-areal recordings can be envisioned, such as HVC/RA, HVC/NIf, HVC/Field L, or maybe even HVC/Field L/Area X. Owing to the length of the Neuropixels shank, one can even envision to reach deep structures in the thalamus (e.g. DLM) or in the midbrain (e.g. VTA), thus opening a wealth of possibility to investigate neurophysiological mechanisms across brain areas. Here, we focused on LMAN and Area X. We were able to simultaneously record for several hours dozens of neurons in LMAN and Area X while birds were resting, singing alone or in the presence of a female.

Area X includes several neuron types (Goldberg et al., 2010b; Goldberg & Fee, 2010b). Here, we were able to identify cells with properties corresponding to the different striatal and pallidal Area X cell types. Surprisingly, while MSN neurons are supposed to be the most frequent cell type in Area X, only a few examples were detected in our dataset. One reason might be related to the sparse firing that MSN neurons exhibit when the birds are singing and to the way the spike sorting algorithm handles sparseness of spikes occurrence. The Kilosort algorithm that we used allows adjusting the minimum firing rate, but it increases the number of false detections. So there was a need to perform manual curation of the data, which on the one hand is time consuming but on the other hand is the sole possibility to merge or split units. Improving spike sorting algorithms or curation methods were beyond the topic of the current study, but future research into providing reliable and reproductible spike sorting is still required (Buccino et al., 2020).

Insights into the neural processing within the AFP

Although our main aim was to design and adapt chronic recordings with Neuropixels probes to songbird, our data might provide some insights on the neural dynamics within the AFP underlying socially induced song modulation. Zebra finch songs are more variables when a bird sings alone (UD) than when he sings for a female (FD). While modulations of neural variability in the AFP reflect the song-to-song spectro-temporal variability, the respective contribution of LMAN and Area X, and how activity in one area impacts activity in the other, remains open to debate. Here, we confirmed that, as populations, most of LMAN and Area X neurons tend to exhibit both song and socially-modulated activity (Hessler & Doupe, 1999d; Kao et al., 2008b; Kojima et al., 2018b). We also corroborated that LMAN neurons exhibit more bursts of spikes during UD than during FD (Hessler & Doupe, 1999b; Kao et al., 2008b; Moorman et al., 2021b; Ölveczky et al., 2005a). Results are less clear regarding each putative Area X neuron type. Putative striatal neurons did not exhibit clear changes between FD and UD. The absence of social modulation of putative LTS

and FS during FD and UD might be related to the low number of each neuron type. It is worth noticing that only putative MSNs tend to fire less and that the example neuron exhibits a clear increase in sparseness during FD than UD. It is consistent with previous report showing that viral manipulations that reduce Area X MSN activity yield to decrease in song variability (Heston et al., 2018). On the other hand, both types of putative pallidal neurons (GPe and GPi) fired less with slightly lower burstiness (increased median ISI) during FD than UD. These results support the hypothesis that increased activity of pallidal neurons contribute to song variability (Kojima et al., 2018b).

Considering the local network within Area X, with all the different cell types, and the nature of the projections from one area to another within the AFP loop, various scenarii have been suggested to account for socially-modulated song variability (Heston et al., 2018). LMAN-Area X projection neurons are mostly glutamatergic, Area X-DLM projecting neurons are mostly GABAergic and DLM-LMAN projecting neurons are mostly glutamatergic (S. W. Bottjer et al., 1989; Farries & Perkel, 2002; Goldberg et al., 2010b; Livingston & Mooney, 1997a; M. Luo & Perkel, 1999a, 1999b). Within Area X, evidences support the existence of direct and indirect pathways between striatal and pallidal neurons (Gale & Perkel, 2010). The direct pathway supports the hypothesis that striatal neurons send GABAergic projections to pallidal neurons thus contributing to an inhibition of the pallidal neurons and thus to a disinhibition of DLM and in turn LMAN neurons. On the other hand, the indirect pathway would induce opposite effects: activation of striatal neurons contributes indirectly to a disinhibition of pallidal neurons, which in turn would yield to an increase of DLM and LMAN neuronal activity (Gale & Perkel, 2010). But since LMAN neurons send recurrent projection to Area X, it is also possible that it exerts a neural drive on Area X.

Our comprehensive dataset with simultaneous recordings of several LMAN and Area X neurons, provide us a unique starting point from which to decipher the intricate dynamics at play within this neural network. The analysis we conducted, employing canonical correlation analysis (CCA), revealed that the hypothesized prominent peaks surpassing the shuffle predictor's range were largely absent. This was true for all but one bird, which notably had a higher number of renditions, suggesting that an ample dataset may be crucial for uncovering significant neural interactions.

Our findings are based on analysis from a single recording session per bird. However, CCA's strength lies in its capacity to abstract from individual instances, focusing instead on low-dimensional patterns within the population activity. Thus, expanding our analysis to cover multiple sessions may provide a more robust framework for detecting subtler patterns of connectivity, potentially overcoming the constraints imposed by a limited dataset.

An alternative interpretation of our results might suggest that, in mature zebra finches, the signal transmission within the recurrent loop of the AFP could be predominantly characterized by stochastic fluctuations, which may also influence vocal output but not biasing it. This notion aligns with findings in rodents, where the coupling between interconnected striatal and cortical areas tends to solidify or become more pronounced throughout the learning trajectory (Kondapavulur et al., 2022; Lemke et al., 2019). Therefore, it is conceivable that during periods of active vocal learning or during adult reinforcement learning, the AFP may engage in a more coordinated

activity between LMAN and Area X. In the absence of such learning-related tasks, this functional coherence may recede into the background.

Indeed, in addition to the undisturbed singing condition, pitch-shifting experiments were undertaken with a subset of the birds while performing simultaneous neural recordings from LMAN and Area X with our NIA (detailed in Table 2.2). Preliminary examination of this data revealed that merely one bird consistently altered its vocal pitch in response to the contingent aversive feedback. Nonetheless, this observation positively reinforces the notion that the songbirds, when carrying a NIA, can modulate their vocal pitch.

Taken together, our study demonstrates a significant advancement in neural recording techniques for small, freely moving songbirds using Neuropixels probes. Our lightweight implant design has facilitated the collection of an extensive neurobehavioral dataset. Furthermore, our primary results reveal intricate dynamics within key song-related brain regions. The findings, while replicating many of the observations previously reported, hint at a more variable neural transmission within the AFP than expected, providing a fresh perspective on the neural underpinnings of songbird communication. This foundational work not only enhances our understanding of avian neural networks but also lays the groundwork for future research into the broader implications of neural flexibility and behavioral adaptation.

Chapter 2
Quantitative modeling and neural signatures of motor variability in zebra finches

Introduction

Movements are inherently variable. Even though we can repeat the same movement again and again, it is almost impossible to perform it in precisely the same manner twice. This holds true even for highly stereotyped behaviors such as playing a musical piece that was practiced for many hours, or a sports performance that was trained to perfection (Bisio et al., 2014a; Van Beers et al., 2013).

Indeed, motor variability from one rendition to the next (from here on shortened as motor variability) originates as part of many processes along the motor pathway. For instance, diverse cellular and neural processes, sometimes as early as during sensation, introduce variation due to their inherent stochasticity. This includes the random opening and closing of ion channels that affect the timing of action potentials, or fluctuations at the neuromuscular junction which contribute to variations in muscle-force (Faisal et al., 2008; Lisberger & Medina, 2015).

Assuming that the nervous system is primarily interested in reliably executing stable motor behavior, such variability seems counteractive. However, in absence of supervision, motor variability can be convenient in that it provides the capacity to explore and learn from better outcomes and to adapt to changing conditions.

Adaptive control is a key component of motor learning theories. Behavioral and modelling studies in humans, for example, proposed the concepts of learning-related central planning variability and peripheral execution variability to reconcile incidental and meaningful motor variability (Therrien et al., 2016; Van Beers, 2009; Van Beers et al., 2004). Support for a central source of meaningful motor variability comes from neurophysiological investigations in non-human primates and rodents. These studies identified correlates of movement variations in neural population dynamics in motor specific cortical regions (Churchland et al., 2006, 2010; Kondapavulur et al., 2022). Cross-areal recordings and perturbation studies further indicated that the central contribution is represented in a distributed network spanning across cortical, striatal and thalamic regions (Kondapavulur et al., 2022; Peters et al., 2014; Sauerbrei et al., 2020). Still, a direct link between latent, central factors and neural substrates of exploration remains elusive.

Zebra finches (*Taeniopygia guttata*) provide insight into this matter. Adult male zebra finches repeat their song plentiful throughout the day; however, each rendition features subtle variations.

Birdsong variability largely stems from the anterior forebrain pathway, notably its output LMAN. LMAN's contribution to vocal variability has been firmly established through lesioning, temporal inactivation, or experiments that disrupt synaptic transmission (Hampton et al., 2009; Kao & Brainard, 2006; Ölveczky et al., 2005b; Stepanek & Doupe, 2010; Warren et al., 2011). These

interventions consistently lead to reduced spectral variability mostly tested on the pitch of a syllable. Additionally, targeted electrical microstimulation of LMAN subregions disrupts and alters the vocal output transiently after a short latency (40-50 ms) indicative of a direct contribution to the ongoing vocal performance (Giret et al., 2014a; Kao et al., 2005a; Kojima et al., 2018a).

The variability that the AFP introduces plays a crucial role in vocal learning, especially during the initial song ontogeny (S. W. Bottjer et al., 1984; Ölveczky et al., 2005b; Scharff & Nottebohm, 1991). During development, the neural functionality of the AFP appears to shift, with LMAN initially driving RA and later modulating (Aronov et al., 2008). After song crystallization, a remnant of this capacity to adapt features persists as evident in classical reinforcement learning paradigms. In these paradigms, birds receive feedback, usually a loud white noise burst, if they sing a particular syllable rendition higher (or lower) than an experimentally set threshold (commonly based on the prior feature distribution). The birds adapt their pitch and systematically shift in the direction that escapes the aversive stimulus. Such behavior is often referred to as '*pitch-shifting*' and relies on the functionality of the AFP (Ali et al., 2013; Andalman & Fee, 2009; Tumer & Brainard, 2007; Zai et al., 2020).

It has led to the idea that the AFP generates explorative variability (Fee & Scharff, 2010) that is reinforced contingent on reward through phasic dopamine release in the basal ganglia (Duffy et al., 2022; Gadagkar et al., 2016; Hisey et al., 2018). Nevertheless, LMAN does not account for all observed variability as residual variability remains even after complete lesion (Kao et al., 2005a; Kao & Brainard, 2006). Similarly, diurnal fluctuations in pitch have been observed in zebra finches (Wood et al., 2013) and more recently in entropy (Brudner et al., 2023). The origin of these fluctuations are yet unclear but similar circadian correlates have been described in human sports performance with peaks of performance in the afternoon (e.g. López-Samanes et al., 2017) and correlating with endogenous processes such as body temperature (reviewed in Atkinson & Reilly, 1996). It remains an open question how songbirds control adaptation in the midst of the conglomeration of learning-related and unrelated effects and the magnitude of the AFP's impact on the observed vocal variability.

In this chapter, we build on the work of Zai (2019) to discern and quantify the precise role of LMAN in shaping vocal motor variability and adaptation. Similar to prior birdsong studies (Hampton et al., 2009; Kao & Brainard, 2006; Zai et al., 2020), we adopt an atomistic approach, focusing on the pitch of a distinct syllable with clear harmonic structure (atomic pitch). By applying a dynamic latent variable model to atomic pitch trajectories, we aim to unveil the AFP's influence by comparing model parameters before and after LMAN lesion. Linking the identified noise source to learning factors, we analyze data from a set of birds engaged in the pitch-shifting task. Lastly, we test the model by correlating LMAN spiking activity with observed atomic pitch trajectories and latent LMAN pitch contribution, respectively.

The findings suggest that LMAN introduces noise in a random manner, thereby accounting for 70% of the observed atomic pitch variability on average across tested animals. Together with excellent fits of atomic pitch adaptation trajectories during pitch-shifting, this supports the notion

that LMAN variability is fueling adaptive control by integrating explorative and evaluative signals independently.

Methods

Animals

21 adult male zebra finches were subjected to the study in this chapter. The animals were bred and raised either in the avian facility at the *Institute of Neuroinformatics*, Zurich, Switzerland or in the avian facility of the *Institut des Neurosciences Paris-Saclay*, France. The animals' age ranged between 90 and 200 dph at the beginning of the experiment. They were housed individually in sound-attenuated recording chambers with access to food and water *ad libitum* throughout the experiment. The day/night cycle was set to 14 hours of day mimicked by artificial light and 10 hours of night.

All experimental procedures were performed in accordance with the Veterinary Office of the Canton Zurich, or the French Ministry of Research and the ethical committee of Paris-Sud et Centre.

Song recording

Audio signals were recorded with a wall-attached microphone (Audio-Technica PRO42). The signal was amplified, filtered, and digitized at 32 kHz, and continuously analyzed by a custom written LabView (National Instruments, Inc.) program (Herbst et al., 2023) that triggered saving upon the detection of harmonic sounds.

Automatic syllable detection and pitch assessment

Like in Canopoli et al. (2014) and Zai et al. (2020), an additional module of the custom written software in LabVIEW (National Instruments, Inc.) was employed that includes two features. The first feature consists of a two-layer neural network for automatic online detection of specific audio events. The network was pre-trained to detect a pre-selected target syllable. The second feature applies an adaptation of the Harmonic Product Spectrum algorithm (Noll, 1970) to the incoming audio signal to provide a pitch estimation of the last 16 ms every 4 ms. The detections were manually curated at the end of the experiment.

Operant conditioning of pitch

After a minimum of seven days of habituation and experimentally undisturbed singing, we started to play back a loud sound of white noise (WN) contingent on the pitch of the pre-selected target syllable. Upon detection of the target syllable, pitch was assessed and WN was played immediately (20 ms) for 40 ms through a loudspeaker (Harman/Kardon HKS 4BQ) as a response to renditions that were below (or above) a set threshold. This well-established conditioning protocol has revealed a form of adult song plasticity that allows birds to adapt their pitch to avoid the aversive sound.

The threshold was updated daily to the median pitch of the previous day. The sign, i.e., whether WN was played on high or low renditions, was determined randomly at the beginning of the experiment and remained constant throughout the conditioning period.

Lesion surgery

Lesion surgeries were performed under general anaesthesia (0.7-1.2% isoflurane, induction at 2%). Pain was managed by analgesia administered both globally through subcutaneous injection (0.05 ml lidocaine solution) and locally along the planned skin incision (0.05 g Emla® creme). Craniotomies were performed based on stereotactic coordinates (5.3 mm anterior to the confluence of sinuses and 1.7 mm lateral to the midline, at a 35-degree angle of the flat anterior part of the skull). In each hemisphere, the location and edges of LMAN were confirmed electrophysiologically using a single Tungsten wire electrode of 0.8-1.2 MΩ impedance (Micro Probe, Inc.).

Lesions were induced with ibotenic acid (Sigma-Aldrich I276), diluted and pressure injected (Picosprizter® III, Parker) into LMAN at three sites along the coronal plane using a micropipette. Pipettes were custom-made out of a borosilicate glass rods (BF-120-69-10, Sutter Instrument) that were pulled with a micropipette puller (P-97, Sutter Instrument).

The penetrations targeted the center of LMAN and positions 250 µm medially and laterally to it to account for the oval shape of the nuclei that stretches along the coronal plane.

Histology

At the end of the experiment, birds were euthanized and subsequently perfused intracranially with phosphate-buffered saline (PBS) followed by 4 % paraformaldehyde. The cerebrum was recovered and submerged in 4% paraformaldehyde for at least 24 hours before further histological preparation.

The cerebrum was then cut along the midsagittal plane and each hemisphere was processed individually. The medial part of the hemisphere was glued to a metal platform and subsequently embedded in 4% agar (Sigma Aldrich) to fixate the tissue and ensure even slicing. The specimen was cut in 80-µm thick sections along the sagittal plan using a vibratome. All slices were then Nissl stained with a 0.3 % cresyl violet acetate solution.

Lesion volume quantification

Two zebra finch brain atlases (Lovell et al., 2020; Nixdorf-Bergweiler & Bischof, 2007) were considered for the estimation of LMAN's location in sagittal slices. In general, LMAN can be identified by its large cell nuclei and its location in the nidopallium, dorsal to Area X and located between the lamina frontalis superior and the lamina pallio-subpallialis as shown in Figure 3.1A (Lovell et al., 2020; Nixdorf-Bergweiler & Bischof, 2007). The lesioned LMAN volume was estimated relative to the average intact LMAN volume in unmanipulated (control) animals. In control brains, the area of LMAN was quantified in each sagittal section (ImageJ) and multiplied by the section thickness. The resulting volumes were averaged across animals (n = 8 birds) and hemispheres to result in a unihemispheric reference volume of 0.12 +/- 0.02 μm^3 [3].

[3] Performed by Anja T. Zai.

In experimental birds with lesion (n=21), the unlesioned LMAN volume was quantified in the same fashion as in controls[4]. The fraction of lesioned volume in a hemisphere was then calculated as the difference between the remaining volume and the reference volume and normalized by the reference volume:

$$V_{lesion} = \frac{V_{control} - V_{remaining}}{V_{control}}$$

The final fraction of lesioned LMAN volume in each bird was averaged across the right and left hemispheres:

$$V_{lesion} = \frac{V_{lesion\,right} + V_{lesion\,left}}{2}$$

Dynamic latent variable model of pitch[5]

In order to estimate the impact of unknown underlying factors that affect syllable pitch, a stochastic latent variable model was applied (Zai, 2020). The model assumed that the atomic pitch p_t of a syllable rendition t is a linear combination of independent motor primitives Y_t and an independent and identically distributed (iid) source of noise, characterized by a Gaussian distribution with zero mean $\epsilon_t \sim N(0, \sigma_\epsilon^2)$

$$p_t = Y_t + \epsilon_t$$

Four different primitives were included in the final model to account for diverse effects in atomic pitch production: the song memory, slow fluctuations, circadian patterns, and history dependence.

*The **Song memory p**** was assumed to be the fixed atomic pitch component that birds sing ad libitum and was estimated as the mean pitch that birds sing under unconstrained conditions.

Slow fluctuations $\boldsymbol{c_t}$ were considered to affect pitch deviations like a random walk. They were modeled with a temporal autocorrelation structure that decays exponentially conditional on the time constant τ, in combination with a Gaussian noise term $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

$$c_t = (1 - \tau) c_{t-1} + \varepsilon_t$$

*The **circadian pattern*** $\boldsymbol{D_t}$ was empirically motivated based on the re-occurring pitch trajectories in a 24-hour cycle that have previously been shown (Wood et al., 2013; Zai, 2019) and likewise observed in the recorded data. These macroscopic fluctuations were approximated by a piecewise linear function. First, timestamps within a day were converted to relative numbers so that individual timepoints $h(t)$ of a rendition t ranged between 0 and 1 ($0 \leq h(t) < 1$). Next, N_D intervals $[H_i, H_{i+1}]$ were set with the constrain that each interval covered the same number of renditions on average across days. This was important in order to reduce the impact of fluctuations in the diurnal singing rate as birds tend to sing more in the morning. $h(t)$ was then

[4] Eight birds were analyzed by the author, the remaining animals were previously published (Zai, 2019).
[5] The model was developed and applied by Anja T. Zai. A more detailed description of a former but comparable version of the model is available in Zai (2019).

associated with the interval $j \in [1,\ N_D]$ it occurred in $H_j \leq h(t) < H_{j+1}$. Finally, the relative timepoint of the rendition within this interval $\theta^j_{h(t)}$ was multiplied by the clock parameter d_j. Formally, this can be expressed as

$$D_t = \sum_{j=1}^{N_D} d_j \theta^j_{h(t)}, \quad \text{where} \quad \theta^j_{h(t)} = \begin{cases} 1 - \frac{h(t)-H_j}{H_{j+1}-H_j} & \text{if } H_j \leq h(t) < H_{j+1} \\ 0 & \text{otherwise} \end{cases}$$

Five intervals returned good fits without substantial improvement thereafter and was therefore kept for all fits ($N_D = 5$).

*A **history dependence B_t*** was assumed to bias pitch so that the time difference from the current to the previous syllable ($h(t) - h(t-1)$) would have an added effect on the current pitch, thus, accounting for differences between syllables within a song bout, for instance. The effect was modeled as a step function that built upon the inter rendition interval distribution (IRI) of the relative timepoints $h(t)$. The effect was thought to be universal across birds, thus, the IRI was split into N_B periods with edges Δ_j, $j = 1, \ldots N_B + 1$ so that all periods had to contain the same number of renditions averaged across all birds from all experiments. The individual timepoint $h(t)$ associated with the j-th period was then multiplied by the history parameter b_j. This is expressed formally as

$$B_t = \sum_{j=1}^{N_B} b_j \delta^j_t, \quad \text{where} \quad \delta^j_t = \begin{cases} 1 & \text{if } \Delta_j \leq h(t) - h(t-1) < \Delta_{j+1} \\ 0 & \text{otherwise} \end{cases}$$

The IRI was well-approximated by five splits and subsequently used as parameter for all birds ($N_B = 5$).

*The **Final composition Y_t*** was approximated as follows:

$$\begin{aligned} Y_t = c_t + D_t + B_t + p^* &= (1-\tau)(Y_{t-1} - D_{t-1} - B_{t-1} - p^*) + \varepsilon_t + D_t + B_t + p^* \\ &= (1-\tau)Y_{t-1} + \tau D_{t-1} + D_t - D_{t-1} + \tau p^* + B_t - (1-\tau)B_{t-1} + \varepsilon_t \\ &\approx (1-\tau)Y_{t-1} + \tau(D_t + p^*) + B_t - B_{t-1} + \varepsilon_t \end{aligned}$$

Model extension to reinforcement learning

Zebra finches can learn to bias their pitch when they receive aversive or appetitive pitch contingent feedback (Andalman & Fee, 2009b; McGregor et al., 2022; Zai et al., 2020). To account for this reinforcement related motor bias in the operant pitch conditioning paradigm (see above), the model was extended and another independent latent variable r_t was added:

$$p_t = \epsilon_t + Y_t + r_t$$

The bias was modelled as a stochastic leaky integrator that steadily decays over time while integrating upon rewarded renditions. The decay was parametrized by δ, and the reward term $R_t$was scaled by the learning rate α, and $\eta_t \sim N(0, \sigma_\eta^2)$ was a gaussian noise source with unknown variance σ_η^2 :

$$r_t = (1-\delta)r_{t-1} + \alpha R_{t-1} \cdot \epsilon_{t-1} + \eta_t$$

The reward term was set to $R_t = 0$ on escape trials (no WN played) and $R_t = -1$ on hit trials (WN played).

Parameter estimation and model constrains

The latent variable model was interpreted as a linear Kalman filter[6] with time-dependent coefficients and inputs. Parameters were fitted iteratively using the expectation maximization algorithm (Ghahramani & Hinton, 1996; Shumway & Stoffer, 1982), where the objective is to minimize the error in the model prediction by maximizing the log-likelihood of observing the data given the model parameters. More details can be found in Zai (2020).

The song memory term p^* and the history parameters b_j are pure additive factors. To overcome this redundancy in the model and to circumvent the model from being under constrained, the first history parameter was fixed to zero ($b_j = 0$) so that renditions with small IRI are closest to p^* in the absence of conditioning stimuli.

Neurobehavioral analysis

We leveraged data from Neuropixels recordings in LMAN and Area X from Chapter 1 and used one recording session per bird for further analysis in this chapter. For methodological details on the electrophysiological data collection see Chapter 1. Details specific to this chapter are described in the following.

Atomic pitch was computed post-acquisition in target syllables with a stable harmonic stack. We used a custom-written Matlab function[7] of the same algorithm described before (p. 61). This process involved the time-dependent power decomposition of the target syllable's sound waveforms. Specifically, a spectrogram was generated using the discrete Fourier transform applied to a 16 ms sliding window which advanced in increments of 4 ms.

In all except one bird, we directly computed the atomic pitch on the audio recording of the Neuropixels system (see Chapter 1). In one bird (g4r4), we leveraged the continuous monitoring system that recorded vocalizations during all times and computed the atomic pitch of a target syllable from the RecOOrder data (Figure 1.5, SFigure 3.2). To align the timepoints in the neural data stream, we cross-correlated the pseudo-RMS of the RecOOrder spectrogram and the Neuropixels system's (nidq) spectrogram (SFigure 3.3).

The ***detrended atomic pitch*** was calculated by removing trends from the atomic pitch trajectory, utilizing a moving average over 50 samples***.*** We considered this method as an approximation to isolate what we hypothesized to be independent and identically distributed (iid) noise, presumably originating from the LMAN. The last and first 25 samples as well as renditions that were sung towards a female were excluded from the analysis.

[6] The Kalman filter is widely adopted in various disciplines such as control systems, robotics and signal processing because of its ability to produce good estimates of internal states from noisy observations.
[7] The function was written by a former graduate student Joshua Herbst.

Neural premotor activity was derived from LMAN units (see Chapter 1) and counting the number of spikes in the 40 ms window that preceded the atomic pitch assessment by 60 ms. This timeframe was chosen based on findings from previous studies (Giret et al., 2014a; Kao et al., 2005a) which identified a premotor lead time in LMAN of approximately 40-50 ms. It was additionally advanced to account for the pitch evaluation process which is based on the specrogram. Notably, each column of the spectrogram represents the sound waveform of the preceding 16 ms. These two factors culminated in a 60 ms lead time for our assessment of neural premotor activity. Furthermore, we chose a 40 ms window to allow for longer premotor lead times (Kao et al., 2005b).

Correlation analysis

To evaluate the relationship between premotor activity in an LMAN unit and the variability of atomic pitch, we employed Pearson's correlation (Pearson & Galton, 1997). The correlation between the number of spikes S in the premotor window and the observed or detrended atomic pitch P_X, respectively, was calculated using the formula:

$$\rho_X = \frac{cov(S, P_X)}{\sigma_S \sigma_{P_X}}$$

where cov denotes the covariance and σ is the standard deviation.

The explanatory power of the correlation was determined through a simple linear regression (Freedman, 2009) . In this model, S is the predictor variable, while P_X is the dependent variable. with ϵ representing the residuals. The explained variance was calculated using the formula for R^2

$$R^2 = 1 - \frac{\sum_i (P_{x,i} - \bar{P_x})^2}{\sum_i \epsilon_i}$$

This equation quantifies the proportion of the variance in the observed/detrended atomic pitch P_X that is predicted by the number of spikes S.

The same equations were used to compute correlations and corresponding explanatory power for the relationship between lesion extent and changes in σ^2_{pitch} and σ^2_{ε}, respectively (Figure 3.2D). To quantify the relationship between σ^2_{ε} in freely singing and pitch-shifting conditions (Figure 3.3D), a type-II linear regression was fit that takes into account that both variables are dependent (Ludbrook, 2010). R^2 values were derived as stated above.

Results

In this chapter, we built on the thesis of Anja T. Zai (2019). We leveraged a similar lesion approach as in prior birdsong investigations (Hampton et al., 2009; Kao & Brainard, 2006; Thompson et al., 2011) and used pre- and post-lesion observations to fit a dynamic latent variable model that estimates different structures of variability. In combination with atomic pitch trajectories during a reinforcement learning paradigm, this modeling approach allowed us to identify and quantify the induced variability by LMAN and its involvement in adult spectral song plasticity.

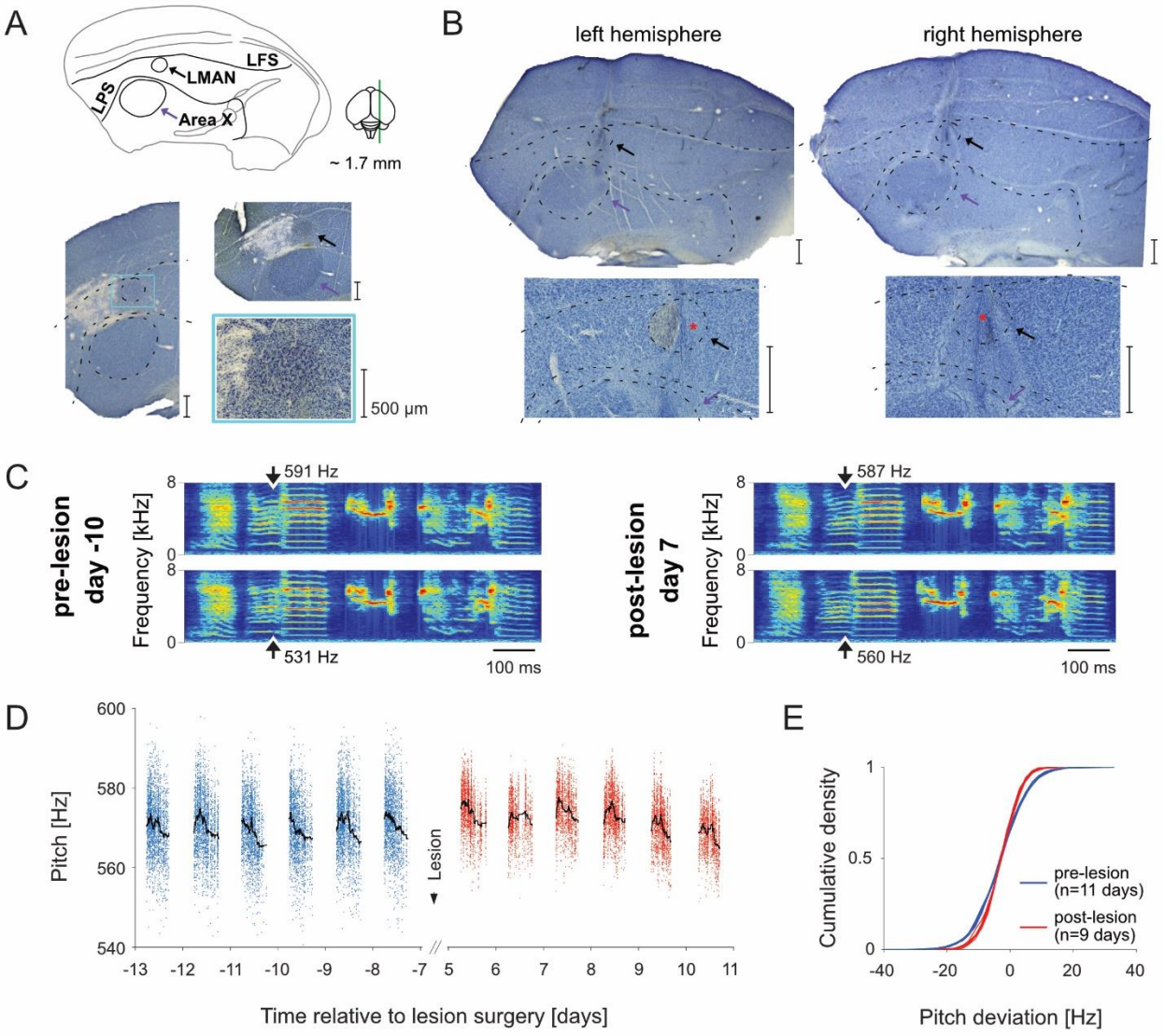


Figure 3.1 LMAN lesion affects atomic pitch variability but not diurnal pattern in the example bird o11y4. A *LMAN's anatomical localization in sagittal brain slices. LMAN is part of the forebrain's nidopallium, located dorsal to Area X, between the lamina frontalis superior (LFS) and the lamina pallio-subpallialis (LPS) (A, top, schematic with relevant structures marked by thick lines). In Nissl-stained brain sections, Area X and LMAN stand out because of the high neural density and large cell nuclei (A, bottom, song nuclei indicated by arrows). Brain slices from animal b2j2.* ***B*** *In lesioned brains, the same markers were used to determine LMAN's former location. Remaining parts of LMAN were quantified in magnified images (bottom, dashed lines corresponding to bold lines A, red crosses indicate parts of LMAN that were lesioned). In this example, 23% of LMAN (30 % left, 16 % right) remained intact.* ***C*** *The overall acoustic and temporal structure of the bird's motif was unaffected by the lesion as shown by two examples before and after the lesion. To test for fine-grained acoustic variability, pitch was obtained at a target syllable with clear harmonic structure (atomic pitch, arrow).* ***D*** *Recordings over several days showed a strong diurnal pattern in this bird with renditions of higher pitch in the morning and lower pitch in the evening. The pattern persisted after the lesion but with an overall lower degree of variability.* ***E*** *The decrease in pitch variability can also be seen in a steepening of the cumulative density functions after the lesion (each day is plotted individually and centred around its average).*

Histology

21 birds were subject to ibotenic acid lesions that targeted LMAN in both hemispheres[8]. The lesioned volume was quantified with respect to healthy controls and varied between 0 and 100 % lesion (61 % +/- 28 %, mean +/- std, Supplementary Figure 3.1). In previous studies, lesion extent and reported song changes after LMAN lesion were uncorrelated (Kao and Brainard et al., 2006, Hampton et al., 2009). However, lesion sizes were already substantial in these studies (> 40% in Kao and Brainard (2006), > 70 % in Hampton et al., (2009)) which could have led to a ceiling effect. Additionally, Thompson et al. (2009) found evidence that mere surgical intervention tends to reduce pitch variability, an effect that can introduce an offset between the relationship between variability drop and lesion extent. Thus, assuming that the extent of LMAN lesion can be positively correlated with the decrease in song variability and to account for effects related to the surgical intervention but unrelated to the lesion extent, we decided to include all birds in our further analyses.

Exemplary evidence for distinct sources of atomic pitch variability

Figure 3.1 shows the histology and behavioral recordings of an example bird with substantial bilateral LMAN lesion (77 %). Like previous studies (Hampton et al., 2009; Kao & Brainard, 2006), the motif structure remained unaltered after the lesion intervention. The holistic, 'gestalt' like appearance was qualitatively comparable between healthy and lesioned conditions (Figure 3.1 C). This allowed us to focus our analyses on the pitch of a selected target syllable with clear harmonic structure (the atomic pitch, Figure 3.1C). We tracked the atomic pitch of this syllable over several days before and after the lesion which gave us access to extensive behavioral time-series data (Figure 3.1D).

This bird exhibited a robust daily pattern, with renditions of higher pitch in the morning than in the evening (Figure 3.1D). While such slow, diurnal fluctuations are widespread in zebra finches (Wood et al., 2013), they are often overlooked in analyses limited to a subset of renditions within a day (e.g. Kao & Brainard, 2006) or analyses that aggregate the data (e.g. Kojima et al., 2018a).

Leveraging our comprehensive time-dependent atomic pitch trajectories, we found evidence that LMAN is not implicated in producing the diurnal pattern; the bird persisted in singing at higher pitch in the morning compared to the evening in the post-lesion condition. Consistent with previous findings (Kao and Brainard, 2006, Thompson et al., 2011), on the other hand, we see a strong decrease in the overall variability after lesioning LMAN (Figure 3.1D, E).

Dynamic model of atomic pitch attributes LMAN to a random noise source

Interested in modeling the different sources of motor variability that accumulate to observable song variations, we extended the thesis work by Anja Zai (2019) with additional animals. We fit a dynamical latent variable model to atomic pitch trajectories of individual birds before and after lesioning LMAN.

The model consists of a linear combination of time- and history dependent factors and a random noise source that supplement the memorized template (see Methods, Figure 3.2A). Each factor

[8] 13 birds were part of Anja Zai's thesis (Zai, 2019).

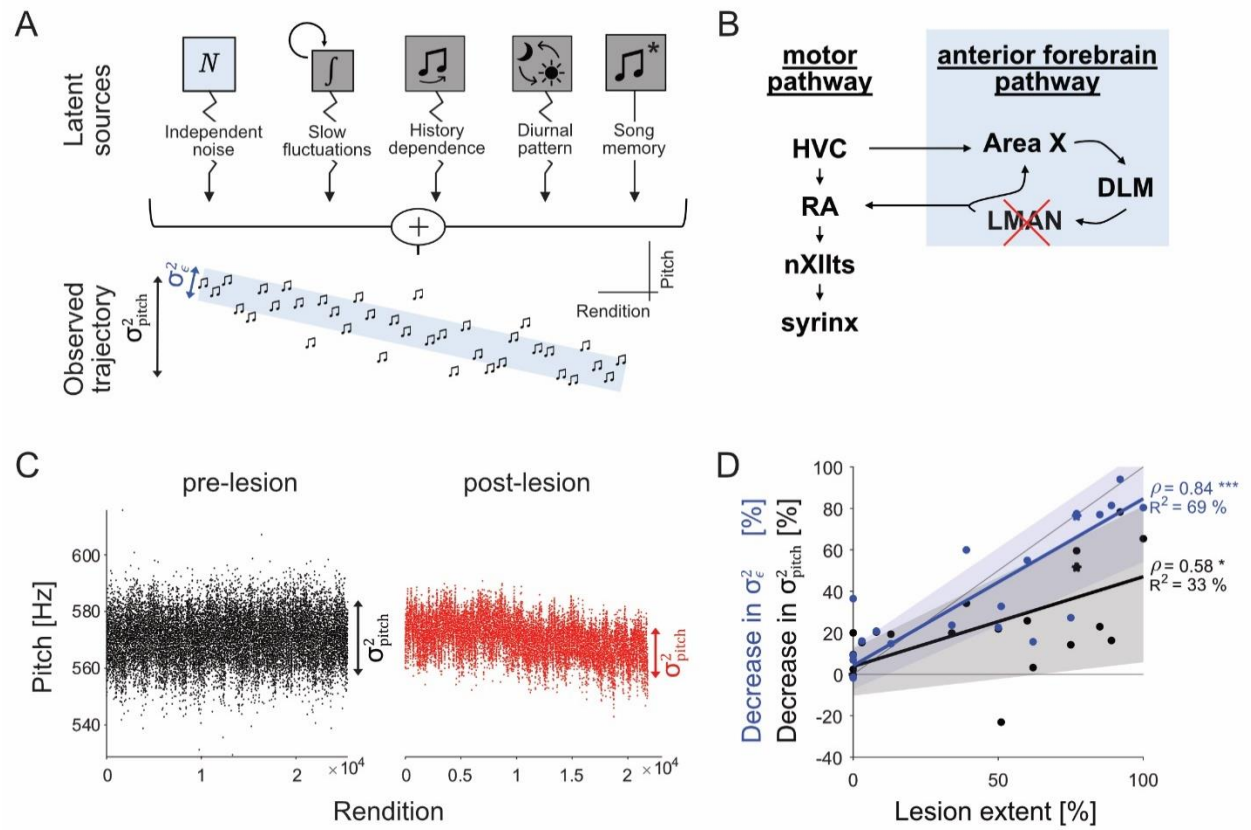


Figure 3.2 Dynamic pitch modelling associates LMAN with iid gaussian noise A *We modeled pitch trajectories as a linear combination of four different, latent sources (jagged arrows) that cause fluctuations around the fixed song memory (straight arrow).* ***B*** *Distinct brain circuitries previously identified to control the production of pitch. Variability can arise at any stage and propagate down to the syrinx but with various structure and behavioral relevance. We hypothesized that the anterior forebrain pathway acts as a random noise source and that lesioning LMAN will affect the corresponding model parameter ϵ (light blue color in A and B).* ***C*** *Over 20000 renditions before and after LMAN lesion were used to differentiate and quantify the variability induced by LMAN per bird. Data shown for the example bird o11y3 in Figure 3.1.* ***D*** *The lesion extent accounts for a large fraction of variance of the decreased random noise variability ($\sigma_\epsilon{}^2$, blue) and only a smaller fraction of the directly observed variability (σ^2, black). Shown is the type-1 linear regression with 95 % confidence interval (100 bootstrapped replicates) and annotated statistics (ρ Pearson correlation coefficient, R^2 percent explained variance, * p<0.05, *** p<0.001). Data from the example bird in C and Figure 1 is highlighted with a star.*

adds relevant power to the model and the final combination can faithfully simulate bird-specific pitch trajectories (see Zai, 2019).

We hypothesized that in each rendition, the AFP drives the atomic pitch away from its memorized template in a random fashion (Figure 3.2A). Indeed, comparing parameter fits of atomic pitch before and after lesion (Figure 3.2B, C), we found that the lesion extent is an excellent predictor of the decrease in independent and identically distributed gaussian noise variance $\sigma_\epsilon{}^2$ (n = 17, $R^2 = 68\%$) and both variables are highly correlated ($\rho = 0.84, p < 10^{-5}$, Figure 3.2D). In comparison, a fit between the LMAN lesion extent and the decrease in observed atomic pitch

variance σ^2 was less effective ($R^2 = 33\%$) and the correlation between both variables was weaker ($\rho = 0.58, p < 0.05$).

We also observed that the parameter c_t (reflecting slow fluctuations) became larger after lesioning LMAN. However, compared to ${\sigma_\epsilon}^2$, the change of c_t was unrelated to the lesion extent. The parameters that represent the history-dependent source and to the diurnal pattern in the pre-lesion condition were similar in comparison to their estimates in post-lesion conditions.

Reward associated with random noise source can account for non-monotonic pitch adaptation.

We trained 24 adult zebra finches[9] in the pitch-shifting paradigm where renditions of low (or high) atomic pitch triggered a white noise burst. Adjusting the threshold incrementally, we observed that birds systematically shifted away from their learned target to escape the aversive feedback consistent with previous reports (Ali et al., 2013; Andalman & Fee, 2009; Tumer & Brainard, 2007).

Like the example bird in Figure 3.3A, we observed that roughly half the birds produced atomic pitch trajectories with non-monotonic shifts. These birds would change their atomic pitch to escape white noise feedback over the course of several days, but their within-day change followed the opposite direction (Figure 3.2B, upper left quarter). Consequently, within a day, these birds would sing so to receive more white noise in the evening than in the morning (Figure 3.2A, B). This behavior seems counterintuitive and suboptimal in theory but could be reconciled if the diurnal pattern (and other noise sources acting against the learning gradient) were independent of the learning-related explorative noise source. Consequently, diurnal patterns that are orthogonal to the learning gradient remain unchanged and instead mask the adaptive changes within a day, as observed in our subset of birds with non-monotonic shifts.

To test this hypothesis, we asked whether an extension of the dynamic latent variable model that incorporates reinforcement signals is sufficient to explain these behavioral trajectories during pitch-shifting. Figure 3.3C shows a schematic of the amended model. Notably, the model is constructed so that the reward is associated only with the random noise source that was previously attributed to LMAN and is unrelated to the remaining behavioral primitives. In this sense, the learned bias r_t is the discounted bias accumulated in the past plus an update that scales with the injected random noise if the rendition was punished.

Estimating parameters of the extended model in the context of our atomic pitch adaptation data provided good fits and was able produce adaptive shifts over the course of multiple days with non-monotonic behavior within days. Furthermore, the derived σ_ϵ^2 during pitch-shifting was almost perfectly correlated with the parameter derived from the model that was independently fit on data during free singing (Figure 3.3E, n=19, rho = 0.99, $p < 10^{-14}$). This suggests that the underlying noise structure attributed to LMAN is comparable during free singing and under conditions of externally driven adaptive control. In combination with the validity of the model, this supports the idea that LMAN drives adaptive changes while retaining the same magnitude of exploration. In

[9] 16 birds were trained and recorded by Anja Zai or Alessandro Canopoli.

other words, independent forces of exploration and adaptation are sufficient to explain our empirical atomic pitch-adaptation data.

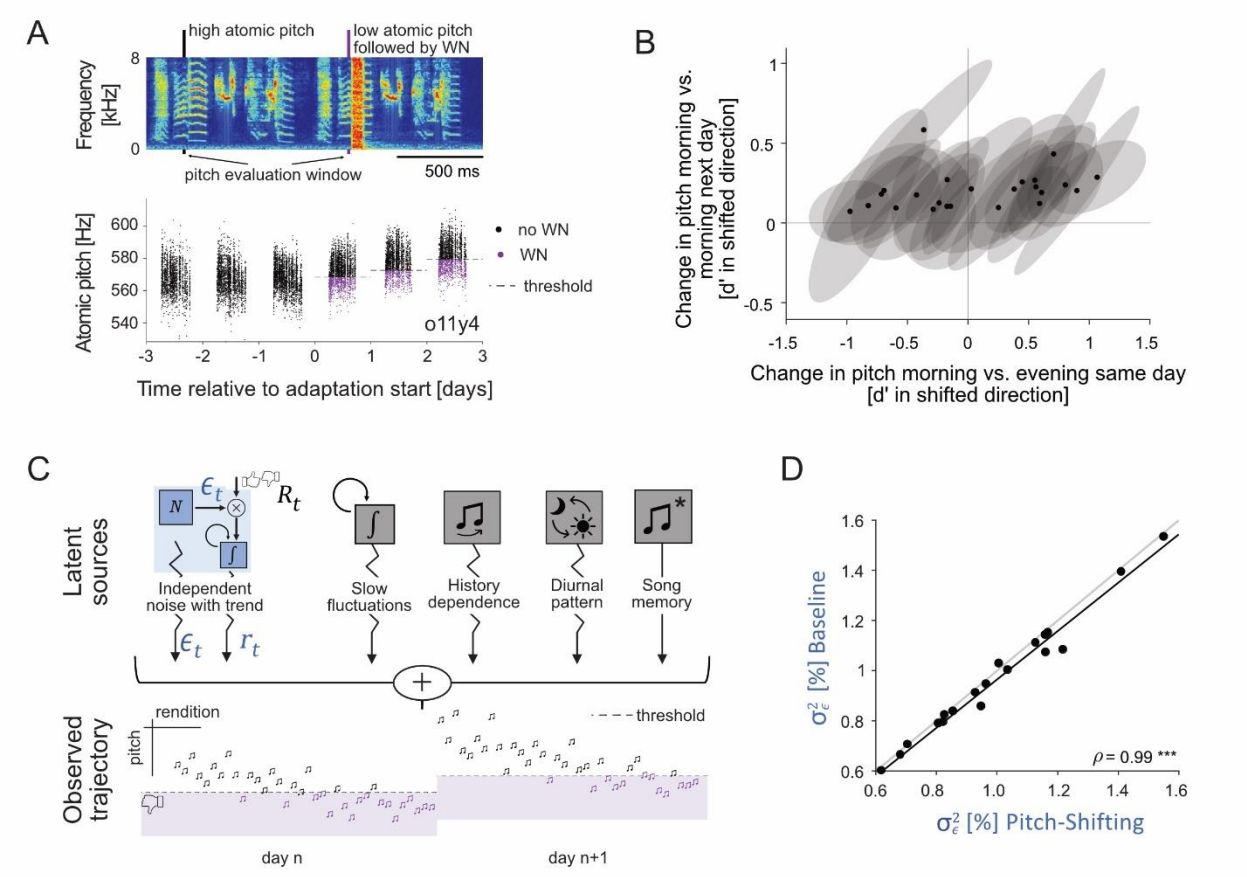


Figure 3.3 Associating reward with latent iid gaussian noise can explain non-monotonic pitch adaptation. A *Shown is data from the example bird from Figure 3.1 during the pitch shifting paradigm. The top shows an excerpt of the bird's song with two renditions of the target syllable. In this bird, low pitch renditions triggered an aversive white noise burst. The first rendition escaped the punishment since the pitch was higher than the threshold. The second rendition was produced with pitch lower than the threshold which triggered the white noise burst. Below is the atomic pitch trajectory during days of undisturbed singing followed by the first three days of pitch-shifting.* ***B*** *Like the example bird, half of the tested animals (n = 24) performed better (receiving less white noise) in the morning than in the evening. This is reflected in negative changes between morning and the evening performance (x-axis). On the other hand, all birds including those with non-monotonic daily learning, adapted their atomic pitch over a larger timescale as evident from consistent positive changes in direction of the shift from one morning to the next (y-axis). Plot by A. Zai* ***C*** *The amended model integrates a reward signal R_t in the rendition t with the output of the independent noise source ϵ_t to simulate a behavioral bias r_t as in the pitch-shifting paradigm. This model can simulate adaptation trajectories with non-monotonic diurnal features. See Methods for details on the model.* ***D*** *The variance of the independent noise source estimated from data during pitch-shifting almost perfectly matches the parameters derived from undisturbed, baseline data (n = 19 birds, Pearson correlation coefficient ρ= 0.99, p< 10^{-6}, bold black line shows type-II linear regression fit).*

Neural correlates between LMAN premotor spiking and atomic pitch variability

If LMAN introduces random iid gaussian noise among various sources of atomic pitch variability, a fundamental question persists regarding how this influence is represented in neural activity. A prior study in the downstream recipient of LMAN and main driver of vocal behavior, RA, revealed correlation between RA spiking activity and variations in of syllable specific features including pitch, entropy and amplitude on a rendition-to-rendition basis (Sober et al., 2008). We applied a similar analysis to the neurobehavioral recordings of five birds that carried a chronically implanted Neuropixels probe in LMAN (see Chapter 1).

All five birds produced syllables with at least one flat harmonic stack that allowed us to derive atomic pitch trajectories over the course of the neural recording (Appendix II). Given the comparably short recordings, we approximated the latent influence of LMAN on atomic pitch variations by determining the residuals from the smoothed trajectory (detrended atomic pitch).

We assessed premotor neural activity by counting spikes in LMAN units within a 40 ms premotor window, set 60 ms before the evaluation of atomic pitch. Panel A and B in Figure 3.6 show data from an example bird (j8v8), in which we succeeded in recording 28 spiking units in LMAN (Figure 3.6A) during a five-hour long session that contained 1375 syllable renditions (Figure 3.6B, green points).

The two fundamental questions we aimed to address here were:

1. Can (part of) the atomic pitch variation be explained by the variation in premotor neural activity in LMAN?
2. Is the prediction better for the detrended atomic pitch attributed to LMAN?

To do so, we computed correlations between 1. premotor neural activity and observed atomic pitch and 2. premotor neural activity and detrended atomic pitch and compared the prevalence and strength of significantly correlated units. The example unit in the top panel of Figure 3.4C yielded a weak, but significant ($p = 0.015$) negative correlation with observed atomic pitch. In contrast, the correlation analysis between this unit and detrended atomic pitch was weaker and not significant ($p = 0.89$). Similarly, R^2 values were small for observed atomic pitch variations ($R^2 = 0.004$, meaning 0.4 % of the variance can be explained by spiking variation in this premotor time window of this unit) and absent for detrended atomic pitch variations ($R^2 < 10^{-5}$).

We repeated the analysis for all units in this bird (Figure 3.4C, D). Out of the 29 units, seven were significantly correlated with the observed atomic pitch variation with almost equal number of positively and negatively correlated units (four positive vs three negative significant correlations). In contrast, only three units were significantly correlated with the detrended atomic pitch and all of them yielded positive correlation with comparable strength to correlations with observed atomic pitch. Notably, there was no overlap in units being significantly correlated with observed atomic pitch variations and units that were significantly correlated with variations in the detrended atomic pitch (Figure 3.4C).

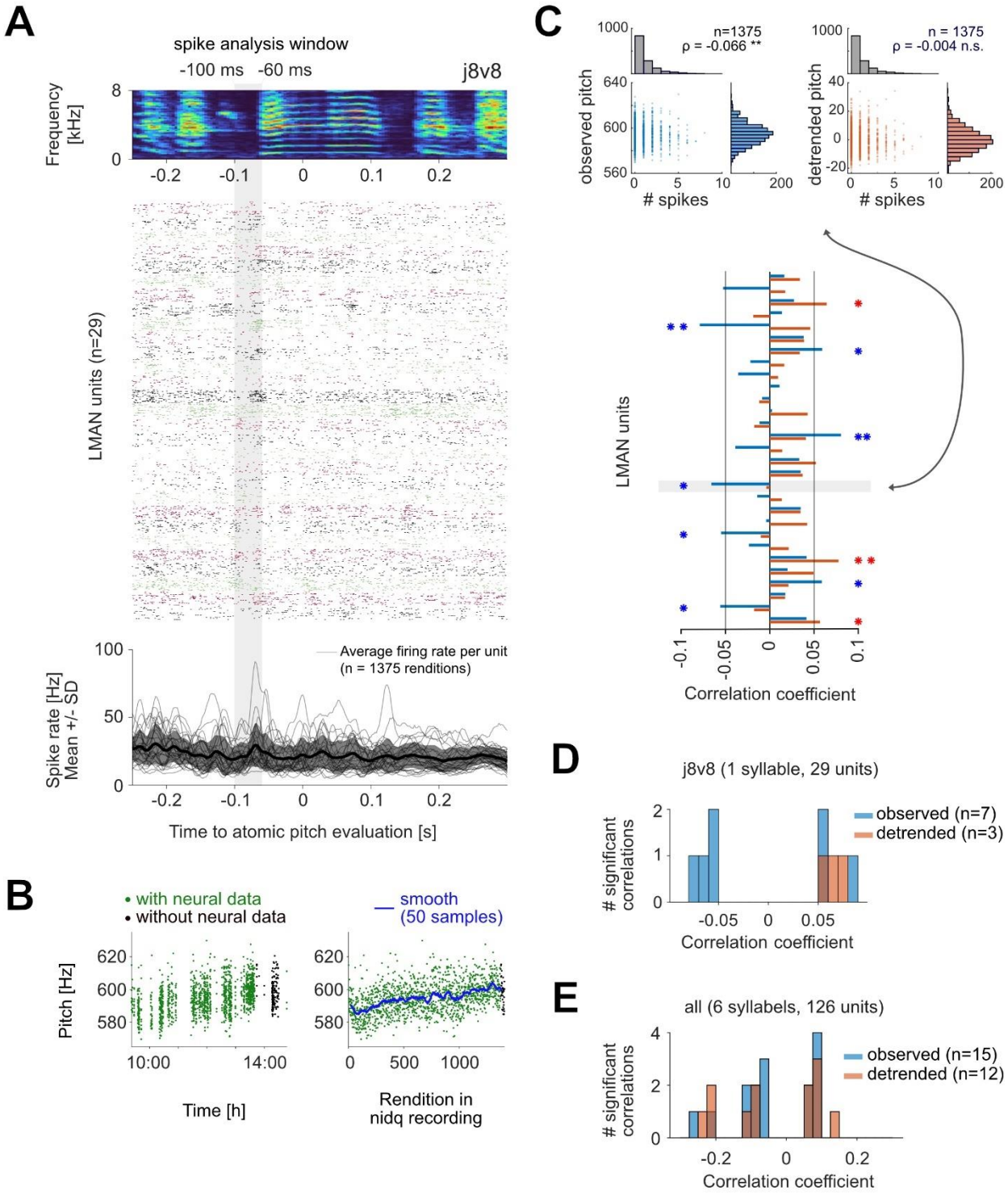

A
spike analysis window
-100 ms -60 ms
j8v8
Frequency [kHz]
8
0
-0.2 -0.1 0 0.1 0.2
LMAN units (n=29)
Average firing rate per unit
(n = 1375 renditions)
Spike rate [Hz]
Mean +/- SD
100
50
0
Time to atomic pitch evaluation [s]
B
• with neural data
• without neural data
— smooth
(50 samples)
Pitch [Hz]
620
600
580
10:00
14:00
Time [h]
0 500 1000
Rendition in
nidq recording
C
n=1375
ρ = -0.066 **
n = 1375
ρ = -0.004 n.s.
observed pitch
detrended pitch
1000
640
600
560
40
20
0
-20
0 5 10 200
spikes
LMAN units
-0.1 -0.05 0 0.05 0.1
Correlation coefficient
D
j8v8 (1 syllable, 29 units)
significant correlations
2
1
0
observed (n=7)
detrended (n=3)
-0.05 0 0.05
Correlation coefficient
E
all (6 syllabels, 126 units)
significant correlations
4
2
0
observed (n=15)
detrended (n=12)
-0.2 0 0.2
Correlation coefficient

Figure 3.4 Correlation analysis of LMAN spiking and atomic pitch variation A *Multisite raster of LMAN units aligned to the onset of the atomic pitch evaluation for one example bird Top: Spectrogram of the target syllable. Centre: Raster plot of 29 LMAN units depicted in alternating colors, each illustrating 20 randomly selected rendition. Bottom: Average spike rate for each unit, and average across all units (bold, +/-SD) across 1375 renditions. The grey area highlights the premotor analysis window.* ***B*** *The atomic pitch trajectory as a function of daytime or rendition count since the start of neural recording. The latent contribution of LMAN was estimated as the detrended atomic pitch calculated as the deviation from the smoothed trajectory (blue). Note that in this recording, the neural data acquisition was interrupted while the system continued saving the auxiliary audio input (black dots).* ***C*** *Correlation coefficients for one example unit (top) and summary of all units (bottom) in relation to both observed and detrended atomic pitch from the recordings shown in A and B. Significant correlations are marked with stars (** p<0.01, *p<0.5).* ***D*** *Aggregated significant correlations for bird j8v8.* ***E*** *Cumulative histogram summarizing all significant correlations for the entire dataset of 126 LMAN units and six syllables.*

We extended our analysis to include data from four additional birds, pooling a total of 126 LMAN units from five birds. Each bird produced at least one syllable with a stable harmonic component, adding up to atomic pitch trajectories from six syllables. We set a significance level at 5% and anticipated that a similar percentage of units might demonstrate significant correlations merely by chance. Our results exceeded this threshold: 12 % of the units showed a significant correlation with the observed atomic pitch, while 10% correlated significantly with the detrended atomic pitch (Figure 3.4E).

Among the significantly correlated units, six were present in both groups (significantly correlated with both observed and detrended atomic pitch), with comparable correlation strength and direction. In the other units, correlation directions were generally consistent with one exception where the direction inverted (Figure 3.4C, fifth unit from the top, $\rho_{observed}$ = -0.079, p = 0.003, $\rho_{detrended}$= 0.046, p = 0.09).

This uniformity in correlation direction might result from the fact that observed and detrended atomic pitch variations were highly correlated in all syllables. This observation could account for the similar proportions of significant correlations, while also suggesting that any differences might manifest more in the strength of these correlations. However, our data did not support this hypothesis. The absolute coefficients of significant correlations were overall small but comparable ($|\rho_{observed}|$ = 0.098 ± 0.059, n = 15; $|\rho_{detrended}|$ = 0.122 ± 0.066, n =12, mean±SD; two-sided Wilkoxon rank sum test p = 0.32).

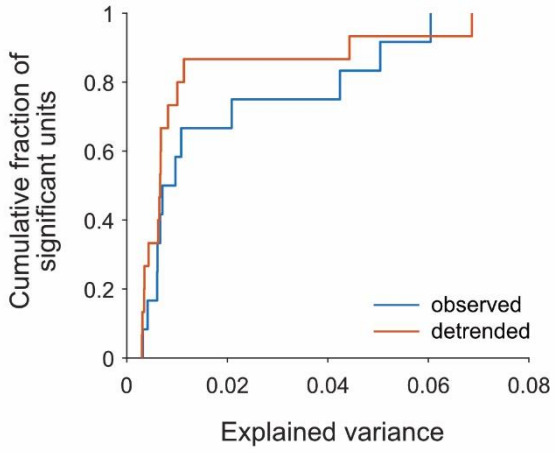


Figure 3.5 Explanatory power in significant neurobehavioral correlations. *Empirical cumulative distributions of the explained variance (R^2 values) for LMAN units with neural activity that significantly correlates with observed atomic pitch (in blue, n = 15) and detrended atomic pitch (in orange, n = 12). The explained variance reflects the proportion of pitch variability that can be accounted for by the neural activity.*

The explained variance for significant correlations was generally small. Figure 3.5 shows the empirical cumulative distributions for the two groups. For the significant correlations between LMAN units and observed atomic pitch, the R^2 ranged between 0.3 and 6% (1.9 ± 0.2 %, n = 15, mean±SD). In the case of significant correlations between LMAN units and detrended atomic pitch, the range was comparable, spanning from 0.3% to 7%, and on average slightly but not significantly smaller (1.3 ± 0.2 %, n = 12, mean±SD; two-sided Wilkoxon rank sum test p = 0.32).

Discussion

In this chapter, we assessed the contribution of the AFP, and particularly LMAN, to the acoustic variability of adult zebra finch song. We focused on the atomic pitch of a selected target syllable with clear harmonic structure as a statistically amenable and observable acoustic feature of song variability.

LMAN's contribution to atomic pitch variability

Atomic pitch variability has previously been reported to originate in part in LMAN in studies that lesioned LMAN (Kao and Brainard, 2006, Hampton et al., 2009, Stepanek and Doupe, 2010, Thompson et al., 2011) or impeded the synaptic transmission in the downstream RA (Stepanek & Doupe, 2010). We used a similar approach and compared pre- and post-lesion conditions to fit a dynamic latent variable model. Our quantitative model shows that LMAN's effect to atomic pitch variability is well-explained by a random noise source with iid properties. This was possible because of the extensive recordings over several days and parameter comparisons between pre- and post-lesion conditions.

A latent learning strategy accounts for non-linear adaptation patterns

LMAN, as part of the AFP, has also been shown to be necessary to adapt atomic pitch to external feedback as in the pitch-shifting task (Andalman & Fee, 2009b; Tumer & Brainard, 2007; Zai, 2019). Contrary to previous assumptions, our research reveals that pitch adaptation in songbirds does not always adhere to a linear pattern. About half of our birds adapted their pitch in a non-monotonic way, i.e. they sang to receive more punishment in the evening than in the morning but systematically shifted away from the punishment over the course of several days (also reported in Zai, 2019). This stands in contrast to prior work that came to the conclusion that songbirds adapt their pitch by reinforcing instances that avoided punishment (Charlesworth et al., 2011) and challenges classical reinforcement learning models that primarily associate learning with reward history (Williams, 1992).

To better understand this phenomenon, we amended the dynamic latent variable model by integrating reward signals and fitting behavioral trajectories during pitch-shifting. By linking the reward signal exclusively to the noise attributed to LMAN, while considering other sources independent, we were able to account for the observed non-linear pitch changes. This approach also reconciles the empirical data with traditional reinforcement learning theories, suggesting a more complex, latent learning strategy in songbirds.

Our modeling during pitch-shifting indicates that the exploratory behavior underlying learning remains constant throughout the process. This finding contrasts with those from human motor

learning studies, where motor variability is typically higher in beginners who just began acquiring a new skills such as playing tennis or throwing darts, and diminishes as they become more skilled (Bisio et al., 2014b; Van Beers et al., 2013). However, it must be acknowledged that these tasks are more analogous to the initial song learning phase in songbirds, during which the variability in vocal output decreases as the development progresses (Fee & Goldberg, 2011), than the pitch reinforcement learning studied here. Nevertheless, it is worth noting that the near-perfect alignment of parameters derived from model fits to pitch adaptation and freely singing trajectories (Figure 3.3D) indicates that in adult songbirds, motor variability is not modulated based on learning progress. This suggests a possible difference in how variability is regulated during initial learning compared to vocal plasticity in adults.

Sources of behavioral primitives

The origin of motor variability other than that stemming from LMAN remain elusive at this moment. Previous research by Wood et al. (2013) investigated diurnal pitch variations in zebra finches but did not observe any central modulation through serotonergic pathways that are implicated in sleep and other circadian behaviors (Monti, 2010). An alternative agent essential for regulating sleep-wake cycles and circadian rhythms in vertebrates is Melatonin (Cassone, 1990). It could exert an indirect effect of diurnal pitch variation by altering core body temperature (Cagnacci et al., 1997), which in turn affects neural signal transmission and was shown to impact motor performance (Atkinson & Reilly, 1996). Melatonin might also have a direct, central effect. The presence of melatonin receptors in the primary song control nuclei (Bentley & Ball, 2000; Gahr & Kosar, 1996) indicates that melatonin could establish a daily rhythm through internal brain regulatory processes. While our results suggest that diurnal patterns are not under the direct control of the AFP, the hypothesis that melatonin may affect atomic pitch by modulating spike activity within the song motor pathway remains an intriguing avenue for future research.

Neural correlates of LMAN's contribution to atomic pitch variability

If LMAN causes variations in syllable pitch from one rendition to the next in an iid manner, a fundamental question remains how this is represented in neural activity. Leveraging the neural spiking data recorded in Chapter 1, we asked if variations in premotor neural activity in LMAN can generally explain atomic pitch variations, and 2) whether the prediction is stronger or more prevalent for the detrended atomic pitch which approximates the distinct contribution of to LMAN.

Our analysis identified a fraction of LMAN units whose spiking activity was significantly correlated with atomic pitch variations. The prevalence these significant correlations were half as high compared to those observed in the RA in Bengalese finches (12 % vs 26 %), and the predictive power was more than four times lower (1.9 % vs 8 %) (Sober et al., 2008). Yet, the discovery of significant correlations between LMAN spiking and atomic pitch variation is a crucial first step in linking acoustic variability to a premotor area.

Consistent with Sober and colleagues' findings in RA (2008), we found both positive and negative correlations. Given that neurons in LMAN that innervate RA exert an excitatory influence on downstream neurons through NMDA receptors (Mooney & Konishi, 1991) and considering that the projections from LMAN to RA are organized topographically (M. Luo et al., 2001), it is possible

that these varying directions of correlation are a reflection of this organizational structure. Another possibility is that our recordings were made from LMAN interneurons, which are known to inhibit projection neurons (Boettiger & Doupe, 1998; Livingston & Mooney, 1997b). This mechanism could influence RA neurons and their subsequent pathways in a contrary manner. However, as Kao et al. (2008a) noted, recording from these interneurons, both in anesthetized adult birds and in slice preparations, has proven difficulty (Livingston & Mooney, 1997b; Rosen & Mooney, 2000) and to our knowledge, there are no reports of recordings of LMAN interneurons during wakefulness or singing.

The function of recorded several units, which were not significantly correlated with atomic pitch variation, is still not well understood. Recent insights from neuron population analysis discovered that neurons, even those not specifically tuned to or associated with a certain motor behavior, can provide valuable insights through the dynamics of neural populations (Gallego et al., 2017). It indicates that the influence and function of individual neurons on specific behaviors, such as atomic pitch variation, might be better understood and appreciated when analyzed in the broader context of their collective neural activity.

Our results showed comparable outcomes between observed and detrended atomic pitch. We can think of several technical reasons explaining this observation. Firstly, unlike the initial phase of our study that examined behavioral trajectories across several thousand renditions recorded over many days, the analysis of atomic pitch trajectory in the latter analysis was confined to a few hours in the day when we performed neural recordings. Secondly, there is a high correlation between observed and detrended atomic pitch. This, coupled with the limited number of renditions, could obscure small differences. Thirdly, due to the restricted number of renditions, we estimated the latent contribution of LMAN by detrending the time series using a moving average filter. There is a chance that employing the latent variable model, even with this smaller dataset, might reveal distinctions between correlations between LMAN spiking and observed or LMAN-influenced variations in atomic pitch, respectively.

Chapter 3
Interactive extraction of diverse vocal units from a planar embedding without the need for prior sound segmentation

Summary

Research on animal vocal communication heavily relies on the analysis of vocal signals. However, to dissect, categorize, and proofread an individual's vocal output still requires intense human labor, especially when the vocabulary is unknown and there are background noises that must be ignored. These challenges are usually met by preprocessing steps such as sound segmentation, annotation of training data sets for machine learning, and/or by visual inspection of a representative subset of data, which all introduce additional challenges.

We propose a simple and efficient semi-automated clustering approach aiming to speed up and automize part of this time-consuming pipeline. We segment the audio recordings into small, overlapping snippets and bypass the segmentation problem by projecting all segments independent of their content into the plane. In this plane, the different vocalization types tend to appear as distinct elongated structures. From these structures, we extract vocal instances by defining pairs of characteristic regions, one region related to vocal onset and the other related to the vocal offset. We present a simple graphical user interface (GUI) for manual tuning of these vocalization-defining regions which allows a trained user to cluster the vocal output of a zebra finch from a day-long recording within a few minutes. The clustering is precise and requires little human post processing. With this simple and intuitive method for extracting arbitrary vocal units from large data sets, we hope to increase the convenience of analyzing longitudinal recordings of animal vocalizations.

The study has been published in Frontiers of Bioninformatics in 2023 (Lorenz et al., 2023). I contributed to the concept and the design of the study, created visualizations and took part in the writing of the paper. I also refined the graphical user interface and its underlying code and created tutorials on its usage (https://gitlab.switch.ch/hahnloser-songbird/published-code/automtedclustering/2n-extraction).

The following is a direct copy of the original publication (Lorenz et al., 2023) with figure and section numbering adjusted to match the document style.

Interactive extraction of diverse vocal units from a planar embedding without the need for prior sound segmentation

Corinna Lorenz[1,2], Xinyu Hao[1,3], Tomas Tomka[1], Linus Rüttimann[1], Richard H.R. Hahnloser[1]*

[1] Institute of Neuroinformatics and Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland

[2] Université Paris-Saclay, CNRS, Institut des Neurosciences Paris-Saclay, 91400, Saclay, France

[3] Tianjin University, School of Electrical and Information Engineering, Tianjin, China

* Corresponding Author

Keywords: visual analytics, acoustic signal processing, dimensionality reduction, segmentation problem, birdsong

Abstract

Annotating and proofreading data sets of complex natural behaviors such as vocalizations are tedious tasks because instances of a given behavior need to be correctly segmented from background noise and must be classified with minimal false positive error rate. Low-dimensional embeddings have proven very useful for this task because they can provide a visual overview of a data set in which distinct behaviors appear in different clusters. However, low-dimensional embeddings introduce errors because they fail to preserve distances; and embeddings represent only objects of fixed dimensionality, which conflicts with vocalizations that have variable dimensions stemming from their variable durations. To mitigate these issues, we introduce a semi-supervised, analytical method for simultaneous segmentation and clustering of vocalizations. We define a given vocalization type by specifying pairs of high-density regions in the embedding plane of sound spectrograms, one region associated with vocalization onsets and the other with offsets. We demonstrate our two-neighborhood (2N) extraction method on the task of clustering adult zebra finch vocalizations embedded with UMAP. We show that 2N extraction allows the identification of short and long vocal renditions from continuous data streams without initially committing to a particular segmentation of the data. Also, 2N extraction achieves much lower false positive error rate than comparable approaches based on a single defining region. Along with our method, we present a graphical user interface (GUI) for visualizing and annotating data.

Introduction

Many real-world bioinformatics problems are best dealt with using semi-supervised approaches (Käll et al., 2007; Wrede and Hellander, 2019; Peikari et al., 2018), in particular when supervised and unsupervised methods are either unfeasible or unsuitable. For example, supervised learning is impractical when for a given task no training data is available or when the task definition is unclear, e.g., during explorative analysis. On the other extreme, unsupervised learning may provide interesting information, but perhaps not in a way that best suits a user, e.g., the components discovered in a data set may not be of the granularity of interest. The goal of semi-supervised learning is to combine the best of both supervised and unsupervised approaches to provide maximally useful insights with minimal human effort.

Animal behavior is an example domain well suited for semi-supervised learning, as each individual animal exhibits its own repertoire of complex actions. The task of segmenting behavioral sequences into their constituent parts is particularly challenging in the vocal space, because many vocal behaviors such as birdsongs consist of re-occurring elements that tend to be hierarchically organized (Sainburg et al., 2019) and that contain long-range structure (Markowitz et al., 2013). Here we set ourselves the goal of rapidly clustering the different types of vocalizations emitted by an individual songbird. For this type of problem, dimensionality reduction techniques come in handy because they allow to display even high-dimensional data points such as complex vocal utterances on a two-dimensional computer screen (Kollmorgen et al., 2020; Sainburg et al., 2019, 2020). However, in distance-preserving embeddings such as t-SNE (Maaten and Hinton, 2008) or UMAP (McInnes et al., 2018), the distance between two points in the plane only approximates the true distance between the pair of vocalizations in the higher-dimensional space of the original data (Kollmorgen et al., 2020). In fact, embedding distances are not perfectly preserved because local neighborhoods in two dimensions are much smaller than the true neighborhoods in the high-dimensional space. How to efficiently deal with such embedding distortions remains a bottleneck in data browsing, proofreading, and annotation tasks (Chari et al., 2021).

Moreover, natural vocalizations tend to have variable durations, which clashes with the rigid dimensionality of embeddings. Although there are workaround techniques such as zero padding, these depend on segmenting the signal into foreground and background as a preprocessing step, which tends to introduce errors caused by background noises. For example, when a background noise occurs just before or after a vocalization, that vocalization might be missed or inferred as being too long; and similarly, when a noise happens between two vocalizations, these might be interpreted as a single vocalization instead of as a pair. In general, to deal with the segmentation problem as a pre-processing step acts against end-to-end extraction of vocalizations from raw data. All these caveats and challenges limit the widespread adoption of dimensionality reduction techniques for annotating and proofreading vocalizations.

In general terms, the goal of our data annotation task is to extract flexibly defined and variably sized events from a continuous data stream. Our approach to vocal clustering is somewhat orthogonal to previous approaches where automated classifiers are optimized for the assignment of pre-computed segments to labels, either in a supervised (Tachibana et al., 2014; Nicholson,

2016; Cohen et al., 2022; Goffinet et al., 2021; Steinfath et al., 2021) or unsupervised manner (Goffinet et al., 2021; Sainburg et al., 2020). The goal there is to identify an efficient workflow that minimizes human involvement. In contrast, our aim is to place the human expert in in the center of the process along the lines of modern visual analytics (Thomas and Cook, 2006; Cui, 2019). We want to provide the flexibility of exploring and navigating audio data while visually extracting diverse vocalizations in a fast and intuitive workflow.

We address the problem of clustering zebra finch vocalizations, which is to distinguish the diverse vocalizations they produce from all other sounds in the environment. In essence, the task is to correctly identify in midst of noise, all vocalizations including their types, their onset times, and their offset times. To robustly extract variable vocalizations from possibly distorted embeddings, we introduce for each vocalization type a pair of distinguishing feature sets, one anchored to the onset of the vocalization and the other to the offset. Hence, unlike traditional approaches, where vocalizations are represented by single dots in the embedding plane (Goffinet et al., 2021; Sainburg et al., 2020; Steinfath et al., 2021; Sainburg et al., 2019; Kollmorgen et al., 2020), in our workflow, vocalizations are represented by pairs of dots. Also, because we extract vocalizations without segmentation as a pre-processing step, our definition of vocalizations by their onset- and offset-anchored feature sets implicitly solves the segmentation problem.

Methods

Datasets and sound preprocessing

Our data stem from single-housed birds (n=2) recorded with wall-mounted microphones as described in Canopoli et al. (2014) or from a pair (n=1) of birds which wore harnesses carrying accelerometers that signal body vibrations stemming from self-produced vocalizations (Anisimov et al. (2014). Although all birds have been recorded in acoustically isolated environments, the extraction of vocal units problem is in species such as the zebra finch that produce not just harmonic sounds but that also emit broadband vocalizations which can resemble non-vocal sounds (Figure 3.1, Supplementary Figure 3.1).

We sampled sound (and vibration) signals at 32 kHz and computed log-power spectrograms using the short-time Fourier transform in 512-sample Hamming windows and hop size among adjacent windows of 128 samples (i.e., 4 ms). We pre-segmented the data into sound intervals (assuming that without sound there is no vocalization) by thresholding the spectral power of microphone signals in the range 312 Hz to 8 kHz and of accelerometer signals (illustrated in Figure 3.3D) in the range 312 Hz to roughly 4 kHz. The threshold for sound interval extraction (Figure 3.1) was set to 5 standard deviations above the average spectral power calculated during periods of silence.

Neighborhood extraction from 2d-embedded spectrogram snippets

2N extraction was performed by dividing the spectrograms (within sound intervals) into spectrogram snippets of fixed width in the range of 12-16 columns (corresponding to snippet

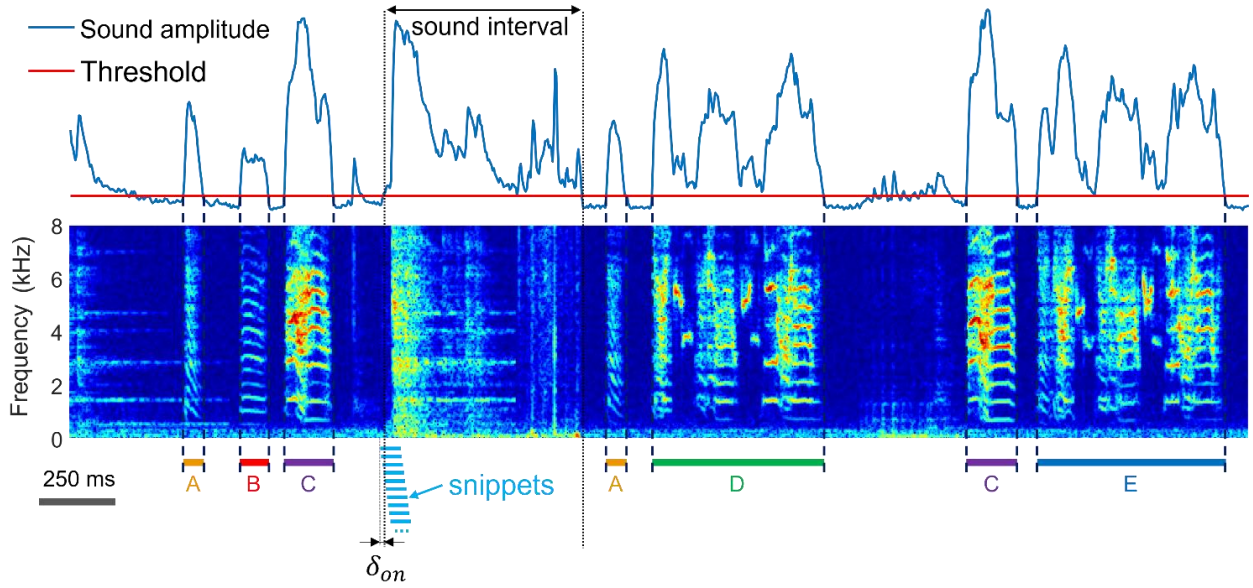


Figure 3.1: The task of extracting vocalizations is to correctly detect the onsets and offsets (dashed lines) of vocalizations and to determine their type (here A to E) amidst diverse non-vocal sounds (noise). Shown is a time-frequency log spectrogram of adult zebra finch song. Our approach to extraction of vocalizations is to first extract sound intervals (based on threshold crossings of sound amplitude) and to dissect these into 64-ms long snippets (light blue bars) with 60 ms overlap among adjacent snippets. To achieve robustness to segmentation errors, we consider each snippet a potential onset or offset of a vocalization. The first snippet associated with an amplitude threshold crossing precedes the threshold crossing by a small margin δ_{on} (free parameter). The shown spectrogram was produced by concatenating two recording segments (at the second black dotted line), chosen to illustrate the diversity of vocalizations and noises.

durations of 48-64 ms). The snippet duration is a variable that a user can adjust; in general, snippets should be large enough to yield robust separation of different vocalization types in the embedding plane and small enough to be able to cleanly slice a vocalization without cutting into adjacent vocalizations. The hop size between adjacent segments was given by one column (i.e., 4 ms), Figure 3.1. To the first data snippet associated with a sound interval, we ascribed a time lag $d = 1$. The onset time δ_{on} of that snippet precedes the sound interval onset by $\delta_{on} = 4$ spectrogram columns (i.e., 16 ms, Figure 3.1). Similarly, to the last snippet of a sound interval we associated a time lag $d = -1$; the offset time of that snippet exceeds the sound interval offset by $\delta_{off} = 6$ (i.e., 24 ms). These choices ensured that brief silent gaps before and after syllables were included in the defining characteristics of a vocalization. By definition, the second snippet of a sound interval had a time lag of $d = 2$ (i.e., 4 ms) and the second-last snippet had a time lag of $d = -2$ (i.e., -4 ms), etc. This dissection of the data into snippets resulted in a total of 581k snippets for the day-long recording of the bird shown in Figure 3.1 (given the chosen snippet duration of 64 ms).

We then embedded the snippets into the plane using UMAP (McInnes et al., 2018) and visualized high-density neighborhoods associated with a given time lag d, as follows: We drew a small disk of radius r around every point (Figure 3.2A, red dots) in the embedding plane that had a time lag

of either d or $d+1$ if d was positive (onsets), and either d or $d-1$ if d was negative (offsets). Thresholding the number of overlapping disks per pixel with a density threshold ϑ revealed supra-threshold (high-density) regions of points, which we refer to as neighborhood-defining blobs (Figure 3.2A, outlined in blue). The blobs for $d=1$ are shown in Figure 3.2B.

By changing the lag value d, blobs were interactively moved to unique regions along the 1d manifold of a vocalization type where no points from other vocalizations could be found (the latter we verified by visually inspecting spectrograms associated with a given neighborhood using our GUI). The points falling into a selected onset-related blob (Figure 3.2A, filled symbols) were then associated with that vocalization type. The vocalization onset times were given by the timestamp of the chosen snippets (i.e., the earliest snippet in case several adjacent points were found) minus the chosen time lag d (the smaller lag in case blobs were defined by two time slices). The time differences between the extracted vocalization onsets and the start times of the underlying sound intervals are shown as a cumulative density in Figure 3.3A. The analogous density for vocalization offsets relative to sound-interval endings is shown in Figure 3.3B and Figure 3.3C,D show the population averages.

2N extraction algorithm

In the following, we describe the detailed extraction procedure of vocalizations, starting with their onsets. Extraction is parameterized by three variables: a time lag d, a radius r, and a density threshold ϑ. The 2N extraction algorithm consists of the following steps:

1. Define an integer time lag $d>0$ as small as possible (start with $d=1$).
2. Identify all points in the embedding plane associated with this lag d.
3. Replace each identified point with a disc of radius r and sum up these discs, yielding a 2d density.
4. Identify the regions where the 2d-density exceeds a threshold ϑ. These regions we refer to as blobs, they are the defining characteristics of vocalization onsets.
5. Change $d, r,$ and ϑ to place and shape the blob such that it defines a uniquely characteristic region of the vocalization of interest. Ideally, choose d close to zero such that the blob is close to the onset.
6. Repeat steps 1-5 up to K times to define diverse onset blobs for a given vocalization type (typically, $K=1$ because there is a unique blob for each vocalization type).
7. Identify all points inside the K blobs. These points uniquely define the onsets of the extracted vocalizations given by their timestamps minus the anchoring time lag d_i of the underlying blob ($i=1,\dots,K$).

In this procedure, the optimal choices of radius r and threshold ϑ depend on the local density of points and should be individually chosen for each vocalization type and each blob. Essentially, the radius should be chosen as large as possible to not miss any onsets and likewise the threshold should be as low as possible to make the blob as large as possible to maximize the number of points harvested. However, a blob should not extend into points associated with confounding vocalizations, so a bit of manual fine-tuning is required for each vocalization type.

When all onsets of a vocalization type are defined, we perform the analogous definition of offsets; the lags of offset-defining blobs satisfy $d<0$; the parameters $d, r, \vartheta,$ and K of offset blobs needed

similar fine-tuning in practice. Once both onset and offset blobs are defined for all vocalizations, we extract the vocalizations as the spectrogram regions delimited by an onset and a subsequent offset. Cases in which an onset was followed by an offset of another type were discarded. Also discarded were onsets following an onset with missing intervening offset (and vice versa, discarded were offsets following an offset with missing intervening onset). In the three birds analyzed, we extracted 5, 7, and 9 vocalization types per bird, respectively. We then visually inspected the extracted spectrograms and manually corrected segmentation and clustering errors. That is, onset and offset times were manually adjusted by an expert in 4-ms steps to the nearest true onset or offset; and, misclassified vocalizations were assigned to the correct vocalization type or to noise (we observed neither, see Section 2.3).

Practical notes In all adult zebra finches examined, we managed for each vocalization type to find distinct onset and offset blobs. We carefully selected each blob as close as possible to its extreme position, i.e., as close as possible to the onset resp. offset of the associated sound interval. In nearly all birds, we found at least one point cloud in the embedding plane corresponding to non-vocal noise; in this cloud it was impossible to select both onset and offset blobs, presumably because noise tends to be unstructured, i.e., for noise there were no distinct time lags to amplitude threshold crossings at which noise snippets appeared more similar with each other than with snippets at other lags. We therefore extracted noise segments as the time intervals from the extracted onset until the ending of the underlying sound interval or the next vocalization onset, whichever came first.

Evaluation measures

In all birds, all 2N-extracted vocalizations were correctly classified into their types, as revealed by visual inspection of spectrograms. However, we observed occasional segmentation errors, where syllable onsets or offsets slightly deviated from the assessment by an expert. To calculate the precision of 2N extraction, an expert determined for all 2N-extracted vocalizations of a given type the fraction f of correctly classified 4-ms time bins.

The goodness of the implicit segmentation obtained with 2N extraction was evaluated by comparing the onset times of the extracted sound intervals to the corresponding onset times of the manually curated vocalizations, shown as a cumulative density in Figure 3.3A. The same procedure was performed for offsets, shown in Figure 3.3B. Averages of onset and offset cumulative densities are shown in Figure 3.3C,D, the less deviation between 2N-extracted and human-annotated segments, the better the segmentation performance.

To quantify the benefits of defining vocalizations from both ends instead of just one, we repeated the same calculation for 1N-extracted vocalizations that were defined by considering only the onsets blobs or only the offset blobs rather than both. In the onset-anchored 1N baseline, we extracted a vocalization from the time difference $\Delta t = t_a - t_b$, where t_b is the timestamp of a point in the onset blob minus the selected time lag d and t_a is the end of the underlying sound interval or the timestamp of the next point in an onset blob, whichever came first. Analogously, in the 1N-offset baseline, we extracted a vocalization from the timestamp of a given point in the offset blob minus the selected (negative) time lag d, backwards, until the previous point in an offset blob or

the beginning of the underlying sound interval, whichever came first (going backwards in time). The corresponding extraction errors $\epsilon = 1 - f$ are shown in Figure 3.4 for diverse vocalization types and birds.

Results

The first step of our extraction method is to detect sound intervals (as opposed to intervals of silence). In line with other approaches (SAP http://soundanalysispro.com, Avisoft http://www.avisoft.com), we assume that very often (but not necessarily always), a vocalization onset corresponds to a lower (low-to-high) threshold crossing of sound amplitude, and an offset

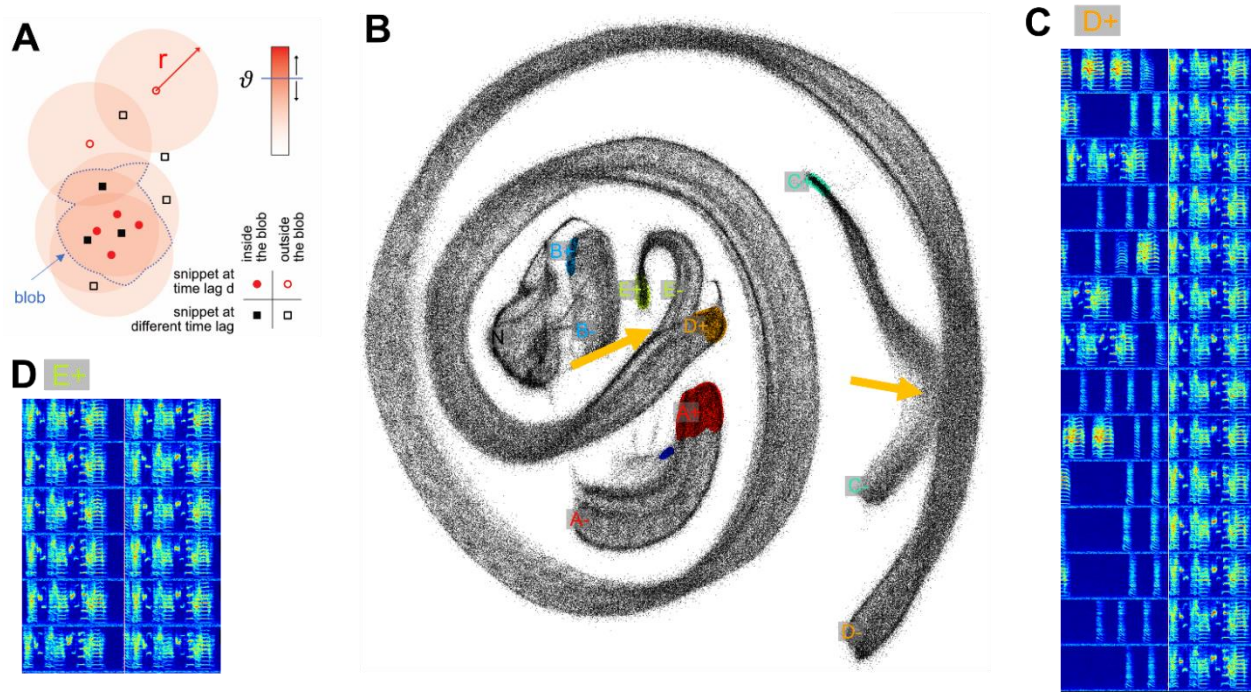


Figure 3.2: Two-neighborhood (2N) extraction of vocalizations from dense spectrogram embeddings. A) *Schematic illustrating the definition of a blob. After projecting all data snippets into the plane using UMAP, we replace the points corresponding to upward threshold crossings ($d = 1$, red circles, first light blue bar in Figure 3.1) by large, filled disks (light red) of radius r that we sum up. All pixels at which the sum exceeds a given threshold ϑ (blue horizontal bar) are grouped into a blob (delimited by blue dashed line). All points that fall into this blob correspond to extracted vocalization onsets, including points that were dissected at a lag different from d (black filled squares).* ***B)*** *Projected data snippets (black dots) from a one-day long recording. The onset blobs corresponding to the time slice $d = 1$ are shown in color (different colors for different vocalization types). The yellow arrows point to indistinguishable snippet embeddings stemming from different vocalization types. The letters A-E indicate manually chosen onset slices '+' and offset slices '-' for each vocalization type (same lettering as in Figure 3.1). The cluster labelled 'N' (gray) is a noise cluster without distinct onset and offset behavior, this cluster was ignored. The small blue blob next to the A+ blob is an onset variant of the introductory note, which can either be included in the definition of A+ (introductory note) or excluded.* ***C), D)*** *Spectrograms of example syllables taken from blob D+ (C) and blob E+ (D). Same bird as in Figure 3.1.*

corresponds to an upper (high-to-low) threshold crossing. In other words, we assume that either before or after a vocalization there are brief periods of silence, implying that sound intervals often begin and end with vocalizations. We then densely dissect the sound spectrograms associated with sound intervals into overlapping snippets (Figure 3.1). By considering each snippet as a potential vocalization onset or offset, the vocal segmentation problem remains unresolved at this processing step (it will be resolved at a later step).

We then project all spectrogram snippets into a plane, similar to the continuous UMAP embeddings in Sainburg et al. (2020). However, in our case, the first snippet of a sound intervals protrudes the sound-interval onset by δ_{on} and the last snippet protrudes the offset by δ_{off} (Figure 3.1), which makes vocalizations appear as distinct 1D structures in the embedding plane with clear boundaries rather than as a single excessively long 1d structure as in Sainburg et al. (2020). Provided that for a given vocalization type, there are sufficiently many vocal renditions that are precisely segmented by sound amplitude, the corresponding snippet embeddings will lie close to each other in the embedding plane and form a dense cloud of points. In general, we expect to find an extended cloud of points in the embedding plane for each vocalization type that ranges from self-similar onset snippets on one end to self-similar offset snippets on the other.

Among the cloud of onset-related points, we will also find points that display a nonzero lag to the nearest threshold crossing of sound amplitude. Namely, due to noise, we expect that some vocalizations are not cleanly segmented by sound amplitude but that instead are preceded or followed by a suprathreshold noise, which means that the time lag to an amplitude threshold crossing can be arbitrarily large. Such noisy vocalizations can nevertheless be correctly extracted with our method because for extraction we rely not on prior segmentation but on similarity with cleanly segmented renditions.

In short, to extract vocalizations of a given type, we define in the embedding plane a dense region of points (a blob) associated with the onsets of this vocalization type, and a blob associated with the offsets of that type. Blobs are defined by points of a given lag value *d*, see Figure 3.2A ($d = 1$ for onsets and $d = -1$ for offsets). To minimize effects of embedding distortions and to disambiguate confounding vocalization types, we set a blob's lag variable to a suitable value. We set the size of a blob as a function of the local density of points by adjusting two parameters: a radius r and a threshold ϑ. The radius r sets the size of the disks that are placed at the locations of the embedding points and the threshold ϑ sets the height that the summed disks must exceed for a pixel to be included in a blob (Figure 3.2A), for details, see Methods Section 2.2.1.

The extracted vocalizations are then defined as the spectrogram chunks that start at the timestamp of a point in an onset blob minus the blob's lag value d and end at the timestamp of the first subsequent point within an offset blob minus the blob's lag value, i.e., we extract vocalizations simply within the shortest (lag-corrected) time intervals between pairs of points in an onset and an offset blob. Importantly, we harvest all points within a blob, including points that were sliced at different time lags than the blob defining lag. Therefore, although the definition of vocalizations depends on the time lags to amplitude threshold crossings, the harvesting is

oblivious of these lags, and so our method can correctly extract vocalizations that are not cleanly segmented by sound amplitude.

When during this extraction process, an onset of a given type is followed not by an offset of the same type but by another type of event such as either an onset, the end of the sound interval, or the end of the file, we simply extract a vocalization from the onset until one time bin before the said event. Alternatively, when the onset is followed by an offset of another type, we extract no vocalization (to maximize precision[10]).

We illustrate our extraction method on a one-day-long recording of an isolated male zebra finch. We sliced the recorded sounds intervals into more than half a million snippets of 64 ms duration each that we embedded into the (2d) plane using UMAP. As can be seen in Figure 3.2B, snippets from different syllables can appear indistinguishable in the embedding plane, either 1) because a bird repeats an indistinguishable sub-syllable or note in a different context (i.e., as part of a different syllable as illustrated in Figure 3.2: parts of syllables E and D), or 2) because of UMAP projection errors (i.e., when nearest neighbors are hallucinated (Kollmorgen et al., 2020) — the latter we visually found to be quite common). When such non-discriminability occurs at either an onset or an offset, the respective snippet loses its distinguishing characteristic for that syllable (Figure 3.2B, yellow arrows). This situation is quite common in zebra finches that tend to sing different syllable types with indistinguishable endings. This ambiguity implies that the spectrogram snippets near an upper threshold crossing do not uniquely define the ending of that syllable type (presumably the same is also true for some syllable onsets). As a workaround to such repetitive structure inherent in birdsong (and language for that sake), our method provides the freedom to define syllable endings and beginnings at fixed time lags away from threshold crossings, at places within a syllable where the defining snippet becomes unique for that syllable.

Using our GUI (Supplementary Material), users can increase and decrease the lag variable d to observe the blobs move around in the embedding plane until they reach a region in the plane where there are no confounding points from other vocalization types. Such confounding points can be recognized thanks to the elongated 1d-structure of vocalizations in the embedding plane (Sainburg et al., 2020): the confounding points are the ones where two different 1d-structures come too close to each other (see Figure 3.2B, yellow arrows). For example, the bird in Figure 3.2B produced two very long and complex song syllable types in rapid succession, whereby the second type (E) displayed an additional small down sweep at the syllable beginning, making it clearly distinct from Syllable D only by virtue of this down sweep (Figure 3.2C, D). As a result, the endings of syllables D and E in the embedding plane coincided with each other, which is why we had to define the offset-anchored blob E- not far from the onset-anchored blob E+ in a region where it was distinct from any neighborhood of D-, to make sure the endings of syllable D are not confounded with parts of syllable E.

[10]If one wants to optimize also recall, one could extract two different types of vocalizations, one starting with the onset and ending after the median duration of that type, and the other ending with the offset and starting the median duration before (but that would correspond to 1N extraction, which is not the aim of this work).

For the bird shown in Figure 3.1 and 4.2B, about 96.6% of extracted vocalizations had cleanly segmented onsets and about 98.7% had cleanly segmented offsets (i.e. onsets and offsets coincided with sound amplitude threshold crossings). In another bird, also recorded with a microphone and kept alone in a soundproof box, clean segmentation was even more frequent (99.9% for onsets and 99.0% for offsets, respectively). However, in birds housed in pairs and recorded with a wireless accelerometer mounted to their back, the fraction of cleanly segmented vocalizations was much lower (down to 72% for onsets and 89% for offsets, respectively). Thus, the level of noise depends strongly on the recording method, but our method allows harvesting vocalizations even in noisy situations.

Performance evaluation

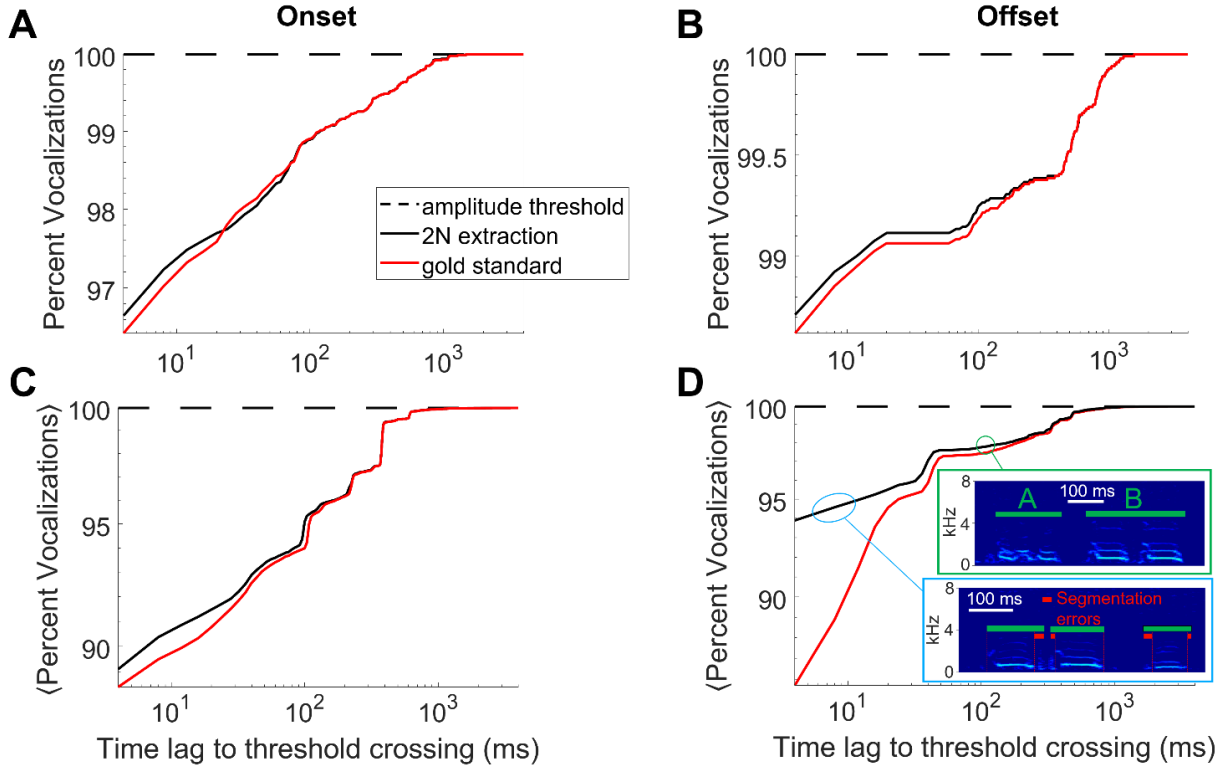


Figure 3.3: 2N-extracted vocalizations (black curve) are similarly segmented as human-extracted vocalizations (red). A) *Shown is the cumulative percentage of vocalizations with an onset that falls within a given time lag (x-axis) following a sound amplitude threshold crossing.* ***B)*** *Same for offsets that are within a given time lag preceding a sound-amplitude threshold crossing. Same bird as in Figure 3.1.* ***C,D)*** *Cumulative percentage averaged across n=4 birds. The insets in D show typical errors made by 2N extraction (green bars), which is to interpret a string of two calls as a single call (here labelled 'B', green bounding box) or to introduce segmentation errors (red bars) from too generous inclusion of surrounding noises (blue bounding box), data taken from an accelerometer-recorded bird. (A-D) Segments extracted by amplitude threshold crossings (dashed) trivially display no time lag whatsoever.*

How good are the extracted vocalizations? We computed two performance measures associated with the extraction procedure: 1) The *quality of the segmentation* in terms of the time differences of extracted onsets and offsets relative to gold standard human annotations; and 2) the *clustering*

performance in terms of the false-positive error rate of misclassified time bins, again assessed by human experts (see Methods).

With regards to 1) the quality of the segmentation, for the bird shown in Figures 1 and 2B and for both onsets and offsets, the time lags to threshold crossings were similarly distributed for 2N- and for human-extracted vocalizations (Figure 3.3A, B), suggesting that 2N extraction extracts vocalizations from background noise in similar manners as humans do. For both onsets and offsets in this bird, the largest time lags to amplitude-threshold crossings were up to 1 s long. In all birds tested (n=2 mic and n=2 accelerometer birds), we found similarity between 2N-extracted vocal segments and the human gold-standard counterparts, constituting a big improvement over simple sound amplitude thresholding (i.e., defining onsets and offsets as lower and upper crossings of sound amplitude thresholds, respectively, Figure 3.3C, D).

Some disagreements were seen when data was noisy; namely, we observed a tendency in human evaluators to segment the offsets earlier, which we found was often due to double misses that occurred in strings of calls where both an offset and the following onset were missed, leading to the hallucination of a much longer call than there actually was (Figure 3.3D). We do not evaluate workarounds for such problems but propose to fix them by detecting for each vocalization type the outlier renditions of excessively long durations and by discarding these. The shorter segmentation errors within 10-20 ms of true syllable offsets were almost always caused by inclusion of respiratory or movement artifacts near syllable boundaries (Figure 3.3D) and very rarely were they caused by truncations of parts of syllables, which would be more detrimental for subsequent feature-based syllable analysis. In summary, our method provides improved vocal segmentation compared to simple sound amplitude thresholding, in particular when recordings are noisy as in pair-housed birds recorded with animal-borne sensors.

With regards to 2) the clustering performance, an expert evaluated the precision of 2N extraction in terms of the fraction of 4-ms time bins that were assigned to the correct vocalization type. We were particularly interested in comparing our findings to a baseline of extracting vocalizations not from two defining (sets of) regions in the embedding plane, but from only a single region, either anchored to the onset or the offset, but not both. In these 1N extraction baselines, we extracted vocalizations from a point in a blob until either the next point in the blob or until the end of the sound interval, whichever came first (see Methods).

We found that 2N extraction outperformed 1N extractions by a large margin, achieving 3-6 times fewer extraction errors (Figure 3.4). The superior precision of 2N extraction came only at a minimal cost of lower recall. Namely, for the bird shown in Figure 3.1, 2N extraction retrieved almost as many time bins as did 1N extraction, namely 99.6%. On average (n=4 birds), the fraction of time bins retrieved with 2N extraction was 96.9% relative to the mean number of bins retrieved with 1N extractions (averages across onset- and offset-based 1N extraction methods). Thus, the added benefit of much lower extraction error came only at a minimal cost of potentially retrieving fewer vocalizations. Thus, in terms of extraction performance, it pays off to extract vocal units in terms of two sets of defining characteristics, one anchored to the onset and the other to the offset. In terms of manual processing time, 2N extraction comes at the obvious cost of twice

the workload compared to 1N extraction. However, given that 2N extraction can be routinely done within less than five minutes for an experienced user, irrespective of the size of the data set, this overhead seems negligible in practice.

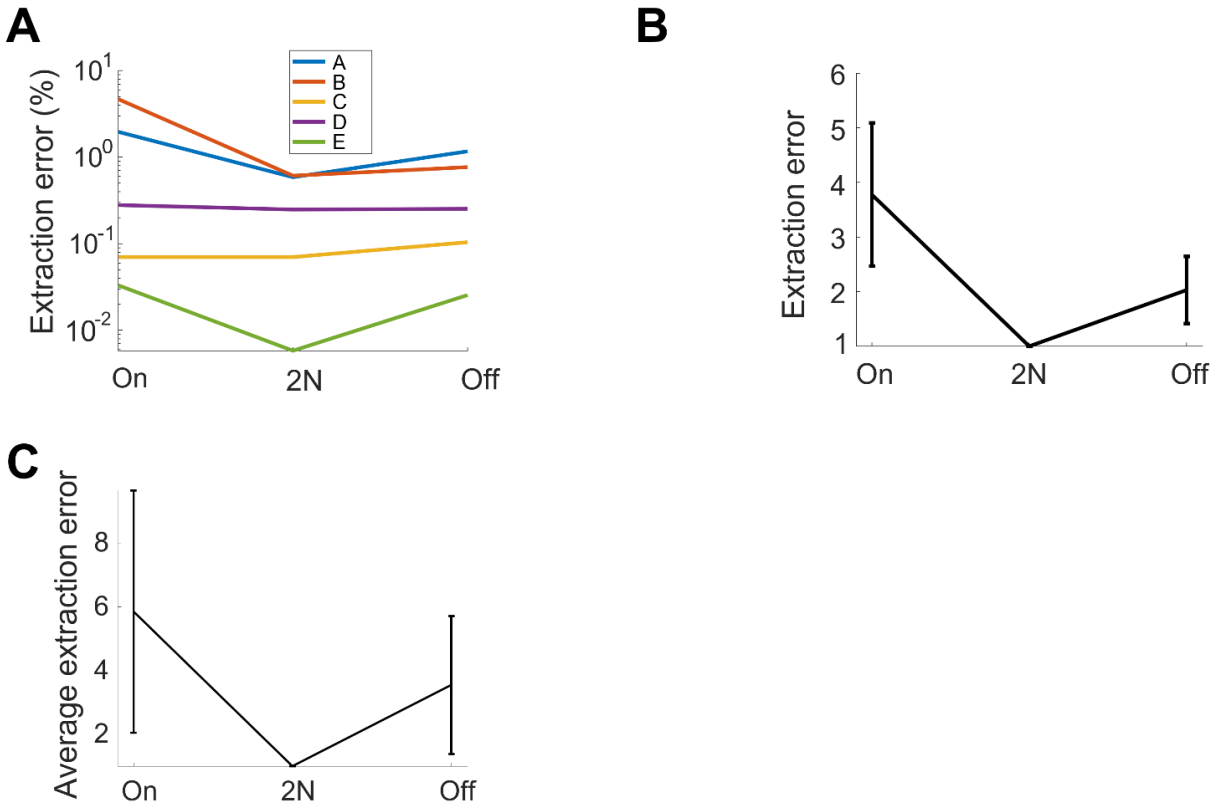


Figure 4: 2N-extracted vocalizations achieve higher precision than their 1N-extracted counterparts*. The fraction of extracted (4-ms) time bins that are misclassified (extraction errors) is shown for diverse methods: 2N extraction and either onset-anchored (on) or offset-anchored (off) 1N extraction.* ***A)*** *For each vocalization type, the error is lower when vocalizations are extracted from two neighborhoods than when extracted from one neighborhood. Same bird as in Figure 3.1 and 3.2.* ***B)*** *Same data, averaged over all syllable types and after normalizing by the error rate of 2N extraction (shown is average ± std across vocalization types). The retrieval error of 1N extraction is 2-4 times higher than that of 2N extraction.* ***C)*** *Normalized relative extraction error (average ± std across 4 birds). 2N extraction achieves 3-6 times fewer errors.*

Discussion

We presented a simple and intuitive method for extracting arbitrary vocal units in embeddings of a continuous stream of data. Embedding methods such as UMAP and t-SNE have been criticized for the distortions they can create, especially in genomic data (Chari et al., 2021)(Chari et al., 2021)(Chari et al., 2021). While we find similar distortions in vocal data, our workaround is to flexibly define vocal units based on regions in the embedding plane that are far from ambiguities and presumably also from distortions.

Our key contribution is to identify vocalizations via two sets of characteristics near the onsets and offsets, rather than through a single set of characteristics tied to either the onset, the offset, or even the entire vocalization. The benefits of this dual recognition are better segmentation and higher clustering performance, because 2N extraction is designed to suppress errors resulting from vocal ambiguities, cage noises, and embedding distortions. The method works best when onset and offset defining blobs are sufficiently far apart such that there is no overlap between them.

2N extraction is flexible and can be tailored to meet specific user requirements such as correctly detecting syllable variants even when they are composed of sub-syllables forming a small unwanted silent gap, which birds sometimes produce. Short gaps can be ignored for example by smoothing sound amplitudes before computing sound intervals (which our GUI allows). With smoothed amplitudes, split syllables are robustly extracted when the defining neighborhoods are tied to the stable sub-syllable parts (rather than the gap).

2N extraction makes most sense on large data sets because there is only a time penalty for computing the embedding but virtually no overhead for defining the onset and offset blobs. We routinely calculated UMAP embeddings of up to 1 million sound snippets using standard desktop PCs with 32 GB of RAM. Since our method is not tied to a particular embedding method, we expect it to work also on other planar embedding types (we obtained similar results on t-SNE (Maaten and Hinton, 2008) embedded data).

2N extraction is a very flexible annotation method, providing 4+4 degrees of freedom for each vocalization type: the radius, the threshold, the lag, and the number of blobs (for each onset and offset). As noted, the defining regions of a vocalization type need not constitute a connected set. For example, we could have chosen to combine syllables D and E in Figure 3.1 into a single syllable type by defining its onset characteristic in terms of the two blobs labeled D+ and E+ in Figure 3.2. Thus, our GUI provides the user with high flexibility of defining vocal units, which minimizes the need for postprocessing including the correction of segmentation errors.

Our human-centered workflow is in line with other semi-supervised methods in bioinformatics (Wrede and Hellander, 2019) that focus on reducing knowledge-requiring and time-consuming algorithmic optimization. For the problem of extracting vocalizations, we see it as an advantage that users can resolve ambiguous situations by making an informed decision after exploring the full vocal repertoire (such as deciding whether some vocalizations belong to the same type or not

as in Figure 3.1, syllables D and E). Such fine decisions are part of critical data assessment (Thomas and Cook, 2006; Cui, 2019) and are common when working with animal data.

Currently, our method requires a pre-segmentation into sound intervals, as otherwise we do not obtain the time lag d needed for defining blobs and for extracting vocal units. In data that is so noisy that there are barely any vocalizations that are cleanly segmented by sound amplitude, our method is not trivially applicable. We would recommend trying to use another sound feature than sound amplitude to obtain blobs as in in Figure 3.2A: Provided the feature identically dissects a significant number of vocalizations, high-density regions of dots should emerge in the embedding plane, which would make our method applicable.

We imagine that our approach to extraction of vocalizations can generalize to biological and physical processes other than vocalizations. Namely, we believe that our approach will work well for the extraction of units in natural processes that contain rigid elements that sequentially unfold in variable sequences and at variable speeds. The duration of entities of interest should be typically longer than the snippet size. When overlapping data snippets from such processes are projected onto the plane, elongated structures will result, ideally displaying uniquely defining beginnings and endings as in Figure 3.2B. Our method might also work for spatial rather than temporal data, provided that the same requirement of repetitive sequential structures applies. We hope that our GUI can be of use to researchers wanting to adopt our methods for their work and as a basis for further developments.

Data availability statement

The GUI presented in this study together with instructions and tutorials on how to use it can be found in the ETH Research Collection https://doi.org/10.3929/ethz-b-000582761 or on https://gitlab.switch.ch/hahnloser-songbird/published-code/automtedclustering/2n-extraction. Further inquiries can be directed to the corresponding author.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contribution

RHRH, XH, TT, LR and CL contributed to the concept and design of the study. LR recorded data from backpack experiments. RHRH implemented the algorithm and initial GUI and performed the analysis. RHRH and CL created visualizations and wrote the manuscript. CL refined the GUI. XH, TT, and LR edited and provided feedback on early versions of the manuscript and the GUI.

Acknowledgement

We thank Anja Zai for performing additional song recording experiments and Heiko Hörster for his help thereof. All experimental procedures were approved by the Veterinary Office of the Canton of Zurich.

Funding

This study was partly funded by the Swiss National Science Foundation (Grant 31003A_182638; and the NCCR Evolving Language, Agreement No. 51NF40_180888) and by the China Scholarship Council (Grant No. 202006250099).

Anisimov, V. N., Herbst, J. A., Abramchuk, A. N., Latanov, A. V., Hahnloser, R. H. R., and Vyssotski, A. L. (2014). Reconstruction of vocal interactions in a group of small songbirds. *Nat. Methods* 11, 1135–1137. doi:10.1038/nmeth.3114.

Canopoli, A., Herbst, J. A., and Hahnloser, R. H. R. (2014). A higher sensory brain region is involved in reversing reinforcement-induced vocal changes in a songbird. *J. Neurosci.* 34, 7018–7026. doi:10.1523/JNEUROSCI.0266-14.2014.

Chari, T., Banerjee, J., and Pachter, L. (2021). The Specious Art of Single-Cell Genomics. *BioRxiv*. doi:10.1101/2021.08.25.457696.

Cohen, Y., Nicholson, D. A., Sanchioni, A., Mallaber, E. K., Skidanova, V., and Gardner, T. J. (2022). Automated annotation of birdsong with a neural network that segments spectrograms. *eLife* 11. doi:10.7554/eLife.63853.

Cui, W. (2019). Visual analytics: A comprehensive overview. *IEEE Access* 7, 81555–81573. doi:10.1109/ACCESS.2019.2923736.

Goffinet, J., Brudner, S., Mooney, R., and Pearson, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *eLife* 10. doi:10.7554/eLife.67855.

Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 4, 923–925. doi:10.1038/nmeth1113.

Kollmorgen, S., Hahnloser, R. H. R., and Mante, V. (2020). Nearest neighbours reveal fast and slow components of motor learning. *Nature* 577, 526–530. doi:10.1038/s41586-019-1892-x.

Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 2579–2605. https://www.jmlr.org/papers/v9/vandermaaten08a.html

Markowitz, J. E., Ivie, E., Kligler, L., and Gardner, T. J. (2013). Long-range order in canary song. *PLoS Comput. Biol.* 9, e1003052. doi:10.1371/journal.pcbi.1003052.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*. doi:10.48550/arxiv.1802.03426.

Nicholson, D. (2016). Comparison of machine learning methods applied to birdsong element classification. in *Proceedings of the 15th Python in Science Conference* Proceedings of the python in science conference. (SciPy), 57–61. doi:10.25080/Majora-629e541a-008.

Peikari, M., Salama, S., Nofech-Mozes, S., and Martel, A. L. (2018). A Cluster-then-label Semi-supervised Learning Approach for Pathology Image Classification. *Sci. Rep.* 8, 7193. doi:10.1038/s41598-018-24876-0.

Sainburg, T., Theilman, B., Thielk, M., and Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. *Nat. Commun.* 10, 3636. doi:10.1038/s41467-019-11605-y.

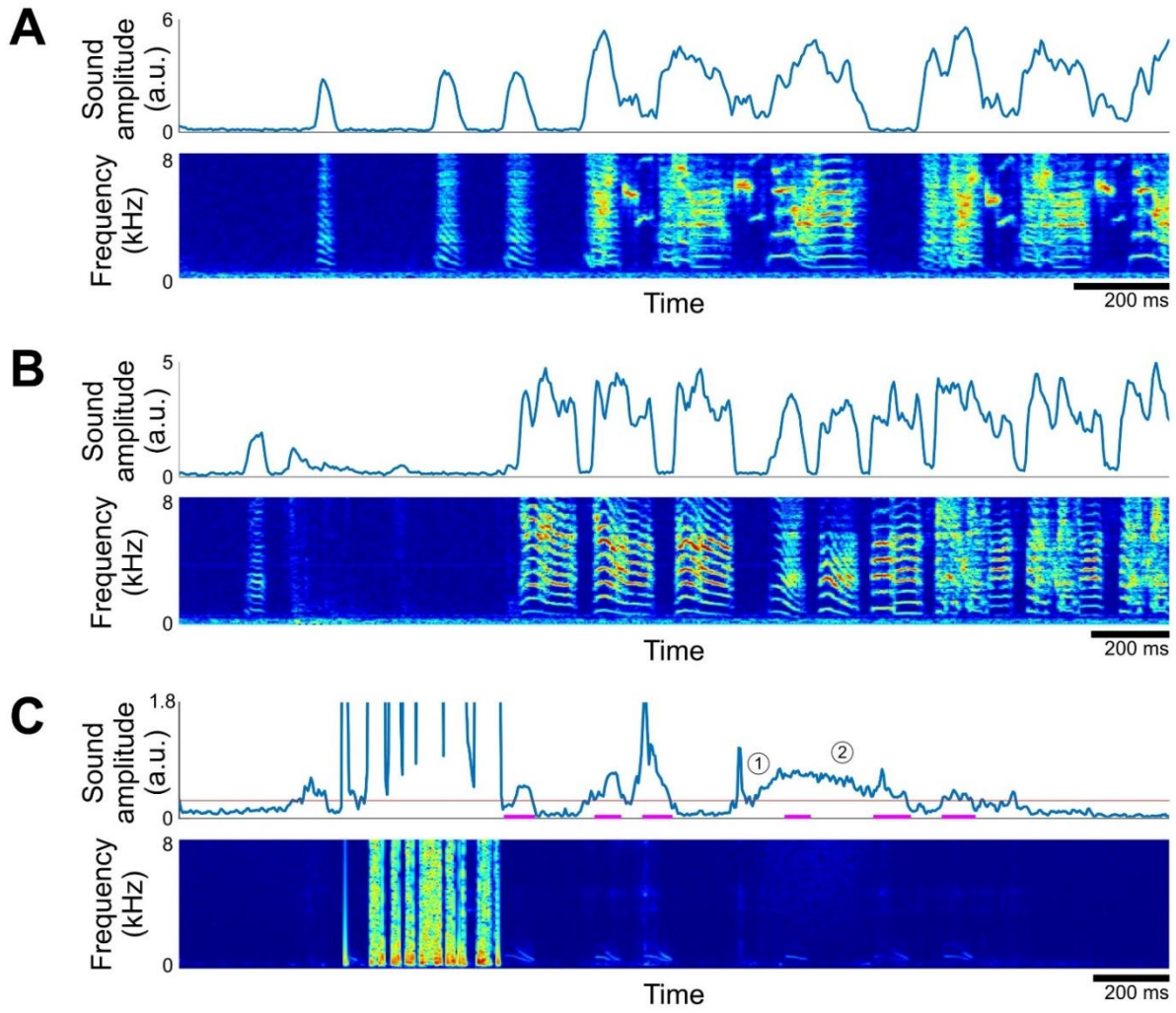
Sainburg, T., Thielk, M., and Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Comput. Biol.* 16, e1008228. doi:10.1371/journal.pcbi.1008228.

Steinfath, E., Palacios-Muñoz, A., Rottschäfer, J. R., Yuezak, D., and Clemens, J. (2021). Fast and accurate annotation of acoustic signals with deep neural networks. *eLife* 10. doi:10.7554/eLife.68837.

Tachibana, R. O., Oosugi, N., and Okanoya, K. (2014). Semi-automatic classification of birdsong elements using a linear support vector machine. *PLoS ONE* 9, e92584. doi:10.1371/journal.pone.0092584.

Thomas, J. J., and Cook, K. A. (2006). A visual analytics agenda. *IEEE Comput. Graph. Appl.* 26, 10–13. doi:10.1109/mcg.2006.5.

Wrede, F., and Hellander, A. (2019). Smart computational exploration of stochastic gene regulatory network models using human-in-the-loop semi-supervised learning. *Bioinformatics* 35, 5199–5206. doi:10.1093/bioinformatics/btz420.



Supplementary Figure 3.1: Illustration of data sets. *Shown are example sound amplitudes and spectrograms of bird vocalizations. Signals were acquired with microphones from single-housed birds (A, B) or with wireless accelerometers from a pair of birds (C). In (C), soft calls (highlighted in magenta) are poorly segmented by a simple amplitude threshold (red line): co-occurring noises can cause premature onset detection (1) or push the vocal signal above threshold during pauses so that multiple vocalizations are extracted as one (2).*

Conclusion and outlook

In this thesis, we set out to bring high-yield electrophysiology to birdsong research (Chapter 1) with the goal to understand neural processes that feature intrinsically motivated singing (Chapter 1 & Chapter 2) and vocal adaptation (Chapter 2). Complementing this approach, we introduce a human-in-the-loop strategy specifically designed to streamline the vocal annotation process in extensive vocal recordings (Chapter 3). In the following, we summarize the main findings, highlight their relevance, outline limitations, and engage in a discussion regarding future steps that we envision at this stage.

Large-scale electrophysiology in freely moving songbirds

Ever since the first intracortical recording of action potentials (Renshaw et al., 1940), much insight has been gained into how individual neurons enable a faithful interpretation of the world and coordinate bodily actions. In the last decades then, we have seen neural recording technologies advance and reach unprecedented granularity and scale (Hong & Lieber, 2019). Thanks to advances in micro- and nano-circuit technology, the capacity for recording sites per penetrating shank dramatically increased enabling the simultaneous recording of significantly larger numbers of neurons (Hong & Lieber, 2019; Steinmetz et al., 2018). Fast forward to the present day, studies such as those by Steinmetz et al. (2021) and Zhao et al. (2023) show the capability of capturing the simultaneous activity of over a thousand neurons, offering a much more comprehensive understanding of the neural processes in the brain that enable complex behaviors.

The advantages of these technological advancements are twofold. First, they allow for the acceleration and scaling of neuronal recordings, which traditionally has been an expensive and time-consuming endeavor. Second, by capturing the activity of numerous neurons at once, researchers can identify patterns and features that might not be apparent when observing individual neurons in isolation. For instance, this holistic view of neural activity has provided a new understanding of how populations rather than individual neurons encode complex behaviors within and distributed across brain areas (Fortunato et al., 2023; Kondapavulur et al., 2022; Lemke et al., 2019; Mante et al., 2013; Semedo et al., 2022; Steinmetz et al., 2019; Stringer et al., 2019).

However, the implementation of high-yield electrophysiological techniques in naturalistic settings presents its own set of challenges. The equipment required for scaled recordings can be cumbersome, often necessitating experiments to be conducted with subjects in a head-fixed position. The restricted mobility can be particularly limiting since it suppresses behaviors that are inherently motivated. For example, song production in zebra finches serve as an ideal model for the study of complex motor behaviors, owing to the ease of collecting extensive behavioral data and the accessibility of their underlying neural circuitry. Zebra finches have been shown to sing in a head-fixed setting, but this appears only possible by externally incentivizing them with the presentation of a female (Moll et al., 2023; Picardo et al., 2016) which contrasts the self-motivated nature of the behavior and, more importantly, is produced in a different brain state than undirected

singing (see Chapter 1, Kao et al., 2008a; Kao & Brainard, 2006; Stepanek & Doupe, 2010; Woolley et al., 2014a). In contrast, their natural singing behavior demands experimental designs that accommodate the bird's mobility and welfare (Yamahachi et al., 2020) clashing with the rigidity and weight of common large-scale neural recording tools.

To address these challenges, we adopted a large-scale recording approach and tailored it for the use in small songbirds. We developed a custom chronic implant device equipped with a state-of-the-art microelectrode array: the Neuropixels (Jun et al., 2017). This device stands out for its lightweight and its capability to target brain regions at more extreme angles of entry (Table 2.1 and Figures 2.3), offering improvements over previous designs (Juavinett et al., 2019; T. Z. Luo et al., 2020b; Steinmetz et al., 2021; Van Daal et al., 2021). Zebra finches adapt well to this implant, and resume singing few days after the surgery. Together, this enabled us to compile a comprehensive dataset of neural activity in song-relevant brain regions in the anterior forebrain during intrinsically motivated singing.

Comparing our approach for large-scale neural recordings in freely moving songbirds

Our approach introduces a novel aspect that we believe complements alternative techniques recently seen arising in birdsong research. For instance, calcium imaging is increasingly employed nowadays also in freely moving songbirds thanks to miniaturized microscopes (Alvarado et al., 2021a; Cohen et al., 2020; Daliparthi et al., 2019; Düring et al., 2020; Roberts et al., 2017; Scherrer et al., 2023). One of the key advantages of calcium imaging is that it captures high-yield neural activity in superficial brain areas and that it can be employed to target and study specific cell types (Alvarado et al., 2021a; Düring et al., 2020). However, this method faces challenges in accessing deeper brain structures without the use of penetrating lenses, requires viruses to deliver fluorescent indicators which take a significant lead time to express, and the signal kinetic is slower than the underlying dynamic of an action potentials (Grienberger et al., 2022).

In contrast, electrophysiological methods are readily usable and capable of reaching deep brain structures, enabling high-yield recordings with exceptional temporal resolution. The advantages have led to a surge in recent studies that have successfully utilized multielectrode arrays (MEAs) for chronic neural recordings in areas such as RA, HVC or in subthalamic regions (Arneodo et al., 2021; Das & Goldberg, 2021; Egger et al., 2020; Elmaleh et al., 2021). Notably, these studies employed passive MEAs with electrodes situated only at the tip of the shank, thereby missing the chance to sample activity along the shank and perform multi-areal recordings. In contrast, there have been instances of simultaneous cross-areal recordings in LMAN and RA, yet these were confined to single -electrode or tetrode setups, lacking the scalability and spatial coverage offered by MEAs (Darshan et al., 2017a; Tian et al., 2023). Our adoption of an active MEA (the Neuropixels probe) allowed the deployment of 384 electrode sites along the entire penetrating area of the shank, facilitating access to neural tissue with precise spatial resolution across both striatal and pallidal structures including two brain regions relevant for song learning and production.

Adaptability of our device to other songbirds and small animals

While our device has been tested for multi-areal recordings in the LMAN and Area X, its design promises wide-ranging adaptability for various experimental setups. The Neuropixels' long shank and the flexibility of our device to enter the cranium at angles different from that of the headstage (Figure 2.3), support a multitude of multi-areal recording strategies in the songbird brain, including combinations like HVC with RA, NIf, or potentially Area X. While our work concentrated on zebra finches, these experiments can also be performed in on other songbirds like Bengalese finches (*Lonchura striata domestica*) or canaries (*Serinus canaria domestica*), that are established subjects in neuroscience because they produce statistically amenable songs with variable sequences (Cohen et al., 2020; Koparkar et al., 2024; Veit et al., 2021).

Our device equally offers the opportunity to investigate unique vocal behaviors in other songbirds whose neural mechanisms are yet to be fully understood. For example, nightingales (*Luscinia megarhynchos*) can instantly mimic the pitch of a whistle they just heard, indicating a precise and seamless translation from the auditory perception to motor execution (Costalunga et al., 2023). With our device, it would be possible to record neural activity in auditory (e.g., Field L, Nif) and motor areas (e.g., HVC) simultaneously, facilitating the exploration of neural processes during pitch matching within and across areas. Similarly, territorial birds such as European robins (*Erithacus rubecula)*, offer valuable insights into context- and season-specific singing behaviors, from defending their territory and repelling intruders all-year round to attracting mates during courtship in breeding seasons (Lack, 1965; Young, 1955). Investigating whether neural changes in the AFP during territorial singing correspond with those during female-directed singing (see Chapter 1) could reveal functional nuances and regulatory mechanisms of song production. Such research, facilitated by the adaptability of our device for use in small, freely moving songbirds of any kind, could significantly enhance our understanding of vocal control mechanisms and enrich our knowledge of the neural foundations of communication and social interaction.

Overall, we believe that the wide-ranging utility of our device is not limited to small songbirds. While we have not directly tested its application, the device might also be suitable for larger songbirds, such as starlings or corvids. Potential challenges posed by their aversion to chronic implants could be mitigated by their capability to accommodate additional weight. Enhancing the device with supporting materials and protective layers, such as a protective cylinder, could improve the implant's durability and foster greater acceptance.

In addition to avian research, the flexibility of our device's design may also extend to applications in small mammals, including species such as mice, rats, or even small primates such as the mouse lemur (*Microcebus*). For instance, the lightweight nature of our device is an improvement over previous designs, potentially reducing the burden even on small animals like mice. Should our device not align with certain experimental requirements, we are confident that the distinct aspects, including the minimalistic design, the detachability of the headstage and probe or the selection of materials, could serve as a source of inspiration for researchers, encouraging them to adapt and innovate based on our work.

Challenges and decisions related to the preprocessing of our Neuropixels recordings

The extensive access to neural tissue enabled by Neuropixels recordings comes with the challenge of managing large datasets. A single hour of recording with 384 active electrodes accumulates ~60 GB of raw data and around ~30 GB during preprocessing. Because continuously recordings provide more stable spike sorting results and allow for the detection of mechanical drift, we decided to acquire neural data in session that usually lasted between one and six hours. With one or two sessions per day in eight birds, in different experimental conditions (undirected singing, female directed singing, naturally sleeping, exposed to playback), this led to the collection of over 350 hours of neural and behavioral data which altogether generated a massive and complex dataset of more than 30 TB.

Managing, storing, and analyzing such a vast amount of data necessitates efficient data processing tools and substantial computational resources. Fortunately, we were able to leverage existing solutions for most tasks. For instance, the detection and clustering of action potentials were facilitated by an automated spike sorting algorithm (Kilosort, Pachitariu et al., 2023). Nonetheless, manual curation of the spike sorting results and handling of individual files (e.g. cutting, concatenating, moving[11]) was necessary, and in particular because of the large size of the files, added substantial time needed during preprocessing.

We observed early on that recordings shorter than 30 minutes did not provide enough data to the Kilosort algorithm to produce reliable spike sorting outcomes. Therefore, we opted to record a minimum of one hour and only spike sorted sessions that lasted at least ~60 minutes and featured at least ~100 renditions of a bird's motif. This resulted in the spike sorting and manual curation of over 140 hours of recorded data from 91 sessions. Including the screening of all audio recordings for song time and the time spent developing scripts and graphical user interfaces (GUIs) to facilitate the data management, completing this step required multiple months.

Summary and availability of the recorded Neuropixels data

In line with our dedication to open science, we intend to publish and make the data set from Chapter 1, including the results from spike sorting and vocal analysis, publicly available (data publication in preparation). The dataset will comprise a consolidated version of the data combining all raw signals and all preprocessing results. In addition to the extensive recordings conducted during undirected singing, we anticipate that the availability of this data set will inspire further analysis of the data captured under various experimental conditions.

The chronic nature of our recordings facilitated the data collection not only during instances of free singing but also across multiple behavioral states, without necessitating additional effort. Our primary focus was on capturing free and undirected singing, yet we also introduced a female zebra finch occasionally to evoke social context modulation. Our analysis of the spiking activity under these varied conditions not only confirmed but also extended previous findings regarding

[11] Raw files from individual session had to be concatenated for the spike sorting and copied to the computer that ran the algorithm on a fast GPU. Applying the algorithm to data on the mapped network server would have extended the duration vastly because the algorithm loads batches of the data multiple times.

changes in spiking within Area X and LMAN in different social contexts (Figure 2.10, Figure 2.10 Suppl Fig 2) and their cross-areal communication (Figure 2.12).

Furthermore, to delve into the neural complexities of the songbird brain, particularly within the AFP, we performed recordings during natural sleep. During these recording sessions, we repeatedly observed the complete cessation of ongoing spiking activity in the raw signal of all electrode channels covering LMAN. Analyzing neural population activity, we could confirm that during sleep, LMAN spiking fluctuates between phases of synchronized firing and silence, and states of desynchronization, akin to the sleep modulation of cortical activity observed in mammals (Chauvette et al., 2010; Kobak et al., 2019; Steriade et al., 1993; Xu et al., 2019). Although this data is not elaborated upon within this thesis, it is worth noting that our findings reveal a significant and global sleep-related modulation of spiking and local field potential (LFP) signals, with these modulations being more marked in the pallium than in the striatum. A publication investigating this topic together with Janie Ondracek's group at the Technical University in Munich, the group of Arthur Leblois at the University of Bordeaux, and the group of Nicolas Giret at the University Paris-Saclay, together with the publication of a combined dataset, is in preparation (Appendix III).

Additionally, we exposed the birds in some sessions to auditory stimuli, such as playback of their own song or white noise bursts. The latter were triggered contingent on the pitch of a specified syllable with the goal of inducing pitch shifting in a subset of the subjects. Despite only one bird demonstrating consistent pitch shifting away from the normal pitch, we believe that this diverse data set holds an added value for other researchers interested in investigating questions related to auditory processing and reinforcement learning.

Evidence of vocal exploration and adaptation originating in the AFP

The songbird's AFP is a thalamo-pallio-striatal circuit necessary for vocal learning and adaptive control (Ali et al., 2013b; Andalman & Fee, 2009b; S. W. Bottjer et al., 1984; Scharff & Nottebohm, 1991; Sohrabji et al., 1990; Tachibana et al., 2022; Zai et al., 2020). Using our custom-designed Neuropixels recording device, we were the first to record and report simultaneous activity in two brain areas part of this circuit: the cortical-like premotor LMAN and Area X, which is part of the basal ganglia (Chapter 1). We were able to replicate findings from previous intra-areal studies, including increased spiking activity during singing (Figure 2.9), the identification of different cell types from song-related spiking patterns in Area X (Figure 2.10), and the social modulation observed in these neurons when the otherwise isolated bird sings towards a female (Figure 2.10 Suppl. Figure 2)(Goldberg et al., 2010a; Goldberg & Fee, 2010a; Hessler & Doupe, 1999c; Kao et al., 2008b; Woolley et al., 2014b).

The results from our cross-areal analyses revealed a lack of significant patterns in functional neural connectivity between neuron populations in LMAN and Area X during singing when using CCA, with one notable exception. One explanation for this inconclusive finding is that signal transmission within the AFP of mature zebra finches primarily involves stochastic fluctuations that affect vocal output without introducing bias. This observation is consistent with studies in rodents showing that neural connections between striatal and cortical areas strengthen with learning (Kondapavulur et al., 2022; Lemke et al., 2019), implying that the AFP's neural coordination

between LMAN and Area X might intensify during vocal learning or adaptation but diminishes in the absence thereof.

The hypothesis is in line with proposed frameworks of reinforcement learning theory in the songbird literature where synaptic connectivity (and therefore information transmission) is strengthened upon rewarded actions (Fee & Goldberg, 2011; Woolley & Kao, 2015). Reward signals in Area X have been reported in the form of transient increases in DA release by neurons projecting to Area X from the ventral tegmental area (VTA) (Gadagkar et al., 2016). Dopaminergic modulation in Area X is also a necessary and sufficient driver of pitch shifting (Hisey et al., 2018; Hoffmann et al., 2016; Xiao et al., 2018) and has been shown to facilitate synaptic plasticity in slice preparations (Ding & Perkel, 2004).

Motor exploration or variability is another key component of reinforcement learning theory. The AFP, and notably LMAN as its output, is a source of vocal variability (Hampton et al., 2009; Kao & Brainard, 2006; Thompson et al., 2011). Our analysis on different sources of vocal variability in Chapter 2, reveals that LMAN's contribution to pitch variability aligns with a random noise source with independent and identically distributed (iid) properties. Additionally, we discovered that LMAN neural spiking can, to some extent, predict pitch variations, with about 12% of LMAN units demonstrating predictive capability. Despite the modest and limited explanatory power of these correlations, they importantly suggest that a part of acoustic variability in song is centrally generated by LMAN activity.

Evidence from modeling studies suggest that the DA-mediated reward signal in Area X from VTA is tuned around the average acoustic variations during undisturbed singing but re-tunes when the bird is trying to bias its output, for instance in presence of a new vocal target (Duffy et al., 2022; Toutounji et al., 2022). The adjustment, or motor bias, is the third critical element of the reinforcement learning framework, with evidence showing a decrease of the learned bias upon the transient inactivation of LMAN during pitch-shifting tasks (Ali et al., 2013b; Andalman & Fee, 2009b; Tachibana et al., 2022), with Area X being critical to this facilitate this bias (Ali et al., 2013b; Zai, 2019). In Chapter 2, we expanded the dynamic latent variable model to incorporate both a reward signal and a motor bias in the context of pitch-shifting (illustrated in Figure 3.3C). By specifically associating the reward signal with the noise attributed to LMAN and dismissing other factors like diurnal effects as irrelevant to the learning strategy, our model adeptly captures pitch adaptation trajectories during pitch shifting, even when birds demonstrate non-linear learning dynamics. This finding brings our observations in line with conventional reinforcement learning theories and suggest that songbirds adopt an optimal learning strategy within this framework.

The assumptions of our model are consistent with the anatomical arrangement in which RA-projecting neurons in LMAN extend collaterals to Area X (Vates & Nottebohm, 1995), providing a crucial copy of the motor signal for integration with the dopamine (DA)-mediated reward signal in Area X. The exact neural mechanisms through which synaptic reinforcement in Area X enhances motor bias in LMAN are yet to be determined. Between Area X and LMAN exists a complex network involving both excitatory and inhibitory routed through DLM (see Chapter 1, Figure 2.12A). While it is yet unclear how information is transmitted and processed within this

network, it is clear that it is drives bursting activity in LMAN (Kojima et al., 2013) and is necessary to facilitate the motor bias in pitch-shifting paradigms (Ali et al., 2013b; Zai et al., 2020).

We can draw two important predictions from the reinforcement framework. First, a rewarded rendition should prompt a transient DA increase in Area X, reinforcing synaptic connections and resulting in more consistent activity patterns in LMAN (or Area X) across successive renditions. In contrast, unrewarded renditions are expected to produce more varied activity patterns, signaling a reinforcement deficit. This mechanism should be particularly evident when vocal behaviors are deliberately altered, such as during pitch-shifting experiments (Tumer & Brainard, 2007) or targeted imitation tasks (Lipkind et al., 2017; Toutounji et al., 2023), suggesting that DA-mediated reinforcement reliably induces specific activity patterns in Area X. These, in turn, trigger persistent changes in LMAN that mediate the motor bias. Over time, this process may lead to detectable correlation patterns between these brain regions.

Exploring these predictions offers a compelling direction for future research, potentially bridging the gap between our inconclusive CCA findings and theoretical frameworks, as well as allowing a direct comparison with studies conducted in rodents (Kondapavulur et al., 2022; Lemke et al., 2019) and primates (Toni et al., 2002).

Vocal analysis in the era of big data

In many fields including natural science, there is a growing trend towards amassing large data sets in hope to reveal patterns and insights that remained undetectable before. In birdsong research, collecting vocal recordings, such as those of laboratory songbirds, is cost-effective and easily scalable. In fact, many studies recorded vocalizations over the course of weeks, months or even years, for instance, to investigate the intricacies of vocal learning (e.g., Marler & Peters, 1982). A major challenge arises in the segmentation and clustering of these vocalizations, especially in the presence of background noise or when vocalizations from various animals overlap. Traditionally, this process has been executed manually, demanding considerable time and labor. Because the effort scales linearly with the number of vocalizations, the manual pre-processing of lengthy recordings becomes an arduous undertaking. Consequently, many of the original studies performed sporadic recordings, or, vice versa, focused their analysis on a sparse subset of vocalizations.

In Chapter 2, we took a different approach and recorded the song of zebra finches over the course of several weeks before and after brain manipulation and while they performed in a reinforcement learning paradigm. We focused on one specific feature of a particular syllable in a bird's motif. Thanks to an automatic detection of this syllable, the pre-processing was minimal and subsequently facilitated the atomic pitch extraction that resulted in extensive behavioral time series. Our results show that observed pitch variability can be approximated by a model of different variability generating sources. Most sources are independent of the central AFP system that is involved in song learning and adult plasticity (Ali et al., 2013b; Andalman & Fee, 2009b; S. W. Bottjer et al., 1984; Scharff & Nottebohm, 1991; Zai et al., 2020).

It is likely that other song features are also influenced by diverse sources of motor variability. While the pitch is a representative acoustic feature of song, it would be interesting to test if the

identified behavioral primitives (daily pattern, slow fluctuations, and history dependence) exhibit consistent patterns across syllables and features. For instance, a recent publication suggested diurnal changes in syllable entropy (Brudner et al., 2023). However, to corroborate this statement, a more comprehensive analysis would be necessary including extensive segmentation and clustering efforts. fitted of atomic pitch trajectories

Modern machine learning techniques promise to facilitate this facet of vocal analysis especially in the context of extensive datasets. Automatic classifiers optimized for the assignment of pre-computed segments to labels, either in a supervised (Tachibana et al., 2014; Nicholson, 2016; Cohen et al., 2022; Goffinet et al., 2021; Steinfath et al., 2021) or unsupervised manner (Goffinet et al., 2021; Sainburg et al., 2020) have been put forth as a solution for the song (and call) clustering problem. Some of these indirectly find vocalization on- and offsets because of the matching between training and test data sets, thus, implicitly solving the segmentation problem. Yet, in most cases, either preprocessing a training set or manual post-curation or both are required for clean results. In contrast, we propose an alternative approach in Chapter 3 that places the human expert in the center. We embed acoustic events in a plane that allows users to visually inspect and browse the content to make an informed decision (Lorenz et al., 2023).

Our approach might still be of interest in the light of new approaches that rely on better algorithms. For instance, colleagues recently adopted the Whisper Transformer Model (Radford et al., 2023) to detect animal vocalizations based on a small training set, reaching excellent performance values for segmenting vocal activity in different animal species (Gu et al., 2023). Given the significant improvements and heightened performances seen by deep-learning models (LeCun et al., 2015), this promises a road map to a fully end-to-end annotation tool for vocal recording. However, because deep neural networks are susceptible to short-cut learning (Geirhos et al., 2020), and optimizing a training set or network can be time-consuming and requiring in-depth knowledge about the architecture and training procedure, a human-in-the-loop approach could be a suitable and efficient compromise. For instance, users can correct biases and resolve ambiguous situations (such as deciding whether some vocalizations belong to the same type or not as in Figure 4.1, syllables D and E). Such fine decisions are part of critical data assessment (Thomas and Cook, 2006; Cui, 2019) and are common when working with animal data.

Machine learning tools and algorithms that can detect patterns in large and high-dimensional data sets also give rise to new interpretations of birdsong. Embedding the raw or spectro-temporal representation of the sound waveform in a lower dimensional space by means of statistical structure analysis, has led to new insights into developmental processes (Brudner et al., 2023; Kollmorgen et al., 2020) or neurobehavioral correlates of fine-grained variations from one rendition to the next (Alvarado et al., 2021a).

Taken inspiration from these more holistic approaches, we can envision a structure analysis detached from predefined spectral features. For our Neuropixels data, we implemented an analysis along those lines, where we tested whether the relationship between LMAN population spiking and the spectro-temporal variation in song are continuous (Appendix IV). We found initial evidence for such a structured mapping where small changes in neuron population space relate

to small changes in vocal space. This suggest a promising avenue for future research, where the nuanced interplay between neural dynamics and vocal output can be further explored, potentially uncovering new dimensions of how brain activity orchestrates complex behaviors.

Summary

The contributions of this thesis are threefold. First, we introduce a novel chronic implant device that enables the recording of high-yield electrophysiological data with a Neuropixels probe in freely moving birds. The chronic nature of the implant allowed for the acquisition of an extensive data set of simultaneous neural activity in LMAN and Area X during singing.

Second, we employed a dynamic latent variable model and identified sources of pitch variability with different stochastic properties. The strong correlation between LMAN lesion extent and changes in parameters from the independent noise source suggest that LMAN produces variability that is independent from one rendition to the next. Neural spiking activity in LMAN was significantly correlated with pitch variability, which corroborates the notion that LMAN is a central source of motor exploration.

Third, we complement our work by proposing a user-guided segmentation and clustering approach that takes advantage of a structured low-dimensional embedding of vocalizations.

In conclusion, this thesis has advanced the methodological arsenal of neural recording techniques in freely moving songbirds and vocal data analysis, and elucidated the central contribution to motor exploration and learning originating in the AFP. Songbirds provide a suitable model system for further studies to understand how neural communication within and across the nuclei of the AFP implicates vocal learning and adaptation.

Bibliography

Adrian, E. D. (1926). The impulses produced by sensory nerve endings: Part I. *The Journal of Physiology*, *61*(1), 49–72. https://doi.org/10.1113/jphysiol.1926.sp002273

Adrian, E. D., & Bronk, D. W. (1928). The discharge of impulses in motor nerve fibres: Part I. Impulses in single fibres of the phrenic nerve. *The Journal of Physiology*, *66*(1), 81–101. https://doi.org/10.1113/jphysiol.1928.sp002509

Adrian, E. D., & Zotterman, Y. (1926a). The impulses produced by sensory nerve endings: Part 3. Impulses set up by Touch and Pressure. *The Journal of Physiology*, *61*(4), 465–483. https://doi.org/10.1113/jphysiol.1926.sp002308

Ali, F., Otchy, T. M., Pehlevan, C., Fantana, A. L., Burak, Y., & Ölveczky, B. P. (2013a). The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong. *Neuron*, *80*(2), 494–506. https://doi.org/10.1016/j.neuron.2013.07.049

Alvarado, J. S., Goffinet, J., Michael, V., Liberti, W., Hatfield, J., Gardner, T., Pearson, J., & Mooney, R. (2021a). Neural dynamics underlying birdsong practice and performance. *Nature*, *599*(7886), 635–639. https://doi.org/10.1038/s41586-021-04004-1

Alvarado, J. S., Hatfield, J., & Mooney, R. (2023). Divergent projections from locus coeruleus to the corticobasal ganglia system and ventral tegmental area of the adult male zebra finch. *Journal of Comparative Neurology*, *531*(8), 921–934. https://doi.org/10.1002/cne.25474

Andalman, A. S., & Fee, M. S. (2009a). A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proceedings of the National Academy of Sciences*, *106*(30), 12518–12523. https://doi.org/10.1073/pnas.0903214106

Arneodo, E. M., Chen, S., Brown, D. E., Gilja, V., & Gentner, T. Q. (2021). Neurally driven synthesis of learned, complex vocalizations. *Current Biology*, *31*(15), 3419–3425. https://doi.org/10.1016/j.cub.2021.05.035

Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A Specialized Forebrain Circuit for Vocal Babbling in the Juvenile Songbird. *Science*, *320*(5876), 630–634. https://doi.org/10.1126/science.1155140

Atkinson, G., & Reilly, T. (1996). Circadian Variation in Sports Performance: *Sports Medicine*, *21*(4), 292–312. https://doi.org/10.2165/00007256-199621040-00005

Beer, R. D. (1990). *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroethology*. Academic Press.

Bentley, G. E., & Ball, G. F. (2000). Photoperiod-Dependent and -Independent Regulation of Melatonin Receptors in the Forebrain of Songbirds. *Journal of Neuroendocrinology*, *12*(8), 745–752. https://doi.org/10.1046/j.1365-2826.2000.00523.x

Bisio, A., Avanzino, L., Ruggeri, P., & Bove, M. (2014a). The tool as the last piece of the athlete's gesture imagery puzzle. *Neuroscience*, *265*, 196–203. https://doi.org/10.1016/j.neuroscience.2014.01.050

Boettiger, C. A., & Doupe, A. J. (1998). Intrinsic and Thalamic Excitatory Inputs Onto Songbird LMAN Neurons Differ in Their Pharmacological and Temporal Properties. *Journal of Neurophysiology*, *79*(5), 2615–2628. https://doi.org/10.1152/jn.1998.79.5.2615

Bolhuis, J. J., & Gahr, M. (2006). Neural mechanisms of birdsong memory. *Nature Reviews Neuroscience*, *7*(5), 347–357. https://doi.org/10.1038/nrn1904

Bottjer, S., Miesner, E., & Arnold, A. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*, *224*(4651), 901–903. https://doi.org/10.1126/science.6719123

Bottjer, S. W., Halsema, K. A., Brown, S. A., & Miesner, E. A. (1989). Axonal connections of a forebrain nucleus involved with vocal learning in zebra finches. *The Journal of Comparative Neurology*, *279*(2), 312–326. https://doi.org/10.1002/cne.902790211

Bottjer, S. W., Miesner, E. A., & Arnold, A. P. (1984). Forebrain Lesions Disrupt Development But Not Maintenance of Song in Passerine Birds. *Science*, *224*(4651), 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. *Nature*, *404*(6779), 762–766. https://doi.org/10.1038/35008083

Brudner, S., Pearson, J., & Mooney, R. (2023). Generative models of birdsong learning link circadian fluctuations in song variability to changes in performance. *PLOS Computational Biology*, *19*(5), e1011051. https://doi.org/10.1371/journal.pcbi.1011051

Buccino, A. P., Hurwitz, C. L., Garcia, S., Magland, J., Siegle, J. H., Hurwitz, R., & Hennig, M. H. (2020). SpikeInterface, a unified framework for spike sorting. *eLife*, *9*, e61834. https://doi.org/10.7554/eLife.61834

Budzillo, A., Duffy, A., Miller, K. E., Fairhall, A. L., & Perkel, D. J. (2017). Dopaminergic modulation of basal ganglia output through coupled excitation–inhibition. *Proceedings of the National Academy of Sciences*, *114*(22), 5713–5718. https://doi.org/10.1073/pnas.1611146114

Cagnacci, A., Kräuchi, K., Wirz-Justice, A., & Volpe, A. (1997). Homeostatic versus Circadian Effects of Melatonin on Core Body Temperature in Humans. *Journal of Biological Rhythms*, *12*(6), 509–517. https://doi.org/10.1177/074873049701200604

Canopoli, A., Herbst, J. A., & Hahnloser, R. H. R. (2014). A Higher Sensory Brain Region Is Involved in Reversing Reinforcement-Induced Vocal Changes in a Songbird. *The Journal of Neuroscience*, *34*(20), 7018–7026. https://doi.org/10.1523/JNEUROSCI.0266-14.2014

Carlson, D., & Carin, L. (2019). Continuing progress of spike sorting in the era of big data. *Current Opinion in Neurobiology*, *55*, 90–96. https://doi.org/10.1016/j.conb.2019.02.007

Cassone, V. M. (1990). Effects of melatonin on vertebrate circadian systems. *Trends in Neurosciences*, *13*(11), 457–464. https://doi.org/10.1016/0166-2236(90)90099-V

Cazala, A., Giret, N., Edeline, J.-M., & Del Negro, C. (2019). Neuronal encoding in a high-level auditory area: From sequential order of elements to grammatical structure. *The Journal of Neuroscience*, *39*(31), 6150–6161. https://doi.org/10.1523/JNEUROSCI.2767-18.2019

Chari, T., Banerjee, J., & Pachter, L. (2021). The Specious Art of Single-Cell Genomics. *BioRxiv*, 1–25.

Charlesworth, J. D., Tumer, E. C., Warren, T. L., & Brainard, M. S. (2011). Learning the microstructure of successful behavior. *Nature Neuroscience*, *14*(3), 373–380. https://doi.org/10.1038/nn.2748

Chauvette, S., Volgushev, M., & Timofeev, I. (2010). Origin of Active States in Local Neocortical Networks during Slow Sleep Oscillation. *Cerebral Cortex*, *20*(11), 2660–2674. https://doi.org/10.1093/cercor/bhq009

Chen, J. R., Stepanek, L., & Doupe, A. J. (2014). Differential contributions of basal ganglia and thalamus to song initiation, tempo, and structure. *Journal of Neurophysiology*, *111*(2), 248–257. https://doi.org/10.1152/jn.00584.2012

Chen, X., Wang, F., Kooijmans, R., Klink, P. C., Boehler, C., Asplund, M., & Roelfsema, P. R. (2023). Chronic stability of a neuroprosthesis comprising multiple adjacent Utah arrays in monkeys. *Journal of Neural Engineering*, *20*(3), 036039. https://doi.org/10.1088/1741-2552/ace07e

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences*, *20*(12), 553–557. https://doi.org/10.1016/s0166-2236(97)01149-1

Chung, J. E., Joo, H. R., Fan, J. L., Liu, D. F., Barnett, A. H., Chen, S., Geaghan-Breiner, C., Karlsson, M. P., Karlsson, M., Lee, K. Y., Liang, H., Magland, J. F., Pebbles, J. A., Tooker, A. C., Greengard, L. F., Tolosa, V. M., & Frank, L. M. (2019). High-Density, Long-Lasting, and Multi-region Electrophysiological Recordings Using Polymer Electrode Arrays. *Neuron*, *101*(1), 21-31.e5. https://doi.org/10.1016/j.neuron.2018.11.002

Churchland, M. M., Afshar, A., & Shenoy, K. V. (2006). A Central Source of Movement Variability. *Neuron*, *52*(6), 1085–1096. https://doi.org/10.1016/j.neuron.2006.10.034

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I., & Shenoy, K. V. (2010). Cortical Preparatory Activity: Representation of Movement or First Cog in a Dynamical Machine? *Neuron*, *68*(3), 387–400. https://doi.org/10.1016/j.neuron.2010.09.015

Cohen, Y., Shen, J., Semu, D., Leman, D. P., Liberti, W. A., Perkins, L. N., Liberti, D. C., Kotton, D. N., & Gardner, T. J. (2020). Hidden neural states underlie canary song syntax. *Nature*, *582*(7813), Article 7813. https://doi.org/10.1038/s41586-020-2397-3

Costalunga, G., Carpena, C. S., Seltmann, S., Benichov, J. I., & Vallentin, D. (2023). Wild nightingales flexibly match whistle pitch in real time. *Current Biology*, *33*(15), 3169-3178.e3. https://doi.org/10.1016/j.cub.2023.06.044

Crandall, S. R., Adam, M., Kinnischtzke, A. K., & Nick, T. A. (2007). HVC neural sleep activity increases with development and parallels nightly changes in song behavior. *Journal of Neurophysiology*, *98*(1), 232–240. https://doi.org/10.1152/jn.00128.2007

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*(11), 1500–1509. https://doi.org/10.1038/nn.3776

Daliparthi, V. K., Tachibana, R. O., Cooper, B. G., Hahnloser, R. H., Kojima, S., Sober, S. J., & Roberts, T. F. (2019). Transitioning between preparatory and precisely sequenced neuronal activity in production of a skilled behavior. *eLife*, *8*, e43732. https://doi.org/10.7554/eLife.43732

Darshan, R., Wood, W. E., Peters, S., Leblois, A., & Hansel, D. (2017a). A canonical neural mechanism for behavioral variability. *Nature Communications*, *8*(1), 15415. https://doi.org/10.1038/ncomms15415

Das, A., & Goldberg, J. H. (2021). Songbird subthalamic neurons project to dopaminergic midbrain and exhibit singing-related activity. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00254.2021

Datta, S. R., Anderson, D. J., Branson, K., Perona, P., & Leifer, A. (2019). Computational Neuroethology: A Call to Action. *Neuron*, *104*(1), 11–24. https://doi.org/10.1016/j.neuron.2019.09.038

De La Crompe, B., Schneck, M., Steenbergen, F., Schneider, A., & Diester, I. (2023). FreiBox: A Versatile Open-Source Behavioral Setup for Investigating the Neuronal Correlates of Behavioral Flexibility via 1-Photon Imaging in Freely Moving Mice. *Eneuro*, *10*(4), ENEURO.0469-22.2023. https://doi.org/10.1523/ENEURO.0469-22.2023

Ding, L., & Perkel, D. J. (2004). Long-Term Potentiation in an Avian Basal Ganglia Nucleus Essential for Vocal Learning. *Journal of Neuroscience*, *24*(2), 488–494. https://doi.org/10.1523/JNEUROSCI.4358-03.2004

Doupe, A. J., & Kuhl, P. K. (1999). BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms. *Annual Review of Neuroscience*, *22*(1), 567–631. https://doi.org/10.1146/annurev.neuro.22.1.567

Du Bois-Reymond, E. (1843). Vorläufiger Abriss einer Untersuchung über den sogenannten Froschstrom und über die elektromotorischen Fische. *Annalen der Physik*, *134*(1), 1–30. https://doi.org/10.1002/andp.18431340102

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

Düring, D. N., Dittrich, F., Rocha, M. D., Tachibana, R. O., Mori, C., Okanoya, K., Boehringer, R., Ehret, B., Grewe, B. F., Gerber, S., Ma, S., Rauch, M., Paterna, J.-C., Kasper, R., Gahr, M., & Hahnloser, R. H. R. (2020). Fast Retrograde Access to Projection Neuron Circuits Underlying Vocal Learning in Songbirds. *Cell Reports*, *33*(6), 108364. https://doi.org/10.1016/j.celrep.2020.108364

Egger, R., Tupikov, Y., Elmaleh, M., Katlowitz, K. A., Benezra, S. E., Picardo, M. A., Moll, F., Kornfeld, J., Jin, D. Z., & Long, M. A. (2020). Local Axonal Conduction Shapes the Spatiotemporal Properties of Neural Sequences. *Cell*, *183*(2), 537-548.e12. https://doi.org/10.1016/j.cell.2020.09.019

Elmaleh, M., Kranz, D., Asensio, A. C., Moll, F. W., & Long, M. A. (2021). Sleep replay reveals premotor circuit structure for a skilled behavior. *Neuron*. https://doi.org/10.1016/j.neuron.2021.09.021

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303. https://doi.org/10.1038/nrn2258

Farries, M. A., & Perkel, D. J. (2002). A Telencephalic Nucleus Essential for Song Learning Contains Neurons with Physiological Characteristics of Both Striatum and Globus Pallidus. *The Journal of Neuroscience*, *22*(9), 3776–3787. https://doi.org/10.1523/JNEUROSCI.22-09-03776.2002

Fee, M. S., & Goldberg, J. H. (2011). A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience*, *198*, 152–170. https://doi.org/10.1016/j.neuroscience.2011.09.069

Fee, M. S., & Leonardo, A. (2001). Miniature motorized microdrive and commutator system for chronic neural recording in small animals. *Journal of Neuroscience Methods*, *112*(2), 83–94. https://doi.org/10.1016/S0165-0270(01)00426-5

Fee, M. S., & Scharff, C. (2010). The Songbird as a Model for the Generation and Learning of Complex Sequential Behaviors. *ILAR Journal*, *51*(4), 362–377. https://doi.org/10.1093/ilar.51.4.362

Finger, S. (2005a). Chapter 8 Luigi Galvani: Electricity and the Nerves. In *Minds Behind the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195181821.001.0001

Finger, S. (2005b). Chapter 15 Edgar D. Adrian: Coding in the Nervous System. In *Minds Behind the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195181821.001.0001

Finkelstein, G. (2013). Chapter 4: Science. In *Emil du Bois-Reymond: Neuroscience, Self, and Society in Nineteenth-Century Germany*. The MIT Press. https://doi.org/10.7551/mitpress/9543.001.0001

Fortunato, C., Bennasar-Vázquez, J., Park, J., Chang, J. C., Miller, L. E., Dudman, J. T., Perich, M. G., & Gallego, J. A. (2023). *Nonlinear manifolds underlie neural population activity during behaviour* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.07.18.549575

Freedman, D. A. (2009). *Statistical Models: Theory and Practice* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511815867

Funabiki, Y., & Konishi, M. (2003). Long Memory in Song Learning by Zebra Finches. *The Journal of Neuroscience*, *23*(17), 6928–6935. https://doi.org/10.1523/JNEUROSCI.23-17-06928.2003

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

Gahr, M., & Kosar, E. (1996). Identification, distribution, and developmental changes of a melatonin binding site in the song control system of the zebra finch. *Journal of Comparative Neurology*, *367*(2), 308–318. https://doi.org/10.1002/(SICI)1096-9861(19960401)367:2<308::AID-CNE11>3.0.CO;2-M

Gale, S. D., & Perkel, D. J. (2010). Anatomy of a songbird basal ganglia circuit essential for vocal learning and plasticity. *Journal of Chemical Neuroanatomy*, *39*(2), 124–131. https://doi.org/10.1016/j.jchemneu.2009.07.003

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural Manifolds for the Control of Movement. *Neuron*, *94*(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025

Galvani, L. (1791). D viribus electricitatis in motu musculari: Commentarius. *Bologna: Tip. Istituto Delle Scienze*.

Gasser, H. S., & Erlanger, J. (1922). A study of the action currents of nerve with the cathode ray oscillograph. *American Journal of Physiology-Legacy Content*, *62*(3), 496–524. https://doi.org/10.1152/ajplegacy.1922.62.3.496

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence*, *2*(11), Article 11. https://doi.org/10.1038/s42256-020-00257-z

Gerfen, C. R., & Bolam, J. P. (2010). Chapter 1—The Neuroanatomical Organization of the Basal Ganglia. In H. Steiner & K. Y. Tseng (Eds.), *Handbook of Behavioral Neuroscience* (Vol. 20, pp. 3–28). Elsevier. https://doi.org/10.1016/B978-0-12-374767-9.00001-9

Ghahramani, Z., & Hinton, G. E. (1996). *Parameter Estimation for Linear Dynamical Systems* (Technical Report CRG-TR-96-2).

Girardeau, G., Inema, I., & Buzsáki, G. (2017). Reactivations of emotional memory in the hippocampus–amygdala system during sleep. *Nature Neuroscience*, *20*(11), 1634–1642. https://doi.org/10.1038/nn.4637

Giret, N., Kornfeld, J., Ganguli, S., & Hahnloser, R. H. R. (2014a). Evidence for a causal inverse model in an avian cortico-basal ganglia circuit. *Proceedings of the National Academy of Sciences*, *111*(16), 6063–6068. https://doi.org/10.1073/pnas.1317087111

Gokcen, E., Jasper, A. I., Semedo, J. D., Zandvakili, A., Kohn, A., Machens, C. K., & Yu, B. M. (2022). Disentangling the flow of signals between populations of neurons. *Nature Computational Science*, *2*(8), 512–525. https://doi.org/10.1038/s43588-022-00282-5

Goldberg, J. H., Adler, A., Bergman, H., & Fee, M. S. (2010a). Singing-Related Neural Activity Distinguishes Two Putative Pallidal Cell Types in the Songbird Basal Ganglia: Comparison to the Primate Internal and External Pallidal Segments. *The Journal of Neuroscience*, *30*(20), 7088–7098. https://doi.org/10.1523/JNEUROSCI.0168-10.2010

Goldberg, J. H., & Fee, M. S. (2010a). Singing-Related Neural Activity Distinguishes Four Classes of Putative Striatal Neurons in the Songbird Basal Ganglia. *Journal of Neurophysiology*, *103*(4), 2002–2014. https://doi.org/10.1152/jn.01038.2009

Goldberg, J. H., & Fee, M. S. (2012). A cortical motor nucleus drives the basal ganglia-recipient thalamus in singing birds. *Nature Neuroscience*, *15*(4), Article 4. https://doi.org/10.1038/nn.3047

Grienberger, C., Giovannucci, A., Zeiger, W., & Portera-Cailliau, C. (2022). Two-photon calcium imaging of neuronal activity. *Nature Reviews Methods Primers*, *2*(1), Article 1. https://doi.org/10.1038/s43586-022-00147-1

Gu, N., Lee, K., Basha, M., Ram, S. K., You, G., & Hahnloser, R. H. R. (2023). *Positive Transfer of the Whisper Speech Transformer to Human and Animal Voice Activity Detection* [Preprint]. Animal Behavior and Cognition. https://doi.org/10.1101/2023.09.30.560270

Guitchounts, G., Markowitz, J. E., Liberti, W. A., & Gardner, T. J. (2013). A carbon-fiber electrode array for long-term neural recording. *Journal of Neural Engineering*, *10*(4), 046016. https://doi.org/10.1088/1741-2560/10/4/046016

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and Inaccurate Vocal Imitation after Knockdown of FoxP2 in Songbird Basal Ganglia Nucleus Area X. *PLoS Biol*, *5*(12), e321. https://doi.org/10.1371/journal.pbio.0050321

Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002b). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, *419*(6902), 65–70. https://doi.org/10.1038/nature00974

Hampton, C. M., Sakata, J. T., & Brainard, M. S. (2009). An Avian Basal Ganglia-Forebrain Circuit Contributes Differentially to Syllable Versus Sequence Variability of Adult Bengalese Finch Song. *Journal of Neurophysiology*, *101*(6), 3235–3245. https://doi.org/10.1152/jn.91089.2008

Herbst, J. A., Kotowicz, A., Wang, C. Z.-H., Rychen, J., Canopoli, A., Narula, G., Zai, A. T., & Hahnloser, R. H. (2023). *RecOOrder–A modular recording and controlling environment for songbirds in LabVIEW*.

Hesse, J. K., & Tsao, D. Y. (2020). A new no-report paradigm reveals that face cells encode both consciously perceived and suppressed stimuli. *eLife*, *9*, e58360. https://doi.org/10.7554/eLife.58360

Hessler, N. A., & Doupe, A. J. (1999a). Singing-Related Neural Activity in a Dorsal Forebrain–Basal Ganglia Circuit of Adult Zebra Finches. *The Journal of Neuroscience*, *19*(23), 10461–10481. https://doi.org/10.1523/JNEUROSCI.19-23-10461.1999

Hessler, N. A., & Doupe, A. J. (1999c). Social context modulates singing-related neural activity in the songbird forebrain. *Nature Neuroscience*, *2*(3), 209–211. https://doi.org/10.1038/6306

Heston, J. B., Simon, J., Day, N. F., Coleman, M. J., & White, S. A. (2018). Bidirectional scaling of vocal variability by an avian cortico-basal ganglia circuit. *Physiological Reports*, *6*(8), e13638. https://doi.org/10.14814/phy2.13638

Hisey, E., Kearney, M. G., & Mooney, R. (2018). A common neural circuit mechanism for internally guided and externally reinforced forms of motor learning. *Nature Neuroscience*, *21*(4), 589–597. https://doi.org/10.1038/s41593-018-0092-6

Hoffmann, L. A., Saravanan, V., Wood, A. N., He, L., & Sober, S. J. (2016). Dopaminergic Contributions to Vocal Learning. *The Journal of Neuroscience*, *36*(7), 2176–2189. https://doi.org/10.1523/JNEUROSCI.3883-15.2016

Hong, G., & Lieber, C. M. (2019). Novel electrode technologies for neural recordings. *Nature Reviews Neuroscience*, *20*(6), 330–345. https://doi.org/10.1038/s41583-019-0140-6

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Hyland Bruno, J., Jarvis, E. D., Liberman, M., & Tchernichovski, O. (2021). Birdsong Learning and Culture: Analogies with Human Spoken Language. *Annual Review of Linguistics*, *7*(1), 449–472. https://doi.org/10.1146/annurev-linguistics-090420-121034

Jovalekic, A., Cavé-Lopez, S., Canopoli, A., Ondracek, J. M., Nager, A., Vyssotski, A. L., & Hahnloser, R. H. R. (2017). A lightweight feedback-controlled microdrive for chronic neural recordings. *Journal of Neural Engineering*, *14*(2), 026006. https://doi.org/10.1088/1741-2552/aa5848

Juavinett, A. L., Bekheet, G., & Churchland, A. K. (2019). Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *eLife*, *8*, e47188. https://doi.org/10.7554/eLife.47188

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Häusser, M., Karsh, B., … Harris, T. D. (2017). Fully integrated

silicon probes for high-density recording of neural activity. *Nature*, *551*(7679), Article 7679. https://doi.org/10.1038/nature24636

Kao, M. H., & Brainard, M. S. (2006). Lesions of an Avian Basal Ganglia Circuit Prevent Context-Dependent Changes to Song Variability. *Journal of Neurophysiology*, *96*(3), 1441–1455. https://doi.org/10.1152/jn.01138.2005

Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005a). Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature*, *433*(7026), 638–643. https://doi.org/10.1038/nature03127

Kao, M. H., Wright, B. D., & Doupe, A. J. (2008a). Neurons in a Forebrain Nucleus Required for Vocal Plasticity Rapidly Switch between Precise Firing and Variable Bursting Depending on Social Context. *Journal of Neuroscience*, *28*(49), 13232–13247. https://doi.org/10.1523/JNEUROSCI.2250-08.2008

Keifer, J., & Summers, C. H. (2016). Putting the “Biology” Back into “Neurobiology”: The Strength of Diversity in Animal Model Systems for Neuroscience Research. *Frontiers in Systems Neuroscience*, *10*. https://doi.org/10.3389/fnsys.2016.00069

Kobak, D., Pardo-Vazquez, J. L., Valente, M., Machens, C. K., & Renart, A. (2019). State-dependent geometry of population activity in rat auditory cortex. *eLife*, *8*, e44526. https://doi.org/10.7554/eLife.44526

Kojima, S., Kao, M. H., & Doupe, A. J. (2013). Task-related “cortical” bursting depends critically on basal ganglia input and is linked to vocal plasticity. *Proceedings of the National Academy of Sciences*, *110*(12), 4756–4761. https://doi.org/10.1073/pnas.1216308110

Kojima, S., Kao, M. H., Doupe, A. J., & Brainard, M. S. (2018a). The Avian Basal Ganglia Are a Source of Rapid Behavioral Variation That Enables Vocal Motor Exploration. *The Journal of Neuroscience*, *38*(45), 9635–9647. https://doi.org/10.1523/JNEUROSCI.2915-17.2018

Kollmorgen, S., Hahnloser, R. H. R., & Mante, V. (2020). Nearest neighbours reveal fast and slow components of motor learning. *Nature*, *577*(7791), Article 7791. https://doi.org/10.1038/s41586-019-1892-x

Kondapavulur, S., Lemke, S. M., Darevsky, D., Guo, L., Khanna, P., & Ganguly, K. (2022). Transition from predictable to variable motor cortex and striatal ensemble patterning during behavioral exploration. *Nature Communications*, *13*(1), 2450. https://doi.org/10.1038/s41467-022-30069-1

Konishi, M. (1965). Effects of deafening on song development in American robins and black-headed grosbeaks. *Zeitschrift Für Tierpsychologie*, *22*(5), 584–599.

Koparkar, A., Warren, T. L., Charlesworth, J. D., Shin, S., Brainard, M. S., & Veit, L. (2024). *Lesions in a songbird vocal circuit increase variability in song syntax* (p. 2023.07.25.550508). bioRxiv. https://doi.org/10.1101/2023.07.25.550508

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Lack, D. L. (1965). The life of the robin. *(No Title)*.

Lattenkamp, E. Z., & Vernes, S. C. (2018). Vocal learning: A language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences*, *21*, 209–215. https://doi.org/10.1016/j.cobeha.2018.04.007

Leblois, A., Bodor, A. L., Person, A. L., & Perkel, D. J. (2009). Millisecond Timescale Disinhibition Mediates Fast Information Transmission through an Avian Basal Ganglia Loop. *The Journal of Neuroscience*, *29*(49), 15420–15433. https://doi.org/10.1523/JNEUROSCI.3060-09.2009

Leblois, A., & Perkel, D. J. (2012). Striatal dopamine modulates song spectral but not temporal features through D1 receptors. *European Journal of Neuroscience*, *35*(11), 1771–1781. https://doi.org/10.1111/j.1460-9568.2012.08095.x

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), Article 7553. https://doi.org/10.1038/nature14539

Lemke, S. M., Ramanathan, D. S., Guo, L., Won, S. J., & Ganguly, K. (2019). Emergent modular neural control drives coordinated motor actions. *Nature Neuroscience*, *22*(7), 1122–1131. https://doi.org/10.1038/s41593-019-0407-2

Leonardo, A., & Fee, M. S. (2005). Ensemble Coding of Vocal Control in Birdsong. *The Journal of Neuroscience*, *25*(3), 652–661. https://doi.org/10.1523/JNEUROSCI.3036-04.2005

Lipkind, D., Geambasu, A., & Levelt, C. C. (2019). The Development of Structured Vocalizations in Songbirds and Humans: A Comparative Analysis. *Topics in Cognitive Science*, *0*(0). https://doi.org/10.1111/tops.12414

Lisberger, S. G., & Medina, J. F. (2015). How and why neural and motor variation are related. *Current Opinion in Neurobiology*, *33*, 110–116. https://doi.org/10.1016/j.conb.2015.03.008

Livingston, F. S., & Mooney, R. (1997a). Development of Intrinsic and Synaptic Properties in a Forebrain Nucleus Essential to Avian Song Learning. *J. Neurosci.*, *17*(23), 8997–9009.

Livingston, F. S., & Mooney, R. (1997b). Development of Intrinsic and Synaptic Properties in a Forebrain Nucleus Essential to Avian Song Learning. *The Journal of Neuroscience*, *17*(23), 8997–9009. https://doi.org/10.1523/JNEUROSCI.17-23-08997.1997

Lombardino, A. J., & Nottebohm, F. (2000). Age at Deafening Affects the Stability of Learned Song in Adult Male Zebra Finches. *Journal of Neuroscience*, *20*(13), 5054–5064. https://doi.org/10.1523/JNEUROSCI.20-13-05054.2000

Lopez, C. M., Putzeys, J., Raducanu, B. C., Ballini, M., Wang, S., Andrei, A., Rochus, V., Vandebriel, R., Severi, S., Van Hoof, C., Musa, S., Van Helleputte, N., Yazicioglu, R. F., & Mitra, S. (2017). A Neural Probe With Up to 966 Electrodes and Up to 384 Configurable Channels in

0.13 μm SOI CMOS. *IEEE Transactions on Biomedical Circuits and Systems*, *11*(3), 510–522. https://doi.org/10.1109/TBCAS.2016.2646901

López-Samanes, Á., Moreno-Pérez, D., Maté-Muñoz, J. L., Domínguez, R., Pallarés, J. G., Mora-Rodriguez, R., & Ortega, J. F. (2017). Circadian rhythm effect on physical tennis performance in trained male players. *Journal of Sports Sciences*, *35*(21), 2121–2128. https://doi.org/10.1080/02640414.2016.1258481

Lorenz, C., Hao, X., Tomka, T., Rüttimann, L., & Hahnloser, R. H. R. (2023). Interactive extraction of diverse vocal units from a planar embedding without the need for prior sound segmentation. *Frontiers in Bioinformatics*, *2*, 966066. https://doi.org/10.3389/fbinf.2022.966066

Lovell, P. V., Wirthlin, M., Kaser, T., Buckner, A. A., Carleton, J. B., Snider, B. R., McHugh, A. K., Tolpygo, A., Mitra, P. P., & Mello, C. V. (2020). ZEBrA: Zebra finch Expression Brain Atlas—A resource for comparative molecular neuroanatomy and brain evolution studies. *Journal of Comparative Neurology*, *528*(12), 2099–2131. https://doi.org/10.1002/cne.24879

Ludbrook, J. (2010). Linear regression analysis for comparing two measurers or methods of measurement: But which regression? *Clinical and Experimental Pharmacology and Physiology*, *37*(7), 692–699. https://doi.org/10.1111/j.1440-1681.2010.05376.x

Luo, M., Ding, L., & Perkel, D. J. (2001). An Avian Basal Ganglia Pathway Essential for Vocal Learning Forms a Closed Topographic Loop. *The Journal of Neuroscience*, *21*(17), 6836–6845. https://doi.org/10.1523/JNEUROSCI.21-17-06836.2001

Luo, M., & Perkel, D. J. (1999a). A GABAergic, strongly inhibitory projection to a thalamic nucleus in the zebra finch song system. *Journal of Neuroscience*, *19*(15), 6700–6711.

Luo, M., & Perkel, D. J. (1999b). Long-range GABAergic projection in a circuit essential for vocal learning. *The Journal of Comparative Neurology*, *403*(1), 68–84.

Luo, T. Z., Bondy, A. G., Gupta, D., Elliott, V. A., Kopec, C. D., & Brody, C. D. (2020a). An approach for long-term, multi-probe Neuropixels recordings in unrestrained rats. *eLife*, *9*, e59716. https://doi.org/10.7554/eLife.59716

Luo, T. Z., Bondy, A. G., Gupta, D., Elliott, V. A., Kopec, C. D., & Brody, C. D. (2020b). An approach for long-term, multi-probe Neuropixels recordings in unrestrained rats. *eLife*, *9*, e59716. https://doi.org/10.7554/eLife.59716

Lycke, R., Kim, R., Zolotavin, P., Montes, J., Sun, Y., Koszeghy, A., Altun, E., Noble, B., Yin, R., He, F., Totah, N., Xie, C., & Luan, L. (2023). Low-threshold, high-resolution, chronically stable intracortical microstimulation by ultraflexible electrodes. *Cell Reports*, *42*(6). https://doi.org/10.1016/j.celrep.2023.112554

Lynch, G. F., Okubo, T. S., Hanuschkin, A., Hahnloser, R. H. R., & Fee, M. S. (2016). Rhythmic Continuous-Time Coding in the Songbird Analog of Vocal Motor Cortex. *Neuron*, *90*(4), 877–892. https://doi.org/10.1016/j.neuron.2016.04.021

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology*, *71*(2p2), 1.

Marler, P., & Peters, S. (1977). Selective Vocal Learning in a Sparrow. *Science*, *198*(4316), 519–521. https://doi.org/10.1126/science.198.4316.519

Marler, P., & Peters, S. (1982). Long-term storage of learned birdsongs prior to production. *Animal Behaviour*, *30*(2), 479–482. https://doi.org/10.1016/S0003-3472(82)80059-6

Marler, P., & Tamura, M. (1964). Culturally Transmitted Patterns of Vocal Behavior in Sparrows. *Science*, *146*(3650), 1483–1486. https://doi.org/10.1126/science.146.3650.1483

McGregor, J. N., Grassler, A. L., Jaffe, P. I., Jacob, A. L., Brainard, M. S., & Sober, S. J. (2022). Shared mechanisms of auditory and non-auditory vocal learning in the songbird brain. *eLife*, *11*, e75691. https://doi.org/10.7554/eLife.75691

Menardy, F., Giret, N., & Del Negro, C. (2014). The presence of an audience modulates responses to familiar call stimuli in the male zebra finch forebrain. *The European Journal of Neuroscience*, *40*(9), 3338–3350. https://doi.org/10.1111/ejn.12696

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4. *Neuron*, *55*(1), 131–141. https://doi.org/10.1016/j.neuron.2007.06.018

Moll, F. W., Kranz, D., Corredera Asensio, A., Elmaleh, M., Ackert-Smith, L. A., & Long, M. A. (2023). Thalamus drives vocal onsets in the zebra finch courtship song. *Nature*, *616*(7955), Article 7955. https://doi.org/10.1038/s41586-023-05818-x

Monti, J. M. (2010). Serotonin 5-HT(2A) receptor antagonists in the treatment of insomnia: Present status and future prospects. *Drugs of Today (Barcelona, Spain*, *46*(3), 183–193. https://doi.org/10.1358/dot.2010.46.3.1437247

Mooney, R., & Konishi, M. (1991). Two distinct inputs to an avian song nucleus activate different glutamate receptor subtypes on individual neurons. *Proceedings of the National Academy of Sciences*, *88*(10), 4075–4079. https://doi.org/10.1073/pnas.88.10.4075

Moorman, S., Ahn, J.-R., & Kao, M. H. (2021a). Plasticity of stereotyped birdsong driven by chronic manipulation of cortical-basal ganglia activity. *Current Biology*, *31*(12), 2619-2632.e4. https://doi.org/10.1016/j.cub.2021.04.030

Moshayedi, P., Ng, G., Kwok, J. C. F., Yeo, G. S. H., Bryant, C. E., Fawcett, J. W., Franze, K., & Guck, J. (2014). The relationship between glial cell mechanosensitivity and foreign body reactions in the central nervous system. *Biomaterials*, *35*(13), 3919–3925. https://doi.org/10.1016/j.biomaterials.2014.01.038

Nixdorf-Bergweiler, B. E., & Bischof, H.-J. (2007). Introduction to the atlas. In *A stereotaxic atlas of the brain of the zebra finch, Taeniopygia guttata: With special emphasis on telencephalic visual and song system nuclei in transverse and sagittal sections [internet]*. National Center for Biotechnology Information (US).

Nordeen, K. W., & Nordeen, E. J. (1992). Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behavioral and Neural Biology*, *57*(1), 58–66. https://doi.org/10.1016/0163-1047(92)90757-U

Nottebohm, F., & Arnold, A. P. (1976). Sexual dimorphism in vocal control areas of the songbird brain. *Science*, *194*(4261), 211–213. https://doi.org/10.1126/science.959852

Nottebohm, F., Nottebohm, M. E., & Crane, L. (1986). Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behavioral and Neural Biology*, *46*(3), 445–471. https://doi.org/10.1016/S0163-1047(86)90485-1

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, *Serinus canarius*. *The Journal of Comparative Neurology*, *165*(4), 457–486. https://doi.org/10.1002/cne.901650405

Ölveczky, B. P., Andalman, A. S., & Fee, M. S. (2005b). Vocal Experimentation in the Juvenile Songbird Requires a Basal Ganglia Circuit. *PLoS Biology*, *3*(5), e153. https://doi.org/10.1371/journal.pbio.0030153

Ölveczky, B. P., Otchy, T. M., Goldberg, J. H., Aronov, D., & Fee, M. S. (2011). Changes in the neural control of a complex motor sequence during learning. *Journal of Neurophysiology*, *106*(1), 386–397. https://doi.org/10.1152/jn.00018.2011

Pachitariu, M., Sridhar, S., & Stringer, C. (2023). *Solving the spike sorting problem with Kilosort* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.01.07.523036

Pearson, K., & Galton, F. (1997). VII. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*(347–352), 240–242. https://doi.org/10.1098/rspl.1895.0041

Person, A. L., Gale, S. D., Farries, M. A., & Perkel, D. J. (2008). Organization of the songbird basal ganglia, including area X. *Journal of Comparative Neurology*, *508*(5), 840–866. https://doi.org/10.1002/cne.21699

Peters, A. J., Chen, S. X., & Komiyama, T. (2014). Emergence of reproducible spatiotemporal activity during motor learning. *Nature*, *510*(7504), 263–267. https://doi.org/10.1038/nature13235

Picardo, M. A., Merel, J., Katlowitz, K. A., Vallentin, D., Okobi, D. E., Benezra, S. E., Clary, R. C., Pnevmatikakis, E. A., Paninski, L., & Long, M. A. (2016). Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch. *Neuron*, *90*(4), 866–876. https://doi.org/10.1016/j.neuron.2016.02.016

Prather, J. F., Okanoya, K., & Bolhuis, J. J. (2017). Brains for birds and babies: Neural parallels between birdsong and speech acquisition. *Neuroscience & Biobehavioral Reviews*, *81*, 225–237. https://doi.org/10.1016/j.neubiorev.2016.12.035

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

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. *International Conference on Machine Learning*, 28492–28518.

Ramón y Cajal, S. (1888). Estructura de la retina de las aves. *Revista Trimestral de Histología Normal y Patológica*.

Reiner, A. (2002). Functional circuitry of the avian basal ganglia: Implications for basal ganglia organization in stem amniotes. *Brain Research Bulletin*, *57*(3), 513–528. https://doi.org/10.1016/S0361-9230(01)00667-0

Reiner, A., Medina, L., & Veenman, C. L. (1998). Structural and functional evolution of the basal ganglia in vertebrates. *Brain Research Reviews*, *28*(3), 235–285. https://doi.org/10.1016/S0165-0173(98)00016-2

Reiner, A., Perkel, D. J., Mello, C. V., & Jarvis, E. D. (2004). Songbirds and the Revised Avian Brain Nomenclature. *Annals of the New York Academy of Sciences*, *1016*(1), 77–108. https://doi.org/10.1196/annals.1298.013

Renshaw, B., Forbes, A., & Morison, B. R. (1940). Activity of isocortex and hippocampus: Electrical studies with micro-electrodes. *Journal of Neurophysiology*, *3*(1), 74–105. https://doi.org/10.1152/jn.1940.3.1.74

Roberts, T. F., Gobes, S. M. H., Murugan, M., Ölveczky, B. P., & Mooney, R. (2012). Motor circuits are required to encode a sensory model for imitative learning. *Nature Neuroscience*, *15*(10), 1454–1459. https://doi.org/10.1038/nn.3206

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), Article 7. https://doi.org/10.1038/nn.4563

Rosen, M. J., & Mooney, R. (2000). Intrinsic and Extrinsic Contributions to Auditory Selectivity in a Song Nucleus Critical for Vocal Plasticity. *The Journal of Neuroscience*, *20*(14), 5437–5448. https://doi.org/10.1523/JNEUROSCI.20-14-05437.2000

Rousche, P. J., & Normann, R. A. (1998). Chronic recording capability of the Utah Intracortical Electrode Array in cat sensory cortex. *Journal of Neuroscience Methods*, *82*(1), 1–15. https://doi.org/10.1016/S0165-0270(98)00031-4

Ruder, L., Schina, R., Kanodia, H., Valencia-Garcia, S., Pivetta, C., & Arber, S. (2021). A functional map for diverse forelimb actions within brainstem circuitry. *Nature*, *590*(7846), 445–450. https://doi.org/10.1038/s41586-020-03080-z

Sainburg, T., & Gentner, T. Q. (2021). Toward a Computational Neuroethology of Vocal Communication: From Bioacoustics to Neurophysiology, Emerging Tools and Future Directions. *Frontiers in Behavioral Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnbeh.2021.811737

Salatino, J. W., Ludwig, K. A., Kozai, T. D. Y., & Purcell, E. K. (2017). Glial responses to implanted electrodes in the brain. *Nature Biomedical Engineering*, *1*(11), Article 11. https://doi.org/10.1038/s41551-017-0154-1

Santos, F. J., Oliveira, R. F., Jin, X., & Costa, R. M. (2015). Corticostriatal dynamics encode the refinement of specific behavioral variability during skill learning. *eLife*, *4*, e09423. https://doi.org/10.7554/eLife.09423

Sasaki, A., Sotnikova, T. D., Gainetdinov, R. R., & Jarvis, E. D. (2006). Social Context-Dependent Singing-Regulated Dopamine. *The Journal of Neuroscience*, *26*(35), 9010–9014. https://doi.org/10.1523/JNEUROSCI.1335-06.2006

Sauerbrei, B. A., Guo, J.-Z., Cohen, J. D., Mischiati, M., Guo, W., Kabra, M., Verma, N., Mensh, B., Branson, K., & Hantman, A. W. (2020). Cortical pattern generation during dexterous movement is input-driven. *Nature*, *577*(7790), 386–391. https://doi.org/10.1038/s41586-019-1869-9

Scharff, C., & Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: Implications for vocal learning. *The Journal of Neuroscience*, *11*(9), 2896–2913. https://doi.org/10.1523/JNEUROSCI.11-09-02896.1991

Scherrer, J. R., Lynch, G. F., Zhang, J. J., & Fee, M. S. (2023). An optical design enabling lightweight and large field-of-view head-mounted microscopes. *Nature Methods*, *20*(4), Article 4. https://doi.org/10.1038/s41592-023-01806-1

Schregardus, D. S., Pieneman, A. W., Ter Maat, A., Jansen, R. F., Brouwer, T. J. F., & Gahr, M. L. (2006). A lightweight telemetry system for recording neuronal activity in freely behaving small animals. *Journal of Neuroscience Methods*, *155*(1), 62–71. https://doi.org/10.1016/j.jneumeth.2005.12.028

Semedo, J. D., Gokcen, E., Machens, C. K., Kohn, A., & Yu, B. M. (2020). Statistical methods for dissecting interactions between brain areas. *Current Opinion in Neurobiology*, *65*, 59–69. https://doi.org/10.1016/j.conb.2020.09.009

Semedo, J. D., Jasper, A. I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C. K., Kohn, A., & Yu, B. M. (2022). Feedforward and feedback interactions between visual cortical areas use

different population activity patterns. *Nature Communications*, *13*(1), 1099. https://doi.org/10.1038/s41467-022-28552-w

Sherrington, C. (1906). *The Integrative Action of the Nervous System. Oxford*.

Shumway, R. H., & Stoffer, D. S. (1982). An Approach to Time Series Smoothing and Forecasting Using the Em Algorithm. *Journal of Time Series Analysis*, *3*(4), 253–264. https://doi.org/10.1111/j.1467-9892.1982.tb00349.x

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

Sober, S. J., Wohlgemuth, M. J., & Brainard, M. S. (2008). Central Contributions to Acoustic Variation in Birdsong. *The Journal of Neuroscience*, *28*(41), 10370–10379. https://doi.org/10.1523/JNEUROSCI.2448-08.2008

Sohrabji, F., Nordeen, E. J., & Nordeen, K. W. (1990). Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behavioral and Neural Biology*, *53*(1), 51–63. https://doi.org/10.1016/0163-1047(90)90797-A

Sossinka, R., & Böhner, J. (1980). Song Types in the Zebra Finch Poephila guttata castanotis1. *Zeitschrift Für Tierpsychologie*, *53*(2), 123–132. https://doi.org/10.1111/j.1439-0310.1980.tb01044.x

Sponheim, C., Papadourakis, V., Collinger, J. L., Downey, J., Weiss, J., Pentousi, L., Elliott, K., & Hatsopoulos, N. G. (2021). Longevity and reliability of chronic unit recordings using the Utah, intracortical multi-electrode arrays. *Journal of Neural Engineering*, *18*(6), 066044. https://doi.org/10.1088/1741-2552/ac3eaf

Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., Chen, S., Colonell, J., Gardner, R. J., Karsh, B., Kloosterman, F., Kostadinov, D., Mora-Lopez, C., O'Callaghan, J., Park, J., … Harris, T. D. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, *372*(6539), eabf4588. https://doi.org/10.1126/science.abf4588

Steinmetz, N. A., Koch, C., Harris, K. D., & Carandini, M. (2018). Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Current Opinion in Neurobiology*, *50*, 92–100. https://doi.org/10.1016/j.conb.2018.01.009

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, *576*(7786), 266–273. https://doi.org/10.1038/s41586-019-1787-x

Stepanek, L., & Doupe, A. J. (2010). Activity in a Cortical-Basal Ganglia Circuit for Song Is Required for Social Context-Dependent Vocal Variability. *Journal of Neurophysiology*, *104*(5), 2474–2486. https://doi.org/10.1152/jn.00977.2009

Steriade, M., Nunez, A., & Amzica, F. (1993). A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: Depolarizing and hyperpolarizing components. *Journal of Neuroscience*, *13*(8), 3252–3265. https://doi.org/10.1523/JNEUROSCI.13-08-03252.1993

Stevenson, I. H., & Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience*, *14*(2), 139–142. https://doi.org/10.1038/nn.2731

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361–365. https://doi.org/10.1038/s41586-019-1346-5

Tachibana, R. O., Lee, D., Kai, K., & Kojima, S. (2022). Performance-Dependent Consolidation of Learned Vocal Changes in Adult Songbirds. *The Journal of Neuroscience*, *42*(10), 1974–1986. https://doi.org/10.1523/JNEUROSCI.1942-21.2021

Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H., & White, S. A. (2004). Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(13), 3152–3163. https://doi.org/10.1523/JNEUROSCI.5589-03.2004

Therrien, A. S., Wolpert, D. M., & Bastian, A. J. (2016). Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain*, *139*(1), 101–114. https://doi.org/10.1093/brain/awv329

Thompson, J. A., Basista, M. J., Wu, W., Bertram, R., & Johnson, F. (2011). Dual Pre-Motor Contribution to Songbird Syllable Variation. *The Journal of Neuroscience*, *31*(1), 322–330. https://doi.org/10.1523/JNEUROSCI.5967-09.2011

Thorpe, W. H. (1958). The Learning of Song Patterns by Birds, with Especial Reference to the Song of the Chaffinch Fringilla Coelebs. *Ibis*, *100*(4), 535–570. https://doi.org/10.1111/j.1474-919X.1958.tb07960.x

Tian, L. Y., Warren, T. L., Mehaffey, W. H., & Brainard, M. S. (2023). Dynamic top-down biasing implements rapid adaptive changes to individual movements. *eLife*, *12*, e83223. https://doi.org/10.7554/eLife.83223

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Für Tierpsychologie*, *20*, 410–433.

Tinbergen, N. (1989). *The Study of Instinct: With a new Preface*. Oxford University Press.

Toni, I., Rowe, J., Stephan, K. E., & Passingham, R. E. (2002). Changes of Cortico-striatal Effective Connectivity during Visuomotor Learning. *Cerebral Cortex*, *12*(10), 1040–1047. https://doi.org/10.1093/cercor/12.10.1040

Toutounji, H., Zai, A. T., Tchernichovski, O., Hahnloser, R. H. R., & Lipkind, D. (2022). *Learning the sound inventory of a complex vocal skill via an intrinsic reward* [Preprint]. Animal Behavior and Cognition. https://doi.org/10.1101/2022.09.27.509691

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

Vallentin, D., Kosche, G., Lipkind, D., & Long, M. A. (2016). Inhibition protects acquired song segments during vocal learning in zebra finches. *Science*, *351*(6270), 267–271. https://doi.org/10.1126/science.aad3023

Van Beers, R. J. (2009). Motor Learning Is Optimally Tuned to the Properties of Motor Noise. *Neuron*, *63*(3), 406–417. https://doi.org/10.1016/j.neuron.2009.06.025

Van Beers, R. J., Haggard, P., & Wolpert, D. M. (2004). The Role of Execution Noise in Movement Variability. *Journal of Neurophysiology*, *91*(2), 1050–1063. https://doi.org/10.1152/jn.00652.2003

Van Beers, R. J., Van Der Meer, Y., & Veerman, R. M. (2013). What Autocorrelation Tells Us about Motor Variability: Insights from Dart Throwing. *PLoS ONE*, *8*(5), e64332. https://doi.org/10.1371/journal.pone.0064332

Van Daal, R. J. J., Aydin, Ç., Michon, F., Aarts, A. A. A., Kraft, M., Kloosterman, F., & Haesler, S. (2021). Implantation of Neuropixels probes for chronic recording of neuronal activity in freely behaving mice and rats. *Nature Protocols*, *16*(7), 3322–3347. https://doi.org/10.1038/s41596-021-00539-9

Vates, G. E., & Nottebohm, F. (1995). Feedback circuitry within a song-learning pathway. *Proceedings of the National Academy of Sciences*, *92*(11), 5139–5143. https://doi.org/10.1073/pnas.92.11.5139

Veit, L., Tian, L. Y., Monroy Hernandez, C. J., & Brainard, M. S. (2021). Songbirds can learn flexible contextual control over syllable sequencing. *eLife*, *10*, e61610. https://doi.org/10.7554/eLife.61610

Voigts, J., Siegle, J. H., Pritchett, D. L., & Moore, C. I. (2013). The flexDrive: An ultra-light implant for optical control and highly parallel chronic recording of neuronal ensembles in freely moving mice. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00008

Warren, T. L., Tumer, E. C., Charlesworth, J. D., & Brainard, M. S. (2011). Mechanisms and time course of vocal learning and consolidation in the adult songbird. *Journal of Neurophysiology*, *106*(4), 1806–1821. https://doi.org/10.1152/jn.00311.2011

Watkins, K. E., & Jenkinson, N. (2016). Chapter 8—The Anatomy of the Basal Ganglia. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 85–94). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00008-0

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, *8*, 229–256.

Wood, W. E., Osseward, P. J., Roseberry, T. K., & Perkel, D. J. (2013). A Daily Oscillation in the Fundamental Frequency and Amplitude of Harmonic Syllables of Zebra Finch Song. *PLoS ONE*, *8*(12), e82327. https://doi.org/10.1371/journal.pone.0082327

Woolley, S. C., & Doupe, A. J. (2008). Social Context–Induced Song Variation Affects Female Behavior and Gene Expression. *PLOS Biology*, *6*(3), e62. https://doi.org/10.1371/journal.pbio.0060062

Woolley, S. C., & Kao, M. H. (2015). Variability in action: Contributions of a songbird cortical-basal ganglia circuit to vocal motor learning and control. *Neuroscience*, *296*, 39–47. https://doi.org/10.1016/j.neuroscience.2014.10.010

Woolley, S. C., Rajan, R., Joshua, M., & Doupe, A. J. (2014a). Emergence of Context-Dependent Variability across a Basal Ganglia Network. *Neuron*, *82*(1), 208–223. https://doi.org/10.1016/j.neuron.2014.01.039

Xiao, L., Chattree, G., Oscos, F. G., Cao, M., Wanat, M. J., & Roberts, T. F. (2018). A Basal Ganglia Circuit Sufficient to Guide Birdsong Learning. *Neuron*, *98*(1), 208-221.e5. https://doi.org/10.1016/j.neuron.2018.02.020

Xu, W., De Carvalho, F., & Jackson, A. (2019). Sequential Neural Activity in Primary Motor Cortex during Sleep. *The Journal of Neuroscience*, *39*(19), 3698–3712. https://doi.org/10.1523/JNEUROSCI.1408-18.2019

Yamahachi, H., Zai, A. T., Tachibana, R. O., Stepien, A. E., Rodrigues, D. I., Cavé-Lopez, S., Lorenz, C., Arneodo, E. M., Giret, N., & Hahnloser, R. H. R. (2020). Undirected singing rate as a non-invasive tool for welfare monitoring in isolated male zebra finches. *PloS One*, *15*(8), e0236333. https://doi.org/10.1371/journal.pone.0236333

Yartsev, M. M., & Ulanovsky, N. (2013). Representation of Three-Dimensional Space in the Hippocampus of Flying Bats. *Science*, *340*(6130), 367–372. https://doi.org/10.1126/science.1235338

Yartsev, M. M., Witter, M. P., & Ulanovsky, N. (2011). Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, *479*(7371), 103–107. https://doi.org/10.1038/nature10583

Young, H. (1955). Breeding Behavior and Nesting of the Eastern Robin. *The American Midland Naturalist*, *53*(2), 329–352. https://doi.org/10.2307/2422072

Yuste, R. (2015). From the neuron doctrine to neural networks. *Nature Reviews Neuroscience*, *16*(8), Article 8. https://doi.org/10.1038/nrn3962

Zai, A. T. (2019). *Inferring the trial-by-trial structure of pitch reinforcement learning in songbirds* (p. 171 p.) [ETH Zurich; Application/pdf]. https://doi.org/10.3929/ETHZ-B-000382074

Zai, A. T., Cavé-Lopez, S., Rolland, M., Giret, N., & Hahnloser, R. H. R. (2020). Sensory substitution reveals a manipulation bias. *Nature Communications*, *11*(1), 5940. https://doi.org/10.1038/s41467-020-19686-w

Zhao, Z., Zhu, H., Li, X., Sun, L., He, F., Chung, J. E., Liu, D. F., Frank, L., Luan, L., & Xie, C. (2023). Ultraflexible electrode arrays for months-long high-density electrophysiological mapping

of thousands of neurons in rodents. *Nature Biomedical Engineering*, *7*(4), Article 4. https://doi.org/10.1038/s41551-022-00941-y

Appendix Chapter 1

P1: Neuropixels implant assembly (NIA) construction and preparation

Parts

- Neuropixels 1.0 w/o metallic cap — ***IMEC***
- Torlon® based implant holder — *CNC-milled by the* ***Department of Physics UZH workshop***
- M1 screws — *commercially available*

Consumables

- Fast drying adhesive — ***Hardman*** *Double/Bubble non-sag epoxy*
- Kapton tape
- Deionized water

Tools

- Stereotaxic socket — *CNC-milled by the workshop at the* ***Department of Physics UZH***
- Vice or clamp — ***Proxxon*** *Machine vice*
- ESD safe tweezers — ***Bernstein*** *stainless steel tweezers*
- Soldering iron

Assembling the implant

The first step in the assembly of a new implant involves permanently affixing the Neuropixels probe to the holder. To achieve this, we recommend positioning the holder in the stereotaxic socket and firmly securing it to the workbench, such as with a sturdy vice. Subsequently, apply adhesive to the inner surface of the holder and insert the lower rectangular body of the probe into the inset. The probe is securely in place when the slight widening on the upper end contacts the edges of the holder. Utilizing a helping hand to apply gentle pressure to the glued part ensures that the probe and holder maintain horizontal alignment. It is advisable to adjust the helping hand before applying or mixing the adhesive. In general, conducting a dry run with a dummy probe is recommended to acquaint oneself with the equipment and the assembly procedure.

Placing the screws

The selection of screws should be based on the final implant's configuration. In our case, we utilized a longer screw (M1x10mm) for the anterior thread and a shorter screw (M1x8mm) for the posterior thread, considering the relatively shallow penetration angle. It is important to note that maximizing the stability of a fully tightened screw is advised. For example, instead of inserting a shorter screw halfway into the thread, it is recommended to trim a longer screw to fit. Additionally, when assembling the implant for the first time, we suggest placing the screws before seating the Neuropixels.

P2: Surgical procedures and implantation

Parts

- Neuropixels implant assembly (NIA)
- Implant cover *3D printed by the* ***Department of Biochemistry UZH workshop***

Consumables

- 0.9% saline solution
- Betadine
- Cotton tips
- Kapton tape
- Deionized water
- Dental cements
- Kiwk Sil®
- Soft, two-component silicone gel (artificial dura)
- Superglue
- Toothpicks

Tools

- General surgery setup for small animals (stereotactic and anesthesia system, heating pad, microscope)
- Surgical instruments (scalpel, forceps)
- Dental drill
- Custom-built acute recording equipment for single tungsten electrodes
- Soldering ion

Medical supplies

- Isoflurane and oxygen
- General analgetic (e.g., 2% Lidocaine solution for injection)
- Local analgetic (e.g., Anesderm® or Emla®)

The day of the surgery, food sources are removed from the cage approximately half an hour before the start of the surgery.

Animals are anesthetized with Isoflurane (0.6-1.5% inhalation) and placed in the stereotactic apparatus once the flexion reflex is no longer observable. For general analgesia, Lidocaine (0.05 µl of 2 % diluted lidocaine) is administered subcutaneously. In a next step, local analgesia (Anesderm® or Emla®, 0.5 g) is applied to the exposed head skin along the caudo-rostral axis after removing feathers and disinfecting the area with ethanol and betadine. A long incision along this axis ranging from the beak almost to the neck muscles is made before the exposed skull is prepared for the implant. The skin is lightly pushed to the side and the skull surface is cleaned using saline soaked cotton tips, followed by drilling multiple small holes in the first bone layer that serve as anchor points for the cement later.

A small craniotomy contralateral to the final implant is made to place a silver wire in contact with the dura that serves as electrical ground. The main craniotomy is made on the right hemisphere around 1.7 mm lateral to the superior sagittal sinus and 4.5 mm anterior to lambda, the confluence of sinuses, with the flat part of the anterior skull rotate to a 50-degree angle. The positions of

LMAN and Area X are to be mapped with high-impedance tungsten-electrode recordings, identifying the brain areas based on the firing rate properties thereof. A Neuropixels probe version 1.0 (Jun et al., 2017) is implanted between 4 and 4.8mm deep to penetrate the putative centers of LMAN and Area X. Before the implantation, the probe is covered in fluorescent dye ($DiIC_{18}(3)$, Sigma Aldrich 42364) to facilitate tracing the probe in histological slices.

Next, the custom-made holder that carries the Neuropixels probe is fixated to the skull with dental cement and the craniotomy is sealed with a soft two-part silicone gel before it is enclosed with the custom-made casing. Gaps between the casing and the skull are closed and covered with cement or fast-drying two-component silicon (Kwik Sil). The headstage is put in place such that the flex cable formed a narrow, horizontal ‘S’ that also adds protection from the back. The skull is then tilted to a 35-degree angle which corresponds to the natural position of the birds’ head and the headstage is put in its final position by fixating the two flanking wires to the skull with cement. Finally, the ground wire is connected and open holes along the implant are sealed with fast drying two-component silicon.

Still under anesthesia, a thread is glued to the medial side of the headstage that carries the counterweight and is adjusted immediately after placing the bird in the home cage to help the animal cope with the added weight.

A subset of implants (n=1) was performed with a previous version of the implant design, and the first three implants were covered with a grounded copper shield instead of the 3D printed cover design. Results were similar across birds and therefore combined.

P3: Implant recovery and histological examination

Explantation

To recover the implant assembly with the probe, animals are euthanized and placed in the stereotaxic apparatus. The headstage is removed by cutting the adjacent wires and disconnecting it from the flex cable. Next, the cover is freed from cement and glue and taken off. The bird's head is tilted such that the stereotactical holder can smoothly be brought in laterally to hold the implant assembly. Screws at the stereotactic holder socket are tightened before cement around the screws is removed with a drill or by cutting. In cases with excessive cement, we immediately cut the skull and bring up the implant with part of the skull and the cement. Both are then removed piecewise with clamps or fine scissors.

Removing the screws and cleaning

Directly after explanting the probe, screws are removed from the implant assembly. The probe shank is then placed in a 1% Tergazyme® solution for at least 24 hours. In a final step, the probe shank is dipped in de-ionized water several times before storing it in a clean, enclosed box.

Histology

The cerebrum of each animal was recovered after the explantation and fixated with 4% buffered formaldehyde solution for a minimum of 24 hours. Sagittal or coronal brain slices of 80 um thickness were obtained using a vibratome. The trace of the shank relative to the brain areas of interest was identified under the microscope using bright-field and light of the wavelength that excites the dye (550 nm) and allows for visualizing the position of the probe after removal.

P4: Data acquisition

Tools

- Neuropixels data acquisition system — ***National Instruments*** *and* ***IMEC***
- Microphone, preamplifier and filter
- Data collection and storage computer
- Webcam

Programs

- Custom-written LabVIEW program for experimental studies with songbirds
- SpikeGLX (Bill Karsh, https://github.com/billkarsh/SpikeGLX)
- VLC media player

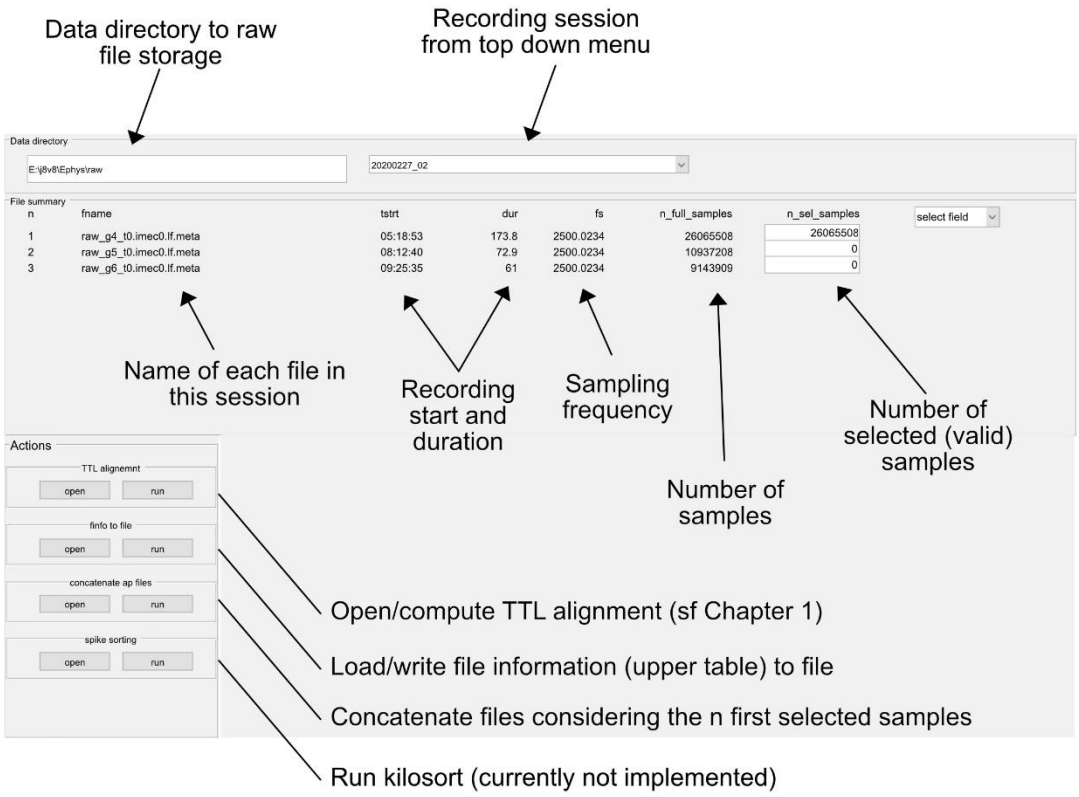
Audio signals are captured with a wall-attached microphone, preamplified, band-pass filtered, and digitized at 32 kHz. The saving is controlled by a custom-written labview program that also implements an online song detection algorithm and pitch calculation, and which controls playback and pitch contingent auditory feedback.

Concurrent audio signals during neural recordings are split before feeding into the recorder and acquired by the nidq system at 20 kHz. SpikeGLX is used to start and stop, as well as monitor and save neural recordings (.ap data sampled at 30 kHz, .lf data sampled at 2.5 kHz).

Birds are routinely connected to the interface cable in the morning and untangled manually if necessary. Neural activity and simultaneous audio signals are recorded in sessions that last between one- and five-hours using Spike GLX. In recording sessions with social modulation, we present a female zebra finch in a separate cage inside the recording chamber. To study effects of vocal and neural plasticity, we play a loud white noise burst sound contingent on the pitch of a stable, harmonic part of a syllable being above or below the median pitch in the preceding day or hours.

Data preprocessing tools

1. **Graphical User Interface for file navigation and pre-processing**: xsummary.mat



Methods

Animals

Eight adult male zebra finches were subjected to the study in this chapter. The animals were bred and raised in the avian facility of the *Institut des Neurosciences Paris-Saclay*, France. The animals' age ranged between 70 and 133 dph at the beginning of the experiment. They were housed individually in sound-attenuated recording chambers with access to food and water *ad libitum* throughout the experiment. The day/night cycle was set to 14 hours of day mimicked by artificial light and 10 hours of night.

All experimental procedures were performed in accordance with the French Ministry of Research and the ethical committee of Paris-Sud et Centre under the license 2017-25-10.

Neurobehavioral analysis

Song related neural activity was analyzed in windows that covered complete motifs offset by -100 ms to account for the premotor nature of the activity in LMAN and Area X. Spontaneous activity was derived in periods of silence in a randomly selected subwindow that matched the median duration of the motif.

Criteria for significant behavioral modulation

Whether a unit was song modulation was assessed by comparison of the song-related firing rate in each rendition against the spontaneous firing rate in motif-matched windows during silent periods. We used an unpaired, non-parametric test (Wilkoxon signrank) and units were classified as song modulated if $p<.001$. Similarly, social modulation was evaluated by comparison of the song-related firing rate between the two conditions using the same test and significance level.

Sparseness Index (Gini Index)

We used the Gini Index (Hurley and Rickard, 2009) as a measure of sparseness of song related spiking activity. We first computed the average song-related activity as a function of time, binning the spike data in 10 ms bins, averaging across bins and smoothing the resulting trajectory with a Gaussian with SD of 15 ms. We then computed the sparseness index as

$$Sparseness\ Index = 1 - 2B$$

where B is the sum of the area under the Lorenz curve, i.e. the sum of the ordered and normalized average song-related activity per bin.

Peak Firing Rate

Peak firing rate was derived from the interspike intervals (isi) that considered all spikes within the song-related analysis window and computed as the 1st percentile of the isi distribution.

Fano Factor

To measure the irregular spiking activity from rendition to rendition, we computed the Fano Factor of the song related neural activity for each unit. We counted the number of spikes ns elicited in the related analysis window of each rendition i and calculated

$$Fano\ Factor = \frac{\sigma_{ns}^2}{\mu_{ns}}$$

Canonical Correlation Analysis

To investigate cross-areal coupling, we compute the canonical correlation analysis (CCA, Semedo et al., 2020). Briefly, CCA is a multivariate linear regression method that finds the linear combination of two neural datasets X of size n×m (n neural units, m renditions) and Y size k×m (k neural units, m renditions) that have maximum correlation ρ with each other. Formally, it can be described as

$$\rho = corr(a^T X, b^T V)$$

where ***a*** and ***b*** are the vectors maximizing this correlation, formally defined as:

$$(a,\ b) = \underset{a,b}{\operatorname{argmax}}\, \rho$$

Our CCA computation focused on neural units within LMAN and Area X. We analyzed the time-warped data using a 40 ms sliding window, advancing in 5 ms increments beginning 50 ms before the motif onset and ending 100 ms before the motif offset. In each 40 ms window, spike times were binned in 5 ms bins and vectorized similar to (Mante et al., 2013). This approach allowed the temporal examination of the coupling between the two populations. To validate the significance of our findings, we employed a shuffle predictor as a control where the same analysis was performed with permutated renditions in the dataset of Area X. This comparison helped in discerning the relevance of the observed effects against chance correlations.

The analysis was conducted using MATLAB's `cancorr.m` function set to find the embedding maximizing Pearson's correlation coefficient (Pearson & Galton, 1997).

Cross-Correlation and Cross-covariance

The cross-correlation (xcor) between the spike trains of two units x and y describes a measure of similarity as a function of displacement. In our analysis, we apply the unbiased sample cross-correlation between the binned spike trains (2 ms bins). For a given temporal lag τ, it is formally described as

$$\hat{r}_{x,y}(\tau) = \begin{cases} \dfrac{1}{N-\tau} \sum_{k=1}^{N-\tau} x(k+\tau)y(k) & if\ \tau \geq 0 \\ \square & \\ R_{y,x}(-\tau) & if\ \tau < 0\,. \end{cases}$$

If the spike trains are taken from i different renditions of the same motor behavior, we average across the rendition-wise cross-correlations

$$\hat{r}_{x,y}(\tau) = \langle \hat{r}_{x,y,i}(\tau) \rangle_i .$$

The cross-covariance is the normalized version of the rendition-average cross-correlation. It describes the momentary similarity that is disconnected from the rendition-average activity by subtracting the shuffle predicted cross-correlation

$$\hat{c}_{x,y}(\tau) = \hat{r}_{x,y,same}(\tau) - \hat{r}_{x,y,shuffle}(\tau) .$$

We evaluated whether peaks in cross-correlation $\hat{r}_{x,y}$ or cross-covariance $\hat{c}_{x,y}$ functions are systematic by determining whether they exceed three standard deviations (SDs). The standard deviation was estimated by Jackknifing renditions (leave-one-out resampling). All functions were smoothed by a gaussian kernel with a 6 ms standard deviation.

Supplementary figures

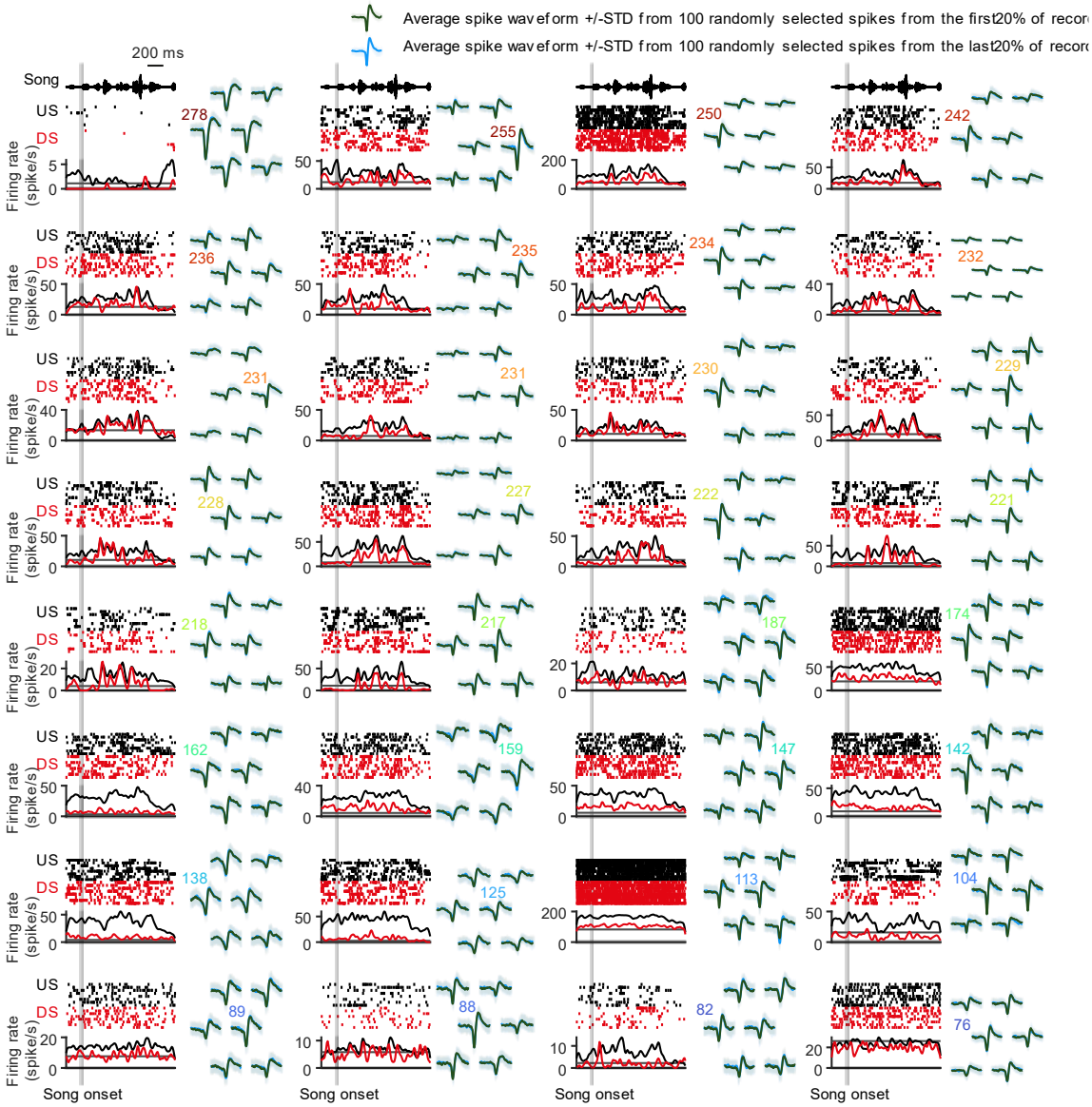


Figure 2.9 Suppl Fig 1: Spike raster and waveform for each unit in LMAN and Area X in Figure 2.9AB. *Each panel depicts the raster during directed (red) and undirected (black) singing and the average firing rate aligned to the motif onset. On the right side of each panel, the average +/- SD spike waveform from 100 randomly selected spikes within the first and last 20 % of the recording is shown. Spike waveforms were stable for all units shown.*

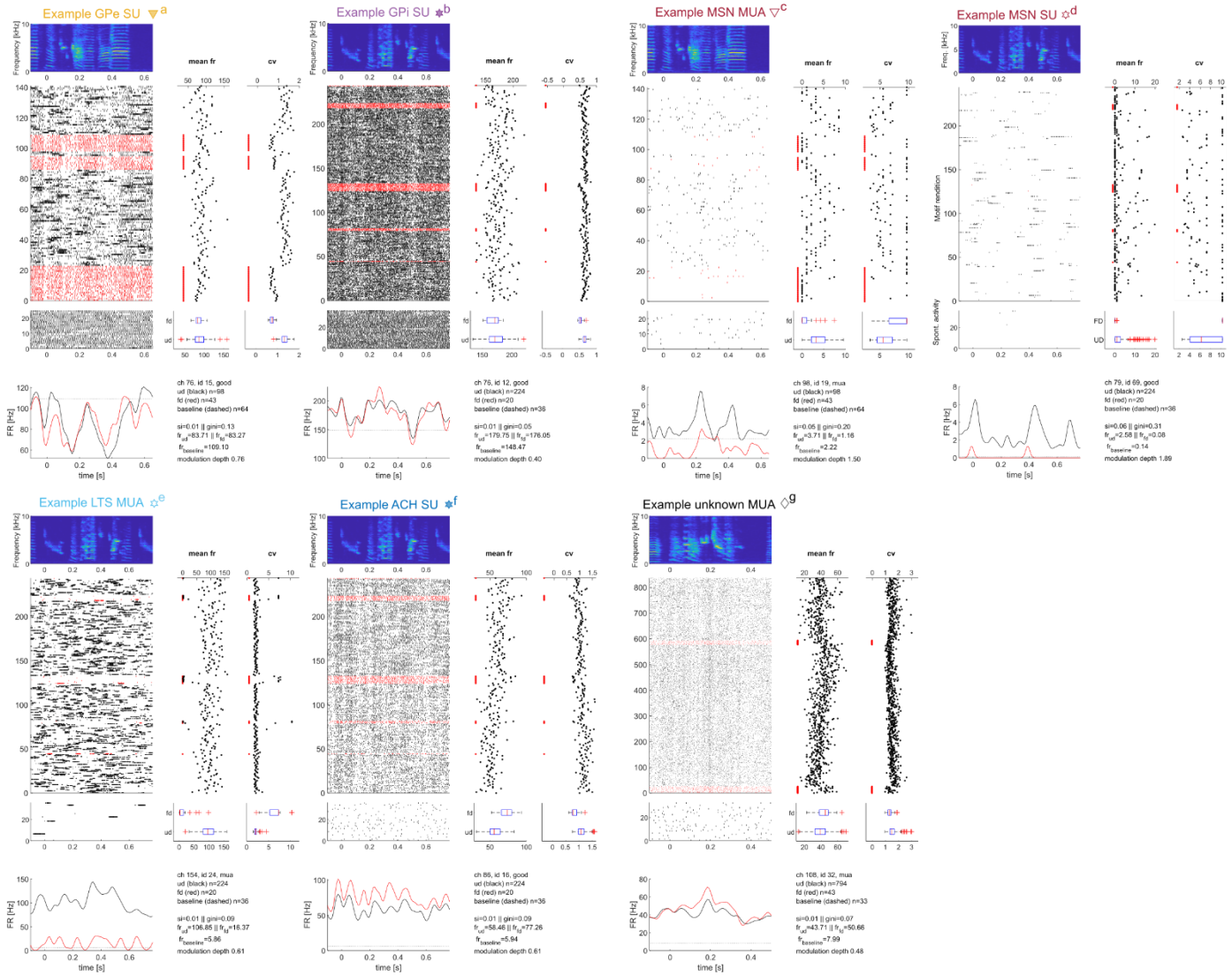


Figure 2.10 Suppl Fig 1: Other examples of putative GPe, GPi, MSN, LTS & ACH single or multiunits *(identified with small letters in the different subpanels of Figure 2.10). One example of unknown cell type is also shown. From top to bottom, spectrogram of a song motif, spike raster plots of consecutive undirected songs (black, UD) and female directed songs (red, FD) aligned to song motif onset, raster of 30 spike trains of spontaneous activity, average firing rate during undirected songs (black), female directed songs (red), and average spontaneous firing rate (dotted line). At the right of each raster plots are shown the average firing rate (in Hz) and coefficient of variation (CV) for the corresponding spike train. Boxplots of firing rate (Hz) and CV for undirected and female directed songs are shown underneath.*

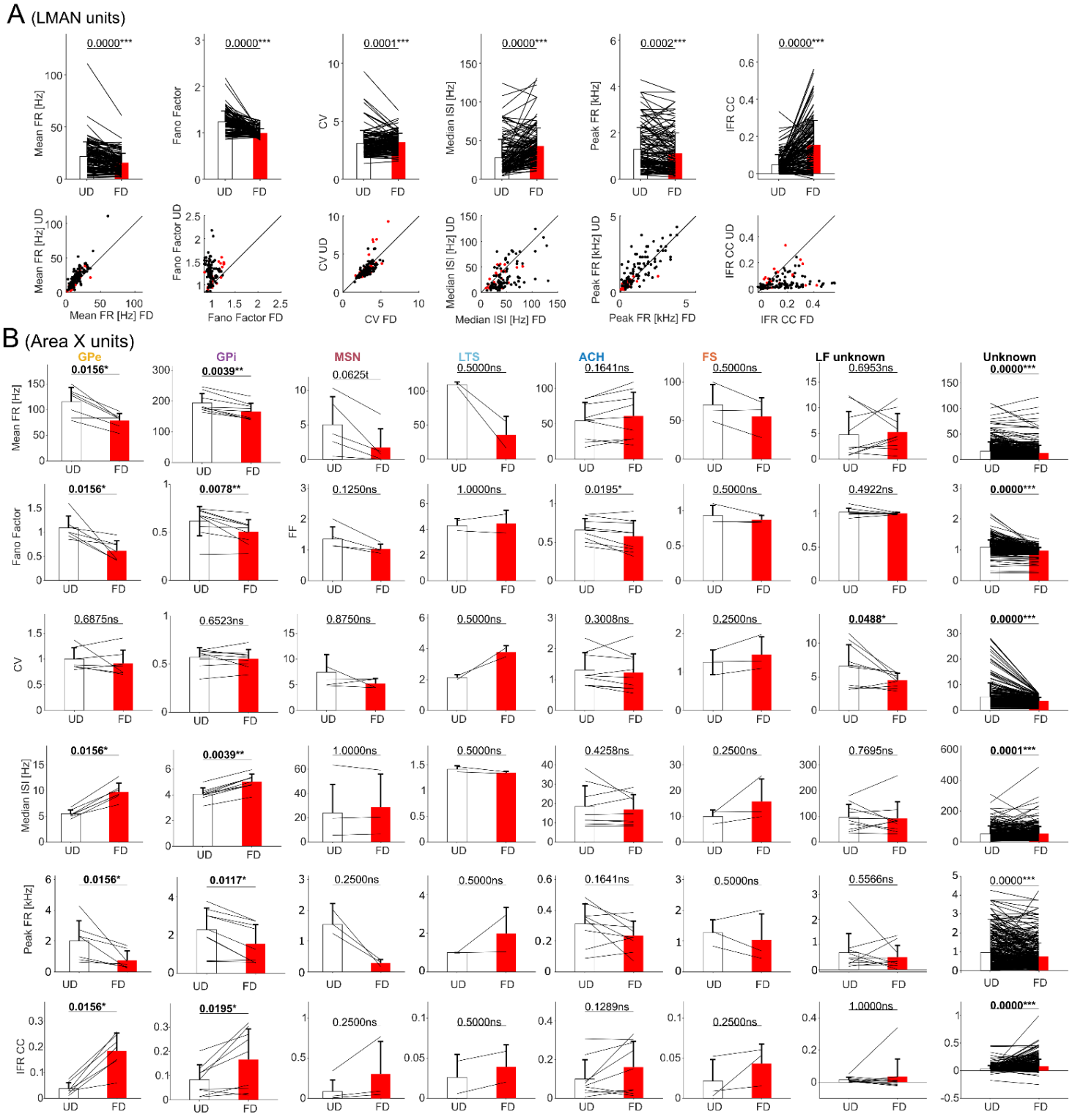


Figure 2.10 Suppl Fig 2. Social modulation in units of LMAN and Area X. *Average motif firing rate, Fano factor, coefficient of variation (CV), median interspike interval (ISI), peak firing rate (FR), instantaneous firing rate coefficient of correlation (IFR CC) during undirected (UD) and female directed (FD) songs for LMAN units (A) and for putative GPe, GPi, MSN, LTS, ACH, FS, LF unknown and other Area X unknown cells (B). Non-parametric paired tests (Wilcoxon signed rank) were applied to the data.*

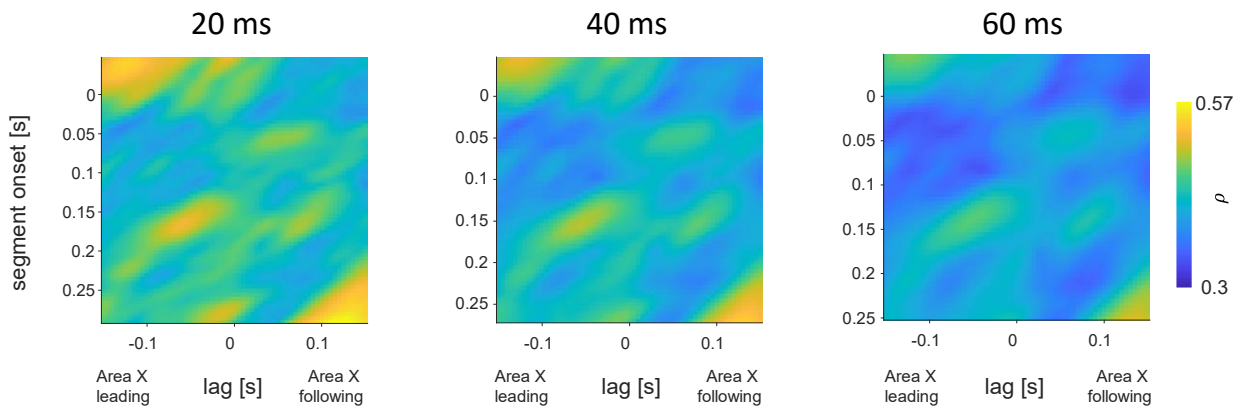
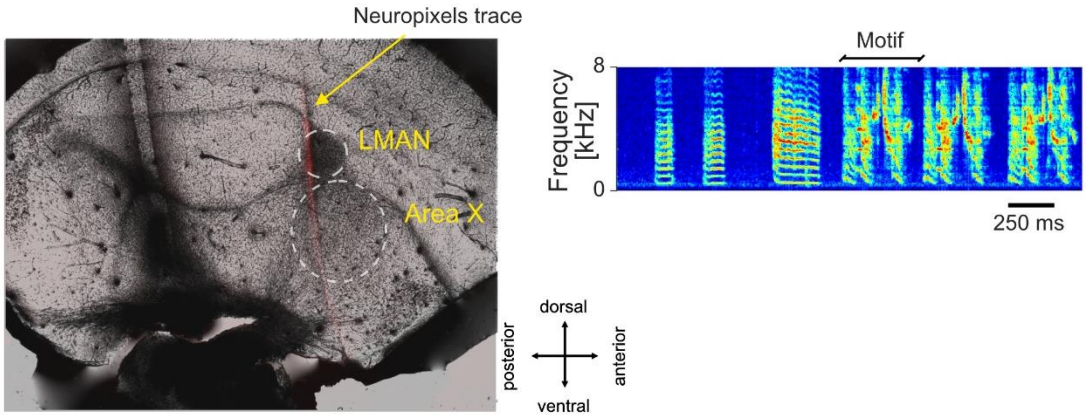


Figure 2.11 Suppl Fig 1 Effect of window size. *The population correlation peak around 150 ms post motif onset for LMAN activity and Area X leading by around 50 ms is stable across different segment lengths of 20 and 40 ms but washes out for larger segments lengths.*

Supplementary data overview

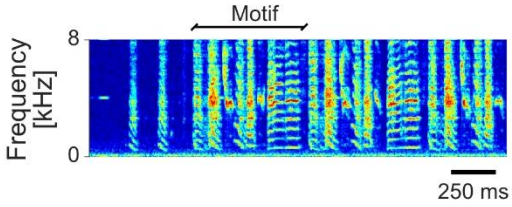
g4r4



---

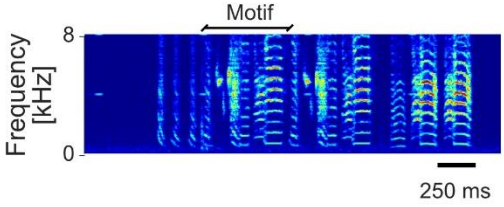
j8v8

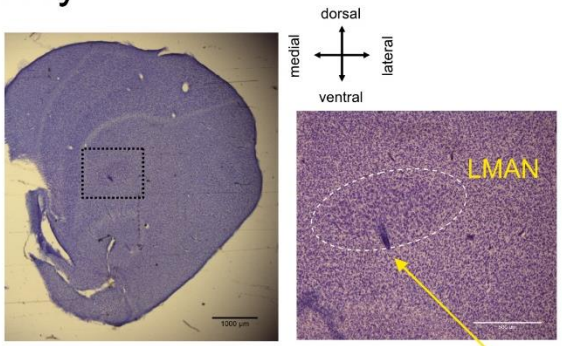
It was not possible to recover the brain from this bird.v

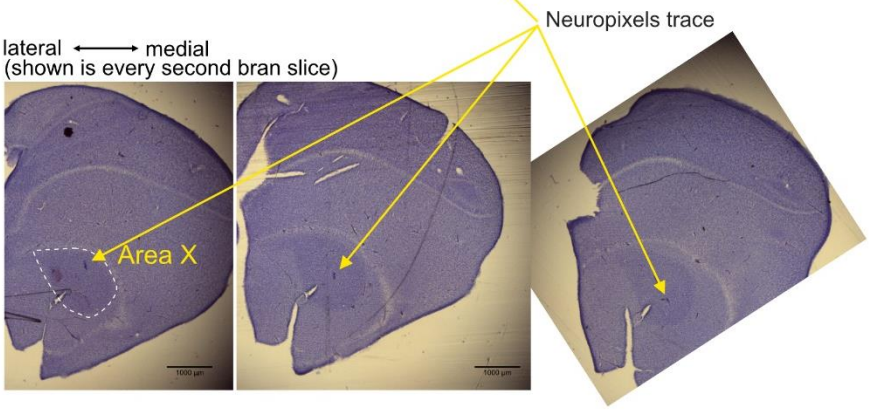


---

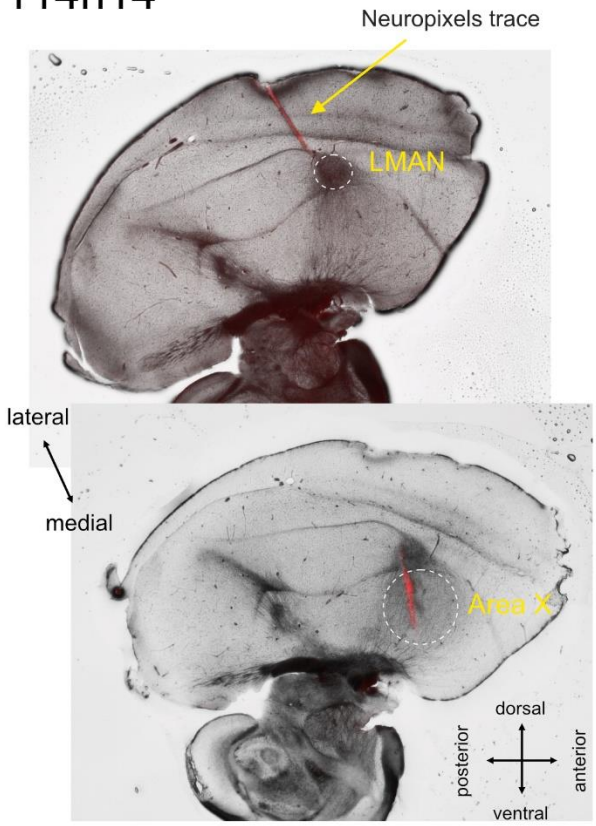
o11y3

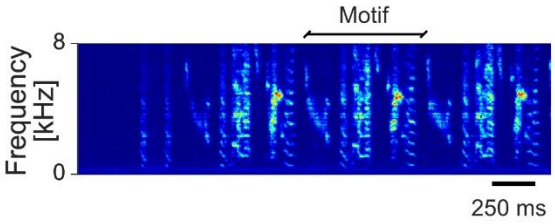




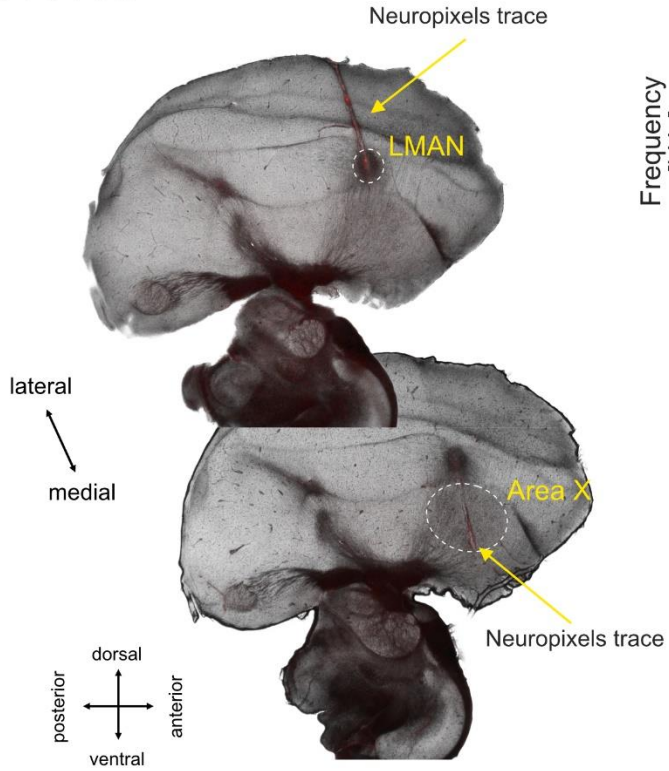


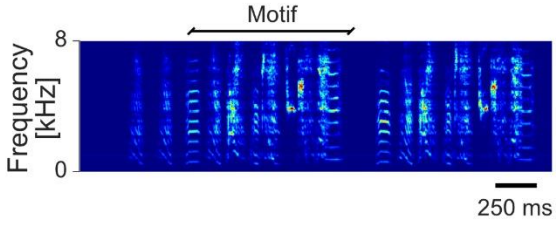
r14n14



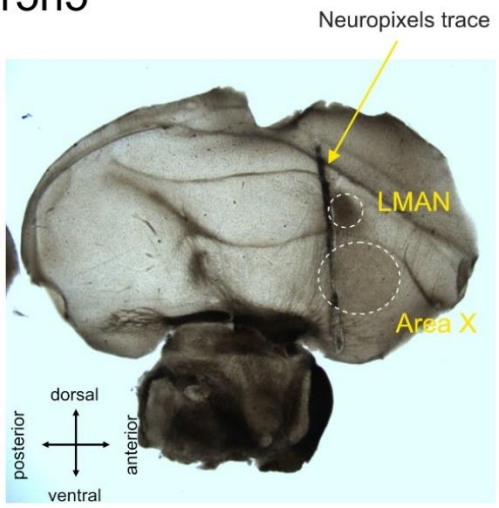


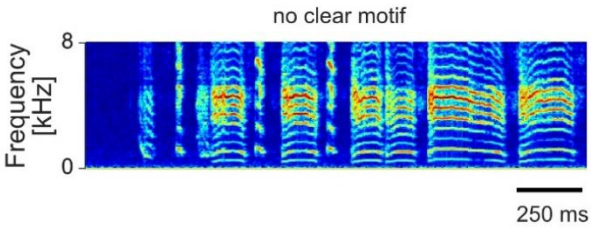
r15v15



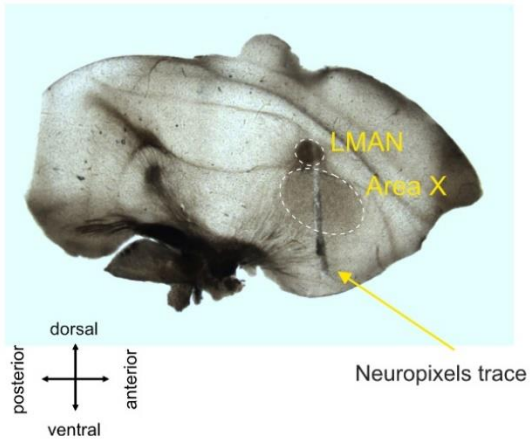


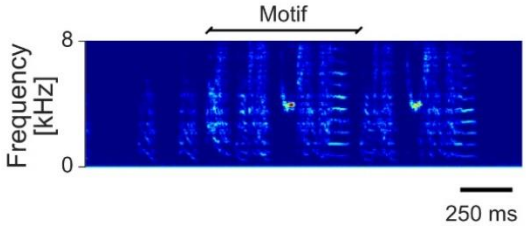
r5n5



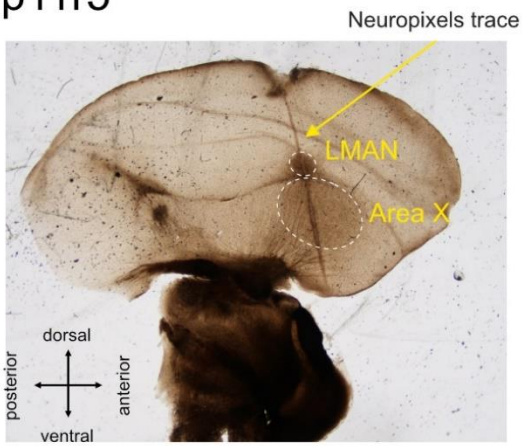


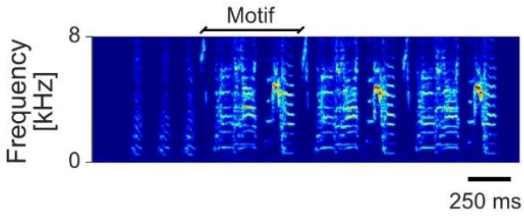
r11n11





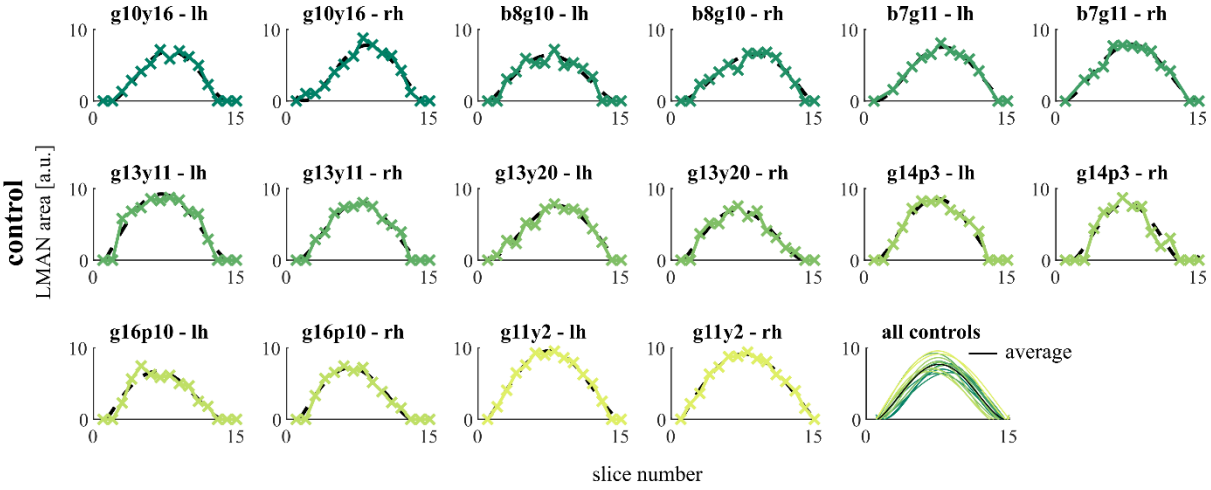
p11r5

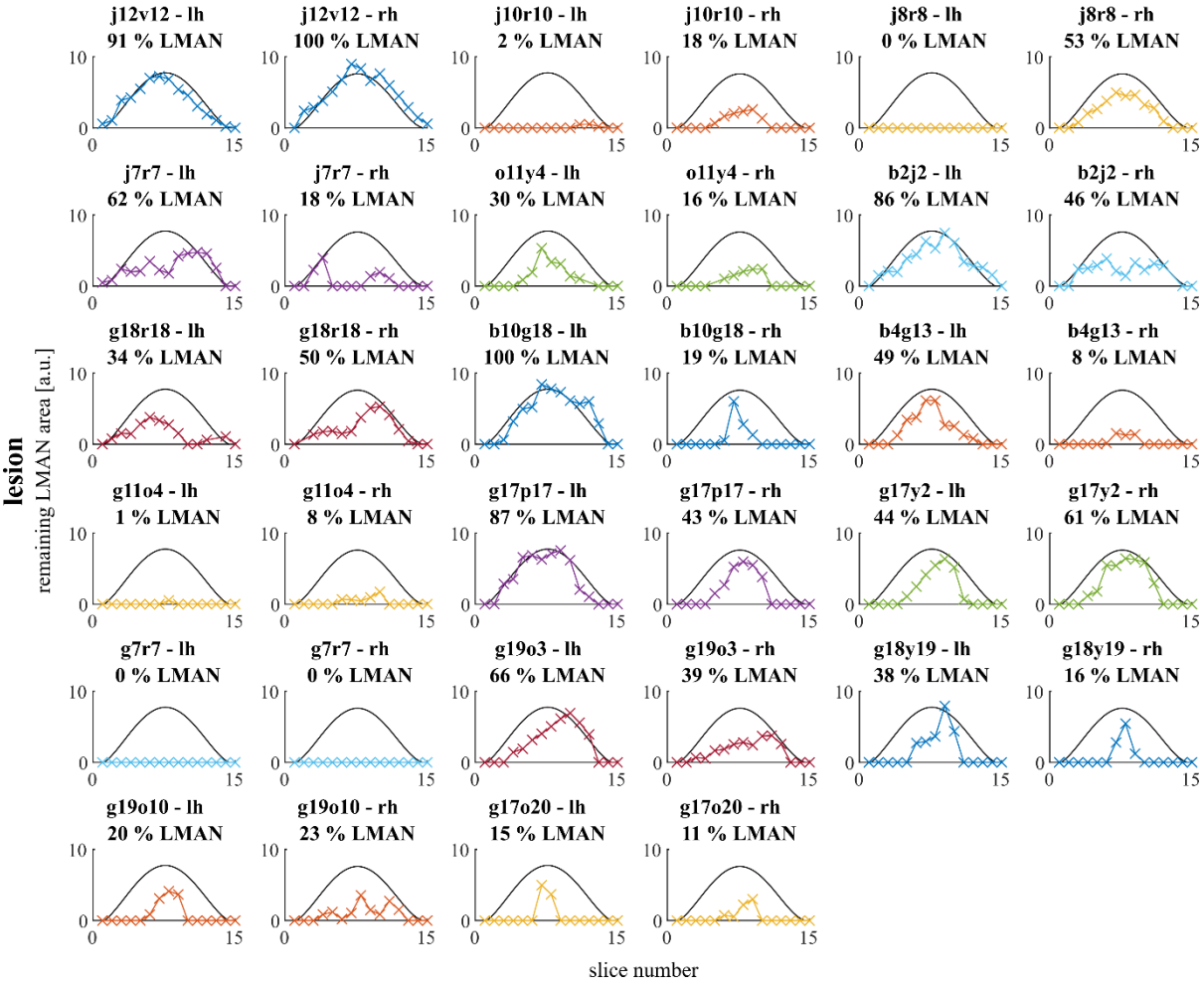




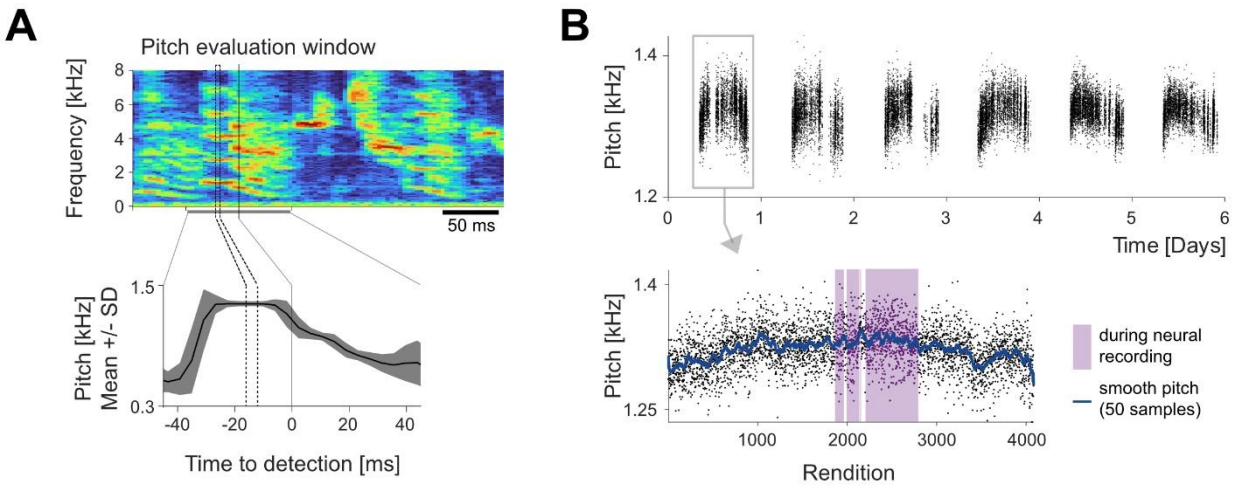
Appendix Chapter 2

Supplementary figures

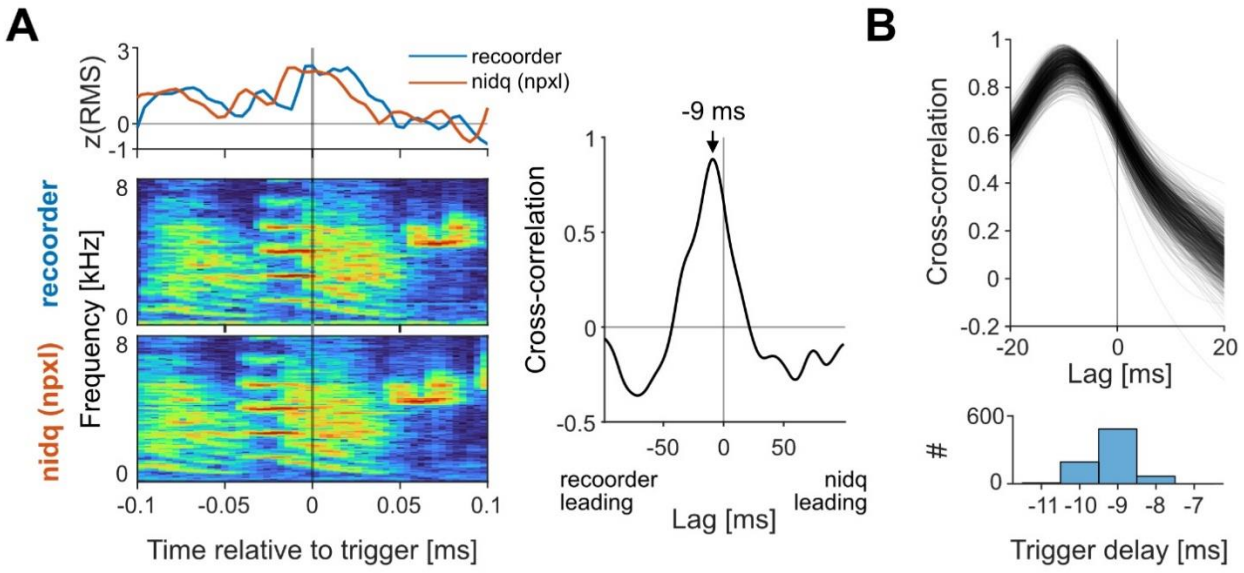




Suppl. Figure 2.1 LMAN extent in slices of healthy and lesioned brains *Shown is the LMAN volume in each hemisphere of the eight controls (top) and seven lesioned birds (bottom panel). The volume is plotted as a function of consecutive sagittal brain slices. The evolution of LMAN in controls was fit with a polynomial function (dashed line). The average fit (upper panel, bottom right) is shown as a reference for remaining LMAN volume in lesioned brains (bottom, solid line). The data from experimental birds was centered to best fit the reference and does not reflect possible lateralization.*



Suppl. Figure 3.2 Atomic pitch evaluation and trajectories for g4r4. A *The atomic pitch was evaluated post-hoc in a 16 ms long window 16 ms before the online detection of the syllable (hop size 4 ms). Shown is the spectrogram of one syllable rendition (top), and the mean pitch and standard deviation across renditions of one day as a function of time (n = 4086).* ***B*** *Atomic pitch trajectories over the course of several days (top) and a zoom in into the first day with highlighted periods of simultaneous neural recording.*



Suppl Figure 3.3 Temporal alignment of nidq recordings to RecOOrder recordings. *The automatic detection of the harmonic stack as well as the pitch computations are performed by custom-written LabView software (RecOOrder, see Methods). Upon detection, it sends a digital signal to the nidq system that records the neural data in combination with the same microphone signal. The alignment of the two microphone signals to the digital trigger shows a temporal delay in the spectrograms and its z-scored root-mean-square (zRMS, A, left). The peak in the cross-correlation function between the zRMS at -9 ms indicates that the detection in the nidq system was 9 ms delayed in this example (A, right). The delay was computed and corrected for each detection (B, top). The temporal lag varied; it was distributed around -9 ms and ranged between -11 and -7 ms (B, bottom).*

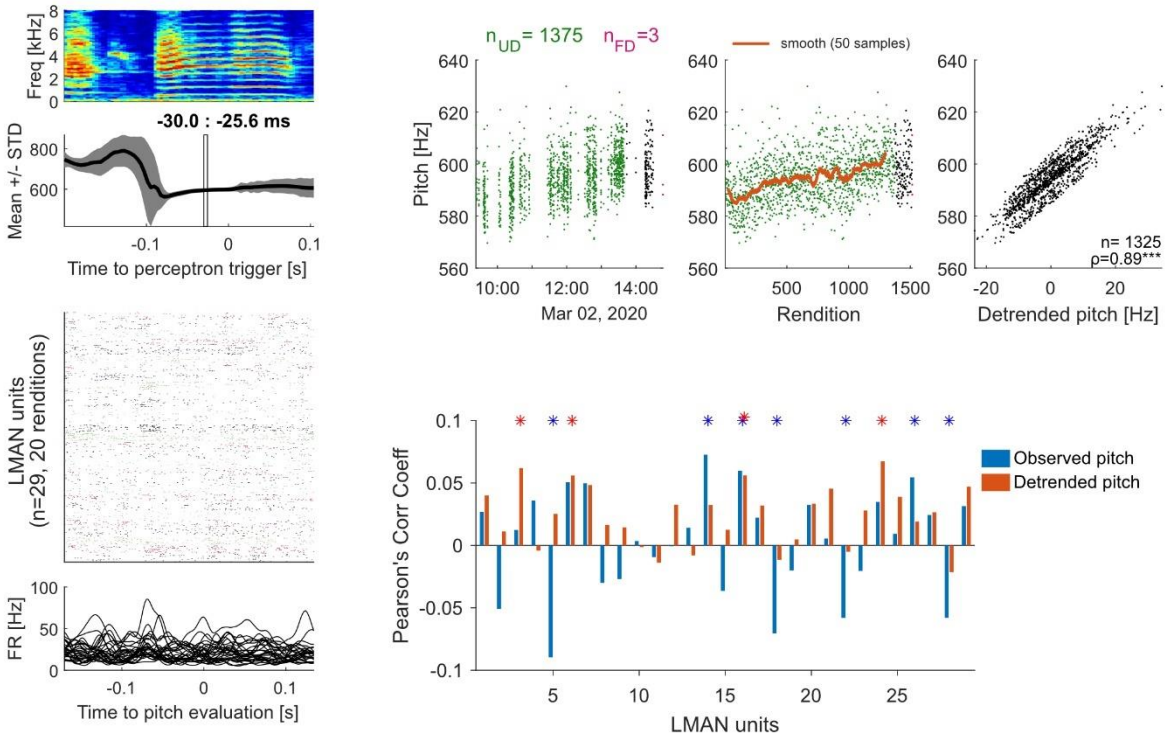
Supplementary data overview

The following shows a data summary for each bird and target syllable evaluated together with results. The upper left panel shows the spectrogram of one syllable rendition together with the pitch statistics as a function of time below. The gray bar highlights the time window of the atomic pitch. On the bottom left spike trains of all LMAN units in 20 renditions are shown, units are shown in different colors, followed by the smoothed histogram (gaussian, 10 ms SD) of all units.

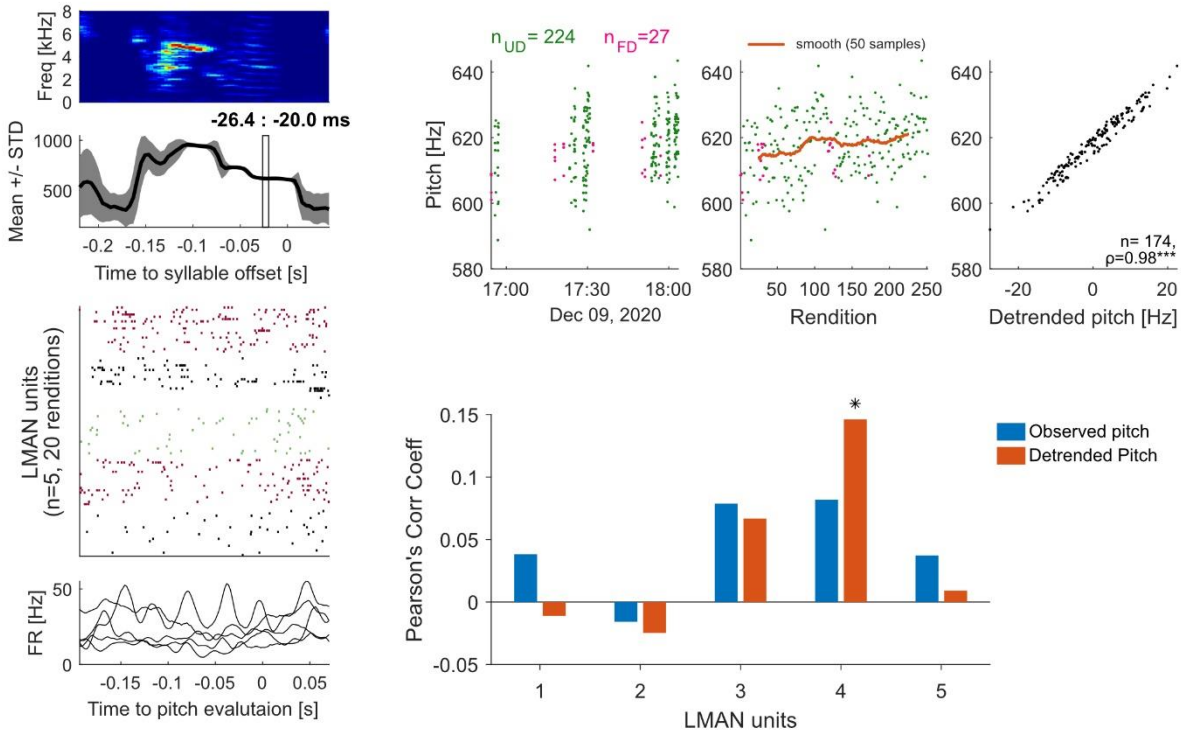
The upper right shows the atomic pitch of each rendition as a function of time (left) and rendition (center) and plotted against the detrended atomic pitch values (right). The detrended atomic pitch was computed as the residual after smoothing the trajectory with a moving averaged over 50 samples (center, orange).

The lower left plot shows Pearson's correlation coefficients between spiking activity in the premotor window and atomic pitch (blue) or detrended pitch (orange), respectively, for each unit. Significant correlations are marked with a star ($p < 0.05$).

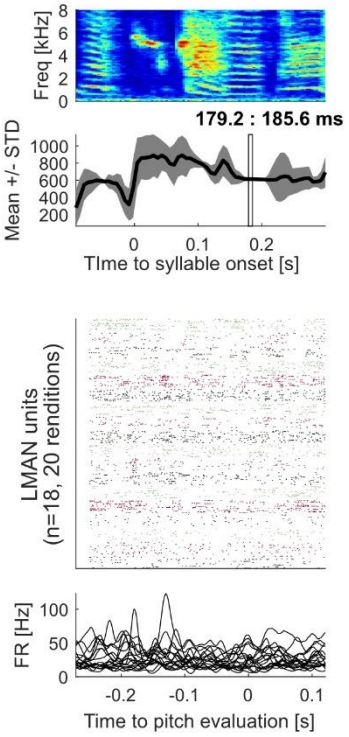
j8v8

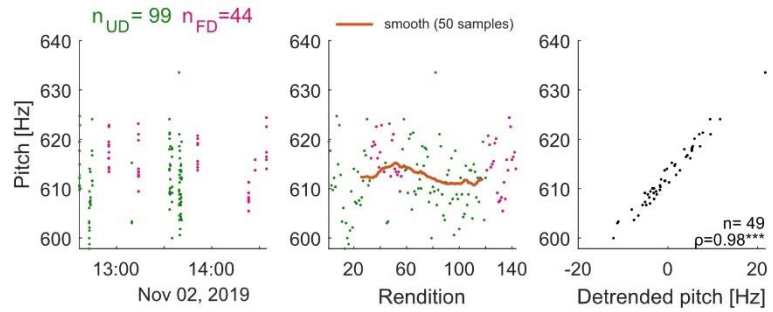

Freq [kHz]
-30.0 : -25.6 ms
Mean +/- STD
Time to perceptron trigger [s]
LMAN units
(n=29, 20 renditions)
FR [Hz]
Time to pitch evaluation [s]
n UD = 1375
n FD =3
smooth (50 samples)
Pitch [Hz]
Mar 02, 2020
Rendition
n= 1325
ρ=0.89***
Detrended pitch [Hz]
Pearson's Corr Coeff
Observed pitch
Detrended pitch
LMAN units

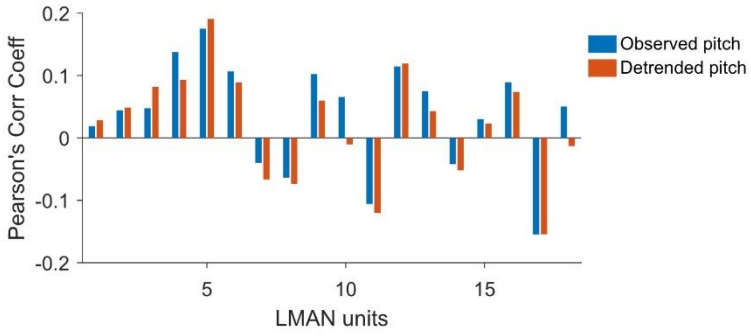
r14n14


Freq [kHz]
-26.4 : -20.0 ms
Mean +/- STD
Time to syllable offset [s]
LMAN units
(n=5, 20 renditions)
FR [Hz]
Time to pitch evalutaion [s]
n UD = 224
n FD =27
smooth (50 samples)
Pitch [Hz]
Dec 09, 2020
Rendition
n= 174,
ρ=0.98***
Detrended pitch [Hz]
Pearson's Corr Coeff
Observed pitch
Detrended Pitch
LMAN units

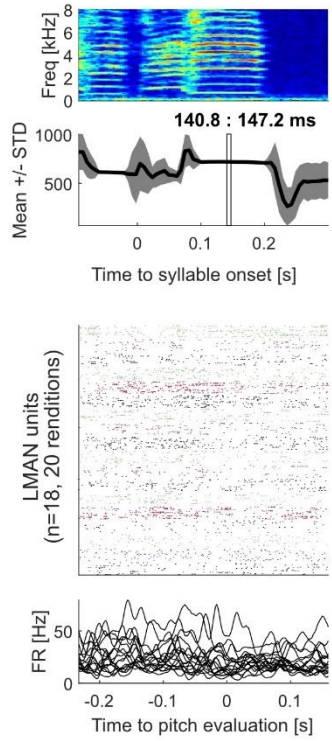
o11y3 - syllable 2

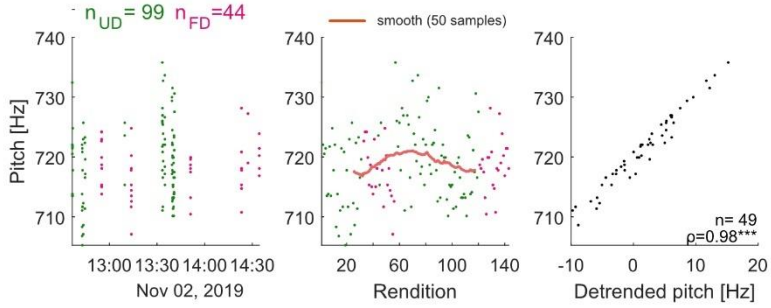


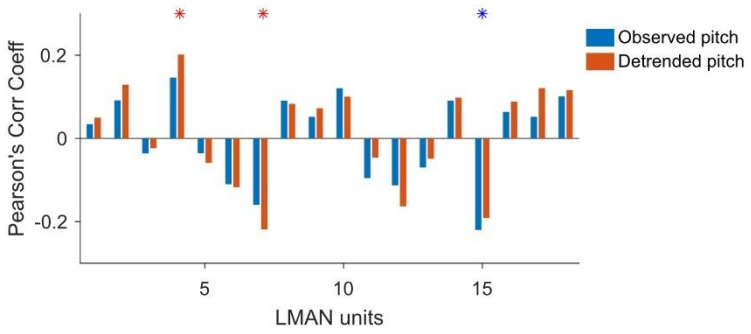




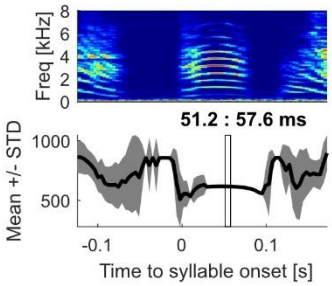
o11y3 - syllable 3

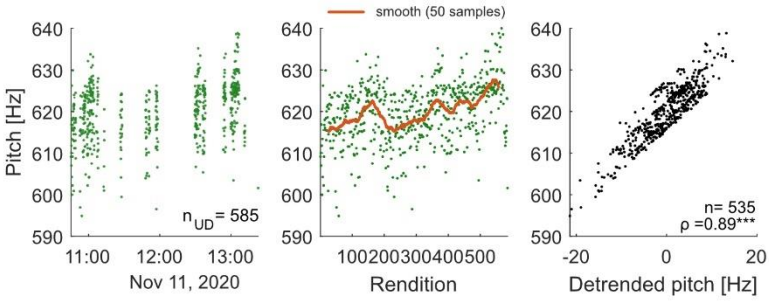


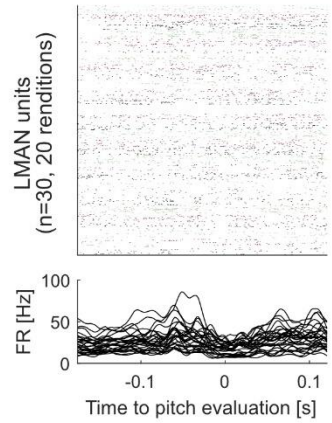


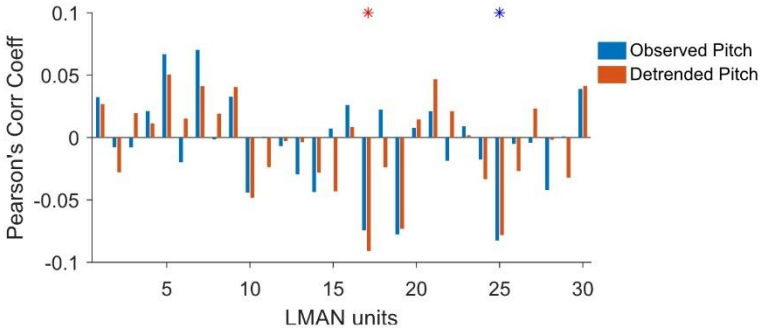


r15v15

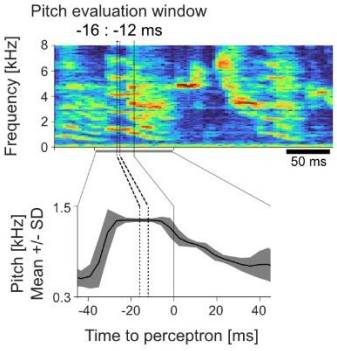

Freq [kHz]
8
6
4
2
0
51.2 : 57.6 ms
Mean +/- STD
1000
500
-0.1
0
0.1
Time to syllable onset [s]

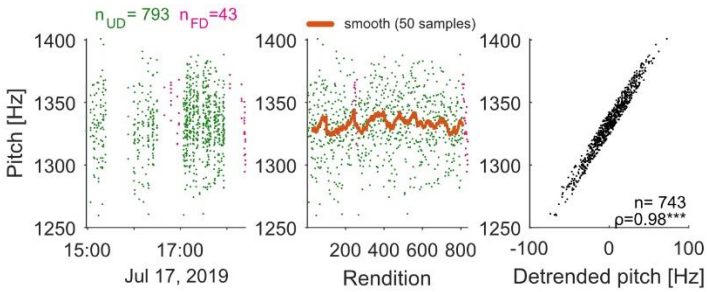

smooth (50 samples)
Pitch [Hz]
640
630
620
610
600
590
n UD = 585
11:00
12:00
13:00
Nov 11, 2020
100200300400500
Rendition
n= 535
ρ =0.89***
-20
0
20
Detrended pitch [Hz]

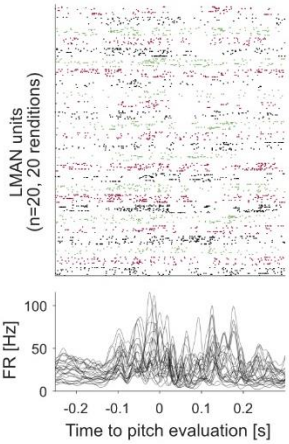

LMAN units
(n=30, 20 renditions)
FR [Hz]
100
50
0
-0.1
0
0.1
Time to pitch evaluation [s]

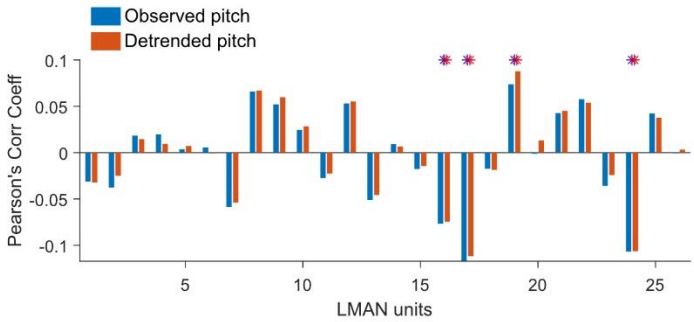

Pearson's Corr Coeff
0.1
0.05
0
-0.05
-0.1
Observed Pitch
Detrended Pitch
5
10
15
20
25
30
LMAN units

g4r4


Pitch evaluation window
-16 : -12 ms
Frequency [kHz]
8
6
4
2
0
50 ms
Pitch [kHz]
Mean +/- SD
1.5
0.3
-40
-20
0
20
40
Time to perceptron [ms]


n UD = 793
n FD =43
smooth (50 samples)
Pitch [Hz]
1400
1350
1300
1250
15:00
17:00
Jul 17, 2019
200 400 600 800
Rendition
n= 743
ρ=0.98***
-100
0
100
Detrended pitch [Hz]


LMAN units
(n=20, 20 renditions)
FR [Hz]
100
50
0
-0.2
-0.1
0
0.1
0.2
Time to pitch evaluation [s]


Observed pitch
Detrended pitch
Pearson's Corr Coeff
0.1
0.05
0
-0.05
-0.1
5
10
15
20
25
LMAN units

Appendix III – Cortico-striatal network dynamics during sleep in songbirds

Cortical-striatal network dynamics during sleep in a songbird

Corinna Lorenz[1,2]*, Anindita Das[2]*, Eduarda Gervini Zampieri Centeno[3,5]*, Hamed Yeganegi[4], Robin Duvoisin[1], Roman Ursu[3], Aude Retailleau[3], Nicolas Giret[2], Arthur Leblois[3], Richard H. R. Hahnloser[1], and Janie M. Ondracek[4]

[1] Institute of Neuroinformatics, University of Zurich and ETH Zurich, 8057, Zurich, Switzerland

[2] Université Paris-Saclay, CNRS, Institut des neurosciences Paris-Saclay, 91400 Orsay, France

[3] Bordeaux Neurocampus, 146 rue Léo-Saignat - 33076 Bordeaux Cedex

[4] Technical University of Munich, TUM School of Life Sciences, Chair of Zoology, Liesel-Beckmann-Straße 4, 85354 Freising, Germany

[4] Amsterdam UMC location Vrije Universiteit Amsterdam, Anatomy and Neurosciences, De Boelelaan 1117, Amsterdam, The Netherlands

(*) Authors contributed equally

Abstract

During vocal learning in songbirds, juvenile songbirds use auditory feedback and motor learning to transition from acoustically simple, highly variable songs to complex and stereotypical adult songs. Similar to skill learning in mammals, vocal learning in songbirds requires a set of interconnected brain areas that make up a basal ganglia-thalamocortical circuit known as the anterior forebrain pathway. Within the AFP, the pallial structure LMAN exerts a direct influence on vocal output though its projection to premotor area RA, and through this influence on the motor pathway, drives variability that can be used for learning. Area X, part of the avian striatum, receives an efference copy of activity in LMAN, thus placing Area X in a position to evaluate exploratory activity in LMAN in the context of the ongoing song. While it is clear that cortico-basal ganglia circuits are required for learning in both birds and mammals, little is known about the specific activity patterns that drive cortico-striatal plasticity, or when they occur during learning. One intriguing possibility is that neural activity patterns during 'offline' periods such as sleep may have a central role in driving cortico-striatal plasticity during learning. We used chronically implanted Neuropixels probes to investigate the cortico-striatal network dynamics during natural sleep in male zebra finches. Zebra finches share many features of sleep with mammals, including transitions through slow wave sleep (SWS) and rapid-eye movement (REM) sleep. We found: (1) Pallial and striatal neurons transition from sparse synchronous firing to asynchronous firing during sleep;(2) Cross-areal (LMAN-AreaX) gamma-band coherence is low and does not change as a function of sleep stage, whereas the cross-areal beta-band coherence is significantly modulated by sleep stage. Intra-areal LFP coherence in delta, theta and beta bands are high and modulated by sleep stage; (3) Phases of high delta activity in the LFP during sleep

are associated with increased spiking variability for LMAN neurons but not Area X; (4) LMAN neurons show reactivation-like bursting activity during REM sleep.

Introduction

Cortical and basal ganglia (BG) circuits regulate behavioral variability, as evidenced by studies of habit and skill learning in mammals. [1–5] In the case of skill learning, the ability to consistently produce a skilled action is accompanied by emerging coordinated neural activity across the motor cortex and striatum during action execution. [5–7]

Vocal learning in songbirds is a developmentally restricted learning paradigm, during which juvenile songbirds use auditory feedback and motor learning to transition from acoustically simple, highly variable “subsongs” to complex and stereotypical adult songs [citation]. Similar to skill learning in mammals, vocal learning in songbirds requires a set of interconnected brain areas that make up a basal ganglia-thalamocortical circuit (Figure 1a) known as the anterior forebrain pathway (AFP).[8–12]

Within the AFP, the pallial structure LMAN (lateral magnocellular nucleus of the nidopallium) exerts a direct influence on vocal output though its projection to premotor area RA (robust nucleus of the arcopallium), and through this influence on the motor pathway, drives variability that can be used for learning.[13] Area X (proper name), part of the avian striatum, receives an efference copy of activity in LMAN,[14] thus placing Area X in a position to evaluate exploratory activity in LMAN in the context of the ongoing song.

While it is clear that cortico-striatal circuits are required for learning in both birds and mammals [citation], little is known about the specific activity patterns that drive cortico-striatal plasticity, or when they occur during learning. One intriguing possibility is that neural activity patterns during 'offline' periods such as sleep may have a central role in driving cortico-striatal plasticity during learning.

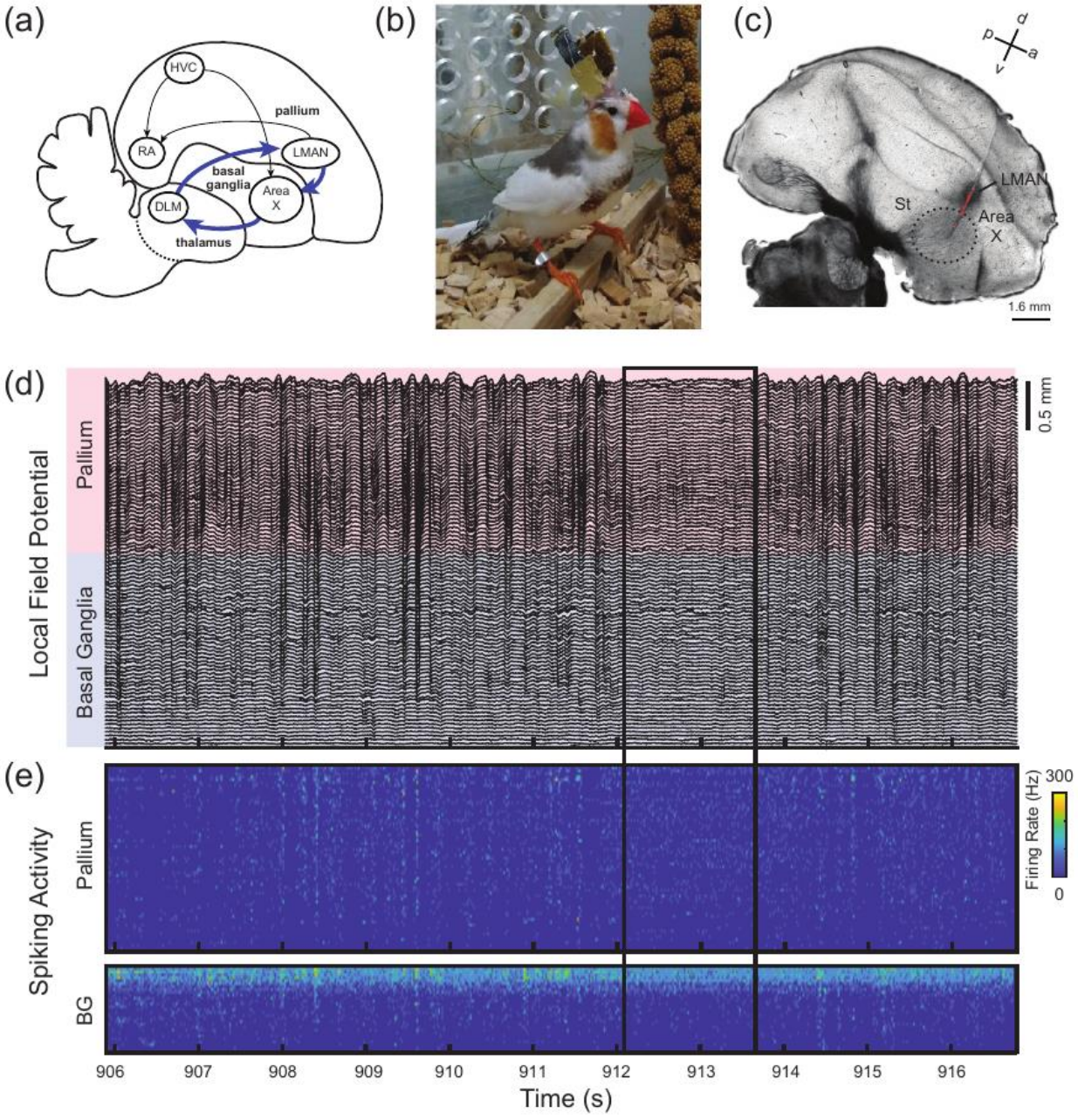
This possibility is motivated by evidence that 'reactivations' of training-related neural activity patterns during sleep promote motor skill learning in mammals [15,16] and songbirds. [17–19] However, the specific activity patterns that may impact cross-area connectivity during sleep in birds remains unknown.

We used chronically implanted neuropixel probes to investigate the cortico-striatal network dynamics during natural sleep in male zebra finches. Zebra finches share many features of sleep with mammals, including transitions through slow wave sleep (SWS) and rapid-eye movement (REM) sleep. [20,21]

Results

Pallial and BG neurons transition from sparse synchronous firing to asynchronous firing during sleep

We investigated cortico-striatal dynamics during sleep in male zebra finches (n = 7) chronically implanted with a Neuropixels probe (Figure 1b). The probe was implanted such that both the pallial structure LMAN and the striatal structure Area X within the avian BG were targeted (Figure 1c).

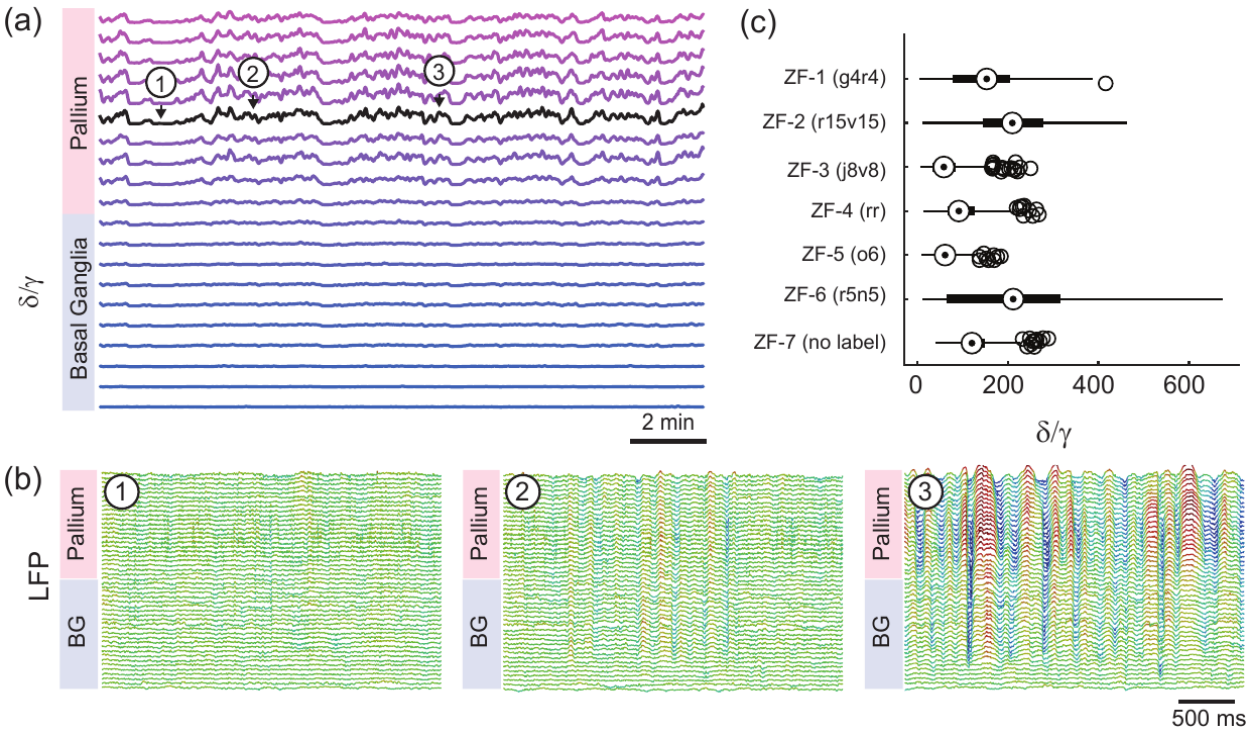


Appendix III Figure 1: Cortical-basal ganglia dynamics during sleep in a songbird. (a) *Schematic depicts the basal ganglia-thalamocortical circuit (blue arrows). Black arrows indicate other important connections of the AFP. HVC, proper name; RA, robust nucleus of the arcopallium; LMAN, lateral magnocellular nucleus of the nidopallium; Area X, proper name; DLM, medial dorsolateral nucleus of the thalamus.* ***(b)*** *Image of a male zebra finch implanted with a Neuropixels probe.* ***(c)*** *Histological brain slice depicting the Neuropixels electrode track (red DiI fluorescence). Area X is indicated with a dotted black line. St, striatum; d, dorsal; a, anterior; v, ventral; p, posterior; for other abbreviations, see (a).* ***(d)*** *10 s segment of local field potentials recorded during sleep. Recording sites were located in the pallium (red shading) and basal ganglia (blue shading) and are organized according to depth.* ***(e)*** *Corresponding spiking activity for identified pallial or striatal neurons. Firing rate is indicated with the color scale, and neurons are sorted according to firing rate. Note how neural activity transitions from sparse and synchronous firing aligned to the LFP troughs to asynchronous firing at 912-914 s (black box).*

We observed that during sleep, slow wave delta activity (1-4 Hz) was widespread throughout the pallium and striatum, spanning more than 4 mm (Figure 1d). Neural spiking was synchronized with the slow wave dynamics and was observable as sparsely bursting patterns of activity in both pallial and striatal neurons (Figure 1d). We also observed fast transitions from periods of slow wave-dominated, bursting activity to short periods of asynchronous firing (Figure 1d, e, gray shading).

We tracked transitions from high delta activity to low delta activity by calculating the ratio of delta activity (δ, 1-4 Hz) to gamma activity (γ, 30-90 Hz). This method is analogous to metrics that have been used with rodents,[22,23] but accounts for the dominant frequency bands present during avian sleep.[20,21,24]

We found that the δ\γ ratio tracked sleep transitions for pallial sites better than for deeper striatal sites (Figure 2a). When using a centrally-located pallial site to calculate δ\γ (Figure 2a, black line), low δ\γ values corresponded to putative REM states and high δ\γ corresponded to putative SWS states (Figure 2b). Median δ\γ values, calculated over the entire duration of sleep, showed some variability across birds (Table 1, Figure 2c) but were within the range of values previously reported using similar metrics.[21].

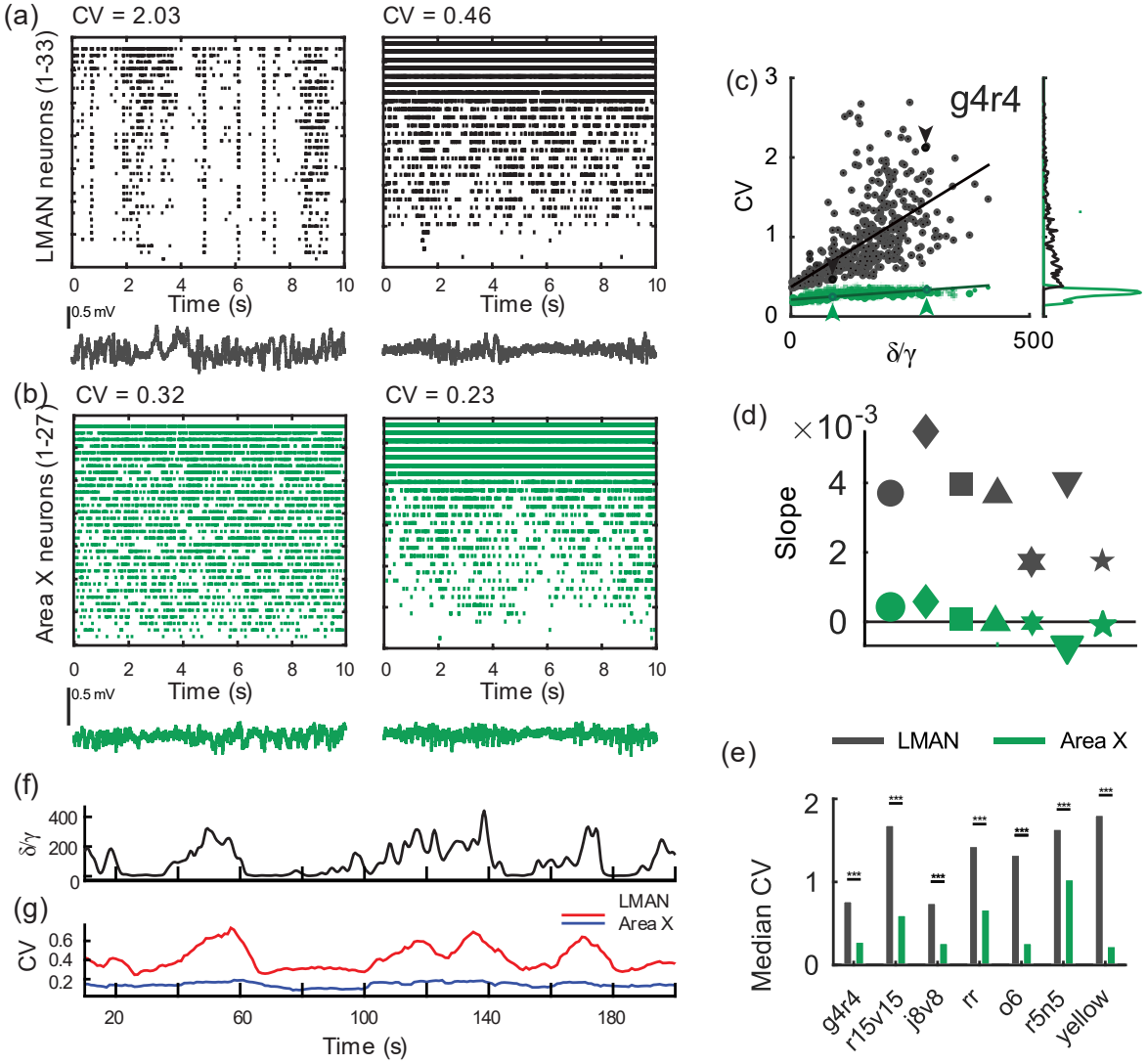


Appendix III Figure 2: Identification of sleep stages and representative neural dynamics*. (a) Ratio of δ/γ power for consecutive depths along the Neuropixels recording electrode. Each trace is separated by 200 microns. Black trace indicates a centrally located pallial site. Numbers and arrows indicate low (1), middle (2), and high (3) δ/γ values. (b) 3-second segments of local field potential activity from pallial and BG sites for the low, middle, and high δ/γ values indicated in (a). Each trace is separated by 80 microns. Note the large amplitude slow wave dynamics present for high δ/γ values indicated in (3). Color scale for visualization purposes only. Data from ZF g4r4. (c) Bar plots indicate average δ/γ values calculated over the duration of the sleep phase per bird. The box bottom and top edges represent the 25th and 75th percentile, respectively, and the middle dot represents the median. Whisker extends to the 25-75 percentile, and outliers beyond this range are indicated as open circles.*

Bird	Median	iqr
g4r4	153.5	127.2
r15v15	210.1	134.3
j8v8	58.84	51.1
rr	92.0	60.4
o6	61.6	33.6
r5n5	212	250.7
Yellow	120	51.8

Appendix III Table 1: *Median δ/γ values for three zebra finches. iqr, inter-quartile range.*

Phases of high delta activity in LFP during sleep associated with increased spiking variability for LMAN neurons but not Area X



Appendix III Figure 3: Spiking variability is significantly higher for LMAN neurons compared to Area X neurons. *(a) Box plots summarize δ/γ ranges for 3 zebra finches, where the box bottom and top edges represent the 25th and 75th percentile, respectively, and the middle dot represents the median. Whisker extends to all points outside of the 25-75 percentile range. (b) Scatter plots indicate the CV versus δ/γ values for 3 different zebra finches (ZF 1, ZF 2, ZF 3) and for LMAN neurons (red dots) and Area X neurons (blue dots). The marginal distributions for the CV values are plotted (right) for LMAN (red line) and Area X neurons (blue lines). Green and black arrows (ZF 1 subplot) indicate the CV and δ/γ values that correspond to the 10 s snippet of data plotted in (c) and (d). (c) Spiking raster plot for 31 LMAN neurons for a low CV value (left plot) and a high CV value (right box). Each tick corresponds to a spike. Data correspond to green and black arrows for red LMAN data in (b). (d) Spiking raster plot for 27 Area X neurons for a low CV value (left plot) and a high CV value (right box). Figure conventions same as in (c).*

We focused our next analysis on two specific structures within the AFP of songbirds: the pallial structure LMAN and a striatal structure within the songbird basal ganglia called Area X. We observed that during sleep, population-wide spiking activity occurred for pallial LMAN neurons that was often aligned to slow wave troughs in the LFP (see Figure 1d, e). In order to quantify the different patterns of spiking activity that we observed during sleep, we calculated the coefficient of variation (CV, standard deviation divided by the mean) for identified LMAN and Area X neuron populations in 20 ms bins (Figure 3b, c). We found that median CVs ranged from 0.23 (Area X) to 1.81 (LMAN; see Table 2 for individual values), and CV values were significantly higher for LMAN neurons compared to Area X neurons in all birds (median LMAN CV over all birds = 1.44 ± 0.16 (s.e.m.); median Area X CV for all birds was = 0.28 ± 0.12).

The CV can be thought of as a measure of spike train variability: CV values for the population rate will approach zero when all neurons are Poisson and statistically independent.[25] High CV values for LMAN neurons compared to lower CV values for Area X neurons suggests that the spike trains for LMAN neurons are more irregular than for Area X neurons. Indeed, when we examined spike rasters corresponding to high and low CV values for LMAN and Area X populations (Figure 3 a-c), spiking patterns were conspicuously irregular - albeit largely synchronized across the population - for LMAN neurons for high CV values. CV values were significantly lower for Area X neurons (Figure 3 d-e), capturing their regular spiking patterns and high firing rates (Figure 3 a-h).

Bird	LMAN CV median ± SEM	Area X CV median ± SEM	wilcoxon ranksum test
g4r4	0.77 ± 0.02	0.28 ± 0.00	1.54E-161
r15v15	1.69 ± 0.04	0.61 ± 0.01	3.23E-72
j8v8	0.75 ± 0.02	0.27 ± 0.00	3.13E-315
redred	1.44 ± 0.06	0.67 ± 0.01	6.01E-56
orange 6	1.33 ± 0.03	0.27 ± 0.00	3.84E-129
r5n5	1.64 ± 0.04	1.04 ± 0.01	1.03E-82
yellow	1.81 ± 0.02	0.23 ± 0.00	2.04E-219
Population	1.44 ± 0.16	0.28 ± 0.12	

Appendix III Table 2: *Median CV values were significantly higher for LMAN neurons compared to Area X neurons.*

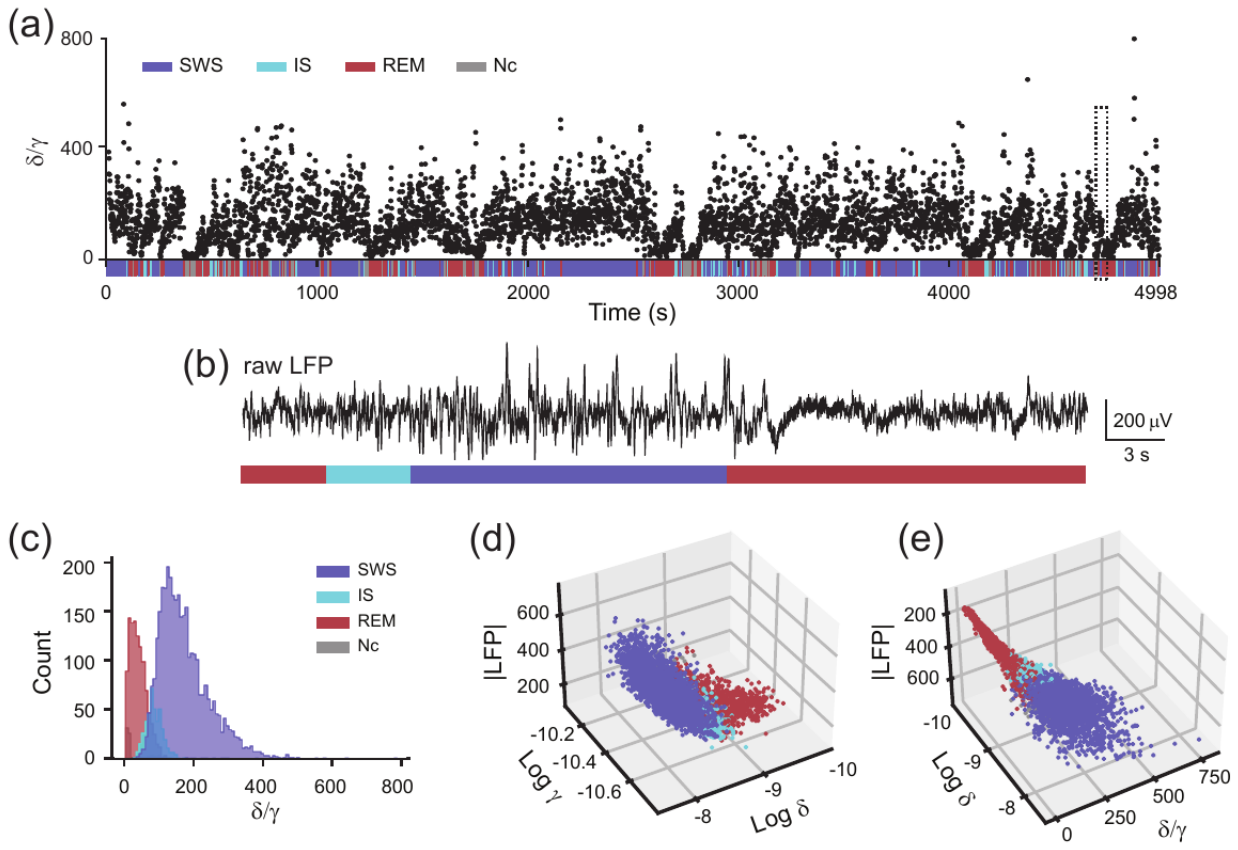
We observed that the CV values fluctuated closely with δ\γ values, such that high CV values were associated with high δ\γ values: this trend was especially obvious for LMAN neurons (Figure 3g). Indeed, we found that CV values for LMAN neurons were highly correlated with δ\γ values for almost all birds (n = 6 out of 7 birds); this was also largely true for Area X neurons as well (n = 4/7 birds).

These findings suggested that LMAN neurons are dynamically driven by the sleep state: Irregular spiking patterns emerge from populations of LMAN neurons during putative SWS (high δ\γ values), whereas regular firing patterns dominate for putative REM sleep (low δ\γ values).

	LMAN: CV versus δ\γ			Area X: CV versus δ\γ	
	R	p		R	p
g4r4	0.64	5.84E-59		0.68	2.46E-68
r15v15	0.50	1.92E-22		0.06	0.2479
j8v8	0.36	5.52E-32		0.25	4.19E-15
redred	0.17	0.0025		-0.28	4.56E-07
orange 6	0.09	0.0814		-0.02	0.6728
r5n5	0.69	1.21E-55		-0.01	0.8525
yellow	0.15	0.0001		-0.13	0.0008

Appendix III Table 3: *CV values were significantly and positively correlated with δ/γ values for 6/7 zebra finches. R, Pearson correlation coefficient.*

LFP coherence in the delta, beta and theta bands are modulated by sleep stage



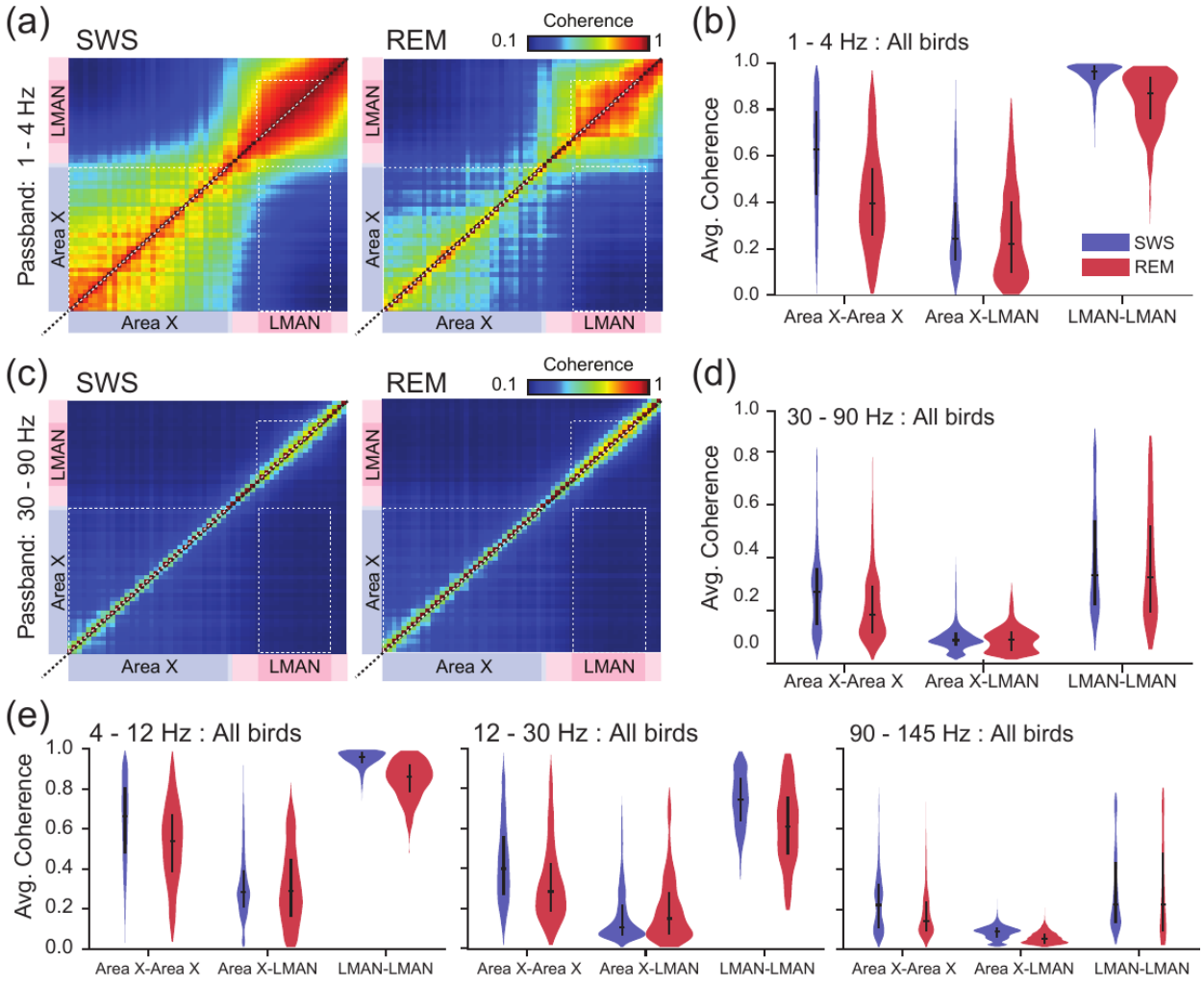
Appendix III Figure 4

Area X receives direct input from LMAN neurons. We were curious whether LMAN neurons affect Area X activity during sleep. In order to investigate functional connectivity between Area X and LMAN during sleep, we first used a clustering-based sleep scoring method [20,26] to categorize sleep into discrete bins of REM sleep, SWS, or intermediate sleep (IS). IS is a transitional stage of sleep that has been compared to the early stages of mammalian non-REM sleep.[20,26] SWS segments generally corresponded to high δ\γ values, and REM sleep generally corresponded to low δ\γ values (Figure 4A).

Then we calculated the magnitude squared coherence (coherence) between all pairs of LMAN and Area X neurons for 6-s long REM and SWS segments (Figure 5a). This technique quantifies the neuronal patterns of synchronicity measured between spatially separated recording sites.[27] We focused on several frequency bands of interest: 1-4 Hz (delta activity), 4-12 Hz (theta activity), 12-30 Hz (beta activity), and 30-90 Hz (low gamma activity), and 90-145 Hz (high gamma activity; Figure 5).

To illustrate the differences in coherence across locations as a function of sleep state, average pairwise cross-coherency plots were made for each passband of interest (Figure 5a,b). We found that pallial areas, and specifically LMAN sites (Figure 5a; pink shading) displayed high intra-area coherence for low frequency passbands during both SWS and REM sleep (Figure 5b; LMAN-LMAN). Inter-area coherence between pallium and BG were highest during putative SWS phases (Figure 4c,d; top left quadrant).

To quantify coherence between LMAN and Area X, we compared two electrodes that were located in the center of these brain areas based on histological reconstructions. The coherence of this pair of electrodes was calculated on 3-s segments of data as a function of the δ\γ value. When pooled over all δ\γ values, coherence values for the 3 zebra finches were significantly higher for the 1-4 Hz and 4-15 Hz frequency bands compared to the 30-90 Hz gamma band (Table 4; Figure 5a).



Appendix III Figure 5: LMAN - Area X coherence is low across birds for high frequency bands independent of sleep state *(a) Box-plots summarize the coherence between a pair of LMAN - Area X electrodes in 3 different birds for low (1-4 Hz), medium (4-15 Hz), and high (30-90 Hz) frequency bands. The box bottom and top edges represent the 25th and 75th percentile, respectively, and the middle line represents the median. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted as individual dots. (b) Box-plots summarize low frequency LMAN-Area X coherence for identified REM, I (Intermediate), and SWS phases. Sleep phases were identified using the IQR method. Color scheme same as in (a). (c) Box-plots summarize the low frequency LMAN-Area X coherence for identified REM and SWS phases using the K-Means segmentation method. Color scheme same as in (a). (d, e) Box-plots summarize middle frequency LMAN-Area X coherence; figure conventions same as for (b) and (c). (f, g) Box-plots summarize high frequency LMAN-Area X coherence; figure conventions same as for (b) and (c).*

Zebra Finch	1-4 Hz	4-15 Hz	30-90 Hz	ANOVA
ZF 1 (g4r4)	$0.59 \pm 1.0e\text{-}3$	$0.57 \pm 8.0e\text{-}4$	$0.25 \pm 2.5e\text{-}5$	$F(2) = 6.37e4, p < 0.001$
ZF 2 (j8v8)	$0.41 \pm 4.1e\text{-}4$	$0.42 \pm 1.6e\text{-}4$	$0.26 \pm 3.1e\text{-}5$	$F(2) = 1.24e5, p < 0.001$
ZF 3 (r5n5)	$0.54 \pm 1.5e\text{-}3$	$0.45 \pm 4.2e\text{-}4$	$0.25 \pm 2.0e\text{-}5$	$F(2) = 2.84e4, p < 0.001)$

Appendix III Table 4: *Mean coherence values were significantly different across frequency bands for three zebra finches. Mean values ± s.e.m. ANOVA, one-sided ANOVA test.*

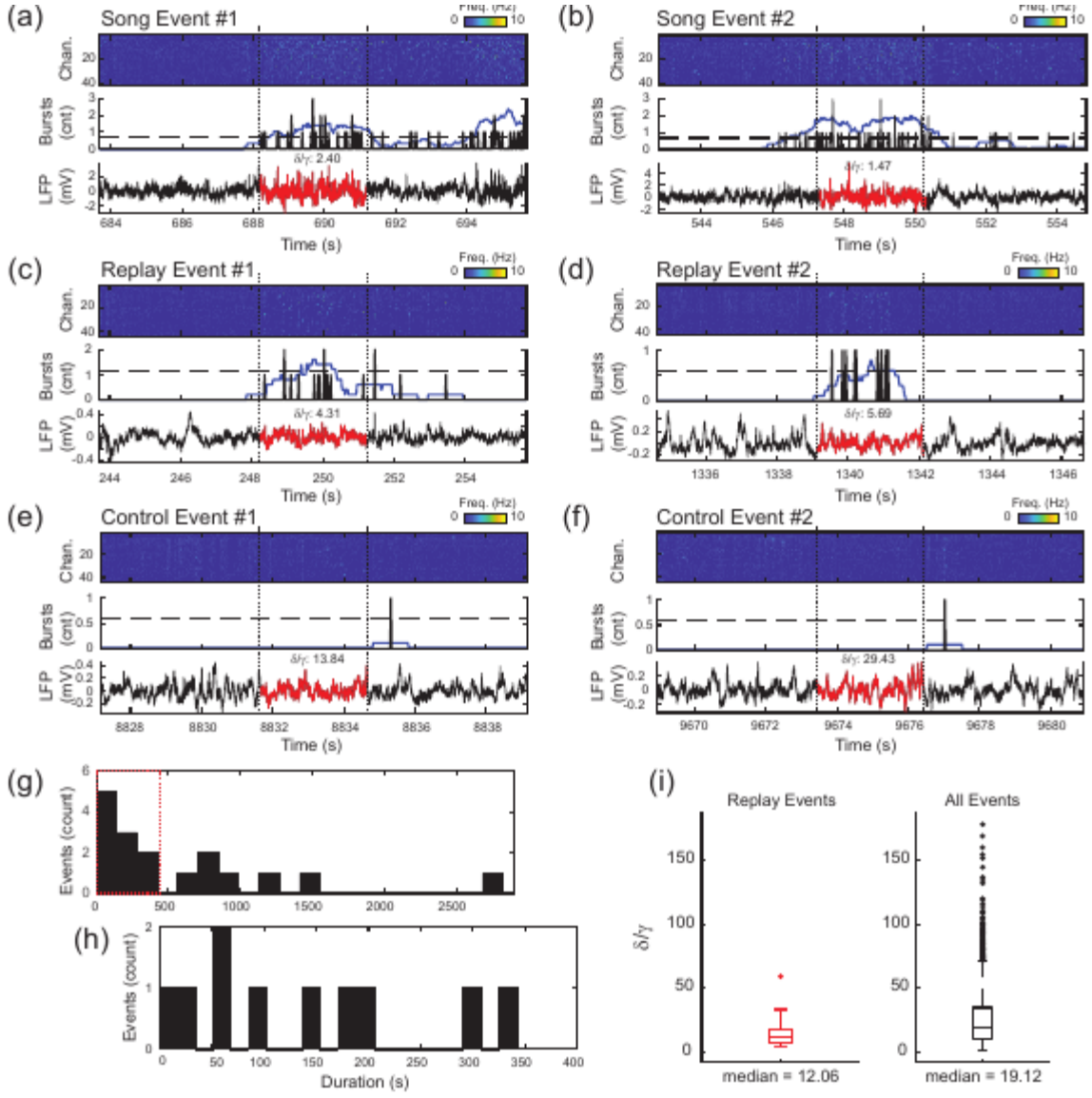
ZEBRA FINCH	1 – 4 Hz	4 – 12 Hz	12 – 30 Hz	30 – 90 Hz	90 – 145 Hz	ANOVA
blackcyan	0.34 ± 0.004	0.31 ± 0.004	0.24 ± 0.003	0.20 ± 0.002	0.21 ± 0.002	F(4) = 405.12 , p < 0.001
extnolabel	0.28 ± 0.006	0.34 ± 0.007	0.18 ± 0.004	0.12 ± 0.003	0.11 ± 0.003	F(4) = 461.63 , p < 0.001
g4r4	0.60 ± 0.003	0.72 ± 0.002	0.44 ± 0.002	0.21 ± 0.002	0.15 ± 0.002	F(4) = 12444.15, p <0.001
j8v8	0.27 ± 0.005	0.39 ± 0.005	0.25 ± 0.004	0.18 ± 0.003	0.14 ± 0.002	F(4) = 668.08 , p < 0.001
nolabel	0.43 ± 0.004	0.53 ± 0.003	0.26 ± 0.003	0.11 ± 0.002	0.08 ± 0.001	
orange6	0.59 ± 0.005	0.66 ± 0.005	0.70 ± 0.003	0.32 ± 0.003	0.14 ± 0.002	
r11n11	0.53 ± 0.002	0.53 ± 0.002	0.31 ± 0.002	0.26 ± 0.001	0.24 ± 0.001	
r15v15	0.51 ± 0.003	0.51 ± 0.002	0.27 ± 0.002	0.14 ± 0.001	0.11 ± 0.001	
r5n5	0.43 ± 0.005	0.48 ± 0.005	0.24 ± 0.004	0.13 ± 0.003	0.12 ± 0.002	
redred	0.44 ± 0.004	0.50 ± 0.004	0.49 ± 0.003	0.20 ± 0.002	0.14 ± 0.002	

Appendix III Table 5: *Mean coherence values were significantly different across frequency bands for all 10 zebra finches. Mean values ± s.e.m. ANOVA, one-sided ANOVA test reported as F statistic (degrees of freedom) and p-values.*

LMAN neurons show bursting reactivation activity during sleep

We recorded neural activity in each of the three zebra finches under slightly different sleeping conditions: ZF-1 (g4r4) was recorded during a sleep session at the beginning of the night (4-5:00 pm); ZF-2 (j8v8) was recorded after a night of ``natural sleep'' at 5:00 in the morning, and ZF-3 (r5n5) was recorded during the day after the lights were turned off.

Interestingly, in ZF-2, but not in ZF-1 and ZF-3, we observed neural activity in LMAN that was highly reminiscent of ``replay activity'' reported in mammalian hippocampus (citations) and zebra finch (Long paper). In order to examine this activity in greater detail, we designed an algorithm that detected unsynchronized bursts of activity during sleep. Unsynchronized bursts of activity were present during bouts of awake singing (Figure 6a, b), and highly similar bursts of activity were present across the population of LMAN neurons during sleep (Figure 6a, b). These activity



patterns were less bursty than for awake singing, but still more patterned than control periods of activity that occurred during sleep (Figure 6e, f).

The occurrence of bursting events followed a long-tail distribution, with many events occurring within 10 minutes of each other (Figure 6g, h) and other events occurring within 30 minutes or more. We were curious whether these unsynchronized bursts occurred more frequently during phases of REM sleep or SWS. For all of the detected unsynchronized bursts, we calculated the δ\γ value associated with the LFP of a reference electrode located in LMAN. On average, the median δ\γ value for all burst detections was lower than for all of the detected events combined (Figure 6i), suggesting that these events are more likely to occur during putative REM periods.

Materials and Methods

Subjects

Eleven male zebra finches (*Taeniopygia guttata*) were used in this study; data were pooled across two laboratories in France (the Institute of Neuroscience in Orsay and the Bordeaux Neurocampus). The age of the animals ranged from 66 days post-hatch (dph) to 232 dph at the time of recording. The animals were bred and housed in the animal facilities of the laboratories. During the experiment, animals were housed individually with water and food *ad libitum* under a 14:10 hr light/dark cycle. The description of the dataset can be found in the repository where the data is stored. All experimental procedures were approved by the French Ministry of Research and the ethical committee ``Paris-Sud et Centre'' (CEEA No. 59, project 2017-25) and “Poissons Oiseaux Nouvelle Aquitaine” (CEEA No. 73, project S73) and performed under the license 2015-25 and APAFIS#13413-2018020713145795.

Surgery and experiments

Surgery

The day of the surgery, food and water sources were removed from the cage approximately 30 minutes before the start of the surgery. Animals were anesthetized with isoflurane (0.6-1.5\% inhalation) and placed in the stereotactic apparatus once the flexion reflex was no longer observable. 0.05 microliters of diluted lidocaine (0.5mg/kg, don’t remember the name of the lidocaine we use) was administered subcutaneously and a local anesthetic (0.5 g Emla® creme) was applied to the exposed head skin along the caudo-rostral axis after removing feathers and disinfecting the area with ethanol and betadine. The scalp was resected and the exposed skull was prepared for the implant by drilling several small holes in the outer bone layer. A small craniotomy contralateral to the final implant was made to place a silver wire in contact with the dura that served as the electrical ground.

The main craniotomy was made on the right hemisphere around 1.7 mm lateral to the superior sagittal sinus and 4.5 mm anterior to lambda, the confluence of sinuses, with the flat part of the

anterior skull rotated to a 50-degree angle. The positions of LMAN and Area X were mapped with high-impedance single-electrodes in order to identify the brain areas based on the firing rate properties thereof.

A Neuropixels probe version 1.0 [28] was implanted 4 mm deep to penetrate the estimated center of LMAN and Area X. The custom-made holder that carried the Neuropixels probe was fixed to the skull with dental cement and the craniotomy was sealed with a two-part silicone gel (Dow Dowsil TM) before it was covered with a custom-made casing. The headstage was fixed at a 35 degree head angle, which corresponded to the natural position of the birds' head. Finally, the ground wire was connected and open holes along the implant were sealed with fast drying two-component silicone gel (Kwik-Sil). Immediately after the surgery, a weight reliever with adjustable counterweight was added to help the animals cope with the additional weight of the implant (1.9 g).

Histology and channel estimation

Post-hoc control of the electrode location in the brain was performed through histological examination of brain tissue. Birds were sacrificed at the end of the experiment with a lethal intraperitoneal injection of Exagon (pentobarbital sodium, 400 mg/mL), perfused intracardially with PBS 0.01 M sometimes followed by 4% paraformaldehyde as fixative. The brain was removed, always post-fixed in 4% for 24 hr, and cryoprotected in 30% sucrose. We then cut 50-micron thick sections in the parasagittal or horizontal (for cerebellum) plane with a freezing microtome. Slices were mounted with Mowiol (Sigma Aldrich) and observed under epifluorescence. Images were analyzed using ImageJ software (Rasband WS, NIH, Bethesda, Maryland, USA). The electrode track as indicated by colored/fluorescent dye, was clearly visible on consecutive brain slices (see Figure 1c).

We defined the boundaries of LMAN and Area X based on darkness of the tissue in regular light or the fluorescent background, as these two brain areas showed a clear contrast with surrounding tissue due to higher myelination, cell density, and size (Niwdorf-Begweiler and Bishof, 2007). The reconstruction of electrode channels to be assigned to LMAN or X was performed based on the location of the electrode tip on the histological pictures, the coordinates of the channels on the electrode (https://www.neuropixels.org/support) and the above-mentioned boundaries of LMAN and X.

Data acquisition

Animals were housed individually in sound-attenuated recording chambers and neural recording sessions were performed once or twice per day with sessions lasting between one and four hours. Sleep-related activity was recorded either during the start or end of natural sleep. Specifically, one bird (j8v8) was recorded at 5:00 in the morning at the end of the ``natural sleep'' phase. The other two birds (g4r4 and r5n5) were recorded at the start of the natural sleep phase, beginning at 16:00 (g4r4) or 17:00 (r5n5). Extracellular activity was recorded with the chronically implanted Neuropixels probe at 30 kHz. Video recordings were acquired in one bird r5n5 to confirm sleeping

behavior. Extensive information on the dataset can be found in the data repository where our data is stored.

Data analysis

Data preprocessing and spike clustering

For neural analysis, spike sorting was performed on all active channels using Kilosort 2.0 and Kilosort 2.5 [29] and manually curated using phy (https://github.com/kwikteam/phy) or automatically curated using SpikeInterface (script available in our data repository). The channels that covered LMAN and Area X were determined based on the characteristics of the raw voltage signal, histological examination of the probe's trace in sagittal brain slices, and spiking behavior of detected spike clusters.

Sleep staging

The sleep staging was performed based on a previously described automated method in songbirds. Since we did not have EEG measurements, we computed the sleep stages from an LMAN LFP channel (identified medially located LMAN channel along the dorsoventral axis of recording) with bins of 3 sec and a window shift of 1 sec. Spectral parameters were computed from the multitaper spectral analysis (Prerau et al. 2016) of the downsampled (250 Hz) LFP. After the classification of the sleep stages, we incorporated an additional step of screening. Based on manual video observation where available and the microphone signal, each bin was classified as wake or movement upon the observation of a sustained artifact in the microphone signal (larger than *n* times the signal's chunk mean noise minus the standard deviation of the same chunk, *n* being an integer defined by visual inspection) of microphone amplitude). We computed the final sleep stages if the microphone signal and/or video also confirmed a lack of movement. Even though this process resulted in classification into 3 sleep stages, REM, SWS, and IS, the number of bins classified as IS was much smaller and shorter in duration. So, for the rest of our analysis, we continued with only the bins identified as either REM or SWS.

Coefficient of Variation

Temporal variability in the spiking patterns was quantified using the coefficient of variation for the spike rate. A sliding window of length 20 s with 50\% overlaps was used to measure the variability in spiking activity. Each window was divided into sub-windows of length 20 ms, and the total number of spikes (from all units either in LMAN or in Area X) were counted in each sub-window. The coefficient of variation (CV) was obtained as the standard deviation of spike counts divided by the mean spike count across all sub-windows in each window. Furthermore, the average firing rate of LMAN and Area X neurons was calculated as the total count of spikes (from all units either in LMAN or in Area X) per second in the same 20-s windows that CV was computed.

The Pearson's correlation coefficient was used to quantify the relevance between Delta/Gamma ratio and the CV of spike activation. In order to align these two variables, Delta/Gamma values were averaged within the corresponding 20-s windows, where CV was computed.

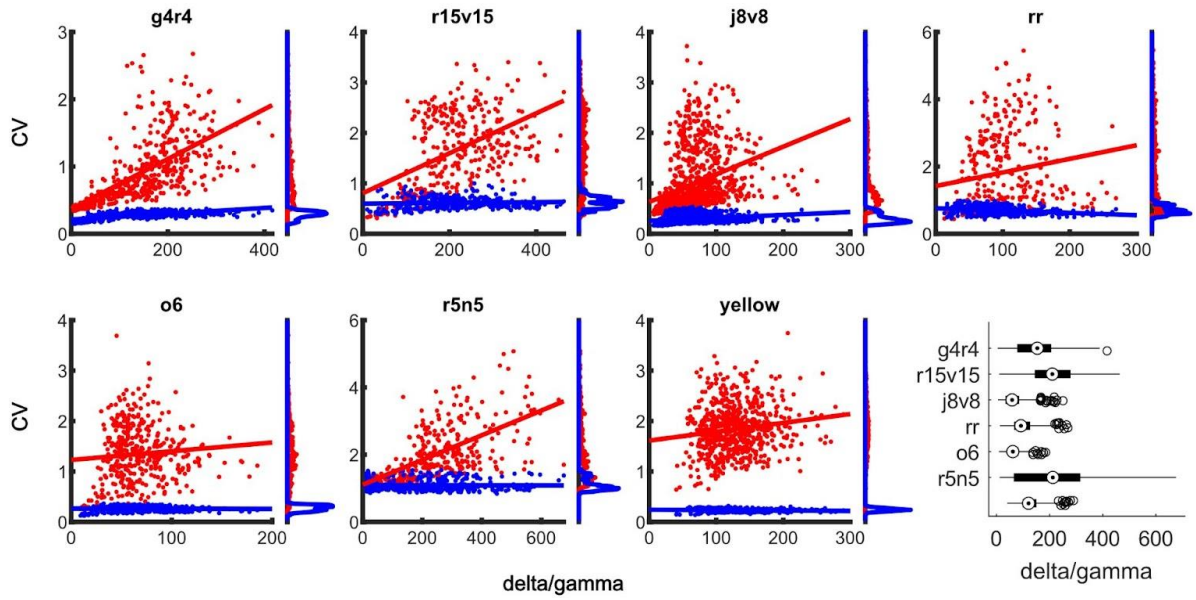
Coherency Analysis

The computation of coherence was done as follows: first, we extracted the LFP data from *.bin* files utilizing SpikeInterface (https://spikeinterface.readthedocs.io/en/latest/overview.html) with phase shift correction using *spikeinterface.preprocessing.phase_shift*. Next, we removed artifacts from the LFP signal by establishing a threshold based on 80% of the maximum absolute trace value and Z-scored the trace using *elephant.signal_processing.zscore*. After that, we identified consecutive six seconds within the same state (*chunk_length*). We used six seconds to guarantee at least six cycles for our lowest frequency of interest (FOI), *i,e,* 1Hz. Afterward, we determined the number of available chunks for each sleep state, and what was the common minimum between states (*n_chunks*). We excluded birds that had less than 15 *n_chunks* to ensure enough computation resolution. We then proceeded by segmenting and concatenating LFP data with the same *n_chunks* for REM and NREM sleep states. Finally, we computed coherence using the *scipy.signal.coherence (https://docs.scipy.org/doc/)* with the number of points per segment (*nperseg)* equal to *chunk_length* multiplied by the signal's sampling frequency, with no overlap, and detrending set to '*constant'*. Considering that NREM's available chunks were often more numerous than REM's, to prevent bias and analyze as much of the data as possible across each recording, coherence was computed *N* times (the integer division of the total NREM available chunks divided by *n_chunks*) using random sampling without replacement. The Python scripts used for this analysis can be found in our data repository.

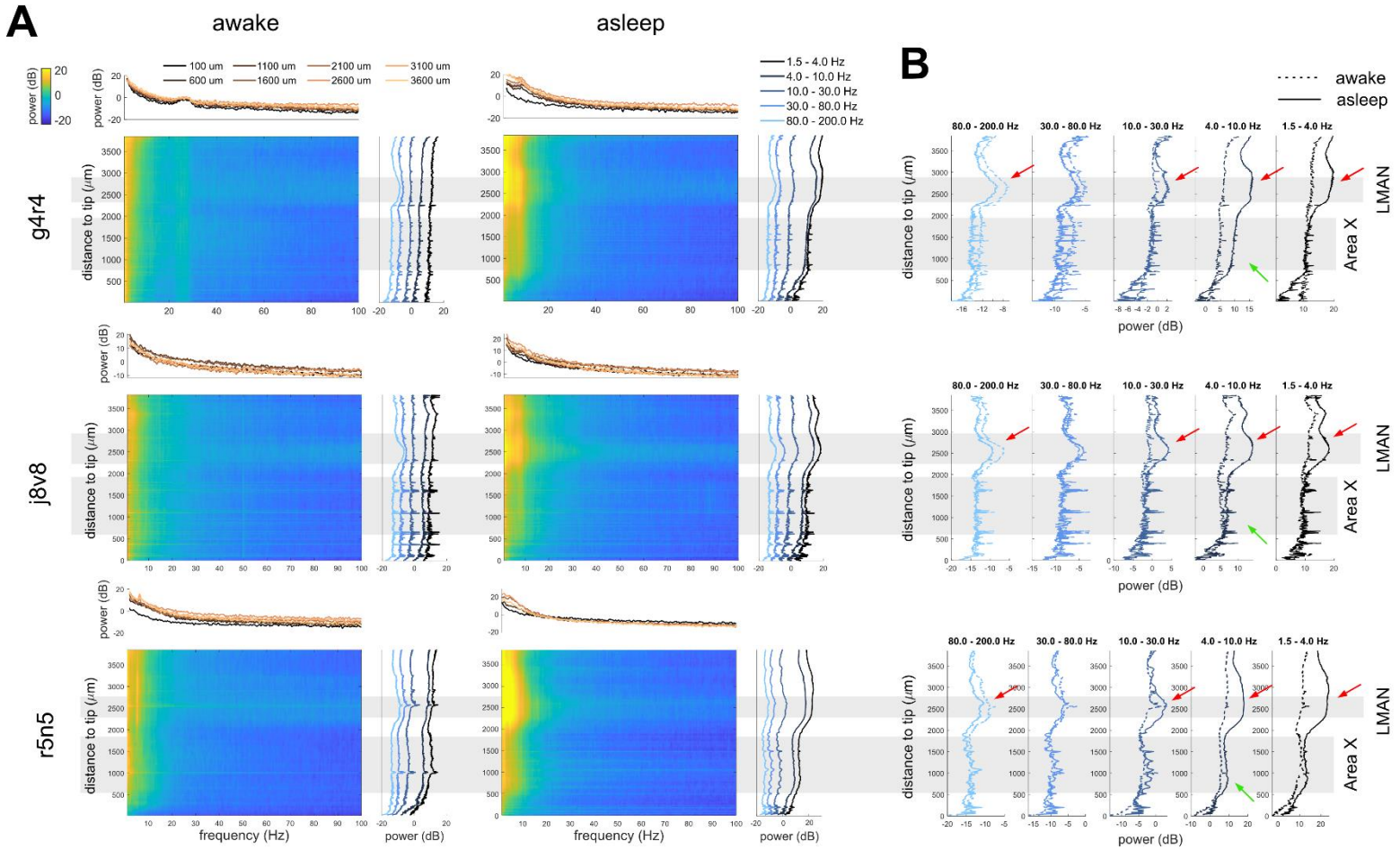
Sleep replay event detection

Sleep replays were observed in one bird as varying, unsynchronized bursts of spiking in the LMAN region similar but slightly weaker to the spiking behavior during singing. To detect replay events during sleep, we introduced a bursting index binning all spikes originating from units in LMAN in XX ms long bins and convolving the discretized signal with a 50ms boxcar kernel to distinguish it from synchronized spiking during deep sleep (Figure 3). Sleep replays were defined as times when the smoothed bursting signals surpassed a threshold of 0.12 and had a prominence of 0.1. The delta/gamma ratio during these sleep events was computed on the central channel in LMAN in a 3 s window around the center of the sleep replay independent of the length thereof.

Supplementary Figures



Appendix III Suppl. Figure 1. *Scatter plot showing CV against delta/gamma and boxplot of delta/gamma distribution (lower right) for each bird.*



Appendix III Suppl. Figure 2 Differential Sleep Modulation of Low-Frequency Power in the Pallium Compared to the Striatum A *Shown is the color-coded power spectrum density (PSD) within the LFP signal for each electrode, organized by depth and contrast states of wakefulness (left) and sleep (right) across birds. Each panel is accompanied by marginal distributions of power across frequencies (top) and depths (right). The analysis is based on the average PSD derived from 10 sec windows, selected randomly, highlighting an increase in low-frequency power and intensified gamma band activity (warmer colors) during sleep.* ***B*** *Power distribution across electrode sites provide a detailed comparison between awake (dashed lines) and asleep (solid lines) in different frequency bands. LMAN exhibits higher gamma band power (>60 Hz) that decreases during sleep, in contrast to the low-frequency bands (<10 Hz), where power increases during sleep (red arrows). In contrast, sleep modulation in Area X is largely absent, except for two instances (g4r4, j8v8), where theta band power, unlike alpha band, shows a slight increase during sleep (green arrows).*

References

1. Knowlton, B. J., Mangels, J. A. & Squire, L. R. A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).

2. Graybiel, A. M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* **31**, 359–387 (2008).

3. Wise, S. P., Murray, E. A. & Gerfen, C. R. The frontal cortex-basal ganglia system in primates. *Crit. Rev. Neurobiol.* **10**, 317–356 (1996).

4. Yin, H. H. *et al.* Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nat. Neurosci.* **12**, 333–341 (2009).

5. Santos, F. J., Oliveira, R. F., Jin, X. & Costa, R. M. Corticostriatal dynamics encode the refinement of specific behavioral variability during skill learning. *eLife* **4**, e09423 (2015).

6. Lemke, S. M. *et al.* Coupling between motor cortex and striatum increases during sleep over long-term skill learning. *eLife* **10**, 1–25 (2021).

7. Koralek, A. C., Costa, R. M. & Carmena, J. M. Temporally Precise Cell-Specific Coherence Develops in Corticostriatal Networks during Learning. *Neuron* **79**, 865–872 (2013).

8. Bottjer, S. W., Miesner, E. A. & Arnold, A. P. Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224**, 901–3 (1984).

9. Scharff, C. & Nottebohm, F. N. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci. Off. J. Soc. Neurosci.* **11**, 2896–913 (1991).

10. Brainard, M. S. & Doupe, A. J. Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* **404**, 762–6 (2000).

11. Farries, M. A. & Perkel, D. J. A Telencephalic Nucleus Essential for Song Learning Contains Neurons with Physiological Characteristics of Both Striatum and Globus Pallidus. *J. Neurosci.* **22**, 3776–3787 (2002).

12. Andalman, A. S. & Fee, M. S. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12518–23 (2009).

13. Tumer, E. C. & Brainard, M. S. Performance variability enables adaptive plasticity of ‘crystallized’ adult birdsong. *Nature* **450**, 1240–4 (2007).

14. Vates, G. E., Vicario, D. S. & Nottebohm, F. Reafferent thalamo- ‘cortical’ loops in the song system of oscine songbirds. *J. Comp. Neurol.* **380**, 275–90 (1997).

15. Ramanathan, D. S., Gulati, T. & Ganguly, K. Sleep-Dependent Reactivation of Ensembles in Motor Cortex Promotes Skill Consolidation. *PLOS Biol.* **13**, e1002263 (2015).

16. Gulati, T., Ramanathan, D. S., Wong, C. C. & Ganguly, K. Reactivation of emergent task-related ensembles during slow-wave sleep after neuroprosthetic learning. *Nat. Neurosci.* **17**, 1107–1113 (2014).

17. Elmaleh, M., Kranz, D., Asensio, A. C., Moll, F. W. & Long, M. A. Sleep replay reveals premotor circuit structure for a skilled behavior. *Neuron* **109**, 3851-3861.e4 (2021).

18. Young, B. K., Mindlin, G. B., Arneodo, E. & Goller, F. Adult zebra finches rehearse highly variable song patterns during sleep. *PeerJ* **2017**, 1–18 (2017).

19. Shank, S. S. & Margoliash, D. Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature* **458**, 73–7 (2009).

20. Low, P., Shank, S. S., Sejnowski, T. J. & Margoliash, D. Mammalian-like features of sleep structure in zebra finches. *Proc Natl Acad Sci U A* **105**, 9081–9086 (2008).

21. Yeganegi, H. & Ondracek, J. M. Multi-channel recordings reveal age-related differences in the sleep of juvenile and adult zebra finches. *Sci. Rep.* **13**, 8607 (2023).

22. Csicsvari, J., Hirase, H., Czurko, A. & Buzsáki, G. Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron* **21**, 179–89 (1998).

23. Siapas, A. G. & Wilson, M. A. Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron* **21**, 1123–8 (1998).

24. Yeganegi, H., Luksch, H. & Ondracek, J. M. Hippocampal-like network dynamics underlie avian sharp wave-ripples. *bioRxiv* 825075 (2019) doi:10.1101/825075.

25. Kobak, D., Pardo-Vazquez, J. L., Valente, M., Machens, C. K. & Renart, A. State-dependent geometry of population activity in rat auditory cortex. *eLife* **8**, (2019).

26. Canavan, S. V. & Margoliash, D. Budgerigars have complex sleep structure similar to that of mammals. *PLOS Biol.* **18**, e3000929 (2020).

27. Bowyer, S. M. Coherence a measure of the brain networks: past and present. *Neuropsychiatr. Electrophysiol.* **2**, 1 (2016).

28. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).

29. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Kenneth D., H. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* (2016) doi:https://doi.org/10.1101/061481.

Appendix IV – Nearest neighbor analysis of neurobehavioral correlates

Birdsong is a complex, learned skill that is controlled by a distinct network of interconnected brain areas. Here, we asked whether the mapping from the spiking activity in the premotor LMAN to the vocal output is continuous, i.e., if small changes in neural space relate to small changes in vocal space similar to findings of V1 responses to a high-dimensional stimulus set in mice (Stringer et al., 2019).

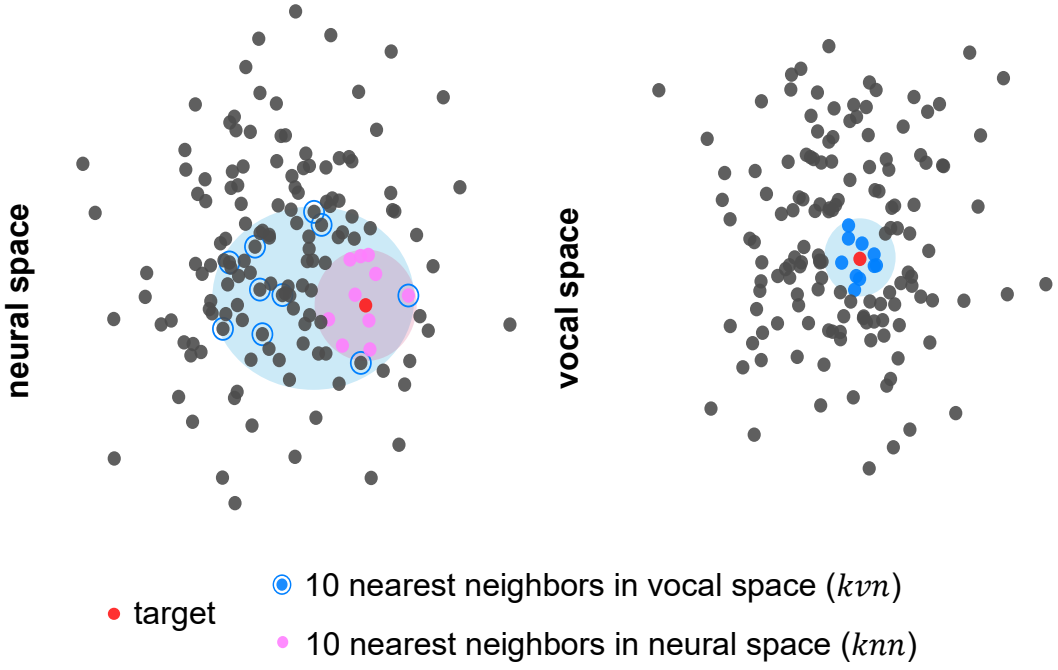


Figure IV.1 Testing for neurobehavioral continuity. *See text for details.*

We used data from chapter 1 and defined continuity in terms of neighborhoods (Kollmorgen et al., 2020). We used the motif spectrograms (see Methods Chapter 2) and LMAN population spiking as readouts that span a high-dimensional vocal and neural space, respectively. For a find grained analysis that considers time-varying effects of LMAN activity (Kao et al., 2005a), we choose a 40 ms window in both domains (step size 20 ms). Spiking times were binned in 4 ms bins.

We defined neighborhoods using the Euclidean distance d and computed the summed distance in neural space between a target rendition i and its k nearest neighbors in vocal space (kvn) relative to the summed distances in neural space to the k nearest neighbors in neural space

(knn). We call the ratio the 'expansion factor' to reflect the inflation from one space to the other (Figure IV.1). It is formally defined as:

$$expansion\ factor\ \ e_i = \ {\sum_{kvn} d\ (x_i,\ x_k)}\Big/{\sum_{knn} d(x_i,\ x_k)}$$

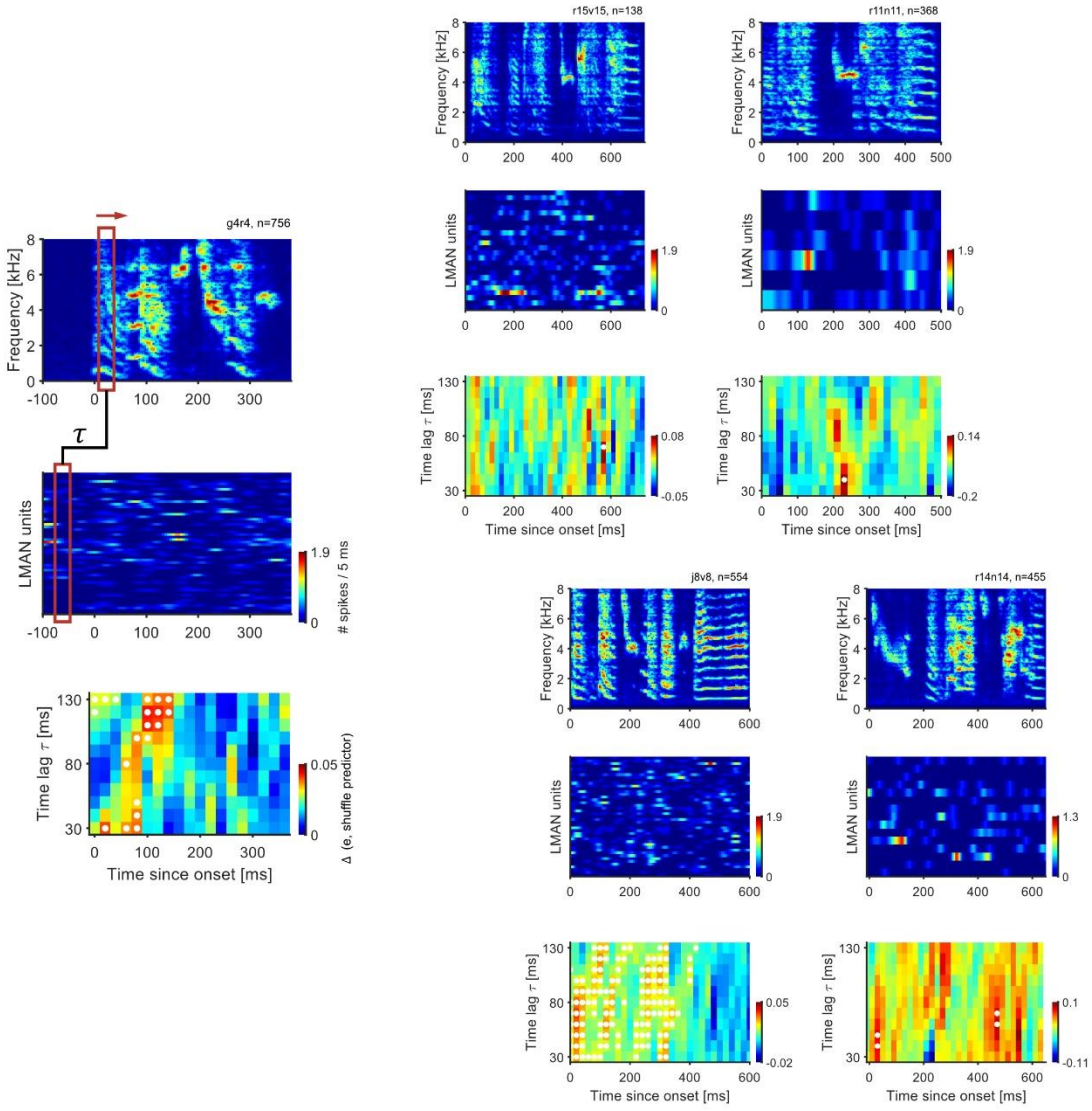


Figure IV.2 Small changes in neural space relate locally to small changes in vocal space *We tested for continuity using a 40 ms sliding window across the motif spectrogram and binned spiking signals with different time lags τ. Shown are the spectrograms and the smoothed neural population spiking of one rendition (name of the bird and number of rendition given on top of the spectrogram). The heatmap on the bottom of each panel depicts the average difference between the shuffled predictor and expansion factor. Significant differences are marked with a white star. The significance level was set to 5% and adjusted for multiple testing using Bonferroni correction.*

For statistical testing, we introduced a shuffle predictor for each neighborhood expansion where the true k neighbors in knn are replaced by k random neighbors.

We chose a neighborhood size of $k = 10$ and computed the expansion factor and shuffle predictor for pairs of LMAN population activity and motif sections at time lags τ ranging from 30 to 130 ms to account for varying premotor time lags (Kao et al., 2005a).

We hypothesized that if e is smaller than the shuffle predictor, there is evidence for neurobehavioral continuity between LMAN premotor activity and vocal output. We applied a one-sided, paired Wilcoxon signed test to assess this statistically (5 % significance level, Bonferroni corrected for each bird).

We show the results individually for each bird (n=5) in Figure IV.2. Using our parameter-free method, we find local evidence for continuity between LMAN spiking activity and vocal output. In each bird, there are moments in the motif where the neighborhood is better preserved than expected by chance (marked by white stars in the heatmap at the bottom of each panel). In some birds like r11n11 or r15v15, there is a very limited number of such moments. In j8v8 and r14n14, on the other hand, the effect is widespread. In j8v8, for instance, a continuous mapping is evident almost throughout the first half of the motif at several time lags.

Could age be a reason for the varying degree of continuity? As zebra finches mature, LMAN's influence on singing behavior diminishes (Aronov et al., 2008; S. W. Bottjer et al., 1984; Scharff & Nottebohm, 1991). Despite both j8v8 and r11n11 being approximately 80 days post-hatch at the recording time, they exhibited distinct continuity patterns. This discrepancy suggests that the direct correlation between chronological age and song development may not fully account for the differential impact of LMAN on song continuity. In line with developmental differentiation of neural behavior found in previous studies (Vallentin et al., 2016) it raises the hypothesis that LMAN's influence is potentially more significant in birds whose song structures are in earlier stages of song learning, indicating that developmental factors could play critical roles in the manifestation (or diminishing) of neurobehavioral continuity. This perspective warrants further investigation into the interplay between song developmental stage and LMAN's role in shaping neurobehavioral outcomes in songbirds.

In conclusion, we derived a parameter-free definition of continuity for neurobehavioral mappings in terms of neighborhoods. Using this method, we find evidence that the neural-to-vocal mapping from the premotor nuclei LMAN to vocal output is continuous at specific moments in the song motif.

Curriculum vitae

Education

since 03/18 **Doctoral student** Institute of Neuroinformatics UZH&ETHZ, Zurich, CH (Hahnloser group) and Institut NeuroPSI, Université Paris-Saclay, Saclay, FR (Giret group)

10/13 - 11/16 **M.Sc. in Cognitive Science**, Eberhard Karls University, Tübingen, DE.

10/10 - 09/13 **B.Sc. in Cognitive Science**, Eberhard Karls University, Tübingen, DE.

Research Experience

11/15 - 07/17 **Research Assistant**
Werner Reichardt Centre for Neuroscience, Tübingen, DE
Neurophysiology of Visual and Decision Processes (Nienborg group).

01/12 - 10/15 **Research Assistant**
Eberhard Karls University, Tübingen, DE,
Chair of Cognitive Modeling (Butz group).

09/11 - 11/11 **Research Assistant**
University Hospital, Ulm, DE, Project *Komm mit ins gesunde Boot*

Publications

Lorenz, C., Arneodo, E. M., Hahnloser, R. H. R. & Giret, N., A chronic Neuropixels probe implant for large-scale recordings in the forebrain-basal-ganglia circuit of singing zebra finches. *In preparation*

Zai, A. T., Lorenz, C., Giret, N., & Hahnloser, R. H. R., Reconciling optimality in reinforcement learning with suboptimal behavior. *In preparation*

Lorenz*, C., Das*, A., Centeno*, E. G. Z., Yeganegi, H., Duvoisin, R., Ursu, R., Retailleau, A., Giret, N., Leblois, A., Hahnloser, R. H. R, & Ondracek, J. M. Cortical-striatal network dynamics during sleep in a songbird. *In preparation*

Lorenz, C., Hao, X., Tomka, T., Rüttimann, L., & Hahnloser, R. H. R. (2023). Interactive extraction of diverse vocal units from a planar embedding without the need for prior sound segmentation. *Frontiers in Bioinformatics*, *2*, 966066.

Yamahachi*, H., Zai*, A. T., Tachibana, R. O., Stepien, A. E., Rodrigues, D. I., Cavé-Lopez, S., Lorenz, C., Arneodo, E. M., Giret, N., & Hahnloser, R. H. R. (2020). Undirected singing rate as a non-invasive tool for welfare monitoring in isolated male zebra finches. *Plos one*, *15*(8), e0236333.

Seillier, L.*, Lorenz, C*, Kawaguchi, K., Ott, T., Nieder, A., Pourriahi, P., & Nienborg, H. (2017). Serotonin decreases the gain of visual responses in awake macaque V1. *Journal of Neuroscience*, *37*(47), 11390-11405.

Butz, M. V., Kutter, E. F., & Lorenz, C. (2014). Rubber hand illusion affects joint angle perception. *PLoS One*, *9*(3), e92854.

---

*Authors contributed equally to this work.