Rockefeller University

Digital Commons @ RU

Student Theses and Dissertations

2024

Influence of Mouse Motor Cortex on Vocal Musculature

César D. M. Vargas

INFLUENCE OF MOUSE MOTOR CORTEX ON VOCAL MUSCULATURE

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
César D. M. Vargas
June 2024

Influence of Mouse Motor Cortex on Vocal Musculature

César D. M. Vargas

Learned vocal communication and spoken language are complex behaviors we have long sought to understand. Studying the neurobiology of speech and language will be advanced by investigating its component traits in model organisms. Most vertebrates share common brainstem circuits for vocalization, and they produce innate vocal repertoires over which the animals have little to no control. Vocal learners can imitate heard sounds, and thus have a high degree of control over their vocalizations. Vocal learning species have a direct projection from the pallial forebrain to the brainstem vocal motor neurons, which facilitate their vocal dexterity. Songbirds have been the standard model of vocal learning, but there are technical limits to our ability to test various hypotheses about the development and evolution of vocal learning circuits. Mice, which are more closely related to humans, have been found to exhibit some rudimentary features that are like those seen in vocal learning species. These include: 1) a direct, but sparse, projection from the primary motor cortex (M1) to the laryngeal motor neurons in the brainstem, termed the laryngeal motor cortex (LMC); 2) cortical lesions that damage this direct projecting population alter the frequency distribution of vocalizations; and 3) change their vocal syntax based on social context. Although mouse ultrasonic vocalizations (USVs) have received a great deal of attention and interest, the role these cortical circuits and the mechanisms by which they might affect USVs have been little explored. Here, we investigate the role of the motor cortex of mice in controlling vocal musculature, and we develop new methods that will allow us to gain more insights into the control mice have over their USVs.

We performed intracortical microstimulation (ICMS) with paired electromyographic (EMG) recordings to test whether the direct projection previously identified in mice can generate laryngeal muscle contractions. Simultaneously, we recorded EMG signals from the anterior digastric, a jaw opening muscle, as a control. We found that the LMC population of neurons can generate laryngeal muscle contractions. We also identified the orofacial motor cortex (OMC) as another region of M1 that can generate laryngeal muscle contractions, although weaker than from LMC. The muscles responded with different latencies from the LMC and OMC stimulations suggestive of both indirect and direct brainstem motor neuron projections, respectively. Using a retrograde transsynaptic virus, we show that the region of M1 containing the LMC has neurons that represent the larynx as well as jaw and forelimb muscles; in a small proportion of neurons, two muscles were represented by single neurons. Using an anterograde transsynaptic tracer from the OMC, we found that there are direct projections to the motor neurons of the jaw. Lastly, chemical lesions of OMC led to a modest change in the number of USVs per sequence.

To test the degree of control mice have over their USVs, we developed a suite of tools to train mice in an operant vocal task. To detect and classify USVs in real-time, we developed a software called Analysis of Mouse VOcal Communication (AMVOC). We combined AMVOC with other open-source hardware and software to design an operant training paradigm to that lets us test the volitional control of mouse vocal behavior. We provide a proof-of-principle application of this system to train mice to increase vocalizations for food reward.

Mice had been assumed to lack a functional cortical representation of the larynx, and similar assumptions have been made about other vocal non-learners. In contrast, the results of my thesis provide evidence that M1 in a vocal non-learner can influence vocal musculature, consistent with the continuum hypothesis of vocal learning. We also demonstrate that the representations of muscles for different behaviors across mouse M1 are highly intermixed, sharing both cortical space and single neurons. These results offer new insights into the origin and evolution of laryngeal control by cortical circuits, suggesting that these circuits are both more commonly distributed across mammalian species and that they may have arisen earlier than was previously hypothesized. Further, the results presented here will allow us to better understand the shared features of vocal production that are integral to better understanding human speech.

Dedication

To my parents. Thank you for all you have done for me.

—

Para mis papas. Gracias por todo lo que han echo por mí.

Acknowledgments

I want to offer a sincere and deep thank you to all the friends, mentors, teachers, and family who have all been a part of my path. In so many notable ways, I can see who I am today because of each person with whom I have had the pleasure to use any of those titles.

Thank you to Dr. Vivien Casagrande for taking an excited undergraduate into the lab and giving me my first true glimpse of what it is like to be a neuroscientist. Your energy and passion were truly infectious, and your support felt immeasurable. I hope, that just like you, I never lose my excitement for the brain. Thank you to Julie Mavity-Hudson for providing me with my first real bench skills and for your mentorship regarding both the lab and life.

My time at Rockefeller has been marked by friendships and interactions I could've only dreamed of, and I'd like to thank many of the people outside the lab who make all our efforts in the lab possible. First, to the Dean's Office, thank you for the support you provide and for making it easy to focus on our endeavors. A special thanks to Cris Rosario, who was the first person I met at Rockefeller, and to Marta Delgado and Andrea Morris, for your personal and professional support. I know whenever I needed any guidance, I could turn to any of you. I want to thank Jim Petrillo and the rest of the PIT team—Peer Strogies, Dan Gross, Nick Belenko, and Michelle Zhang. Although I have known each of you for different amounts of time, there is no doubt that some facet of this work was aided by some insight each of you gave. Thank you to the CBC and all the vets and animal care staff. In a very literal sense, our work depends on you. Without happy and healthy animals, none of this would be possible. Thank you to Allison Richards and Tom Wiley for your help on all things animal care-related. A shout-out to Chaya, Mark, and Cameron for checking-in on our mice in the lab all these years.

Inside the lab, I have had the opportunity to work with and alongside some truly remarkable individuals. Thank you to the postdocs and research scientists in the lab, both past and present—Drs. Lomax Boyd, James Cahill, Irene Ballagh, Chul Lee, Constantina Theofanopoulou, Jean-Nicolas Audet, Grace Smith-Vidaurre, Giulio Formenti, Anna Keyte, and Matthew Biegler. I first got to know Lomax during my rotation in the lab when he was interviewing for a postdoc position. Not too many months later, we were both arriving in the lab as the two newest faces, both taking the plunge into a field that was new to both of us at the time. Your guidance and discussions, especially in those early days, were invaluable. Thank you to Irene for teaching me almost everything I know about evolution. A warm and special thanks to the fellow grad students in the lab—Stephanie Marcus, Matthew Davenport, Brigid Maloney, Elena Waidmann, Peter Schade, Kirubel Belay, and Cassidy Johnson— and to former grad students with whom I overlapped—Drs. Caitlin Gilbert, Greg Gedman, Lindsey Cantin, and Ha Na Cho. Thank you for making the Jarvis lab such a special place to work. Although the science is pretty great, you all make coming into lab, with all its highs and lows, a true pleasure. I especially want to thank Elena for being the first grad student to join the mouse team after me. You make doing science so much fun. To the research assistants in the lab with whom I've worked, Rajvi Agravat and Hector Bermudez, thank you for your help on this project as well as your friendship.

Outside the lab, I have had the great pleasure and honor of getting to be a part of some truly amazing organizations. Namely, the Rockefeller Inclusive Science Initiative (RiSI) and the Summer Neuroscience Program (SNP). At RiSI, we began with a modest goal of creating an environment in which many of us could build a community and share our experiences as minority students. Those first few meetings in the Young lab's conference room felt like the start of something special. This goal quickly expanded, and by the time Josue Regalado and I were co-

presidents, we had a full-fledged organization helping with PhD recruitment, an invited speaker series, and social events, among other activities. It was amazing to help lead an organization full of inspiring friends, and it has been really special to see that momentum continue on with the newest generation of grad students. At SNP, we get to teach neuroscience to, and this is just my opinion, the best high school students in New York City. Watching the joy of an unencumbered and curious teenager as they get to learn about things that really interest them is hard to beat. It was a pleasure to be a co-director alongside Caitlin Gilbert, Amy Dunn, Itzel Ishida, Anna Ryba, Josue Regalado, and Elena Waidmann.

Thank you to Dr. Theodoros Giannakopoulos. It's hard to believe our collaboration began over an email asking a question about one of your tools on GitHub. Even though we have never met in person, I think it is safe to say we have had a great and fruitful collaboration. Thank you for choosing to venture into the world of mouse vocalizations, I am grateful to be able to call you a mentor. A big thank you as well to Vassilina Stoumpou and Christodoulos Bochalis, without whom AMVOC would likely still be just an idea.

To Dr. Erich Jarvis, thank you. I first heard about you the summer before I arrived at Rockefeller. Of course, I knew nothing about your work at the time, but something about your research just sounded cool! Neuroscience, behavior, and evolution? Sign me up! Although my rotation was not the most eventful given that part of it involved helping to set up the lab after your move from Duke, I had a good feeling about the lab. Your kindness and openness are hard to miss, but so is your passion for your research. From day one, I have been absolutely fascinated by all the topics in the lab, be they experiments in mice, birds, marmosets, or even dolphins and humans now. In the end, it is your mentorship that made the sell and what I have appreciated all this time. I have felt lucky to get to be mentored by you both in the lab and outside. In the lab you constantly

push us to be innovative and creative, and as a faculty sponsor for RiSI, you encouraged us to strive for changes we believed in.

Thank you to my thesis committee, Drs. Jim Hudspeth and Priya Rajasethupathy, for your support throughout this project. Thank you both for your enthusiasm towards this project, and for your insights along the way. Thank you to Dr. Hudspeth for helping me record my first EMG, an occasion without which this project would not have gotten to where it is today. Thank you to Dr. Rajasethupathy for her encouragement and mentorship. From the days of my rotation in your lab to today, you have always pushed me to become a better and more critical scientist. Thank you both for always finding the time to meet with me and for offering me your support throughout the years. Thank you to Dr. Yutaka Yoshida for serving as my external examiner. Your work has been an inspiration throughout this project.

To my partner, Marley. Thank you for everything. You light up my days when you aren't even trying. You always have a way of helping ground me and keep me on my proper heading. It's so wonderful to have someone to share my love for nature and all it contains. I can't imagine a better partner, and I'm so thankful that you have been at my side. Lastly, to my parents, (Drs.) Bonnie Gunn and David Vargas (a.k.a. Mamá and Papá). I am who I am today, most of all, because of you. Your unconditional love and support are things I hope everyone can know. Thank you for always letting me explore my passions and never letting me settle. And of course, for instilling in me a joy for life and discovery. I hope to be like you guys one day… I think I still have some time?

Table of Contents

List of Figures

List of Tables

List of Abbreviations

7N- facial nucleus

AAV- Adeno-associated Virus

ACC- anterior cingulate cortex

AFP- anterior forebrain pathway

Amb- nucleus ambiguus

AMVOC- Analysis of Mouse VOcal Communication

AP- anteroposterior

AVA- autoencoded vocal analysis

BSA- bovine serum albumin

CFA- caudal forelimb area

CM- cortical motorneuron

CNN- convolutional neural network

CPU- central processing unit

CT- cricothyroid muscle

DAS- DeepAudioSegmenter

DG- digastric muscle

dLMC- dorsal laryngeal motor cortex

ECR- extensor carpi radialis muscle

EMG- electromyographic

FED3- Feeding Experiment Device 3

GPIO- general-purpose input/output

GPU- graphics processing unit

GUI- graphical user interface

Hyp- hypothalamus

IEG- immediate early gene

IRt- intermediate reticular formation

ISI- inter-syllable interval

LMC- laryngeal motor cortex

M1- primary motor cortex

M1α- primary motor cortex alpha

M1β- primary motor cortex beta

ML- mediolateral

Mo5- motor trigeminal nucleus

MSA1- Mouse Song Analyzer 1

MSA2- Mouse Song Analyzer 2

MSE- Mean Square Error

MUPET- Mouse Ultrasonic Profile Extraction

OMC- orofacial motor cortex

PAG- periaqueductal grey

PCA- Principal Component Analysis

PCRt- parvicellular reticular formation

PFA- paraformaldehyde

PRV- Pseudorabies-Bartha Virus

PyPI- Python Package Index

RF- reticular formation

RFA- rostral forelimb area

S1- primary somatosensory cortex

SBC- single-board computer

SC- superior colliculus

SD- standard deviation

SLN- superior laryngeal nerve

StTA- stimulus triggered average

TA- thyroarytenoid muscle

UMAP- Uniform Manifold Approximation

USV- ultrasonic vocalizations

vLMC- ventral laryngeal motor cortex

WT- wild type

CHAPTER 1. Introduction

1.1 Overview of Vocal Communication and Vocal Learning

Animals have evolved a variety of mechanisms to share information. Some common modalities include chemical cues, visual displays, and emitted sounds (**Figure 1.1A-C**) (Gillam, 2011). Each of these is an effective mechanism for sharing signals across distances. Chemical communication is a potent mechanism, but it occurs on a relatively slow timescale and has a limited range. Visual signals can be transmitted across long distances, but they can be easily occluded. Sound is a form of communication that cannot be easily impeded and can be transmitted rapidly across a variety of spatial scales. Of these three signal types, sound is particularly malleable, even within an individual. Many of its qualities (e.g., duration, frequency, intensity) can be modified by simple changes in the rhythm or structure of the peripheral organs used to produce the sound.

Many animals produce acoustic communication signals in the form of vocalizations. This means that the sound is produced by airflow through a vocal organ (**Figure 1.1D-F**)—the larynx (non-avian vertebrates), syrinx (birds), or nasal phonic lips (cetaceans)—as opposed to mechanical strumming or drumming mechanisms used by crickets and some fish. Vocalizations are used to communicate an individual's internal state, social status, group membership, or reproductive fitness. For the vast majority of species, their vocal repertoire is innate and remains unchanged throughout life. Many of these species can also flexibly change when they use their acoustically innate vocalizations, so-called vocal usage learning. For example, vervet monkeys will learn to use certain calls depending on the type of predator that is present to warn conspecifics; one call signals the presence of eagles, and a different call signals the presence of snakes (Seyfarth et al., 1980). Smaller in number is the group of animals that can learn or improvise new vocalizations,

Figure 1.1: Mechanisms of Communication and Organs Used for Vocalization.
Representation of different communication channels in different animals. **A)** Moths use pheromones to attract mates over large distances. **B)** Poison dart frogs are famous for their brilliant coloration used to signal that they are poisonous. **C)** Wolves use howls as a contact call to inform conspecifics of a variety of information. Vertebrates have evolved different organs for acoustic communication produced by the passage of air (i.e., vocalizations). These specialized organs for vocalization include: **D)** the larynx in non-avian and non-cetacean vertebrates; **E)** the syrinx in birds; and **F)** phonic lips in odontocete cetaceans. Birds have a larynx, but it does not serve any function in vocal production. Similarly, odontocetes (toothed whales) have a larynx and plays a more limited role in sound generation (Madsen et al., 2023; Reidenberg, 2017).

termed vocal production learning, or simply vocal learning. This is a rare behavior that has evolved convergently in only eight groups of animals: five groups of mammals (elephants, pinnipeds, cetaceans, bats, and humans), and three groups of birds (songbirds, parrots, and hummingbirds) (Jarvis, 2004). Vocal learning is considered convergent in each of these groups, as the nearest taxonomic relatives do not exhibit vocal learning (e.g. humans vs other primates).

Despite the gross morphological differences between mammals and birds, there appears to be a core set of neural connections that are shared between distinct functional regions within the brains of each group (Jarvis, 2019; Karten, 2015). For example, both humans and songbirds have: 1) a motor region specialized for vocal control with afferents from auditory brain regions; 2) a cortico-basal ganglia-thalamo-cortical loop used to learn new vocalizations which projects back to the motor region which; and 3) send direct projections to the motor neurons of the vocal organs (larynx in mammals and syrinx in songbirds) (**Figure 1.2A and B**) (Jarvis, 2019). Classically, this direct projection from cortex to vocal motor neurons was considered a hallmark unique to vocal learning species (Fitch et al., 2010). Lesions to this motor pathway remove the ability to produce learned vocalizations, while leaving innate calls like cries intact (Jarvis, 2004). Conversely, lesioning the motor cortical regions of vocal non-learners, like macaques (**Figure 1.2C**), does not have any effect on the animal's ability to produce correct vocalizations (Jarvis, 2004; Jürgens, 2002).

Recent work has shown that these functionally analogous vocal learning regions between human and avian brains have similar specialized up- and down-regulated gene expression patterns (Gedman et al., 2022; Pfenning et al., 2014). The convergence is suggestive of an important, shared role of these genes in vocal learning behavior. Due to technological limitations in songbird

transgenesis, and ethical considerations with human experimentation, the role of these genes in producing vocal learning behavior remains untested. Thus, it is unclear if these shared genes are causal in the formation of vocal learning circuits or if they are simply the end-state of developmental programs. Recent observations in mice (*Mus musculus*) suggests that they may have the neural circuitry and gene expression patterns that makes them an amenable model for the study of vocal behavior and a rich experimental substrate on which to test hypotheses for the evolution of vocal learning (**Figure 1.2D**) (Arriaga & Jarvis, 2013; Arriaga et al., 2012; Wang et al., 2015).

1.1.2 Overview of Mouse Vocalizations

Rodents are the most numerous group of mammals, comprising approximately 40% of the approximate 6,000 species of mammals (Kay & Hoekstra, 2008). The vocal repertoires of a wide variety of rodents have been studied, including those of the gerbil (Kobayasi & Riquimaroux, 2012), lemming, (Brooks & Banks, 1973; Yurlova et al., 2020), rat (Brudzynski, 2007), mouse, hamster (Pierce Jr et al., 1989; Rendon et al., 2015), and guinea pig (Berryman, 1976), among many others. Rodents have evolved a myriad of vocalizations that occur across a vast range of frequencies. Many of us are familiar with the squeak of a mouse or a rat, emitted when they are threatened or harmed (**Figure 1.3A**). Other species, like Alston's singing mouse (*Scotinomys teguina*), use long trilled vocalizations that can be used as territorial signals across the long distances in their mountainous habitat (Pasch et al., 2013). In addition to these human-audible vocalizations (2-20 kHz), many rodents produce ultrasonic vocalizations (USVs) that are above the upper limit of the human hearing range (> 20 kHz). The production mechanism of USVs in rodents has been debated (Haakansson et al., 2021; Mahrt et al., 2016; Riede et al., 2017), but in each explanation the ultimate effect is the production of a pure-tone whistle.
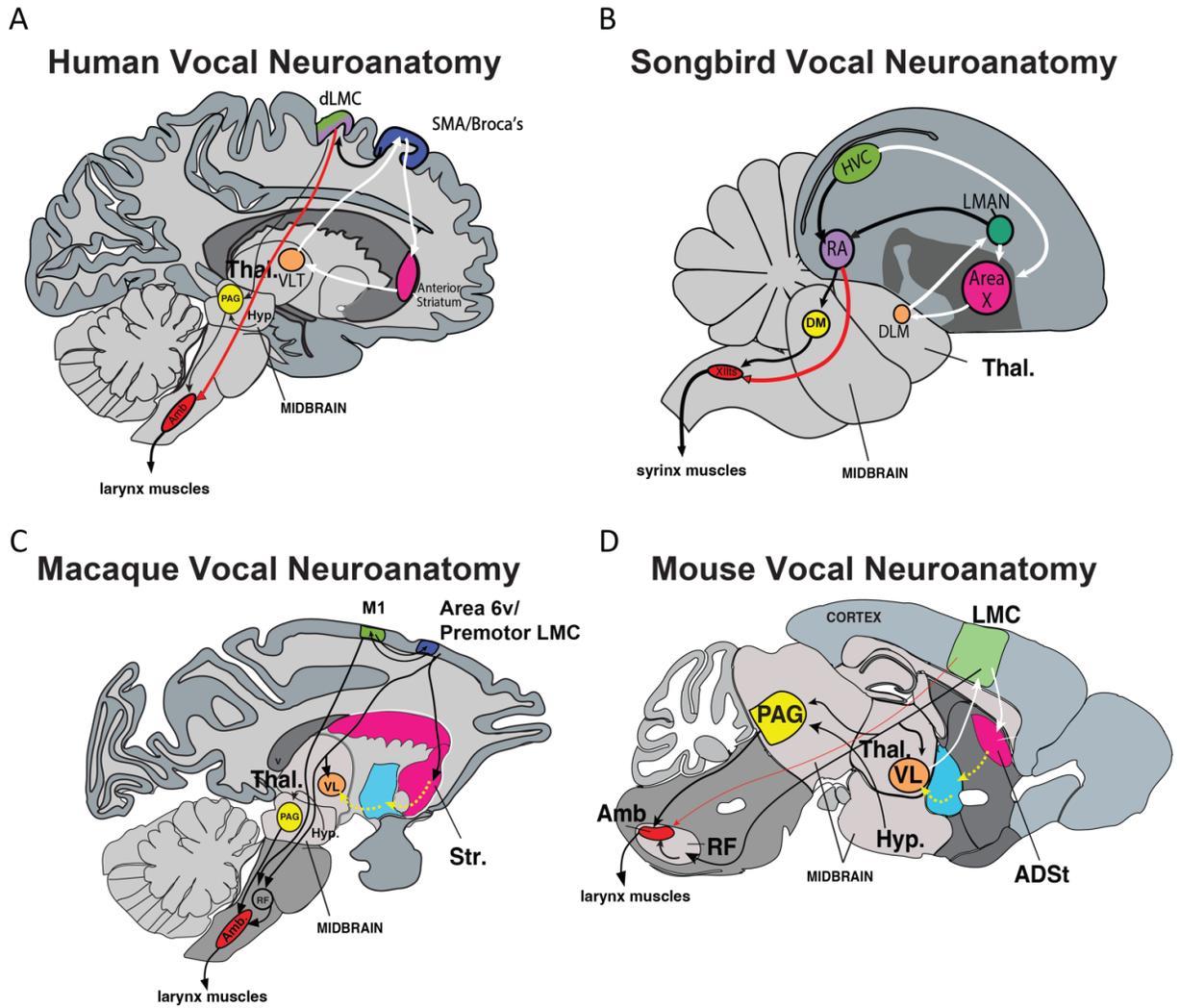
Figure 1.2: Comparative Vocal Neuroanatomy Between Mammals and Songbirds.
The vocal production mechanisms of most vertebrates are conserved at the midbrain and brainstem levels. Mammals and birds have evolved analogous pallial connections that facilitate different levels of control over their vocalizations or features of their vocalizations. Vocal learning species like **(A)** humans and **(B)** songbirds have a direct projection (red arrow) from the motor output regions of the pallium (Layer 5B in M1 and RA in the arcopallium, respectively) to the vocal motor neurons in the brainstem (Nucleus Ambiguus and nXIIts, respectively). Recent evidence in macaques **(C)** demonstrates that there is connection from both M1 and promotor area 6v to the vocal system in the subcortical areas. Our lab's work in mice **(D)** has shown that they also possess a direct projection from M1 to vocal motor neurons in the brainstem, albeit more sparse (thin red arrow) than those seen in vocal learning species. Circuits involved in the production of vocalizations are in black arrows. The cortico-striatal-thalamic-cortical loop is denoted in white. Hypothesized connections in this path are in yellow.

USVs between species appear to have different uses. Rats produce USVs in different frequency ranges to communicate negative (22 kHz) or positive (55 kHz) affective state (**Figure 1.3A**) (Brudzynski, 2013). Mouse vocalizations for negative experiences in adults appear to only occur at sonic frequencies (i.e. "squeaks"; **Figure 1.3B**) (Grimsley et al., 2016). Mouse pups will produce negative-affect USVs when removed from their nest or when they are cold which prompts a parent or adult nest mate to recover the vocalizing pup (Ehret, 2005). Adult mice appear to use USVs primarily for social interactions, including courtship (Egnor & Seagraves, 2016; Portfors & Perkel, 2014). The repertoire (range of possible calls) of adult mice and those from isolated pups are quite similar, with pups producing all but one of the call types seen in adults (Grimsley et al., 2011). Both sexes of mice produce USVs (Maggio, 1985; Neunuebel et al., 2015; Zhao et al., 2021), and machine learning (ML) methods can be used to distinguish USVs produced by either sex (Ivanenko et al., 2020; Steinfath et al., 2021). In the adult animal, males have been the primary focus of study, as they readily vocalize more in the presence of a female, or in response to urine from a female mouse (Nyby et al., 1979).

Female-directed vocalizations were first characterized by Whitney et al. (1973). About 30 years later, Holy and Guo (2005) performed the first rigorous analysis of the spectrotemporal features of USVs. They showed that there is an overarching song-like structure to mouse USVs, including motifs (repeated sequences of syllables or call types) (**Figure 1.3C**). Later work showed that male mice can alter their syntax and vocal repertoire based on social context (Chabout et al., 2015; Chabout et al., 2016) and social experience (Chabout et al., 2012; Zhao et al., 2021). In playback studies used to discriminate stimulus preferences, female mice appear to prefer songs with complex USVs over songs composed of simple USVs (Chabout et al., 2015), and USVs from an adult male compared to pup USVs (Hammerschmidt et al., 2009). These behaviors, along with

their laboratory tractability make mice a promising model for basic principles of vocal production. But are the neural circuits for USV production similar to other animals, or perhaps even like those of vocal learners?

The neural mechanisms used to produce USVs are similar to what has been observed across many vocalizing vertebrates (Bass et al., 2008; Jürgens, 2009). In particular, mice have brainstem circuits that connect the hypothalamus (Hyp), periaqueductal grey (PAG), and reticular formation (RF) to the nucleus ambiguus (Amb) for the production of innate vocalizations (**Figure 1.3D**) (Arriaga & Jarvis, 2013; Gao et al., 2019; Tschida et al., 2018; Tschida et al., 2019). In addition to these highly conserved subcortical circuits for vocalization, previous work from our group used immediate early genes (IEGs) to demonstrate that the primary motor cortex (M1) of mice is active during vocal production (Arriaga et al., 2012). Retrograde transsynaptic tracing from the laryngeal muscles demonstrated a pool of larynx-representing neurons in layer 5B of M1, the layer that projects sub-cortically (**Figure 1.3E**), and these neurons make a direct connection to motor neurons in the Amb (**Figure 1.3F**); lesions of this part of M1 increased the frequency distribution of the USVs (Arriaga et al., 2012). This work provided crucial evidence that M1 is involved in the production, or at least the modulation, of USVs. These results were unexpected. Prior to this data, only vocal learners were predicted to have a larynx representation in M1, with humans being the only mammals yet observed with this projection. Although the M1 laryngeal neurons in mice were not as robust as those of vocal learners, the presence of these neurons in mice required a re-evaluation of hypotheses regarding the cortical control of vocalizations in vocal non-learning species.

A Rat USVs

22kHz   55kHz

B Mouse Alarm "Squeak"

C Mouse USVs

D Mouse Vocal Neuroanatomy

E M1

F Amb Mn   axon
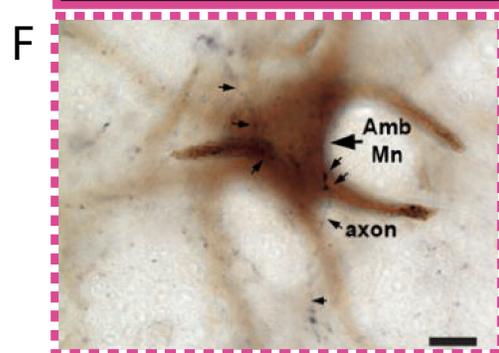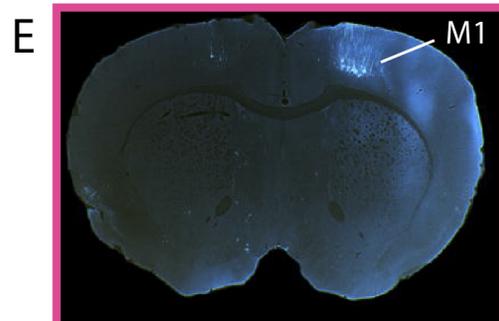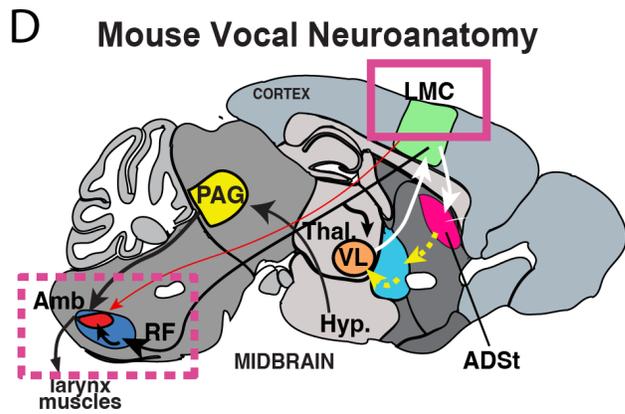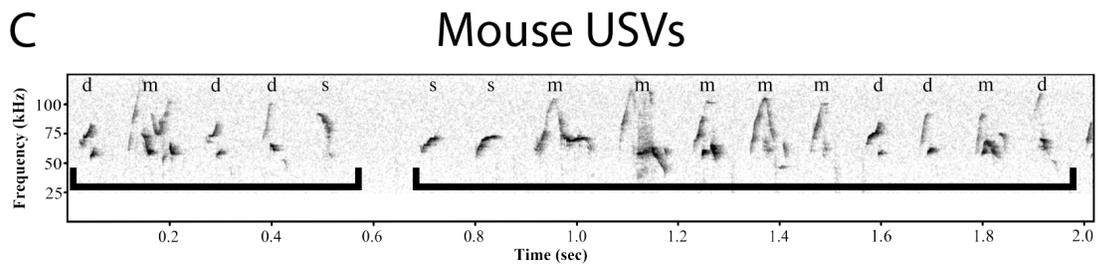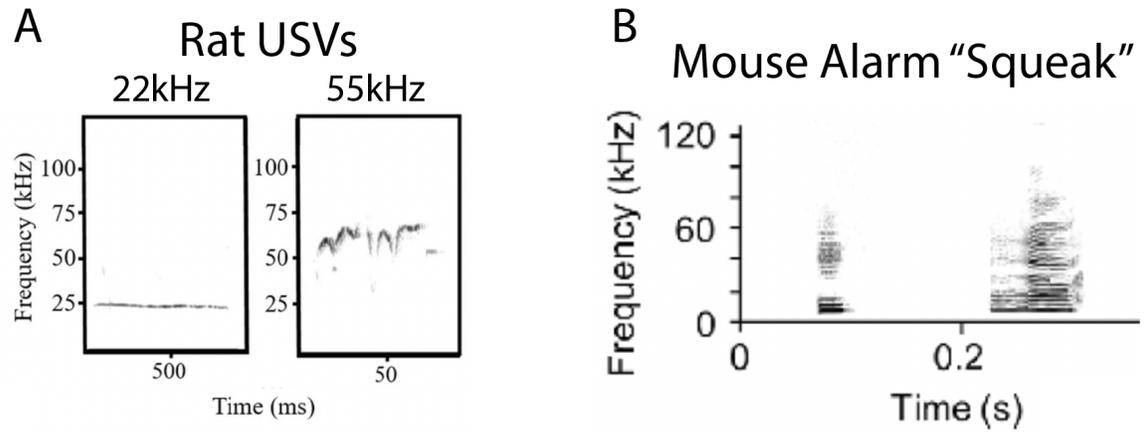
Figure 1.3: Examples of Rodent Vocalizations and Mouse Vocal Neuroanatomy.
A) Example spectrogram of rat USVs that communicate negative affect (22 kHz; top) and positive affect/general social communication (55 kHz; bottom). **B)** Example of a mouse producing an aversive "squeak" often seen when they are threatened or injured. **C)** Example of a mouse USV song, composed of different syllable types (denoted by letters above each call) which in turn are composed into larger groups called sequences. Sequences of USVs are separated by silent periods of at least 250 ms. **D)** General neuroanatomy for laryngeal/vocal control in mice. Highlighted boxes correspond to regions in **(E)** and **(F)** (solid box and dashed box, respectively). **E)** Population of PRV$^+$ neurons in M1 from injections in the larynx. **F)** BDA-labeled afferent axons from M1 seen contacting dendrites of CTB-labeled laryngeal motor neurons in Nucleus Ambiguus. (A) modified from (Simola and Granon, 2019); (B) modified from Niemczura et al. (2020); (E) and (F) modified from Arriaga and Jarvis (2013).

1.1.3 The role of motor cortex in vocalizations

In one of the earliest studies of localized brain function, Paul Broca identified a region in the left frontal area of the human brain that is involved in the production of spoken language (Broca, 1861). Subsequent work by another German neurologist, Carl Wernicke, identified a region in the left posterior part of cortex whose damage limited the patient's ability to understand language (Wernicke, 1874). These observations became foundational, and these eponymous brain regions provided the first anatomical substrates for where humans may derive their linguistic capabilities.

Later work by Wilder Penfield identified a region of primary motor cortex in humans, which, when stimulated, led to a vocal utterance and/or speech arrest (Penfield & Rasmussen, 1949). This region was called the laryngeal motor cortex (LMC), located within the primary motor cortex, M1. There have been many attempts using electrical stimulation to identify similar laryngeal motor region in other mammals. Some evidence for an M1 laryngeal representation has been found in the Egyptian fruit bat (Wirthlin et al., 2022), which is also a vocal learning species. In non-human primates, none of which are known to be vocal learners, the only cortical region that has been identified to induce laryngeal movement when stimulated is in the premotor cortex (Coudé et al., 2011; Hast et al., 1974; Simonyan, 2014). Experiments in cats showed conflicting results as to whether or not M1 had a laryngeal representation (Kirikae et al., 1962; Milojevic & Hast, 1964). Outside the motor cortex, stimulation of the anterior cingulate cortex (ACC) results in vocalizations in a number of species including cats (Hunsperger & Bucher, 1967), rats (Bennett et al., 2019), mice (Gan-Or & London, 2022), bats (Gooler & O'Neill, 1987), guinea pigs (Green et al., 2018), humans, and non-human primates (Jürgens, 2009). This long-standing inability to identify a primary motor representation for the larynx outside of humans and mammalian vocal learners has reinforced a prevalent idea that the anatomical substrates for speech are a unique

feature of the human brain that evolved *de novo* (Belyk & Brown, 2017; Simonyan, 2014). However, recent anatomical evidence has challenged this dominant view.

The findings with viral tracing in mice (see above) showing that M1 contains a population of neurons in layer 5B that are labeled after laryngeal injection (Arriaga et al., 2012), has also recently been shown in *S. teguina* (Zheng et al., 2022), another rodent species. The recent study mentioned above in macaques and marmosets (Cerkevich et al., 2022), showed both species had a pool of larynx representing neurons in M1, among other premotor regions. They observed differences between the two primate species, with marmosets having a higher number of labeled neurons than macaques. This was surprising, as prior experiments in macaques using electrical stimulation had not found a laryngeal representation in this virally traced region of M1 (Coudé et al., 2011; Hast et al., 1974). To explain such findings like these from vocal non-learning species, Arriaga and Jarvis (2013) and Petkov and Jarvis (2012) proposed the continuum hypothesis of vocal learning, where some species have more, or less, robust connections and functions of the brain circuits seen in advanced vocal learners. In this present work, I test the function of these cortical regions for vocalizations in mice, and discuss how our results may influence our understanding of the evolution of these circuits for vocalization in mammals.

In **Chapter 2,** I present experiments using electrical stimulation of the brain with paired muscle recordings to identify if the identified laryngeal representation of M1 in mice can generate laryngeal muscle contractions. I found that there are at least two focal regions that can generate laryngeal muscle contractions. I validate these stimulation experiments with viral tracing methods to further identify the circuits that these two cortical loci impinge upon to produce vocal muscle contractions. Lastly, we predict M1 has a role in the volitional control of behavior and vocalizations. In **Chapter 3,** I investigate this prediction with regard to the syntactic flexibility of

vocalizations that mice exhibit in different social contexts. To facilitate these experiments, I co-developed a new method for analyzing mouse USVs, called Analysis of Mouse VOcal Communication (AMVOC), which can be deployed in real-time to perform online analysis of USVs. I provide an example of how the real-time capabilities of this tool can be used to gain insight into the vocal abilities of mice.

CHAPTER 2. Representation of Laryngeal Musculature in Primary Motor Cortex of Mice

2.1 Introduction

2.1.1 Overview of Mapping the Motor Cortex

In 1870, Gustav Fritsch and Eduard Hitzig published the first documented experiments that used electrical stimulation to identify a motor region in the cortex of dogs (Fritsch and Hitzig, 1870). This provided the first account that stimulating a region of the mammalian cortex which could produce involuntary movements. Over the intervening decades this became a more common technique, and the cortical motor regions of many different species were described. In another seminal study, Leyton and Sherrington (1917) carried out cortical stimulations in orangutans, chimpanzees, and gorillas. Their extensive report notes a somewhat orderly distribution of the body plan along the ventro-dorsal axis of the motor strip, providing an early example of a somatotopic maps (**Figure 2.1A**). Famously, Penfield performed similar experiments on his neurosurgical patients and synthesized this somatotopy into what we now know as the homunculus (Penfield and Rasmussen, 1950). In each of these, there was a localization of function that could be derived from their experiments, with minimal currents being able to elicit the movements of single, or relatively isolated, parts of the body.

The first study to produce a motor map in mice using intracortical microstimulation (ICMS) was by Li and Waters (1991). They found that the representations of movements in mice were broadly similar to those found in rats (Neafsey et al., 1986). Both species had two prominent motor fields for the forelimb, the rostral forelimb area (RFA) and the caudal forelimb area (CFA). By taking the most responsive body part at the lowest effective current, one can still generate what appears to be a general somatotopic organization (Tennant et al., 2011). This somatotopy is

organized along the rostro-caudal axis of the body (**Figure 2.1B**), similar to the human/primate homunculus, but perhaps not as atomized or discrete. In these mapping studies, as well as work from others (Mercer Lindsay et al., 2019; Okobi et al., 2019), has demonstrated the role of the orofacial region, located in the anterolateral portion of motor cortex, for control of jaw and whisker movements, among others. There are similar findings in guinea pigs for rhythmic mastication (Isogai et al., 2012). Notably, none of these studies ever remarked on whether or not vocalizations, or laryngeal contractions in general, could be produced by these motor cortex stimulations. One study in rats noted that swallowing, a movement involving the larynx, could be elicited from medial stimulations, but no EMG signals were demonstrated (Gioanni and Lamarche, 1985).
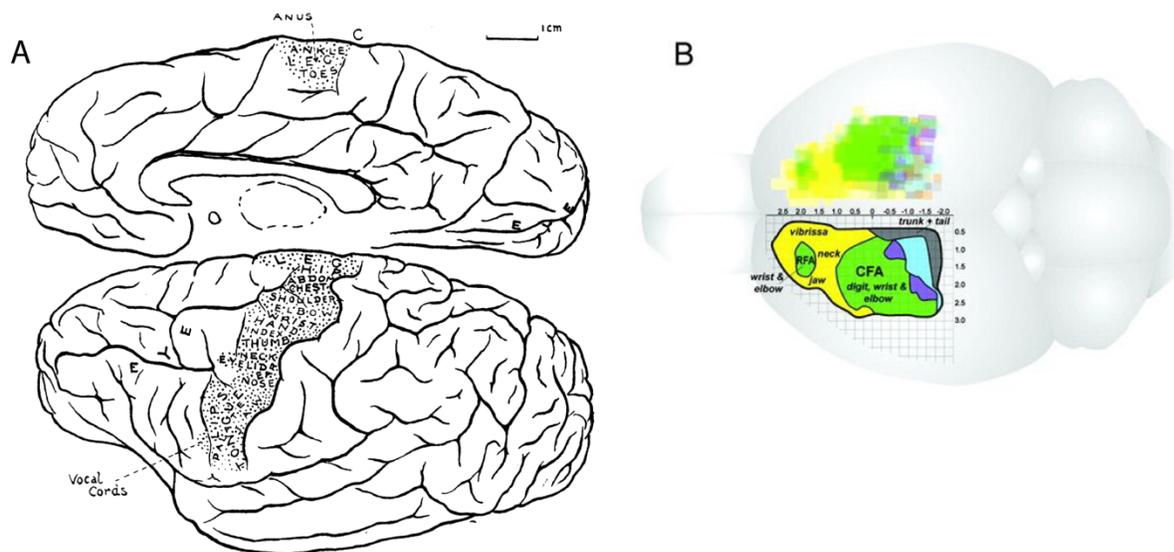
Figure 2.1: Somatotopy in the Gorilla and Mouse Motor Cortex.
A) Motor/movement map of gorillas based on electrical stimulations of the precentral motor strip. A general ventro-dorsal organization is identifiable representing the orofacial regions down to the feed, respectively. **B)** Motor/movement map of mice based on electrical stimulation of primary and secondary motor regions and portions of the primary sensory cortex. **A)** Image used with permission from Leyton and Sherrington (1917). **B)** Image used with permission form Tennant et al. (2011).

2.1.2 The Role of Motor Cortex for Vocal Learning and Vocal Control

Humans, as well as some other animals, have evolved neurons in layer 5 of M1 that form direct, monosynaptic projections with motor neurons in the brainstem or spinal cord. These direct projections are also called cortico-motoneuronal connections (CMs). In humans the laryngeal motor cortex (LMC) has CMs that project directly to motor neurons in the nucleus ambiguus (Amb) which control laryngeal muscles (Kuypers, 1958; Iwatsubo et al., 1990). It is divided into two representations referred to as dorsal LMC (dLMC) and ventral LMC (vLMC). Songbirds have analogous direct projections from the arcopallium to syringeal motor neurons (Nottebohm et al., 1976; Wild, 1993). Greater numbers of CM projections have been hypothesized to endow a species with more dexterity over the muscles of that circuit (Lemon, 2008). These direct-projecting cortical circuits for vocal muscles have been argued to be a unique evolutionary feature of vocal learners (Kuypers, 1958; Jürgens, 2002; Fitch et al., 2010; Petkov and Jarvis, 2012; Jarvis, 2019).

Stimulating the premotor regions of the cortex of various mammals, in particular area 6V in many primates, can induce vocal muscle movement (Leyton and Sherrington, 1917; Hast et al., 1974; Simonyan and Jürgens, 2003). Thus far, only in vocal learners, such as humans, do stimulations of the primary motor cortex (human M1 or songbird RA) lead to vocal muscle contractions (Jürgens, 2002; Simonyan, 2014). Lesions in M1, or analogous motor regions in birds, deteriorates or abolishes the ability to properly produce previously learned vocalizations, while leaving innate vocalizations intact (Jarvis, 2004; Simonyan and Horwitz, 2011; Silva et al., 2022). Conversely, lesions in the vocal motor region (Area 6V) of vocal non-learning primate species have not been found to affect the vocal ability of the lesioned animals (Aitken, 1981; Kirzinger and Jürgens, 1982). Taken together, these findings have been used to argue that vocal non-learners rely on brainstem vocal circuits for the production of vocalizations in the absence of direct

projections from M1 to vocal motor neurons, with M1 of these mammals having a limited, or no role, in vocal behavior (Jürgens, 2002).

This view has recently been challenged by work in mice and several non-human primates (Arriaga et al., 2012; Cerkevich et al., 2022). Prior work from our group in mice (*Mus musculus)* found a region in M1 that showed vocalizing-driven immediate early gene activity, a sparse direct projection to Amb (**Figure 1.3D and E**), and degradation of frequency modulation when lesioned (Arriaga et al., 2012). Based on these findings, they termed this region a putative, but rudimentary laryngeal motor cortex (LMC) in mice. Other work in Alston's singing mouse (*Scotinomys teguina),* showed that manipulations of a different region of M1, the orofacial motor cortex (OMC), can affect the ability of mice to produce vocalizations in a social encounter (Okobi et al., 2019), reminiscent of results of vocal learning species. *S. teguina* also appears to have an LMC similar to *M. musculus,* as demonstrated by retrograde transsynaptic tracing (Zheng et al., 2022). A recent study in marmosets (*Callithrix jacchus*) and two macaque species (*Macaca mulatta* and *Macaca fascicularis)* used a retrograde transsynaptic virus injected into laryngeal muscles and found labeled M1 neurons in these non-human primates (Cerkevich et al., 2022). They inferred the M1 representation to be indirect based on timing of viral tracer transport time. The mouse data and prior primate findings led to the continuum hypothesis of vocal learning. Traditionally, vocal learning is described as an all or none trait. The continuum hypothesis proposes that rodents, non-human primates and the other so-called "vocal non-learning" species display, a subset of abilities and mechanisms present in advanced vocal learners that span a range between the dichotomous options.

2.1.3 Synaptic Order and Response Latency

In humans, stimulating the LMC region in humans, which contains indirect and direct CM projections (Kuypers, 1958; Iwatsubo et al., 1990) (**Figure 2.2A**), can result in laryngeal contractions that are very short. Using transcranial magnetic stimulation (TMS), studies reported latencies in the range of ~4-13 ms, with a mean of ~10 ms (**Figure 2.2B**) (Rödel et al., 2004; Espadaler et al., 2012). Other studies electrically stimulated the surface of dLMC and reported laryngeal muscle responses of ~11-22 ms (Deletis et al., 2009; Dichter et al., 2018). The macaque forearm muscles also receive direct projections from cortex to spinal motor neurons. Electrically stimulating the forearm regions of M1 also results in short, ~10 ms latencies (Cheney and Fetz, 1985). In macaques, the laryngeal circuit from 6V is disynaptic, with the cortical neurons terminating in the RF and then RF neurons projecting to Amb motor neurons (**Figure 2.2C**). When this cortical laryngeal region of the macaque is stimulated, the laryngeal response to the cortical stimulation ranges from ~20 ms to ~40 ms after stimulation, depending on whether the thyroarytenoid muscle (TA) or cricothyroid muscle (CT) was recorded, respectively (**Figure 2.2D**) (Hast et al., 1974). Thus, the difference in laryngeal response latency observed between the two species is likely due to the synaptic order. Although these data add to the hypotheses regarding the direct projections' role in the dexterity of their respective motor systems, it has remained largely unclear if these circuits truly are causal in increase the control/dexterity of a muscle system. Recent experiments have tested this hypothesis in mice.

In adult wild type (WT) mice, during postnatal development, there are a number of cortical neurons that project directly to a number spinal cord motor neuron targets (Gu et al., 2017b; Wang et al., 2017; Murabe et al., 2018). However, these direct projections are pruned and lost into adulthood, leaving only the indirect cortical projection system. In vocal learning species this

elimination of direct projections does not occur to the vocal motor neurons. Consistent with these findings, several studies have found that the axons guidance ligands, and their receptor proteins, that repel some connections from forming are down regulated in both the human layer 5LMC region and songbird RA compared to vocal non-learners (Pfenning et al., 2014; Wang et al., 2015; Gu et al., 2017b). The hypothesis that resulted was that the downregulation of repulsive axon guidance genes in vocal learners, like humans, allowed for layer 5, and similar songbird motor regions, to form a greater number of direct projections to the brainstem/spinal cord motor neurons
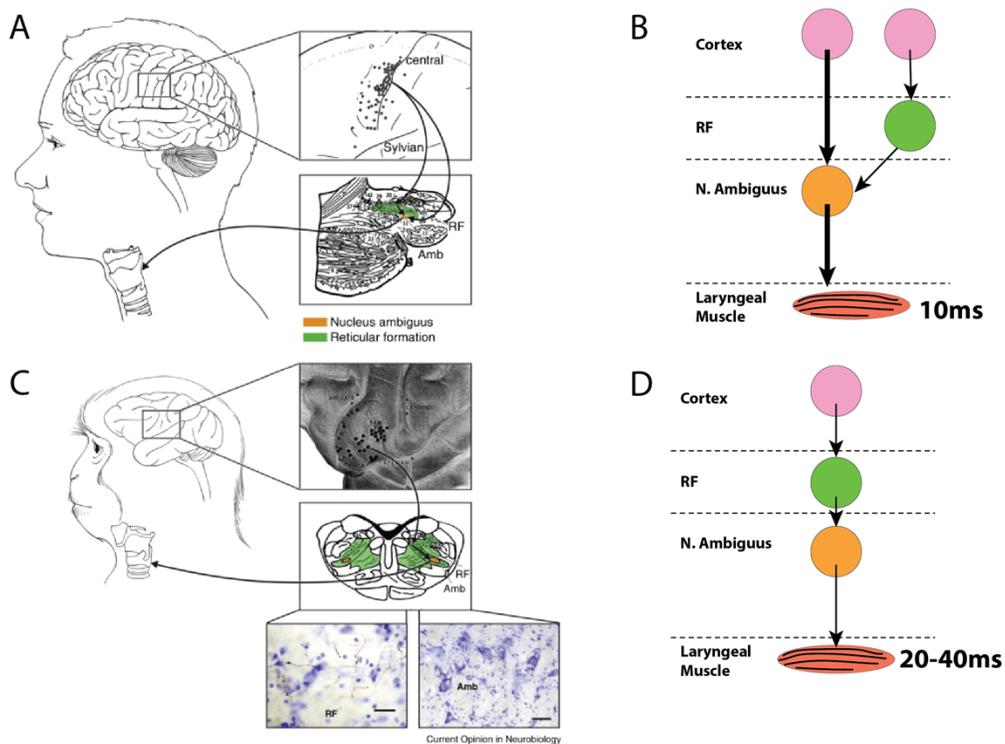
Figure 2.2: Latency of Direct and Indirect Cortical Projections.
A) Representation of the cortical pathways for laryngeal control in humans. Humans have both a direct projection to the Amb. as well the phylogenetically older indirect projection from cortex→reticular formation→Amb. **B)** Schematic representations of the direct and indirect pathways and the experimentally determined latencies in humans. **C)** Representation of the cortical pathway for laryngeal control in macaques. Macaques are only known to have an indirect, poly-synaptic pathway from cortex→reticular formation→Amb. **D)** Schematic representations of the dominant pathways and their experimentally determined latencies in humans. **(A)** and **(C)** used with permission from Simonyan (2014).

(Wang et al., 2015). A recent study experimentally tested this idea, by knocking down expression of PlxnA1 axon guidance receptor involved in the cortex of mice to be more human-like in its expression pattern (Gu et al., 2017b). These mice retain more of their direct-projecting layer 5 projections from M1 to the spinal cord limb motor neurons. Stimulating the CFA region of these PlxnA1 cortical knockdown mice resulted in significantly faster forelimb responses (~10ms) compared to WT mice (~13ms). Gu et al. (2017b) went on to demonstrate that the mice with increased numbers of direct projections could perform fine manipulation tasks better (e.g. handling dried capellini pasta and removing tape from their paws). Thus, the presence of a direct projection directly influences the latency of response and subsequently the fine motor skill of that muscle system.

Our lab's discovery that mice have a sparse direct projection from M1 to vocal motor neurons in Amb even in adults was unexpected (Arriaga et al., 2012). If only vocal learners are predicted to have this direct projection (Fitch et al., 2010; Jarvis, 2019), then are mice vocal learners whose ability we have overlooked? Cross-fostering studies, do not support mice as vocal learners (Kikusui et al., 2011), and genetic studies of different deer mouse species (genus *Peromyscus*) show there is a large genetic contribution to types of vocalizations (Jourjine et al., 2023). Initial experiments on this M1 LMC region in mice showed it may have a role in the control of frequency modulation of USVs (Arriaga et al., 2012), although it was unclear if this was via the loss of direct action on Amb motor neurons or through a different, multisynaptic pathway. Determining whether or not this laryngeal region of mouse M1 can indeed generate laryngeal movements is critical for our understanding of the neurobiology of mouse vocal circuits.

Here we asked if M1 in mice can control vocal muscles (laryngeal and jaw), and if the identified mice LMC connection for laryngeal muscles is similar to what is seen in direct

projections for vocal muscles in vocal learning species. We find that electrical stimulation of LMC generates a range of latencies that are consistent with a small number of direct projections and a large number of indirect projections; surprisingly, we find that stimulation of OMC generates a range of latencies that is enriched for short monosynaptic latencies. Further, both muscles could be activated from the same cortical location. Using retrograde transsynaptic tracing, we show that multiple muscles, including the forelimb, exhibit a large degree of overlapping representations in the LMC region of M1 as well as individual neurons representing multiple muscles of the larynx, jaw, and forelimbs. Lastly, we identify a role for OMC in mouse ultrasonic vocalizations (USV). Based on these results we argue that: 1) cortical control of vocal muscles from M1 is more widely distributed across mammalian species than previously believed; 2) the mouse M1 has much less homuncular organization than larger brained mammals; 3) the continuum hypothesis of vocal learning is supported by our findings.

2.2 Methods

2.2.1 Animals

Adult (greater than 8-weeks old) male and female C57B6/J mice were used for this study (Jackson Laboratories, Bar Harbor, ME). Males and one female were used for stimulation experiments; both males and females were used for tracing experiments in approximately equal numbers; only males were used for lesion experiments; females were used as stimuli to induce male USV song production. Mice were housed socially with a 12-hour light cycle with *ad libitum* food and water. All animal protocols were approved by The Rockefeller University IACUC committee.

2.2.2 ICMS and EMG

Surgery

Prior to the surgical and experimental procedures, mice were anesthetized using a ketamine/xylazine cocktail (100 mg/kg and 20 mg/kg, respectively) given at 0.1 ml/10 g. Mice were anesthetized until they were not responsive to a toe pinch. Throughout the procedure, mice were supplemented with ketamine as needed. The mouse was placed supine on the surgical space. To access the DG muscle of the jaw, a small incision was made at the midline of the neck to expose the trachea and the anterior digastric muscle. Subsequently, to access the CT muscle of the larynx, the sternohyoid muscle was cut at the midline and held aside with retractors. For forelimb EMGs, the mouse was place prone on the surgical area and an incision was made along the length of the forearm to access the extensor carpi radialis (ECR) muscle. A pair of wires, forming a bipolar EMG electrode, were inserted into the CT and secured with a small dab of VetbondTM Tissue Adhesive (3M). Next, a second pair of electrodes was inserted into the belly of the DG. We used a 0.0026" diameter wire for DG and ECR, and a 0.0015" diameter wire for the CT as the larger

diameter could not be easily accommodated in the CT. The tips of the electrodes were deinsulated. All EMG electrodes were made from formvar insulated nichrome wire (A-M Systems, Sequim, WA). Electrode wires were threaded through a pulled glass pipette (Sutter Instruments, Model P-1000). A slight bend was made on the inserted end of the electrode to provide a fishhook that would then remain attached to the muscle tissue.

After EMG wires were inserted into their respective muscles, the incision was closed using VetbondTM. The mouse was the rotated to a prone position and placed in the stereotaxic frame. The fur on the skull was removed and an incision was made along the midline of the skull. Using a dental drill, a large (~3 mmx3 mm) craniotomy window was made to expose the surface of the cortex contralateral to the muscles containing the electrodes. The dura was removed, and the surface of the brain was maintained wet using saline throughout the procedure.

Stimulations and EMG Recordings

For each mouse we performed cortical penetrations at intervals of ~250 μm, with deviations to this dimension to avoid blood vessels. Penetrations were made non-sequentially to limit biasing measurements due to the effects of anesthesia. There was an overall emphasis on the OMC and LMC regions in our sampling, with most experiments beginning in OMC as our control region with a known digastric response in order to verify equipment and recordings were functioning correctly. Coordinates labeled as LMC were between -0.5-1.0 mm anteroposterior (AP) and 1-1.5 mm mediolateral (ML). These were based on PRV tracing (Arriaga et al., 2012). Coordinates labeled as OMC were between 1.5 mm lateral and the lateral edge of the craniotomy, and approximately 1.8 mm to the anterior edge of the craniotomy. These were based on responsive sites obtained for DG by (Okobi, 2016) and descriptions of the ALM regions (Komiyama et al.,

2010). At each stimulation site, microelectrodes (50 µm diameter; FHC, Bowdoin, ME) were lowered to the approximate depth of layer 5B (~800 µm for LMC and ~900 µm for OMC). For stimulations we used an isolated pulse stimulator (Model 2100, A-M Systems), using current in a range of 30 µA-650 µA. For our analyses, stimulations above 600 µA were filtered out, resulting in 525 µA being the highest intensity represented in the analyses. For single-pulse ICMS we used one 0.2ms biphasic pulse provided at a 2 Hz rate. For four-pulse ICMS we stimulated with the following parameters: 8.5ms duration, 0.2 ms biphasic pulse, 2.3 ms inter-pulse interval, provided at 1Hz. Each round of stimulations lasted 30 seconds and constituted one recording. The same current was maintained throughout each 30 second round. Our parameters resulted in 60 single-pulse stimulations per round and 30 four-pulse stimulations per round. Of the 7 mice used all received four- and single-pulse stimulations in at least one site.

EMGs were recorded using a differential amplifier (Model 1800, A-M Systems). All recordings had a system notch filter and signals were bandpassed filtered (10 Hz –5 kHz). Amplifier output was split between an oscilloscope, a speaker, and a recording DAQ (USB-231, Measurement Computing). Data from the amplifier was captured at 15 kHz with the DAQ and streamed to the DAQami™ (Measurement Computing) software on a nearby laptop. One mouse was recorded at 20 kHz per channel, then downsampled to 15 kHz to be comparable with the other recordings.

Data Processing

Signals from each recording were filtered with a 4th-order 1 kHz high-pass Butterworth filter in the forward direction using the sosfilt function in sci-py Python library. After filtering, signals were full-wave rectified by taking the absolute value of the measured voltages. To get a

reliable quantification for stimulus-responses from each cortical site and muscle combination and reduce breathing signal contamination in the CT muscle, for each current tested we calculated the mean of each round (30 seconds) of 60 (single-pulse) or 30 (four-pulse) simulations at a given cortical site within an animal, termed the stimulus triggered average (StTA). One StTA was computed for each level of stimulation at a given site in an individual mouse. For example, if a current was tested more than once at a site, there would only be one StTA.

Each StTA was manually assessed for which sites had positive muscle responses. Responsive sites were considered based on the following two factors: 1) responses were present for at least two different current intensities, and 2) the responses were not exclusive to stimulations above 500μA. Three exceptions were made in the four-pulse sitmulations where there was an apparent response but only one stimulus current was used. Next, for each responsive site, we calculated the standard deviation (SD) of the 30 ms preceding the stimulus artifact. The threshold was then set at 2SD above the pre-stimulation mean. In the post-stimulus period, if the measured voltage in the post-stimulation period crossed the threshold for more than ~0.5 ms, then we considered that to be an EMG contraction. The first point of the EMG response to cross the threshold was used to measure the latency of the response for that cortical site-muscle combination. The latency for responses was determined based on the time between the end of the stimulus artefact and the start of the response as calculated above. Calculations were determined a couple ms after the stimulus artifact in order to prevent a lagging artifact from being considered a responsive value. Maximum amplitude was calculated as the highest voltage value in the post-stimulus phase of the StTA.

All data were processed using Python and other freely available libraries developed for Python, including: numpy, pandas, sci-py, matplotlib, and seaborn. Correlations were determined

using the scipy-py library's Pearson function. Libraries can be downloaded with their respective commands from the Python Package Index (PyPI).

UMAP and K-Means Clustering

All StTAs from mice were used to produce low-dimensional representations using Uniform Manifold Approximation (UMAP). First, we normalized each StTA with a maximum absolute scaler. These outputs were then fed to the UMAP with the following parameters (minimum distanc=0.3, n-neighbors=50, distance metric=Canberra, random state=20). For K-Means clustering, we performed the clustering on the normalized StTAs. This clustering resulted in each point being labeled with a cluster. After the UMAP projection was produced, points were then colored by their respective K-Means cluster label. We performed the K-Means clustering using 4 clusters as that was the number of combinations we had for each muscle (CT and DG) and each region (LMC and OMC) analyzed.

UMAP was performed in Python using the umap-learn library. Normalization and k-means clustering were performed with the sci-kit learn library.

2.2.3 Viral Injections

All viral injections were performed using the Nanoject III (Drummond Scientific). Injection needles were made from borosilicate glass and pulled using a Model P-1000 puller. The needles were backfilled with mineral oil.

Pseudorabies-Bartha Virus (PRV) Injections

We used two transynaptic PRV constructs, PRV-HA-mCherry/PRV-lp298 (titre: 6.8x10^8 pfu/ml) (Schneeberger et al., 2019) and PRV-GFP/PRV-lp297 (titre: 6.6x10^8pfu/ml) (Wang et al., 2020),

obtained from Lisa Pomeranz (Laboratory of Molecular Genetics, The Rockefeller University). For these experiments, both constructs were injected into each mouse, one construct per muscle. We alternated which construct was injected for each muscle between mice.

Injection procedures were similar to those performed previously by our lab (Arriaga et al., 2012; Arriaga et al., 2015). Mice were anesthetized with 1-2% isoflurane with oxygen. Mice were then placed supine with a nose cone continuously providing anesthesia to the mouse. The fur at the site of the incision was removed and betadine was applied to the skin. A midline incision was made at the level of the neck, from between the shoulders up to the interramal whiskers to expose the larynx and digastric muscles in one incision. The skin and fat were separated and maintained in position using retractors. For the DG muscle, injections were made without any further procedures. To access the CT muscle, as with the EMG surgery, the sternohyoid muscle was cut at the midline and held apart with retractors. Injections are then made on the exposed muscle. As the CT is small and mice exhibit deep breathing in the supine position under anesthesia, care was taken to prevent the injection needle from penetrating below the muscle.

For the CT muscle, we injected 50 nl of PRV four times at four different sites at a rate of 15nl/s. For the DG muscle, we injected 125 nl five times at four sites at 15 nl/s. The difference in volume was selected to account for the difference in the size of the muscles. After all the injections were completed, the separated tissues were closed and secured with Vetbond™. Mice were then administered bupivicane subcutaneously (0.25 mg/ml) and meloxicam (2-5 mg/kg). Post-operatively, mice were placed in a recovery cage placed on a heating pad and observed until they ambulated without issue. Once recovered, the mice were returned to their home cages. We used 5 mice for CT and DG co-injections of PRV.

For ECR injections, we performed them in combination with the CT. We used similar procedures as above. After the neck incision was made for the CT injection, the incision was closed as above. Mice were then laid prone in the surgical area. An incision was made on the skin of the forelimb on the dorsal side of the limb. We injected 125 nl five times at four sites at 15 nl/s, as in the DG muscle. The incision on the arm was secured with Vetbond™ and post-operative care was performed as above. We used 3 mice for CT and ECR co-injections of PRV.

Adeno-associated Viruses (AAVs)

Mice were anesthetized using 1-2% isoflurane with oxygen. After being placed in the stereotax, the fur on the scalp was removed. Betadine was applied to the exposed skin. A midline incision was made to expose the skull. The necessary number of craniotomies were made using a dental drill. For cortical injections, the needle was lowered into the brain up to 100 µm past the target and retracted into the target depth. For brainstem injections, the needle was lowered directly to the target depth. After injections were completed, the needle was left in the brain for 5 minutes to prevent injected medium from backflowing out of the brain onto the surface. The needle was retracted, and the skull scalp was closed at the midline with Vetbond™. OMC injections were made at +2 mm ML, +2.2 mm AP, and 0.9 mm below the surface. LMC injections were made at +1.4 mm ML, 0.3 mm AP, and 0.8 mm below the surface. For Mo5 injections, we injected at +1.4 mm ML, -6 mm AP, at an angle of 17° and a travel distance of 4.2 mm from the surface. For anterograde transsynaptic tracing we used AAV1-hSyn-Cre-WPRE. For paired tracing with PRV in the CT we used AAV2-CAG-GFP. For combined OMC and Mo5 injections we used AAV5-DIO-mCherry and AAVrg-GFP-Cre, respectively. We used 6 mice for AAV1-hSyn-Cre-WPRE injections into OMC, and 4 mice for AAV1-hSyn-Cre-WPRE injections into LMC. We used 6

mice for the dual injections of AAV2-CAG-GFP into OMC and PRV into CT. We used 3 mice for dual injections of AAVrg-GFP-Cre in Mo5 and AAV5-DIO-mCherry in OMC.

2.2.4 Perfusion and Cryosectioning

Animals were transcardially perfused using cold 1X PBS and cold 4% paraformaldehyde (PFA) in 1X PBS. Extracted brains were post-fixed in 4% PFA overnight at 4°C. Brains were cryoprotected in 10% sucrose in 1X PBS solution overnight at 4°C, followed by 30% sucrose in 1X PBS at 4°C until the brains sunk. Brains were then placed in tissue molds and frozen with Tissue Plus® O.C.T. Compound (Fisher Scientific). After freezing, the brains were cryosectioned (Leica CM 1950) at 50 μm and stored in 1X PBS with 0.01% sodium azide and stored at 4°C until used.

2.2.5 Histology

For all histology, instead of processing serial sections, for each set of sections we selected every 6th section from the brain. At this interval sections were separated by ~300 μm from one another.

Immunohistochemistry

All immunohistochemistry was performed with free-floating sections. Sections were washed in 1% Tween-20 in PBS for 3x15 min. Sections were then blocked in 1% bovine serum albumin (BSA) with 0.3% Triton-X in PBS for 2 hours. Following blocking, the sections were incubated in primary antibody with 1% BSA, 0.1% Triton-X in PBS overnight. On the following day, the sections were washed with 1% Tween-20 in PBS for 3x15 min and placed in the secondary antibody with 1% BSA and 1% Tween-20 in PBS for ~2-4 hours protected from light. Sections were then washed with 1% Tween-20 in PBS for 3x15 min. Sections were mounted using VECTASHIELD® Antifade Mounting Medium with DAPI (Vector Laboratories) and imaged on

28

an Olympus BX61 fluorescent microscope. Images post-processed in ImageJ. We performed antibody labeling on the following targets: green fluorescent protein (GFP), tdTomato and red fluorescent protein (RFP; same antibody used), and choline acetyl transferase (ChAT). The primary and secondary antibodies we used, and their respective dilutions, are listed in **Table 2.1**.

Nissl

OMC lesions were confirmed either from the naïve tissue, or using Nissl staining. Briefly, sections were washed in 0.1M PB for 10 min, followed by 2 min in ddH$_2$O. Sections were then placed in cresyl violet staining solution for 15-20 min. After staining, sections were placed in ddH$_2$O for 1 min, followed by an ethanol gradient of 50% EtOH and 70% EtOH for 8 min each. Next sections were placed in 95% EtOH for ~1 min with a few drops of acetic acid. Sections were moved to 100% EtOH for 2 min, and a separate fresh 100% EtOH for another 8 min. Sections were then washed in Histo-Clear® (National Diagnostics) for 3x8 min and coverslipped using DPX Mounting Medium (Sigma-Aldrich).

Stereology

To perform cell counts of the double-PRV injected brains (see above), we used the "Cell Counter" plugin in ImageJ. Green, red, and yellow (co-labeled) cell bodies were counted by adding these parameters as 'counters' in the FIJI cell counter plugin. Cell count outputs were analyzed and visualized using custom scripts written in python.

Table 2.1: List of 1°& 2° Antibodies.

1° Antibody	1° Ab Concentration
Rockland Chicken anti GFP	1:1000
Rockland Rabbit anti RFP	1:1000
Rockland Goat anti ChAT	1:100
Rockland Goat anti GFP	1:1000
2° Antibody	**2° Ab Concentration**
Invitrogen 488 Goat anti Chicken	1:500
Invitrogen 568 Donkey anti Rabbit	1:500
Invitrogen 594 Goat anti Chicken	1:500
Invitrogen 488 Donkey anti Goat	1:500
Invitrogen 594 Donkey anti Rabbit	1:500

2.2.6 Cortical Lesions

Vocal Behavior

At 8 weeks of age, group-housed male mice were sexually socialized by placing an adult female in their home cage overnight three days before the start of the experiment. Female-directed recordings were performed as previously described, using similar equipment (Chabout et al., 2015). This equipment includes a cooler for sound isolation (Igloo®). We recorded vocalizations using UltraSoundGateCM16/CMPA ultrasonic microphones which were connected to an

30

Ultrasound Gate amplifier, and digitally recorded using AvisoftRecorderUSG software (Sampling frequency: 250kHz; FFT-length: 1024 points; 16-bits). Microphones, amplifier, and software are all from Avisoft Bioacoustics® (Berlin, Germany). A set of pre-surgical baseline recordings were made for each mouse over three consecutive days. On the fourth day, surgeries were performed on the mice. Mice were singly housed for 4-5 days to allow for recovery from the surgery. After this recovery period, mice were placed back with their original cage mates for the remainder of the experiment. Three consecutive days of recording were repeated at 2 weeks and at 3 weeks post-surgery. Throughout the experiment, care was taken not to expose a male mouse to any one female more than once. At the end of the last recording session the mice were euthanized, transcardially perfused, and brains were extracted for histological processing.

Surgery

Surgery and injections were performed with similar procedures as described above. Mice were injected bilaterally in OMC (+2 mm ML, and +2.2 mm AP) with a total of either 1% ibotenic acid or with saline as a control. Injections were performed at 1200 μm below the cortical surface and 800 μm with 120 nl injected at each level. We injected 40 nl 3 times, at 1 nl/s, with 60 seconds between each injection. After the second set of injections, the needle was removed from the brain. Once all injections were completed, the scalp was closed, and post-operative care was provided.

Acoustic Analysis

Audio files were processed using Analysis of Mouse Vocal Communication (AMVOC) software we developed (Stoumpou et al., 2022). Audio files were processed using AMVOC and clustered into 5 different categories using mini batch K-Means clustering, and a 6th category of silence

syllable (defined as a period of length greater than 250 ms with no vocalizations). Silent "syllables" were used for determining syllable classes that begin and end phrases. We used these silent periods to separate each set of consecutive vocalizations into sequences (Chabout et al., 2015). A sequence was defined by being composed of 2 or more syllables. Acoustic features were measured by providing AMVOC's measured timestamps to Mouse Song Analyzer 2 (Stoumpou et al., 2022). Resultant measurements were analyzed with custom python scripts based on the analyses by Chabout et al. (2015).

Syntax Analysis

For the syntax analysis, we did not include sessions where mice vocalized <20 USVs in 5 min. As found in (Chabout et al., 2015), this type of limited data skewed our syntax results, and so we performed our syntax analyses on the maximum vocalizing day for each week per mouse. To model the syllable transitions within sequences, we used the syllable classes defined above. We represented transitions between each syllable type in a two-dimensional matrix where the rows and columns represent the different syllable categories, and each element equals the absolute probability of the corresponding transition. To examine the similarity between the transitions of two song sequences we applied an elementwise Mean Square Error (MSE) criterion to the corresponding transition matrices with values close to zero implying more similar transitions of syllables. The MSE between each sequence is calculated, and an overall mean MSE is calculated for each mouse in each week. We used this method to examine the changes in temporal structure between the first week and the next two weeks after the surgery.

2.3 Results

2.3.1 Acquiring EMG signals from CT

Prior to this study, there had not been any published reports performing EMG recordings on the CT muscle in mice. To confirm that our method of inserting electrodes to record from this muscle was appropriate for data collection (**Figure 2.3A**), we first tested our signals by stimulating the superior laryngeal nerve (SLN) (**Figure 2.3B**), which is a branch of the vagus nerve that uniquely innervates the CT muscle. After SLN stimulation, we could identify an EMG signal from the CT muscle (**Figure 2.3C**). A good EMG recording from the CT muscle was confirmed by the presence of a contraction rhythm that was synchronized with the mouse's breathing rhythm (**Figure 2.3D**). Next, to confirm that our method for ICMS was functioning, we used the forelimb muscle, extensor carpi radialis (ECR) as a control (**Figure 2.3E**).
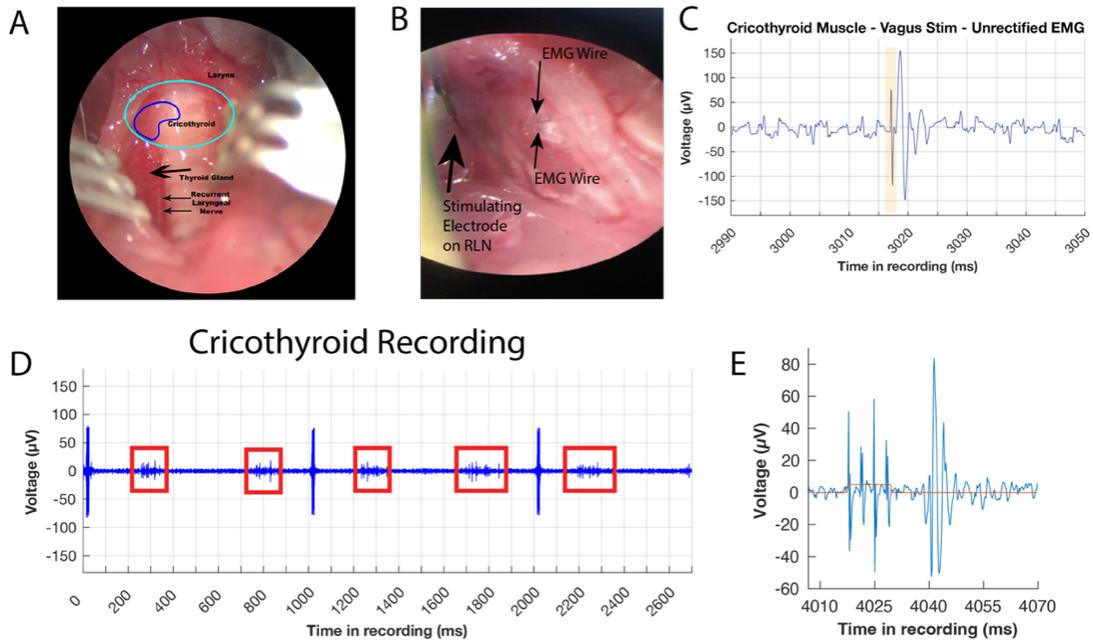
Figure 2.3: Cricothyroid EMG Electrode Location and Signal Acquisition.
A) Image showing the gross anatomy of the larynx and the surrounding tissue and organs. Light blue oval highlights the location of the larynx, which is identifiable by the cricoid cartilage. Dark blue shape outlines the CT muscle. **B)** Image showing the two wires inserted into the CT muscle to create a bipolar EMG electrode. Muscle is opaque due to the dermal glue applied to fasten the wires. The stimulating electrode used to confirm contractile responses in CT. **C)** Example trace of an SLN stimulation (highlighted in yellow) and the subsequent CT contraction. **D)** Example trace showing the rhythmic breathing signal captured by the electrodes in the CT muscle (red boxes). Large peaks not highlighted din red are stimulation artifacts. **E)** Single trace example of ICMS used to generate contraction in the ECR muscle.

2.3.2 Stimulation of Mouse Motor Cortex Induces Vocal and Jaw Muscle Contractions

We performed ICMS across M1 with an emphasis on the putative LMC, as well as OMC (**Figure 2.4A and B**). We repeatedly inserted a stimulating electrode across cortex, while simultaneously recording EMGs from laryngeal and jaw muscles (CT and DG, respectively) (**Figure 2.4C**). We applied ICMS to the cortical hemisphere contralateral to the recorded muscles. From each round of stimulations at each site (comprised of 30 or 60 individual stimulations separated by 0.5 s per site for single-pulse, or 1 s per site for four-pulse, see **Methods** for detail), we calculated the stimulus triggered average (StTA) at that site. This also helped to abrogate noise

created by respiratory activity from the CT. An EMG response was determined for the StTA based on a 2 S.D. threshold above the baseline preceding the simulation (see **Methods**).

Using four-pulse stimulations, we stimulated LMC in 6 mice and OMC in 7 mice. We surprisingly found LMC stimulation caused CT muscle contractions in 5 of 6 mice, as well as causing the DG muscles to contract in 3 of 6 mice. However, the CT muscle could be activated not only from the LMC region, but the OMC in 6 of 7 mice, as well as other portions of M1 (**Figure 2.4D**). The DG muscle could also be activated by stimulation in OMC in 7 of 7 mice (**Figure 2.4E**), as previously described (Mercer Lindsay et al., 2019; Okobi et al., 2019). Some stimulations in the lateral primary somatosensory cortex (S1) could activate the DG while sites in three animals activated the CT. We assessed the distribution of responsive sites by calculating a kernel density estimate that was weighted by the percentage of times that coordinate was responsive to our stimulations across animals (**Figure 2.4F and G**). In both CT and DG, we find that both muscles were generally excitable from both anterior and posterior poles of M1, with CT appearing more readily activated from the more posterior LMC region.

In our single-pulse stimulation trials we stimulated LMC in 4 mice and OMC in all 7 mice. In these stimulations, both muscles responded to fewer cortical stimulations sites, within a smaller anatomical range. These focal regions were primarily restricted to M1 or coordinates near the border of the region (**Figure 2.4H and I**). The CT muscle could be activated from both OMC (4 of 7 mice) and LMC (2 of 4 mice) regions (**Figure 2.4H**). The DG muscle could also be excited from both poles of M1 (OMC: 7 of 7 mice; LMC 1 of 7 mice)(**Figure 2.4I**). Similar to the four-pulse stimulation analyses, we performed a kernel density estimate weighted by percent of times that site was responsive. We found that although CT can be activated from the anterior region, there were more CT sensitive sites in LMC than in OMC (**Figure 2.4J**). Conversely, DG was more

frequently responsive to stimulation in the more anterior OMC region (**Figure 2.4K**). The difference in the maps between four- and single-pulse stimulation suggest that the four-pulse stimuli could be spreading across a wider surface of the cortex. With the single-pulse stimulations we had more certainty that the muscle activations we observed were due to the focal stimulation and not some larger circuit-level effects. Further, the single-pulse responses facilitated our analysis for determining the response latency of the laryngeal and jaw muscles. For these reasons we focus our subsequent analyses and interpretations on the single-pulse data, unless otherwise noted.
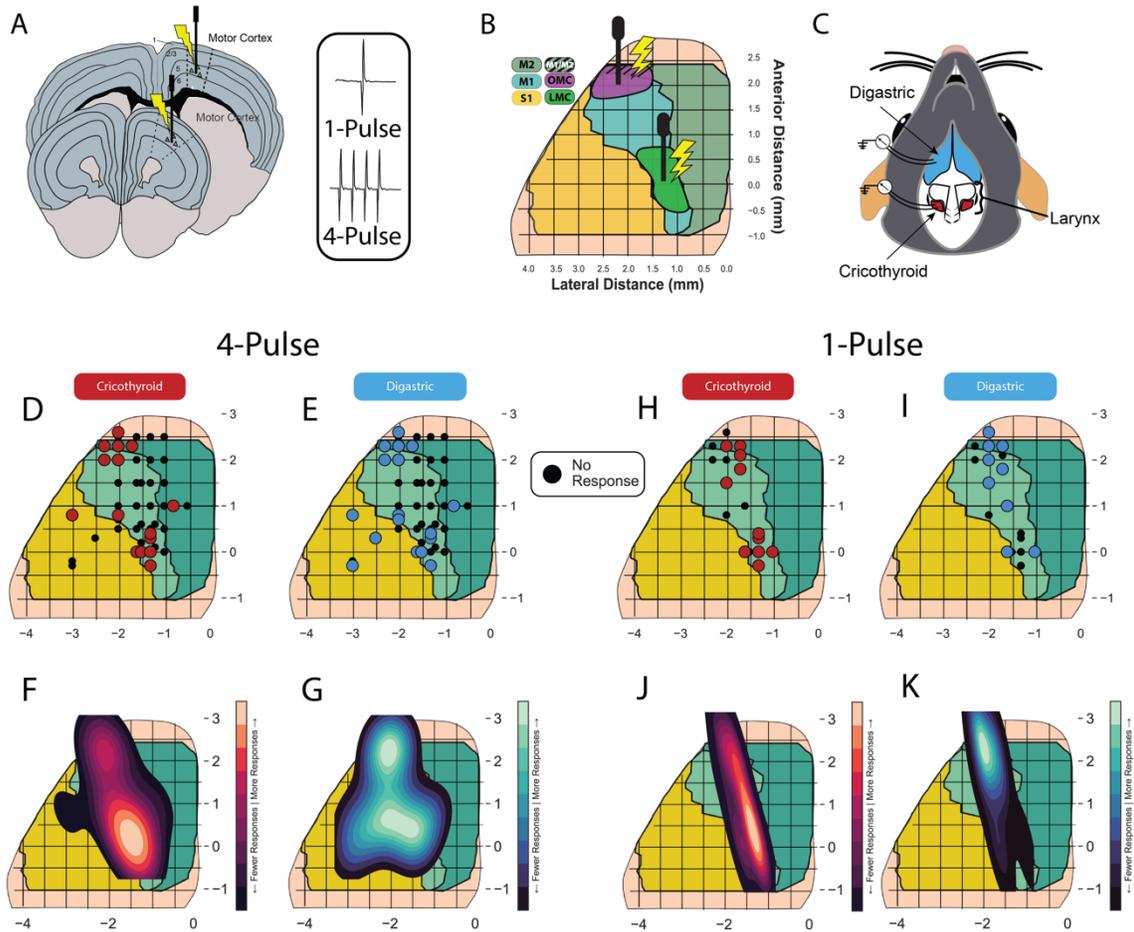
Figure 2.4: Overview of ICMS Technique and Responsive Sites.
A) Example of ICMS in M1 on mouse motor cortex and approximate depth of stimulation at layer 5. **B)** Dorsal view of the surface of mouse cortex, with the LMC (green) and OMC (magenta) regions highlighted. Striped area represents where M1 is less dorsal or below M2 in the dorso-ventral axis. **C)** Overview of the anterior digastric (blue) and cricothyroid (red). **D and E)** Responsive sites to four-pulse stimulations for CT **(D)** and DG **(E)**. **F and G)** Kernel density maps for the distribution of the most responsive sites across mice**. H and I)** Responsive sites CT **(F)** and DG **(G)**. **J and K)** Kernel density maps of responses weighted by the response probability of each site**.**

In our qualitative analysis of example cortical sites and EMG responses, there appeared to be differences in the response properties of each muscle based on the cortical region stimulated (**Figure. 2.5**). Although responses could be generated in both muscles from stimulation in LMC and OMC, the CT had larger responses to LMC stimulation than to OMC stimulation (**Figure 2.5A and B**). Conversely, the DG had larger responses to OMC stimulation (**Figure 2.5C and D**). These patterns also occurred in the four-pulse stimulations (**Figure 2.5E-H**). We noted that LMC stimulations tended to generate faster responses than from OMC stimulation (**Figure 2.5**).

We noted that some of our responsive sites overlapped with previously described representations of forelimb musculature (Tennant et al., 2011). To test whether the laryngeal and forelimb muscles could be activated from similar locations, we again performed ICMS while recording from CT, DG, and the extensor carpi radialis (ECR) simultaneously and found that all three muscles could be activated in different combinations from some of these sites (*n=1* mouse). Using four-pulse stimulations in in M1 we found at least one site that could produce responses from CT, DG, and ECR muscles (**Figure 2.6**).
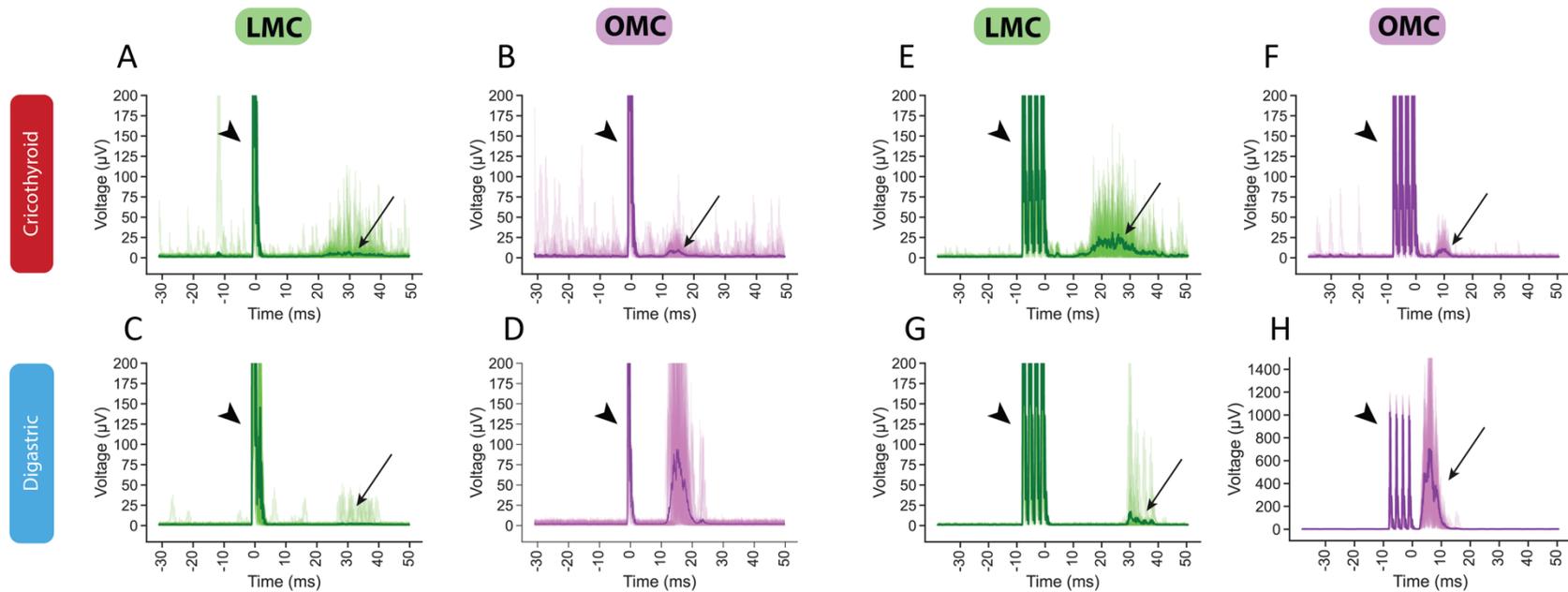
Figure 2.5: Example Responses from CT and DG to Single- and Four-Pulse ICMS.
A and B) Example EMG responses from the CT muscle after a round of single pulse stimulation in the LMC (A) and OMC (B) cortical regions. Shown is a StTA plot in one round of stimulations, at one site, in one animal. Fine traces are overlays on individual ICMS trials, bolded trace is the mean of the individual traces represented in each plot. **C and D)** Example EMG responses from the DG muscle after single pulse stimulation in the LMC (C) and OMC (D) cortical regions. **E-H)** Example EMG responses from CT (E and F) and DG (G and H) after four pulse stimulations. Arrowhead, stimulus artifact. Arrows, EMG response.
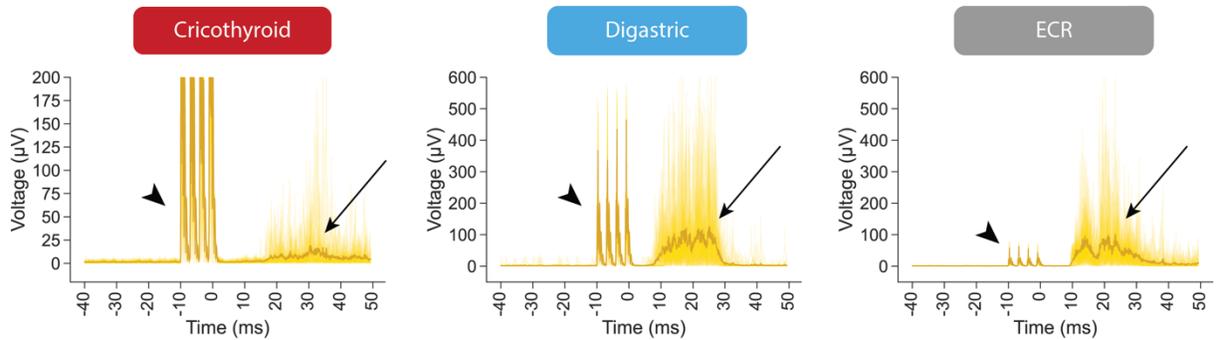
Figure 2.6: Example Cortical Site Producing Responses from CT, DG, and ECR.
Example traces and mean response to stimulation at a coordinate in M1 (1.6L,1.6A) that activated the CT, DG, and ECR muscles simultaneously. Arrowhead, stimulus artifact. Arrows, EMG response.

2.3.3 Quantifying Muscles Responses from ICMS

To quantify the differences in the response properties of the different muscles from either LMC or OMC stimulation, we calculated a response ratio, which we defined as the ratio between the maximum voltage value (μV) of the StTA for each muscle combination (CT/DG and DG/CT) at any site that contracted at least one of the two muscles. The CT muscle amplitude response was ~4 times larger (median) than the DG muscle when the LMC region was stimulated (Mann Whitney U-Test, p=0.0012) (**Figure 2.7A**). The opposite was true for DG responses relative to CT when stimulating the OMC (Mann Whitney U-Test, p=0.032) (**Figure 2.7B**); the median DG response to OMC stimulation was ~12 larger than the CT. We compared the minimum necessary current to detect an EMG in the StTAs for each responsive coordinate as these differences in response may have been due differences in the current used to stimulate the two regions. Although LMC thresholds were generally higher, there was no statistically significant difference in the sensitivity of cortex to generate an EMG in either the CT (t-Test, p>0.05) (**Figure 2.7C**) or in the DG muscle (t-Test, p>0.05) (**Figure 2.7D**).
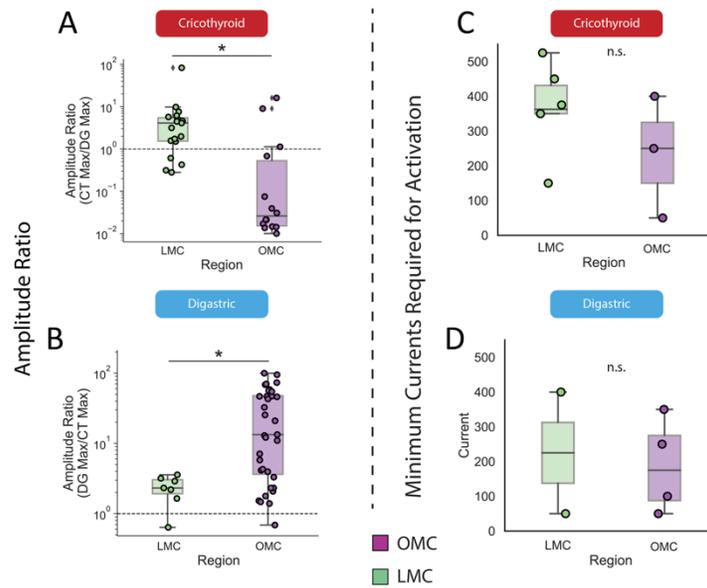
Figure 2.7: Quantification of Response Differences by Region of Stimulation.
A and B) Box plots representing the amplitude response ratio for LMC and OMC for each muscle, respectively. **A)** CT tended to have higher amplitudes relative to DG in LMC (Mann-Whitney, p=0.001). **B)** DG tended to have higher amplitudes relative to CT in OMC (Mann-Whitney, p=0.027). **C and D)** Box plots comparing minimum response currents for StTAs between LMC and OMC. No significant difference between minimum stimulation currents between the two regions of M1 for either muscle (*p>0.05*, Mann-Whitney).

To better understand the differences in the EMG response from ICMS, we performed a low dimensional reduction of all StTAs using uniform manifold approximation (UMAP). The UMAP separated mainly by muscle identity (CT and DG) (**Figure 2.8A**), not cortical region identity (OMC or LMC) (**Figure 2.8B**), and this was the case whether there was a muscle response (black circle edges) or not (no edge). There was only a small cluster of sites in OMC with specific DG responses (**Figure 2.8A and B**, circled). We performed K-Means clustering of the StTAs using n=4 clusters corresponding to the four muscle-cortical region pairs of interest (CT-OMC, CT-LMC, DG-OMC, and DG-LMC) and overlaid cluster identities on the UMAP, showing that two clusters were more associated with DG StTAs (**Figure 2.8C**, each color is a different K-Means cluster). One cluster in particular appeared to be primarily composed of responsive DG sites from

OMC stimulation (circled). These results were similar on the four-pulse stimulations, with muscle identify being separable and the most distinguishable cluster was for a subset of responses from the DG muscle (**Figure 2.8D-F**, circled).

We may not have been able to discretely separate responsive sites between muscles and regions due to number of points relative to the total number of points used, making the particularly strong DG responses from OMC stand out. Alternatively, we may be losing important information for distinguishing responses by using the StTAs which are an averaged response. Overall, our single-pulse stimulations may be more well suited for our goal of determining the minimum region of M1 that can generate a muscle contraction and the latency of these muscle responses than for discerning general relationships of a region to specific behaviors.
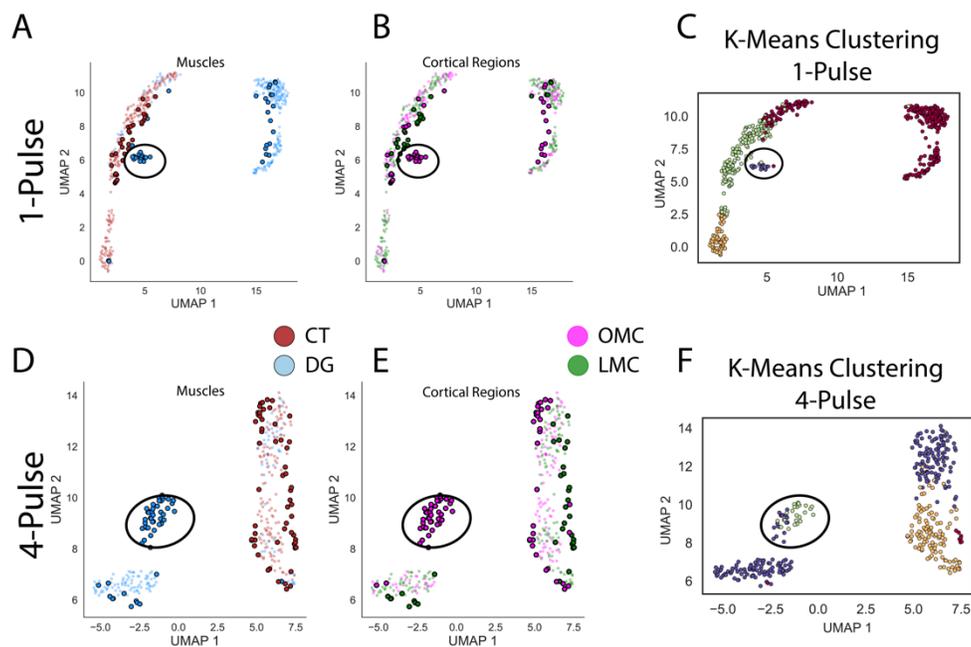
Figure 2.8: Low Dimensional Representations of EMG Traces.
A and B) UMAP projection of EMG StTA data. (**A**) is color coded by muscle and (**B**) is color coded by cortical region. Bolded points are responsive stimulations, pale points are non-responsive stimulations. **C)** K-means clustering of the StTA data overlayed on the UMAP from (**A**) and (**B**). Circled points represent the cluster identified by K-Means that represents the majority of DG responses from OMC in both single-pulse and four-pulse stimulations. **D-F)** Similar representation of data as **A-C** from four-pulse stimulations. Information about muscle is more easily segregated than information about location of cortical stimulation.

42

2.3.4 Synaptic Latencies Indicate Direct and Indirect Connectivity

If the LMC projection in mice is monosynaptic, it would be expected to have short, ~10ms latency responses in the CT muscle contraction (Arriaga et al., 2012). To calculate the latencies, we measured the time between the end of the stimulation and when the EMG event is detected by threshold crossing (**Figure 2.9A**). We found that responses in the CT muscle to single-pulse stimulation in LMC were often slower than responses from stimulation in OMC. Responses latencies from single-pulse stimulations in LMC ranged from ~5 to 27 ms (mean=16.77 ms, n=2 mice responsive), and responses from OMC stimulations ranged from ~5-17 ms (mean=10.51 ms, n=4 mice responsive) (**Figure 2.9B**). The same was true for the DG muscle. DG response latencies to single-pulse stimulation in LMC ranged from ~5ms to 45 ms (mean=23.98 ms, n=1 responsive mouse), and responses from OMC stimulations ranged from ~6 to 18 ms. (**Figure 2.9C**). When we applied the same analysis to our four-pulse stimulations in both muscles, the results were consistent with the patterns observed from single-pulse stimulations. Even with the possibility for summed current from the increased number of pulses per stimulation, the CT responses from LMC stimulation were again slower than those from OMC stimulation (LMC: mean=9.67 ms, n=5 mice responsive; OMC: mean=5.18 ms, n=6 mice responsive) (**Figure 2.9D**). This pattern was also true for DG responses from each area (LMC: mean=21.39 ms, n=3 mice responsive ; OMC: mean=3.23 ms, n=7 mice responsive) (**Figure 2.9E**). The near-zero ms latencies observed are likely due to the first pulses leading to a response before the end of the full four-pulses. The distribution of latencies from both types of stimulations resulted in more narrow latency distributions from OMC, compared to the latency distributions from LMC which had a wider/larger range.

Cumulative distribution plots of the single-pulse responses revealed that in the CT no more than ~50% of the latencies below the ~15 ms, whereas the remaining ~50% were longer. Overall,

the responses from OMC tended to occur earlier, and the distribution of latencies were significantly different between OMC and LMC responses for both the CT (Kolmogorov-Smirnov, $p=6.4\times10^{-4}$) (**Figure 2.9F**) and DG muscle (Kolmogorov-Smirnov, $p=0.031$) (**Figure 2.9G**). These distributions are consistent with a present, but sparse direct projection from the LMC region to Amb, that is dominated by indirect projections (Arriaga et al 2012).

Interestingly there were between-muscle differences we could observe in the StTAs of some coordinates. Notably, with LMC stimulation, although both muscles tended to respond later than from OMC stimulation, the CT muscle could be seen responding before the DG muscle (**Figure 2.10**).

A

Pre-Stimulus Mean
Stimulus Artifact
EMG Response
2-SD Threshold
Detected Latency

1-Pulse

B — LMC / OMC — Cricothyroid

Mean: 16.77ms
Std.: 6.12ms
Median: 17.13ms

Mean: 10.51ms
Std.: 3.54ms
Median: 11.73ms

C — Digastric

Mean: 23.98ms
Std.: 16.61ms
Median: 12.59ms

Mean: 10.35ms
Std.: 2.3ms
Median: 9.86ms

4-Pulse

D — LMC / OMC — Cricothyroid

Mean: 9.67ms
Std.: 6.5ms
Median: 11.33ms

Mean: 5.18ms
Std.: 3.24ms
Median: 4.46ms

E — Digastric

Mean: 21.39ms
Std.: 8.89ms
Median: 24.59ms

Mean: 3.23ms
Std.: 1.76ms
Median: 2.73ms

F Cumulative Distribution – CT

G Cumulative Distribution – DG

OMC
LMC

45

Figure 2.9: Latency of Muscle Responses to Single- and Four-Pulse ICMS.
A) Example StTA showing the pre-stimulus period used to determine the threshold (green), the stimulus artifact (magenta), and the post-stimulus period where a positive muscle response in the StTA may be detected (blue). Black dotted line represent the 2 S.D. threshold determined from the variance of the pre-stimulus period. Red dot indicates where the threshold criteria was met and where the latency of the activation is assigned. **B and C)** Histograms of latencies from combined StTAs for single-pulse stimulations. The muscle latency distribution for stimulation in LMC was much wider than the latency distribution from stimulation in OMC. **D and E)** Histograms of latencies from combined StTAs for four-pulse stimulation. **F)** Empirical cumulative distribution of response latencies for the CT muscle from single-pulse stimulations (Kolmogorov-Smirnov test, $p=6.4 \times 10^{-4}$). **G)** Empirical cumulative distribution of response latencies for the DG muscle from single-pulse stimulations (Kolmogorov-Smirnov test, $p=0.031$). Magenta designates OMC values and green designates LMC values.

Generally, one might expect that a mixed population of direct and indirect projections, the strength (e.g. amplitude) of the muscle response could be stronger in those that receive the direct cortical projection. To test this idea, we plotted the latency against the respective maximum muscle response amplitude for each responsive StTA and determined any possible correlation. We found a trend (LMC p=0.07 and OMC p=0.19) for a negative correlation between latency and muscle response amplitude for the CT muscle's responses from LMC and OMC stimulation where the shortest latencies (indicative of direct projections) tended to be associated with higher amplitude muscle responses (**Figure 2.11A and B**). We noticed there were outlier values in each of these plots. When we recalculated the correlation with these three values removed, the LMC responses were more correlated (**Figure 2.11C**), and OMC responses were more correlated as well (**Figure 2.11D**). When we inspected the StTAs we found many normal responses to a single site (**Figure 2.11E**), whereas an example of one of the outlying values we removed clearly came from an abnormal signal in an otherwise responsive StTA (**Figure 2.12F**). For the DG muscle, responses from LMC
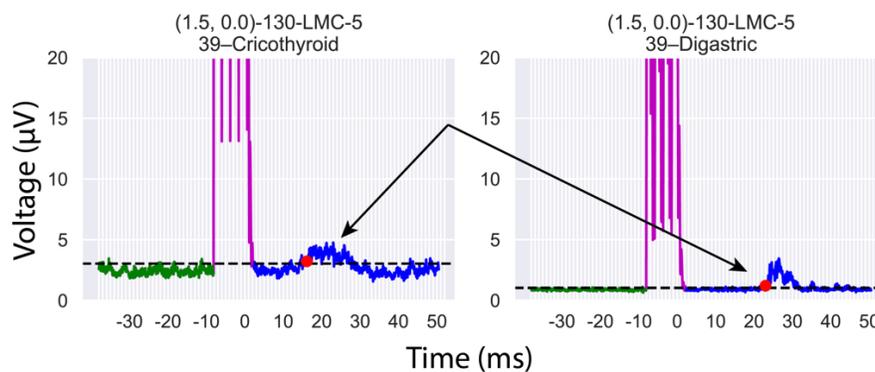
Figure 2.10: Difference in Response Latency Between Muscles.
Example of two StTAs from the CT and DG muscle calculated from the same site and stimulation current in one animal. The mean CT response to these stimulations was earlier than that of the DG responses

appeared to have no correlation between amplitude and latency (p=0.93) (**Figure 2.11G**). The

responses from OMC stimulation had a negative correlation between amplitude and latency, but

did not have a significant correlation (p=0.17) as almost all of the responses were of a similar

latency **(Figure 2.11H)**.
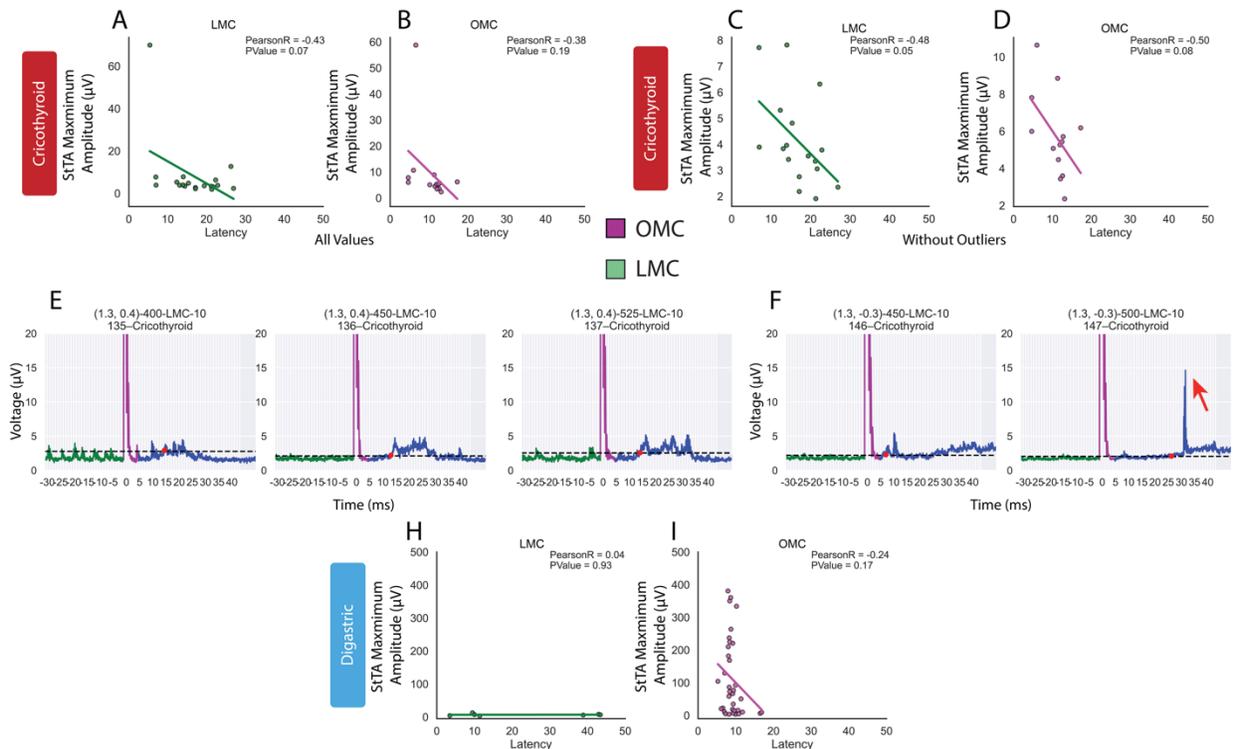
Figure 2.11: Relationship Between Response Latency and Response Amplitude.
A and B) Relationship between latency and maximum amplitude for the CT muscle from
stimulation in LMC (**A**) and OMC (**B**). **C and D)** Same as in (**A**) and (**B**) with the outlying values
removed for the respective region. **E)** Example set of StTAs from one coordinate at different
currents without outlying measurements. **F)** StTAs of a normal response on the left and the
response with a recording artifact on the right that was removed as an outlier (red arrow). **G and
H)** Relationship between latency and maximum amplitude for the DG muscle from stimulation in
LMC (**H**) and OMC (**I**).

We next tested whether latency or muscle amplitude could be affected by changes in the intensity of cortical stimulation. A subtle example of this may be seen in **Figure 2.11F**. Due to the variability in the depth of anesthesia and current intensities used between mice, we performed this analysis per responsive site, rather than in aggregate across mice, as above. We selected sites that had two or more responsive StTAs in the muscles if they were elicited from different current intensities. We measured the overall effect of different stimulation intensities in LMC and OMC by using the slope of the correlations as a measure of the direction in which amplitude or latency tended to change with current intensity. Positive slopes would suggest overall positive correlations; negative slopes would suggest overall negative correlations; slope values near zero indicate no-effect. First, we compared stimulus current with the maximum amplitude of each StTAs. For both the LMC and OMC regions, the CT's slopes of the response amplitude with changing current were not different from zero (Wilcoxon test p>0.05) (**Figure 2.12A**). One site in OMC was an outlier and appeared to decrease in amplitude with increasing current, this site was the same one that was an outlier in **Figure 2.11B**, when removed the result was unchanged. In the DG, increasing currents also did not affect the maximum amplitude of responses to LMC stimulation, but responses to OMC stimulation were significantly positive (Wilcoxon test, p=0.0039) (**Figure 2.12B**). These results suggest that in the CT we are getting the maximum response from that region that is possible with a single-pulse stimulation, whereas the DG has a larger dynamic range of activation and the potential for specialized circuits to do this in OMC.

Next, we performed the same analysis of slope correlations on the latencies as a response to increasing stimulus intensity. We find that correlation slopes for the CT are not significantly different from zero in either the LMC or OMC (**Figure 2.12C**). In the DG we also found that the slopes did not differ from zero (**Figure 2.12D**). The consistency of the latency is important as it
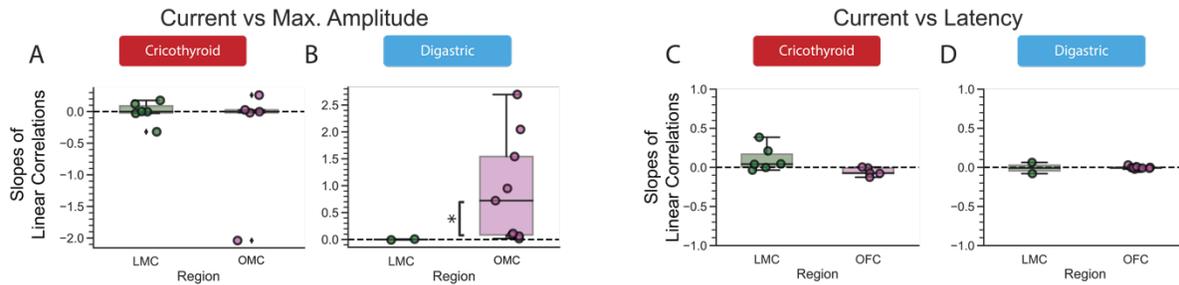
Figure 2.12: Relationship Between Stimulus Intensity and Muscle Response.
A and B) Slopes of the correlations calculated between the stimulus current and maximum amplitude of the StTAs across responsive sites from the CT **(A)** and DG **(B)**. Each point represents the StTAs of that region calculated in individual mice. Only DG correlation slopes were significantly different from zero (Wilcoxon, p=0.0039), other slopes did not statistically differ from zero (p>0.05).**C and D)** Slopes of the linear correlation calculated between the stimulus current and the response latencies from the CT **(C)** and DG **(D)**. For each muscle, none of the distributions were statistically different from zero (p>0.05).

suggests that latency measurements are not based on stimulus parameters but are instead responses

that are bound by the properties of the circuit being activated.

2.3.5 Cortical Neurons Represent Multiple Muscles

In the typical homuncular representation of M1 organization, each region of the cortex is seen as

specialized for the activation of specific muscles, albeit with some inter-individual variability

(Tennant et al., 2011). From our stimulation results we were surprised by the amount of co-

localized muscle activation. Although the OMC region has been shown to control many muscles

in mice—including those for the whisker, jaw, and forelimb (Mercer Lindsay et al., 2019)—, the

overlap in the LMC region was much less expected. To investigate the potential neural substrates

for this co-activation, we injected pseudorabies-Bartha (PRV) encoding eGFP and RFP into the

CT and DG muscles of mice, alternating colors between muscles in different animals, and allowed

the virus to incubate for 96-hour (Chabout et al., 2016; Zheng et al., 2022) (**Figure 2.13A**). We

found some subcortical regions with an overlap of transfected neurons (both single and co-labeled neurons), including the intermediate reticular formation and locus coeruleus (**Figure 2.13B**). Within M1, both CT- and DG-representing neurons infected with PRV were inter-mixed within the same cortical column (**Figure 2.14A**). For both muscles there was a variable density of neurons with a peak density in the middle of the distribution (**Figure 2.14B**). We did not observe any transfected neurons in OMC. Neurons representing each muscles appeared adjacent to each other. Notably, there were double labeled neurons representing both muscles. Across the anterior-posterior axis 6.7-22.2% of M1 neurons were co-labeled from CT and DG muscle injections (**Figure 2.14C**). This finding is consistent with a study conducted in parallel in *S. teguina,* showing co-representation of CT and DG muscles in M1. That study showed a greater percent of co-representation, which could be due to species differences, or degree of synergism between the two muscles (Gu et al., 2017a).
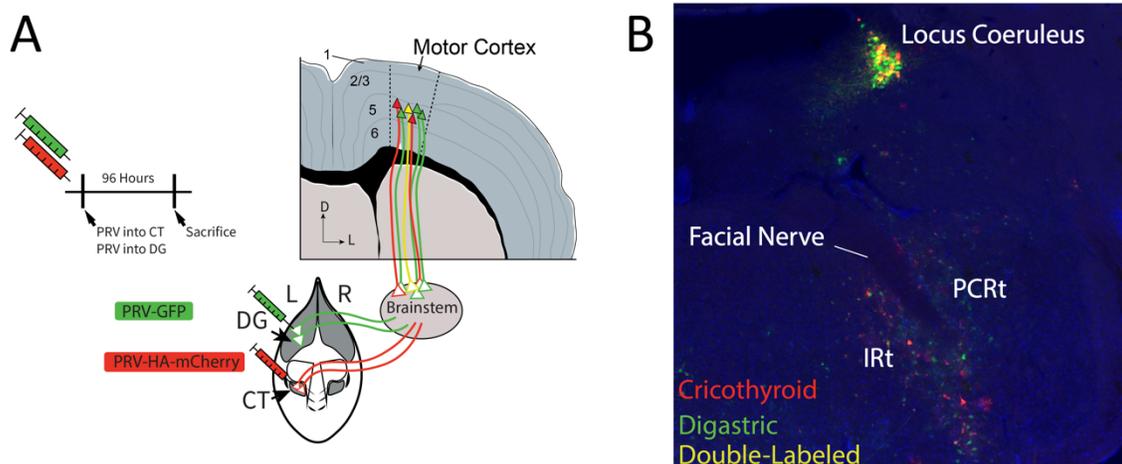
Figure 2.13: PRV Tracing Overview and Co-labeled Neurons in the Brainstem.
A) Overview of experimental procedure to inject PRV tagged with different fluorophores into CT and DG. Virus was alternated for each muscle across mice to balance potential differences in expression. **B)** Backfilled neurons from virus injected into the CT (red) and DG (green) in the intermediate reticular formation (IRt), parvicellular reticular formation (PCRt), and the lateral parabrachial nucleus. Both IRt and lateral parabrachial show co-labeled neurons, while PCRt is primarily single-labeled neurons.
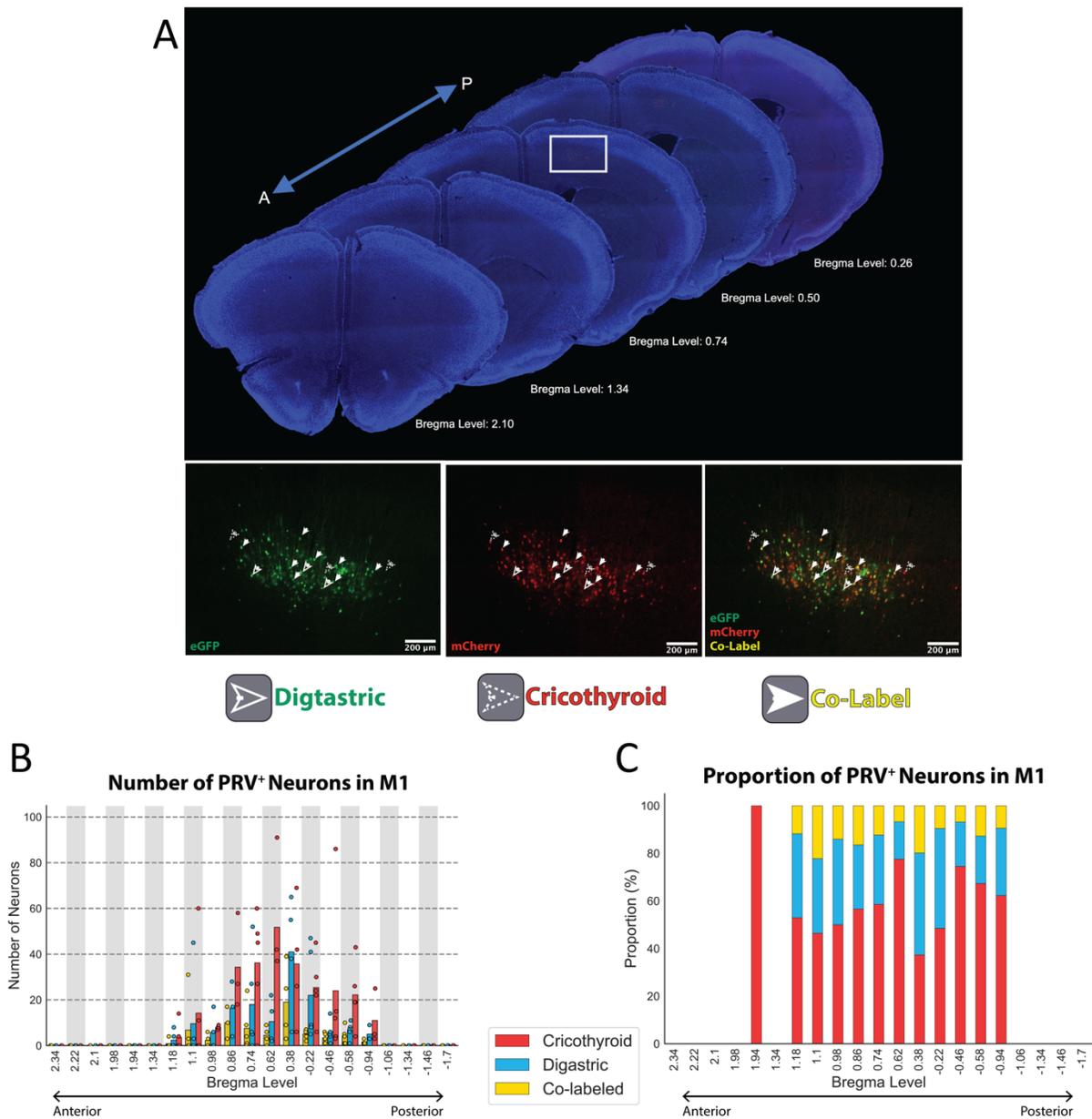
Figure 2.14: PRV Tracing Reveals Overlapping Representations of CT and DG in M1.
A) Coronal slices from a representative animal aligned anterior to posterior and close-ups of region outlined by the white box. Left, green neurons are backfilled from injection into the DG. middle, red neurons are backfilled from injection into the CT; right, merge depicting the single-labeled and co-labeled cells from the two muscles injected. **B)** Mean and raw counts of layer 5 neurons for each mouse with PRV injected into different muscles. **C)** Mean proportion of cells counted to be either single- or co-labeled by PRV injected into CT and DG muscles.

We also saw PRV-labeled neurons in layer 5 of S1 (**Figure 2.15A and B**). Although not described by Arriaga et al. (2012), this discrepancy may be due to the difference between the 96-hour injections used here compared to the 90-hour injections used by Arriaga et al. There were fewer neurons relative to M1, and potentially 3rd and 4th order neurons, we quantified these neurons using a similar approach and found a different pattern of single- and co-labeled neurons. In S1 we found neurons that were at the same anterior-posterior position as the caudal portion of the M1 population we identified (**Figure 2.15C**). Some of these neurons are likely to be from the so-called "overlap zone" between lateral M1 and medial S1 (Tennant et al., 2011; Yamawaki et al., 2021), and some that were more posterior to the M1 populations described above. The proportions of PRV-labeled neurons were more even than in M1. CT-only and DG-only neurons had similar proportions and across our identified population with a slight increase in CT-only neurons in the middle portion of the population (**Figure 2.15D**). Co-labeled neurons comprised 9.4-20.8% of the PRV^{+} population.
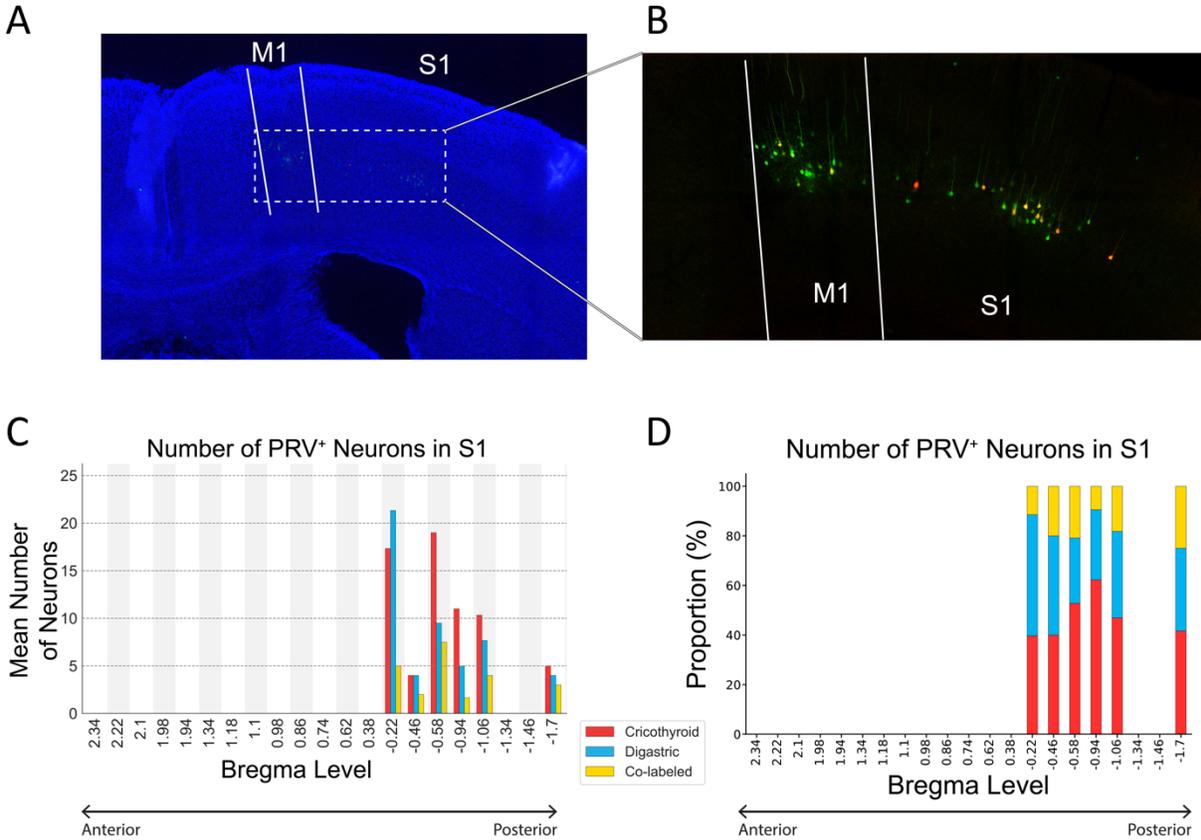
Figure 2.15: PRV Tracing Reveals Overlapping Representations of CT and DG in S1.
A) Coronal section showing location of PRV$^+$ neurons in S1. **B)** Zoom of inset from **(A)**. **C)** Mean number of neurons in S1 labeled from each of the muscles injected. **D)** Proportions derived from means of single- and co-labeled neurons in S1 based on counts from **(A)**.

The region of M1 labeled by PRV injection in CT and DG is what Arriaga et al. (2012) called LMC. Using ICMS this region has been previously identified as a forelimb representing region and called the CFA (Li and Waters, 1991; Tennant et al., 2011), which has been supported by viral tracing (Gu et al., 2017b; Wang et al., 2017). We wondered if the co-representation of muscles was unique to pairs of muscles that are synergistic, like CT and DG, or if this coincidence in cortical location is a broader property of mouse M1. To address this question, we injected PRV encoding different fluorophores into the CT muscle and the ECR of the forelimb (**Figure 2.16A**). Similar to the CT and DG double-PRV injected mice, we found that neurons representing CT and

ECR muscles shared the same cortical space, as well as co-labeled layer 5 neurons that represent both muscles (**Figure 2.16B and C**). Like the CT- and DG-representing neurons, the CT- and ECR-representing neurons had similar distributions in the anterior-posterior axis (**Figure 2.16D**). Although, the relationship between single and co-labeled CT- and ECR-representing neurons was different from the CT-DG injections. We found more single-labeled CT-representing neurons in anterior M1, and more single-labeled ECR-representing neurons more posterior in M1 (**Figure 2.16E**). There were also fewer co-labeled neurons in the more anterior sections. We quantified the cumulative distribution of CT-representing neurons that were transfected in each double-injected group (CT-DG and CT-ECR) (**Figure 2.16F**) and found that there was no significant difference in the anterior-posterior distribution of PRV$^+$ CT-representing neurons (Kolmogorov-Smirnov test, p>0.05) (**Figure 2.16G**). Although there was not a significant difference in the distribution of co-labeled neurons along the anterior-to-posterior axis, co-labeled CT-DG-representing neurons tended to decrease more posteriorly and co-labeled CT-ECR-representing neurons increased along this axis (**Figure 2.16H**). The increase in co-labeled neurons was significantly correlated with anterior-posterior position (Pearson Correlation, p=0.026). These findings suggest that there is not a classical somatotopy or "homuncular" organization in mouse M1, but there are still local, yet overlapping, circuits organized around shared muscles that are organized in a more gradient-like pattern rather than being distinct.

A

Motor Cortex

PRV-GFP

ECR

Spinal Cord

Brainstem

96 Hours

PRV into CT
PRV into ECR

Sacrifice

PRV-HA-mCherry

CT

B

C

Forelimb
Cricothyroid
Co-Label

D

Number of PRV$^+$ Neurons in M1

Number of Neurons

Bregma Level

Anterior   Posterior

E

Proportion of PRV$^+$ Neurons in M1

Proportion (%)

Bregma Level

Anterior   Posterior

Cricothyroid
Forelimb
Co-labeled

F

Mean Counts of PRV+ CT Neurons

Mean Count

Bregma Level

CT-DG
CT-ECR

G

Cummulative Distribution of PRV+ CT Neurons

Cummulative Proportion of Neurons

n.s.

K-S Test
p: 0.987

PRV+ Cricothryoid Neurons

CT-DG
CT-ECR

H

Difference in Direction of Distribution ofPRV+ Co-Labeled Neurons

Proportion of Co-Labeled Neurons (%)

Bregma Position

CT-DG
CT-ECR

56

Figure 2.16: PRV Tracing Reveals Overlapping Representations of CT and ECR in M1.
A) Schematic of ECR forelimb and CT larynx muscle PRV injections. **B)** Coronal slice at the level of LMC of a mouse injected with PRV in CT and ECR muscles. **C)** Magnified view of highlighted region in section. White arrows point to representative double labeled neurons. **D)** Raw neuron count for each mouse injected with PRV in CT and in ECR. **E)** Proportion of counted cells that represent the CT, ECR, or both. **F)** Distribution of mean PRV$^+$ neurons in M1 from a CT injection, between the two double-injection experiments. **G)** Cumulative distribution of transsynaptic filled neurons in M1 representing the CT muscle in both double-injection experiments is not significantly different (Kolmogorov-Smirnov p>0.05). **H)** Distribution of double labeled M1 neurons from PRV injections for CT-DG and CT-ECR groups. Co-labeled neurons for the different muscle pairs exhibit different distribution patterns, with CT-DG representing neurons being more numerous in the anterior part of M1 and CT-ECR representing neurons being more numerous in the posterior portion.

In the songbird song circuit, the lateral nucleus LMAN of the anterior forebrain pathway (AFP), which is known to be important for learning and modulating song, projects to nucleus RA, which is the motor output nucleus to syringeal motor neurons in the brainstem (**Figure 1.2A**). We wondered if the OMC and LMC regions, with their distinct representations of the vocal muscles had a similar relationship. We injected PRV into the CT and AAV2-CAG-GFP into the contralateral OMC region. With this approach we could see where in the brain these two circuits interact.

We found that OMC does project into anterior parts of M1 and adjacent M2, but not into the portion where we find transynaptically backfilled layer 5 neurons from the CT injections (**Figure 2.17A**). We observed very few, if any, AAV-labeled OMC fibers contacting the PRV-labeled laryngeal layer 5 neurons in M1 (**Figure 2.17B**). There could be undetected *en passant* connections, although we did not verify if this was occurring. There was a stronger projection to ipsilateral M1 than to the contralateral side. The OMC neurons projected strongly to S1 but we did not see any PRV-labeled neurons here. In the cortex, we also saw a convergence in the entorhinal cortex. Subcortically, we observed overlaps between the circuits in the accessory nuclei medial to the trigeminal motor nucleus (**Figure 2.17C**). We also noticed some AAV labeled fibers that were located just inside the border of the trigeminal motor nucleus (Mo5) which contains motor neurons for the DG (**Figure 2.17C**). We also saw an overlap of AAV labeled fibers and PRV$^+$ neurons in various parts of the reticular formation. In a few animals we also saw some convergence of the circuits in the superior colliculus which receives strong input from OMC and is involved in some motor behaviors, as well as potential overlap in the red nucleus.
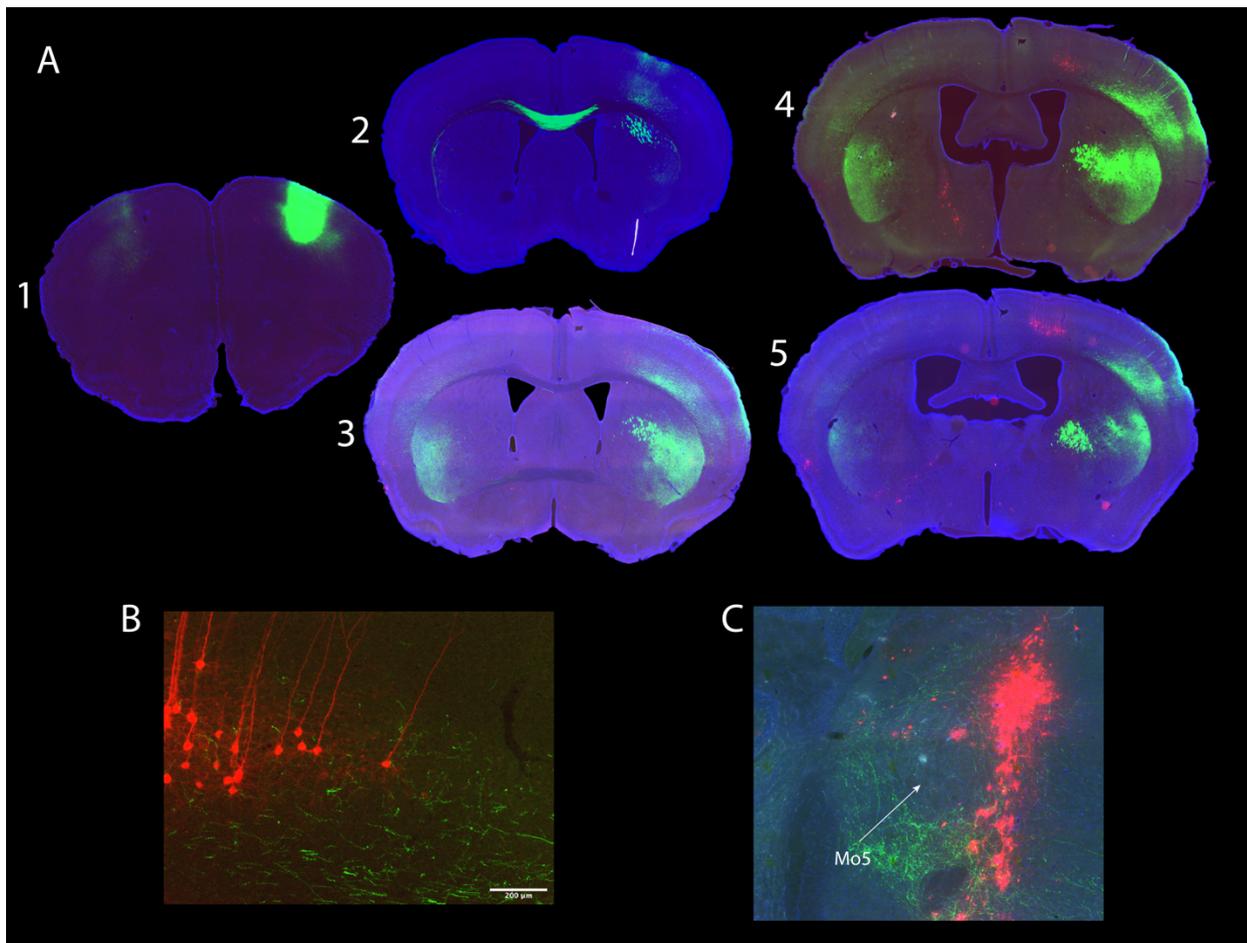
Figure 2.17: OMC Projections to PRV-Labeled Laryngeal Circuitry.
A) Sections in series (~300μm distance) showing the injection site of AAV5-GFP (green) in OMC (1) and layer 5 LMC region filled from PRV-RFP label (red) in CT (5). **B)** Inset of Section 4 showing AAV-labeled fibers passing adject to, and through, a population of LMC neurons in layer 5. **C)** Representative image of the PRV-labeled neurons medial to the Mo5 nucleus. AAV-labeled fibers can be seen surrounding Mo5, with some fibers present inside the nucleus.

2.3.6 CM-like Direct-Projections to Brainstem Motor Neurons from OMC

The lack of transapically labelled PRV neurons in OMC from the CT, DG, or ECR muscles contradicted the finding of the shorter latency responses from stimulating OMC than LMC. Additionally, we are not aware of any study with PRV injected into muscles where OMC was successfully infected (Komiyama et al., 2010). We sought an explanation and surmised that either: a) only an indirect projection exists from OMC to the muscle groups, but that pathway activation is much faster than the direct projection from LCM; or b) projections from OMC to the brainstem have different viral tropism for PRV than the LMC projections and thus is not observed at our timepoint. To identify projections that could be responsible for our short latency measurements, we injected AAV1-hSyn-Cre-WPRE, which functions as an anterograde transsynaptic tracer at high titers (Zingg et al., 2017), into the OMC of Ai14 (pan-cellular Cre-dependent tdTomato reporter) mice (**Figure 2.18A**). In this combination, AAV1 transfected neurons expressing synapsin 1 will undergo CRE recombination to express tdTomato. The AAV1 virus is transported down the axons from OMC neurons and jumps one synapse to their targets.

We found sparse, but clear post-transsynaptic labelled neurons in the Mo5 which contains DG motor neurons (**Figure 2.18B**) and in the facial nucleus (7N) which controls the whiskers (**Figure 2.18C**). We did not find any post-transsynaptically labeled neurons in the Amb (**Figure 2.18D**). We used a ChAT co-label to confirm that these postynsaptic neurons in Mo5 and 7N were in fact motor neurons. In addition to these motor neuron targets, we found few OMC post-transsynaptic neurons in the lower layers of the LMC region (**Figure 2.18E**) and neighboring S1 region. Beyond these regions we found OMC targeted the ventrolateral striatum, thalamus (reticular thalamic nucleus, mediodorsal nucleus, ventromedial nucleus, ventrolateral nucleus, and parafascicular nucleus), zona inserta, substantia nigra, lateral superior colliculus (**Figure 2.19**),
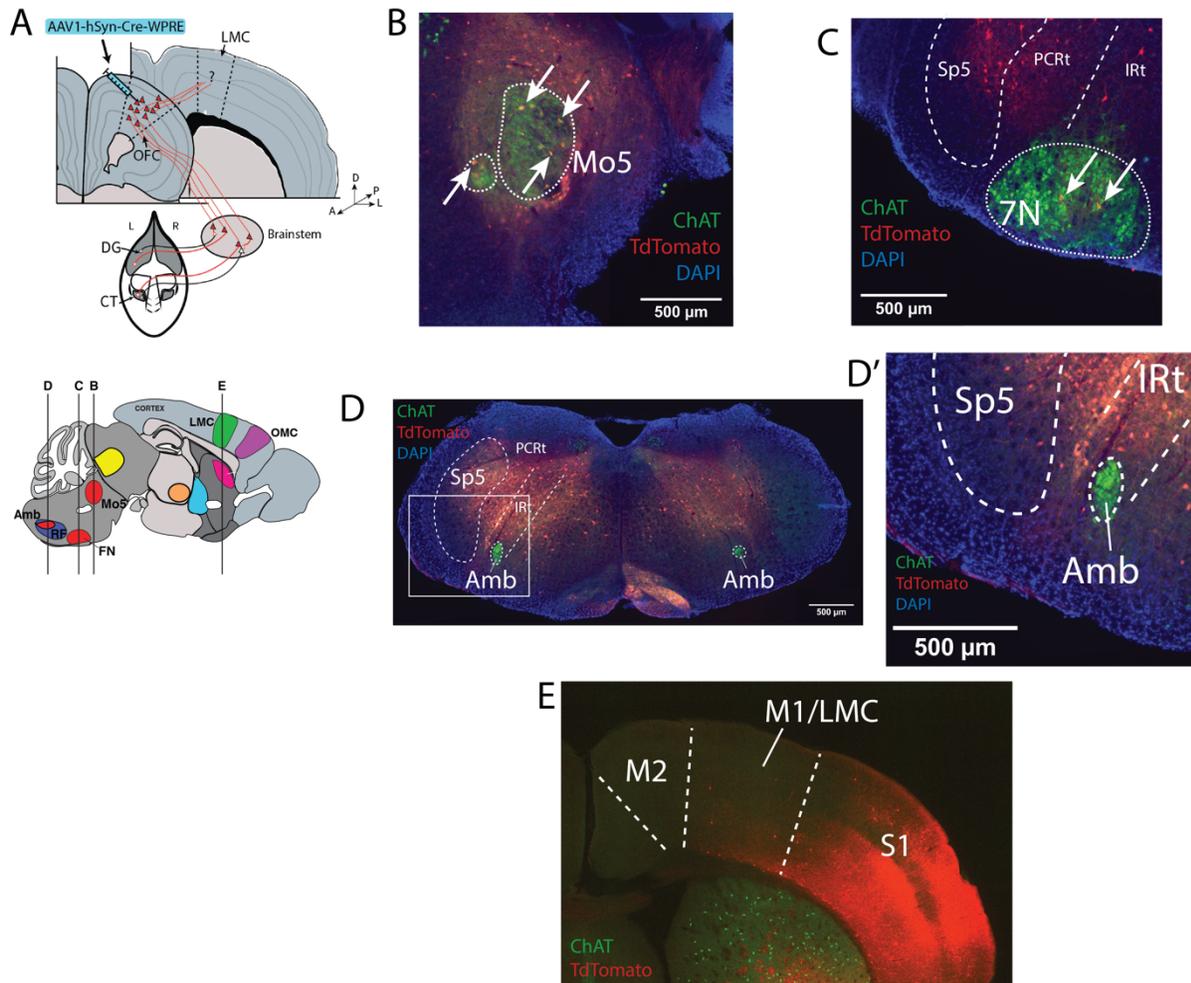
Figure 2.18: OMC Makes Sparse Direct Projections to Brainstem Oral Motor Centers.
A) Top - overview of experimental procedure to inject AAV1 in Ai14 mice. Bottom – Sagittal view of mouse brain with slices representing approximate position of images in B-E. **B)** Example of post-transsynaptically labeled neurons (yellow and arrows) in the trigeminal motor nucleus (Mo5). Verified with ChAT double-label (green). **C)** Example of post-transsynaptically labeled neurons (yellow and arrows) in the facial nucleus (7N). Verified with ChAT double-label (green). **D)** Example showing no post-transsynaptically labeled neurons seen in the nucleus ambiguous (Amb). A substantial number of neurons labeled in the surrounding reticular formation. Highlighted box is shown at higher magnification in **D'**. **E)** Example showing some post-transsynaptically labeled neurons in the LMC region of M1 and S1. Many more projections seen in S1 lateral to LMC. Abbreviations: Sp5, spinal trigeminal nucleus; PCRt, parvicellular reticular nucleus; IRt, intermediate reticular nucleus; M1, primary motor cortex; LMC, laryngeal motor cortex.

61

similar to observations by (Yang et al., 2022). There was also evidence of post-transsyanptic label in the locus coeruleus and numerous projections in the brainstem reticular formation and the dorsal part of the spinal trigeminal nucleus.

We repeated this experiment with injections of the AAV1-hSyn-Cre-WPRE into the LMC region of Ai14 mice. We found post-transsynaptically labeled neurons in many of the projection targets previously described for this region (Arriaga et al., 2012) including the anterior dorsal
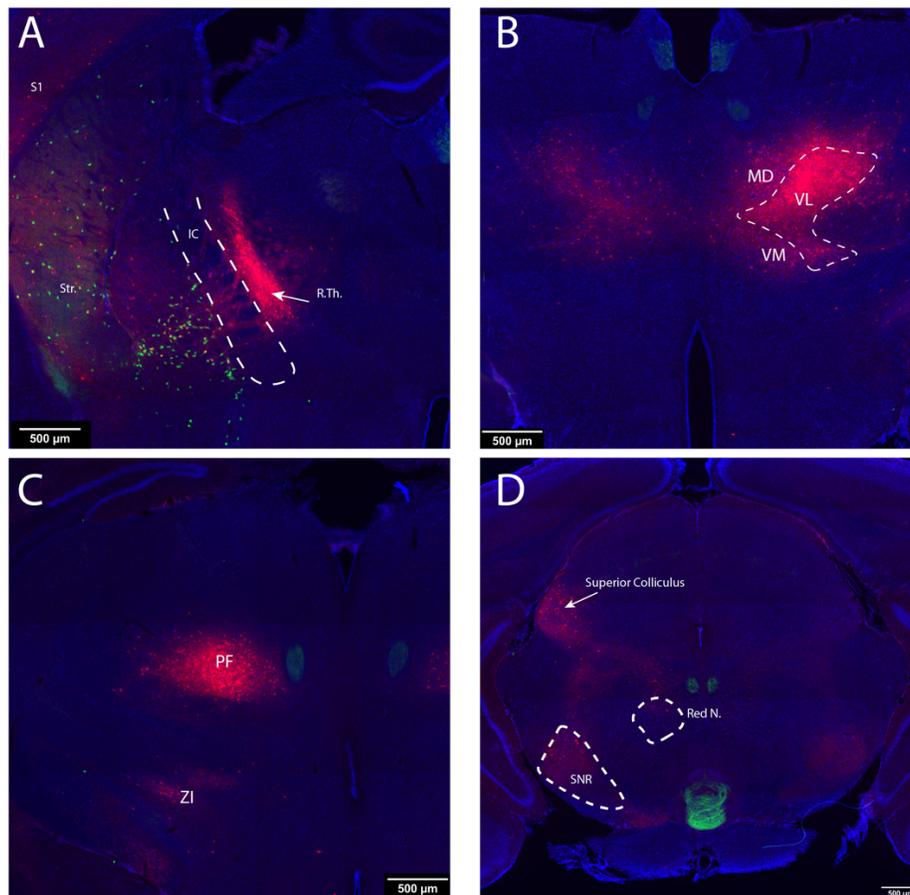
Figure 2.19: Subcortical Targets Of OMC Labeled Post-Transsynaptically with AAV1.
A-C) Example image of post-transsynaptic label in different thalamic nuceli. Other striatal, S1, and zona inserta targets visible. **D)** Example image of post-transsyanptic label to midbrain targets in the superior colliculus, substantia nigra (SNR), and the red nucleus. Anatomical abbreviations: S1, primary somatosensory cortex; Str., striatum; IC, internal capsule; R.Th., reticular thalamic nucleus; MD, mediodorsal thalamic nucleus; VL, ventrolateral thalamic nucleus; VM, ventromedial thalamic nucleus; PF, parafascicular nucleus; ZI, zona inserta; SNR, substantia nigra.

striatum, the auditory cortex, and the ventrolateral nucleus of the thalamus. We also saw the laterodorsal nucleus labeled, likely due to the spread of the virus to adjacent S1. As with the injections to OMC, we did not see any transsynaptically labeled neurons in Amb (**Figure 2.20**). The lack of post-transsynaptic neurons could be due to low efficiency transmission and/or a weak synaptic contact.

Next, to more specifically identify the population that projects to the post-transsynaptic neurons identified in Mo5, we used an intersectional viral approach. We injected an AAVretro-Cre-eGFP into Mo5 as well as an AAV9-DiO-mCherry into the contralateral OMC (**Figure 2.21A**). Neurons infected with just the eGFP virus will remain green, while neurons co-infected with the viruses will undergo a CRE recombination and express mCherry. Axon terminals expressing eGFP from the OMC neurons could be found within, and surrounding, the injected Mo5, as well as a few fibers in the contralateral Mo5 (**Figure 2.21B**). There were many eGFP-expressing neurons throughout the cortex, which may be due to virus spreading beyond Mo5 as
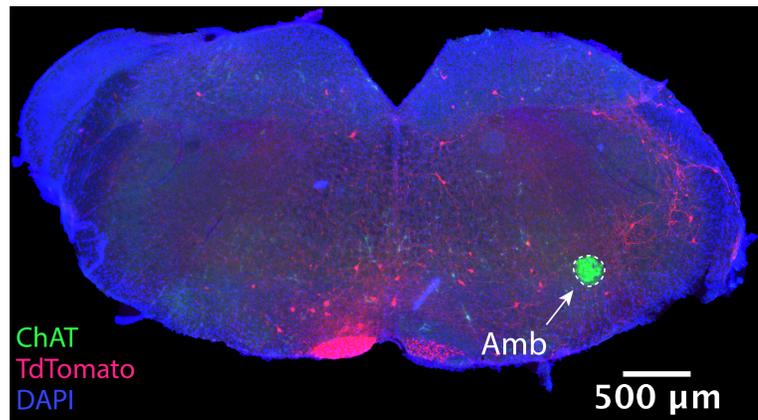
Figure 2.20: LMC Neurons Infected With AAV1 do not Reveal Direct Projection.
Coronal view of the brainstem. No postsynaptically labeled neurons in Amb are visible from an injection of AAV1 in LMC.

the area around Mo5 receives dense motor and sensory input, but the layer 5 OMC neurons that project to Mo5 were clearly visible (**Figure 2.21C**). Interestingly, terminals from the mCherry expressing neurons could also be found in many of the regions noted above from the anterograde transsynaptic-only injections. Some of these regions include the substantia nigra, red nucleus and superior colliculus (**Figure 2.21D**). Other similarly targeted regions include the thalamus (ventromedial, ventrolateral, mediodorsal, and perifascicular nuclei), reticular formation, and SpVO. Among the cortico-cortico projections, we validated the strong projection from OMC to S1 and ventrolateral striatum, but we did not identify any fibers to LMC (**Figure 2.21E-H**). Thus, the cell type that makes a weak projection from OMC to LMC in our AAV1 injections is a different population than the OMC population that projects to Mo5. These findings indicate that both LMC and OMC regions have sparse direct projections to different brainstem motor neuron groups (Amb and Mo5 respectively), but they have different tropism for different retrograde and anterograde transported viruses.
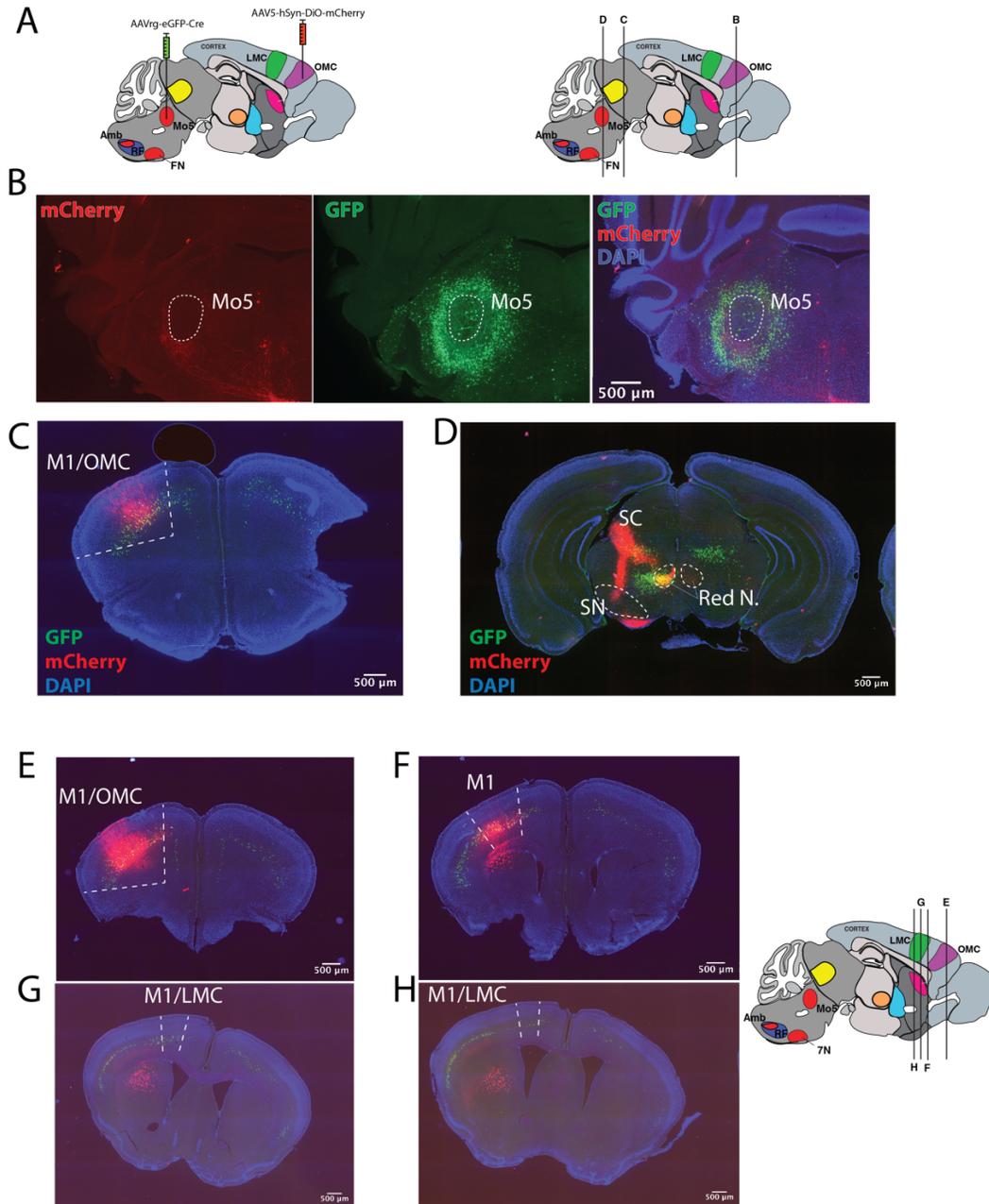
Figure 2.21: Mo5-projecting OMC Neurons Make Many Distal Collateral Connections.
A) Overview of viral strategy. **B)** Example image of terminal fibers originating in OMC (red; Cre-mediated mCherry) and neurons projecting to Mo5 (green; eGFP-Cre). **C)** Mo5-projecting neurons in OMC whose terminals are visible in panel (**B**). **D)** Example of collateral fibers in midbrain superior collicular targets which originate from the same population of OMC neurons that project to Mo5. Some convergence of non-cortical Mo5-projecting neurons overlap with OMC collaterals. **E-H)** Intracortical projection targets of Mo5-projecting OMC neurons to S1 (red). Note lack of fibers in LMC region. Some retrogradely labeled neurons in S1 from Mo5 injection (green). Abbreviations: Mo5, motor trigeminal nucleus; SC, superior colliculus; SN, substantia nigra; M1, primary motor cortex; LMC, laryngeal motor cortex.

2.3.7 Effect of OMC lesions on Vocal Behavior

Lesions of LMC in humans, or the analogous RA in songbirds, leads to difficulties or loss in producing learned sounds (Nottebohm et al., 1976; Jarvis, 2004; Silva et al., 2022). Prior work from our lab demonstrated that lesioning LMC in mice leads to a wider distribution of frequency modulation in USVs, but not loss of production ability (Arriaga et al., 2012). With OMC being able to generate CT vocal muscle contractions, we tested if OMC had a role in USVs. We chemically lesioned OMC bilaterally in male mice (n=4) using ibotenic acid (**Figure 2.22A**). Control mice were injected with saline (n=4). We elicited vocalizations from the experimental male mice by introducing a female mouse into the recording chamber in 5-minute sessions. We performed baseline recordings of the male USVs prior to surgery, and after 2- and 3-weeks post-surgery. We analyzed the data using a combination of Analysis of Mouse Vocal Communication (AMVOC) to get timestamps of the USVs, as we have shown that it is the most accurate detection method available (see **Chapter 3**).

We tested whether there were changes to the overall structure of USVs by analyzing their sequences (groups of two or more USVs). In lesioned mice, the duration of each sequence (measured as the number of syllables per sequence) increased between the baseline week and week 3 (paired t-test, p=0.012) (**Figure 2.22B**). The increase seen from week 2 to week 3 was not significant (paired t-test, p>0.05). Control mice exhibited no change in sequence duration across weeks. There was no significant difference between the control and lesion groups in the baseline week (t-test, p>0.05). By 3 weeks post-surgery sequence length was not significantly different between lesioned and control mice (t-test, p>0.05). This indicates a small but significant difference that can only be detected when comparing within individuals.
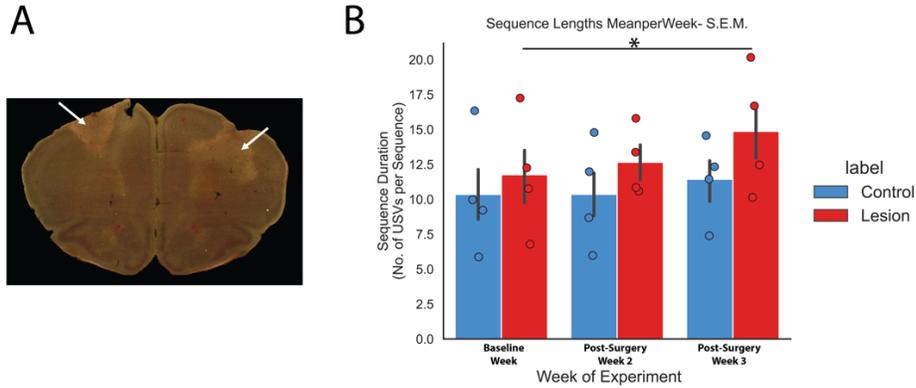
Figure 2.22: Impact of OMC Surgery and Lesion on USV Production.
A) Example darkfield Nissl image of a bilateral OMC lesion from ibotenic acid injections.
B) Length (in syllable number) of USV sequences before surgery, 2 weeks post-surgery, and 3 weeks post-surgery. There was a significant increase in the length of USV sequences in OMC lesioned mice (Paired t-test, p=0.012) and not in the control mice (Paired t-test, p>0.05). After 3 weeks post-surgery the lesioned mice did not have longer sequences than control mice (t-test, p>0.05). Bars indicate standard error of the mean.

Next, we used AMVOC to perform semi-supervised syllable clustering (see **Chapter 3**). We clustered syllables into 5 categories (**Figure 2.23**). We then used Mouse Song Analyzer 2 (MSA2) to measure acoustic features of the vocalizations (Stoumpou et al., 2022). We did not find any changes that broadly affected all the syllable types. When we analyzed the USVs for changes specific to syllable types, we found that mice undergoing the control surgery had some change by in minimum frequency and purity by week 2 but disappeared by week 3 after surgery (**Table 2.2**). In contrast, the lesioned mice show many more changes, and some of these lasted for both weeks after lesion (**Table 2.2**). The fact that we found some differences in the control mice suggest that the surgery may have had some effect, but others are due to the lesion.
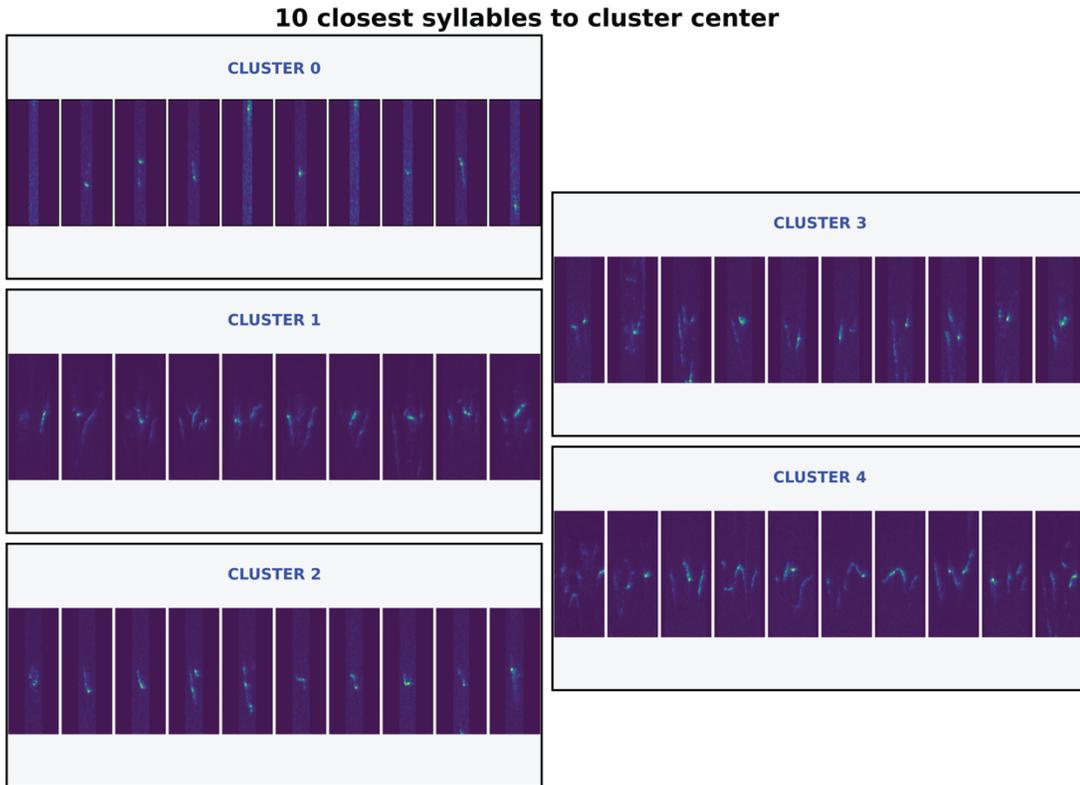
Figure 2.23: Syllable Types Determined by Unsupervised Clustering.
Syllable clusters determined by unsupervised clustering in AMVOC. Each set of images contains the ten syllables closes to the center of the clustering distribution.

	Lesion			Control			Lesion vs Control		
Feature	W1-W2	W2-W3	W1-W3	W1-W2	W2-W3	W1-W3	W1-W1	W2-W2	W3-W3
0Bw	-	-	-	-	-	-	-	-	0.004
0fqEnd	0.014	-	-	-	-	-	-	-	-
0fqMin	-	-	-	0.022	-	-	-	-	0.007
0fqStart	0.022	-	-	-	-	-	-	-	-
0fqVar	-	-	-	-	-	-	-	-	0.006
0Purity	-	-	-	0.021	-	-	-	-	-
1fqEnd	-	0.011	-	-	-	-	-	-	-
1fqMax	-	-	0.041	-	-	-	-	-	-
1fqMean	-	0.042	0.016	-	-	-	-	-	-
2fqMin	-	-	-	-	-	-	-	-	-
3fqMean	-	-	0.049	-	-	-	-	-	-
3fqVar	0.019	-	-	-	-	-	-	-	-
4Bw	0.010	-	-	-	-	-	-	-	-
4fqMin	0.034	-	-	0.037	-	-	-	-	-
4fqStart	-	-	0.032	-	-	-	-	-	-

	W1-W2	W2-W3	W1-W3	W1-W2	W2-W3	W1-W3	W1-W1	W2-W2	W3-W3
Bandwidth Mean	-	0.023	-	-	-	-	-	-	-
Duraiton Mean	0.037	-	-	-	-	-	-	-	-
Fq Mean	-	0.001	0.006	-	-	-	-	-	-
percent composition 0	0.027	-	-	-	-	-	-	-	-
percent composition 1	-	-	-	-	0.035	-	-	-	-
percent composition 2	0.010	-	-	-	-	-	-	-	-
percent starting 1	-	-	0.003	-	-	-	0.013	-	0.049
percent starting 3	-	-	-	-	-	-	-	-	-
percent starting 4	0.034	-	-	-	-	-	0.046	-	-

Table 2.2: Effect of OMC Lesion and Surgery on Acoustic Features
Table contains p-values of paired-tests for the specified comparisons across different features
that were significantly different for different syllable types. Number [0,1,2,3,4] refers to the
syllable types represented in **Figure 2.23.**

We next considered a hypothesis that the mouse cortex may play a role in determining the changes in vocal syntax observed in different social contexts (Chabout et al., 2015). To test if OMC lesioned mice had altered syntax, we used the syllable classifications described above and added an additional "syllable" for transitions to and from silence between sequences. For each mouse we performed a syntax analysis on the maximum vocalizing day for each week, due to some mice having fewer than 100 vocalizations on some days, which would skew results of the transition matrices due to incomplete repertoire data available. We segmented the syllable sequences and determined a transition matrix for each sequence. We then calculated the mean squared error (MSE) between the consecutive syllable transition matrices and calculated an overall mean MSE for all matrix comparisons per week per mouse. There were no significant differences between each week in either condition (paired t-test, $p > 0.05$) or between groups (t-test, $p > 0.05$) (**Figure 2.24**). Overall, these results suggest OMC may have a modest role in regulating vocalizations and vocal structure in *M musculus*.
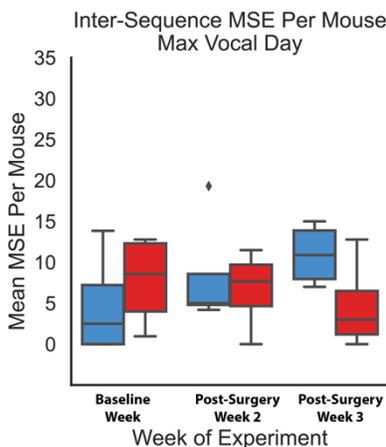
Figure 2.24: Effects of OMC Lesion on Syntax.
Differences in syllable transitions between phrases measured by the mean squared error (MSE) between the transition matrices. Lower MSE indicates more similar matrices. There was no significant difference in inter-phrase transitions between weeks or between groups.

2.4 Discussion

This is the first report we are aware of that stimulation of M1 in a vocal non-learner generates vocal muscle contractions. Stimulation of the PRV-identified LMC region of M1 preferentially drives laryngeal muscle contractions, whereas the OMC region preferentially drives jaw muscle contractions. However, both muscles are, to different degrees, excitable from both regions of cortex. The response latency from the CT to LMC stimulations supports the presence of a sparse, direct projection to vocal motor neurons in the Amb, although this pathway is dominated by the indirect projection. The short latency responses from the DG to OMC stimulation are supported by the direct projections we identified to Mo5 motor neurons. Further, there using PRV, we identified a large representational overlap in the cortical space of M1 for these to muscles involved in vocalization, and also find that laryngeal and forelimb muscles also exhibit overlapping representations. Unlike LMC's impact on frequency modulation, lesioning OMC modestly alters the number of syllables per phrase. Overall, these findings suggest a re-evaluation of prior theories about the role of M1 in vocal non-learning species, as well as the evolution of cortical control of movement, including vocalizations in mammals, which we discuss below.

2.4.1 Cortical Circuits for Vocalizations

M1 is often described as being critical for providing fine, volitional control over movements (Lemon, 2008; Nieder and Mooney, 2020). The discovery that mice have a laryngeal representation in M1, and that lesions to this region affected vocal features, suggested mice may have some degree of cortical control over their vocalizations (Arriaga et al., 2012). One study presented opposing evidence, showing that a mutation preventing most of cortex from developing resulted in the mice still able to produce normal USVs—as well walk and eat on their own

(Hammerschmidt et al., 2015). Later analyses of the vocalizations from these acortical mice found subtle spectral differences detectable with machine learning methods (Ivanenko et al., 2020). Our present study demonstrates that parts of mouse M1 can generate vocal muscle contractions, further supporting the hypothesis that mouse M1 plays some a role in vocal communication. We believe our results call for a close re-examination of claims that other species lack cortical control of laryngeal muscles. It is also critical that we evaluate the accumulated data in light of the continuum hypothesis of vocal learning (Petkov and Jarvis, 2012; Arriaga and Jarvis, 2013). Species have often been considered either vocal learners or vocal non-learners. This continuum hypothesis proposes there is a continuum of phenotypes that can exist as some combination of partial anatomical features (e.g., sparse direct projection) or as partial behavioral phenotypes (e.g., modifying frequency of vocalizations but not imitating entire sounds). Our present data support our previous prediction that mice may be at some intermediate position where they have rudimentary features of vocal learners but do not possess the full behavioral or vocal circuit phenotype. Clarity as to how far away mice are from vocal non-learners will require more behavioral tests that push the limits of mouse vocal behavior, giving us insight into the interactions vis-à-vis the presence of neural circuits and their role in behavioral outcomes.

2.4.2 Comparison to Other Mapping Studies

Prior mapping studies in mice and rats have explored whether activity in certain regions of the cortex has a role in causing movement in certain parts of the body (Neafsey et al., 1986; Li and Waters, 1991). The way a region's function was ascribed was based on what movements occurred at minimal threshold stimulations (Tennant et al., 2011). These prior rodent studies did not mention any laryngeal responses from M1 stimulation. In non-human primates, identification of a

laryngeal representation may have been overlooked by attempts to measure vocal motor representations by observing laryngeal movements with a laryngoscope (Simonyan and Jürgens, 2002; Simonyan and Jürgens, 2003) or by manual probing (Coudé et al., 2011). One study did use EMG recordings from the laryngeal muscles, but they did not fully sample the motor cortex (Hast et al., 1974). Compared to these studies, we used both EMG recordings form the larynx, which is more sensitive, and we had a transsynaptically-identified region of M1 that we could test as a candidate, independent of current stimulation and contractions of other muscles. We believe the combination of cortical stimulation in retrogradely identified regions provides a more unbiased approach to understanding the representational organization of M1. To our knowledge there have not been EMG measurements from the larynx in non-human primates during M1 stimulation. Considering these differences, it is important to know whether stimulating the laryngeal M1 region identified by Cerkevich et al. (2022) using transsynaptic viral methods in marmosets and macaques, two vocal non-learning primates, would also lead to vocal muscle contractions measured with EMGs.

2.4.3 ICMS Estimates of Synaptic Order and Monosynaptic Connectivity

Stimulating regions of the motor cortex with direct projections, such as human dorsal LMC (dLMC) or primate forelimb regions, results in short (~10 ms) latencies for muscle contractions (Cheney and Fetz, 1985; Rödel et al., 2004; Dichter et al., 2018). In this study we found that stimulating the region of M1 with direct projections to Amb resulted in a higher proportion of longer, multisynaptic-like latencies. However, stimulating the OMC region led to mostly short latencies like those seen in other monosynaptic systems. One prior study using optogenetic stimulations with EMG recordings form the forelimb show similar latency differences between a

posterior and anterolateral regions of M1 (Harrison et al., 2012), although they did not discuss the synaptic order of the circuit. Using AAV1 for anterograde transsynaptic tracing, we show that the fast DG response form OMC stimulation likely comes from monosynaptic projections to motor neurons in Mo5 (**Figure 2.25**). The precise circuit responsible for the short latency response of CT resulting from OMC stimulation is still unclear.

The AAV1 injection in OMC did not reveal any direct projections to Amb. Given that AAV1 injections into LMC also did not result in a direct projection to Amb, our negative result from the OMC injections could be due to low rate of transsynaptic transmission in a sparse or weak synaptic connection. Zingg et al. (2020) demonstrated that there is a very low transmission efficiency from cholinergic neurons to their targets. There may be a similar lack of transmission efficiency *to* cholinergic targets like motor neurons. With its much smaller size, the Amb may not be an amenable nucleus for studies using this transsynaptic method. Single-pulse ICMS in the forelimb regions of rats also results in a ~10 ms latency (Liang et al., 1993), but the circuit generating this short-latency response has been demonstrated to be a multisynaptic system (Yang and Lemon, 2003; Alstermark et al., 2004). A similar circuit may be at play between OMC and Amb.

Although CT responses to LMC stimulation tended to be slower than those from OMC, there were a few responses with latencies seen in monosynaptic systems. In macaques, the hindlimb muscles are innervated by direct projecting layer 5 neurons, yet exhibit longer latency responses and show less facilitation from ICMS than their forelimb counterparts, which is hypothesized to be due to the lower synaptic strength (Hudson et al., 2015). In the future, experiments using *ex vivo* slice electrophysiology and optogenetic stimulation could help
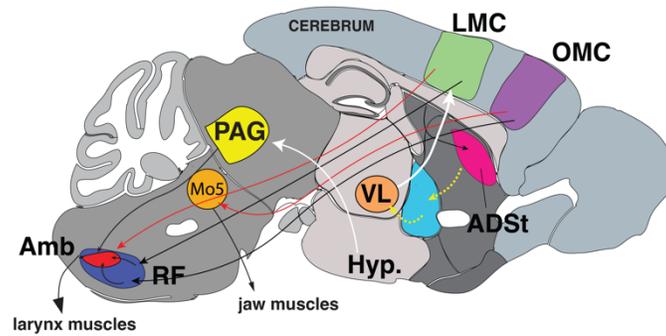
Figure 2.25: Cortical Circuits for the Control of Vocal Muscles and Vocalization.
Updated diagram of the known vocal circuits for USV production in mice. Red arrows indicate direct projections. Dashed arrows indicate predicted connections.

determine the strength of connectivity between the M1 and Amb neurons (Petreanu et al., 2007), and help us better understand the biophysical properties that lead to the slower than expected responses.

2.4.4 Organizing M1 Around Movement

Typical notions of the motor homunculus depict segregated representations of the body. Some overlap between body parts has been found, but these areas are generally considered to be at the boarders of representations of individual muscles (Li and Waters, 1991). Our ICMS and connectivity findings contradict this view. The neurons located in what we termed LMC are within a posteromedial region of M1 that is often labeled as CFA (Tennant et al., 2011). In addition to the larynx and jaw, this portion of M1 in mice has been found to contain neurons representing musculature from the forelimb (Gu et al., 2017b; Wang et al., 2017; Yamawaki et al., 2021b), tongue (Komiyama et al., 2010), and vibrissae (Sreenivasan et al., 2015). Using electrical and optical stimulations, the OMC region has also been shown to represent all four of these muscles (Komiyama et al., 2010; Harrison et al., 2012; Mercer Lindsay et al., 2019; Barrett et al., 2022). Thus, representational overlaps appear common in mice and may not be unique to borders. This

type of organization could facilitate a wide variety of movements that use similar muscles but require distinct muscle ensembles to perform an action.

Work by the Graziano in the early 2000's lab re-kindled debates in the field to about whether the motor cortex was organized around muscles or movements. They used long-train stimulation, compared to the much shorter stimulations typically used for mapping, which Graziano et al. argue is a stimulation at behavioral timescales (upwards of 500 ms) (Graziano et al., 2002). Under these conditions, it was possible to produce repeatable motor behaviors that differed based on where in the motor cortex the stimulation was applied (Graziano et al., 2002). Work in rats (Halley et al., 2020), mice (Mercer Lindsay et al., 2019; An et al., 2022), and tree shrews (Baldwin et al., 2017), among others, have expanded and replicated these results. In macaques, it has been shown that the representation of digits and shoulders do overlap in the cortex (Rathelot and Strick, 2006). Within the arm, cortical neurons have been shown to share inputs to motor neurons of multiple muscles in primates (Cheney and Fetz, 1985) and mice (Gu et al., 2017a). The data presented here from multiple muscles injected with PRV in distinct parts of the body provides an anatomical substrate for how one region of motor cortex could actually engage multiple muscles to shape behaviors requiring more than one body part.

In our present data and previous reports (Arriaga et al., 2012; Chabout et al., 2016), PRV injected into CT (larynx), DG (jaw), and ECR (forelimb) muscles appear to uniquely transfect neurons in an overlapping posteromedial region of M1. We noted that the neurons located in this posteromedial region, which we had termed LMC (Arriaga et al., 2012), are within the posteromedial region of M1 called caudal forelimb area (CFA) (Tennant et al., 2011; Wang et al., 2017; Yamawaki et al., 2021). This area has also been shown to contain representations of the tongue (Komiyama et al., 2010), vibrissae (Auffret et al., 2018), and, in this study, jaw movements.

Due to the superimposed representations of many muscles within this posteromedial cortical column of M1 and its PRV-defined layer 5 neurons from muscle injections, we propose a distinct term for this area as M1 alpha (M1α), and we term the more anterolateral region of M1 that does not exhibit PRV labeling as M1 beta (M1β) (**Figure 2.26A**). Rather than referring to the cortical region by a muscle or body part reference, we propose to use LMC to denote the subpopulation of neurons representing the larynx and CFA to denote the neurons that represent the forelimbs in M1α; OMC will the denote the neurons that represent the jaw, and other orofacial muscles, in M1β. This type of overlapping organization in M1α, and M1β, as suggested by our stimulation results, could facilitate a wide variety of movements that use similar muscles but require distinct ensembles to perform an action. For example, forelimb behavior is known to differ along the anterolateral and posteriomedial axis (Hira et al., 2015)

We hypothesize that greater mixing of muscle groups in the same cortical column of mouse M1α may be the result of the constraints imposed on small brains. Birds, and vocal learning birds in particular, like songbirds and parrots, have evolved smaller neurons that can be packed in high densities in their small brains (Olkowicz et al., 2016). On the other hand, relative to brain size, rodents have a lower neuron density than primates (Olkowicz et al., 2016). We propose that rodents may have adapted to this problem by increasing the representational overlap within and across cortical columns, leading to many body parts represented within the same cortical space, and with some of those projecting to some of the same downstream motor neurons.

2.4.5 M1α and M1β Roles in USVs

Consistent with the segregation of function for multiple representations of other muscles in M1 (Hira et al., 2015), our lesions in M1β (i.e. OMC) did not result in changes in syllable frequency

78

modulation, unlike our prior findings with M1α (i.e. LMC) lesions (Arriaga et al., 2012), but instead result in a modest effect on the duration of sequences. Manipulations of M1β in *S. teguina* resulted in similar, but perhaps more striking, effects on song duration (Okobi et al., 2019). Differences in the role of M1β may be due to differences in how the two species use their jaw during song. Recent work has shown that OMC is also involved in oromanual motor behaviors, such as maintaining a food pellet near the mouth versus adjusting the pellet's position (Barrett et al., 2022). Thus, M1β may be more involved with the broader duration of a behavior rather than fine-scale features within that behavior. The combined findings support a functional segregation of laryngeal and oral muscle representations between LMC and OMC that coordinate with each other during vocalizations.

2.4.6 Evolution of Cortical Control of Vocal Muscles

It has long been argued that direct projections are unique to primates, and, by extension, that humans are unique in having an M1 LMC (Lemon, 2008; Simonyan, 2014). Based on our data, we argue that a rudimentary M1 LMC region evolved at least as early as the last common ancestor of rodents and primates (**Figure 2.26B**). Early in development, mice have many direct cortical projections to spinal cord motor neurons for the forelimb system, which are then pruned postnatally until they are effectively absent in adulthood (Gu et al., 2017b; Wang et al., 2017; Murabe et al., 2018). With the genetic substrate for direct projections available in the ancestor of rodents and primates in M1α, we hypothesize that these transient, or in some cases remaining sparse, direct projections seen in rodents were "enhanced" or expanded upon in the primate lineage and in vocal learners for vocal motor neurons. This in turn led to the more elaborate corticospinal system seen in primates, with the human dLMC and vLMC separating out of M1α and M1β, respectively, as

distinct regions rather than a subpopulation. As with the somatotopy argument above, the so-called "corticalization" of movement (i.e. greater dependance on direct projections from cortex) may also be due to changes in neuron count and density (Herculano-Houzel et al., 2016; Olkowicz et al., 2016). The degree to which a species has these direct projections in their vocal circuitry would partially dictate their position along a vocal learning continuum (Arriaga et al., 2012; Petkov and Jarvis, 2012).

The human dLMC and vLMC regions are separated by jaw and lip motor representations (Bouchard et al., 2013; Pfenning et al., 2014; Gordon et al., 2023). Under the previous assumption that non-human primates did not have an LMC in M1, several hypotheses were developed to explain how humans LMC may have evolved. One hypothesis posited that area 6V is the ancestral vocal region in primates and that in humans it migrated into M1 to become what is now known as vLMC (Simonyan and Horwitz, 2011). This was followed by a duplication of vLMC that migrated dorsally in M1 to become dLMC in humans (Belyk and Brown, 2017). An alternative hypothesis based on our findings here in mice, is that there were two rudimentary laryngeal representations phylogenetically older than primates (e.g. OMC and LMC in M1β and M1α). Primate vLMC and mouse OMC have similar representational overlaps of the orofacial and laryngeal muscles (Hast et al., 1974; Coudé et al., 2011; Mercer Lindsay et al., 2019). The relative position of mouse LMC, human dLMC, and monkey M1 LMC regions (Cerkevich et al., 2022) indicates that they could be homologous rather than the human dLMC arising completely *de novo*. Additionally, like human dLMC (Dichter et al., 2018), mouse LMC may function in pitch modulation of vocalizations (Arriaga et al., 2012; Dichter et al., 2018). Non-human primate Area 6V, in this view, we propose would be the premotor LMC region anterior to vLMC in M1. In this hypothesis, the biggest difference between mice and humans would segregation/specialization of functions of the human

LMC regions from surrounding motor representaitons, aided by the strongly enhanced direct projection to vocal motor neurons. Further functional roles of LMC for vocal learning could occur due to this segregation. For example, human dLMC is also known to be responsive to hearing speech in addition to its role in production (Dichter et al., 2018; Silva et al., 2022). Similarly, a songbird's HVC, a song learning motor nucleus, responds to playback of the bird's own song (Margoliash, 1983; Dave and Margoliash, 2000).

A recent study in humans (Gordon et al., 2023) proposed a major revision to the understanding of homuncular organization. They propose that human and macaque M1 is subdivided into three areas that are local, dual mirrored homunculi for different body sections: 1) face; 2) upper limbs and trunk; and 3) lower limbs. Between these effector regions, are three inter-effector regions that represent movements of many body parts. This work raises two different proposals related to the current study that differ from those above: 1) human dLMC and vLMC are part of the concentric organization around the tongue, also implying non-human primates may have dLMC and vLMC; 2) mouse M1 may be more like the inter-effector regions with broad representational overlaps, though not identical, which remained rudimentary in rodents while the primate linage derived the more specific representations of effector regions through the presence of increased direct projections (**Figure 2.26C**). To further test either hypothesis, it will be necessary to profile the molecular and genetic specializations of these M1 regions across species, to determine the evolutionary relationships between species as well as for the specializations of these laryngeal regions compared to the surrounding motor cortex (Pfenning et al., 2014; Gedman et al., 2022).
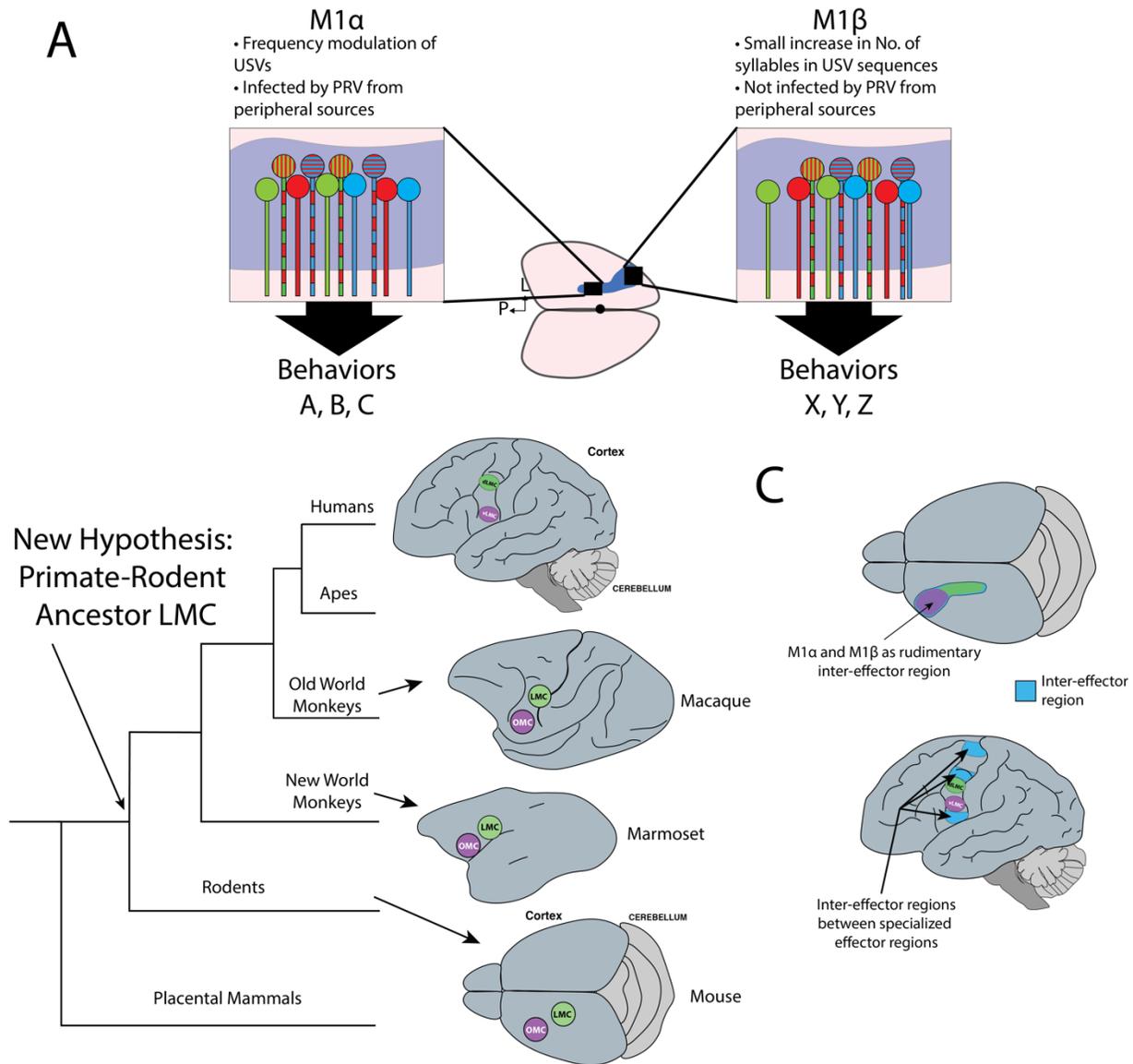
Figure 2.26: Hypotheses for the Evolution of a Cortical Laryngeal Motor Representation. **A)** Schematic of proposed M1α and M1β organization of mouse M1. Circles represent individual muscles and their overlapping representations, while double-colors represent neurons that control multiple muscles at once. M1α and M1β contribute differently to the overall output of M1 by representing different behaviors while representing similar muscles. **B)** Abbreviated phylogenetic tree of mammalian evolution with different hypothesis that would predict the origin of an M1-localized LMC region. This hypothesis predicts the origin of M1 LMC at the last common ancestor of primates and rodents, with a more advanced and functionally distinct LMC in humans. **C)** Potential comparison between the multi-representational organization of the proposed mouse M1α and M1β and its similarities to the recently described inter-effector regions of the so-called "somato cognitive action network" in humans and primates described by Gordon et al. (2023).

CHAPTER 3. New Methods for the Detection of USVs and Their Application to Operant Training of Vocal Behavior

3.1 Introduction

Mice can modify the syntax (i.e. sequencing) of their ultrasonic songs based on social context. For example, male mice sing more complex syllable types and sequences to a live female than to female urine (Chabout et al., 2016). It remains unclear if these changes in syntax are volitionally controlled, or if they are simply innate stimulus-triggered responses. One theory of the function of M1 is that it endows mammals with volitional control of their behaviors (Ebbesen and Brecht, 2017). Direct projections from M1 to brainstem and spinal cord motor neurons are hypothesized to allow volitional modulation of fine movement (Fischer and Hammerschmidt, 2011). Based on the observation that mice have a sparse, but direct projection from M1 to Amb motor neurons (Arriaga et al., 2012), we predict that mice have at least some volitional control over their vocalizations.

Volitional vocalizations have been demonstrated in both vocal learning species (mynah bird, zebra finch, budgerigar, bats, humans, dolphins, elephants, and pinnipeds) (Hake and Mabry, 1979; Richards et al., 1984; Manabe and Dooling, 1997; Shapiro et al., 2004; Lattenkamp et al., 2018; Xiao et al., 2018; Stoeger and Baotic, 2021). and vocal non-learning species (dogs, cats, guinea pigs, macaques, rats, chickens) (Lane, 1960; Salzinger and Waller, 1962; Molliver, 1963; Burnstein and Wolff, 1967; Lal, 1967; Aitken and Wilson, 1979; Johnson et al., 2011). Although volitional vocalizing is not unique to vocal learners, this ability is fundamental to vocal learning. We suggest that the direct projections may aide in increasing the range of features over which an animal has volitional control.

Testing for volitional vocalizations requires a closed-loop operant training mechanism either with human intervention (Johnson et al., 2011) or in a fully automated set up (Manabe and Dooling, 1997; Lattenkamp et al., 2018). In closed-loop operant training, when a behavior of choice is detected some type of reward is provided to the animal. Rewards can be external such as a food (Lattenkamp et al., 2018), removal of an electric shock (Lal, 1967; Aitken and Wilson, 1979), or positive electrical stimulation (Burnstein and Wolff, 1967) — in this case, stimulation of the hypothalamus. To perform these operant training experiments in the context of vocal production, it is necessary to detect and analyze vocalizations in real time. For example, by monitoring the pitch of the vocalizations in real time, animals can be operantly trained to increase or decrease the frequency of their vocalizations for a reward (Xiao et al., 2018; Lattenkamp et al., 2020). By monitoring more global aspects of call structure, vocal learning parrots can also be trained to change acoustic features of a vocalization and even change the type of vocalization entirely (Manabe and Dooling, 1997).

Audio data can be difficult to analyze (e.g., measure acoustic features, or classifying types of vocalizations) in real-time due to the rates of data acquisition combined with the brevity of most vocal behaviors. For example, zebra finches can modulate features of their syllables at the millisecond scale (Fee et al., 1998). One needs high rates of data collection, otherwise these rapid onsets and offsets of features in a vocalization can be missed. Studies in zebra finches have solved some of this difficulty by making use of a unique syllable type produced during song in some birds that has harmonic stacks which makes it easier to identify amongst the other vocalizations (Ali et al., 2013; Xiao et al., 2018). This strategy is not possible with mouse USVs, as most of the spectral content is monotonic which does not allow for the <20 ms detection latency used on zebra finches (Ali et al., 2013). Although there is a plethora of methods for detecting and analyzing mouse

84

vocalizations offline, none of these methods are designed to provide detection and classification in real-time. If we want to test whether or not the direct projection to Amb motor neurons provides mice with vocal skills greater than what has been observed with methods only designed to test innate production, we needed a tool that allows us to analyze USVs quickly and accurately to provide mice contingent behavioral feedback in an operant context.

The most widely used commercial tool available for recording mouse USVs is Avisoft®, which is a closed-source software and hardware system. Avisoft and its associated hardware have a few shortcomings with regards to their use and integration in our desired experimental system. One obstacle was cost (~$3,000 on the low end for a single channel), which would limit the broad adoption of our experimental approach. The second, and more substantial issue, was that we cannot classify individual USVs in real time using Avisoft. The ability to classify USVs and perform template matching, or even detect novel USVs, is necessary if we want to formally test whether mice have bona fide volitional or vocal learning abilities.

To address these issues, here in this Chapter 3, I described a new USV detection system we developed called Analysis of Mouse VOcal Communication (AMVOC). AMVOC can be used in a real-time mode that accurately detects USVs at a rate close to its offline mode with sub-second latencies. Beyond just online detection capabilities, users can train syllable classification models and use these with the online mode to classify USVs as they are being emitted and detected. One of our goals in developing AMVOC was also to ensure that we created a high-performing, open-source method that relied on as many free and other open-source tools as possible. For this reason, AMVOC is written in Python, as opposed to a language like MATLAB. This approach will make it easier to scale the number of simultaneous experiments at a lower cost and will facilitate its adoption by other research groups. We then developed an operant training paradigm using

AMVOC, and we provide a proof-of-principle for our experimental designs for testing the capacity of mice to control their vocalization. We highlight the utility of AMVOC as a tool for advancing our knowledge of mouse vocal production, and how we can improve this species' role as a model for human speech and vocal learning.

3.2 Methods

To develop AMVOC, we used open-source analysis programs and libraries, written in Python 3. Libraries used for processing, visualization, machine learning, and analysis include: numpy, pandas, sci-py, matplotlib, plotly, pytorch, pillow, pyaudio, and tsne. These libraries can be downloaded using their respective instructions from the Python Package Index (PyPI).

3.2.1 Detecting USVs Offline

One of the challenges of identifying and segmenting USVs is their relatively low spectral energy compared to their sonic environment. AMVOC begins its detection by calculating the spectrogram of the recording at a resolution of 0.5 kHz. Next, we apply a spectral energy threshold in two steps (**Figure 3.1A**). The first is to compute a dynamic threshold by determining the average spectral energy across a rolling 2 second window that is convolved with the sum of energy at each time point. Second, we compute the energy of the peak frequency that is between 30 kHz-110 kHz (or some other user defined range). We take the mean energy in a 60 kHz window around this peak frequency (truncated if the range goes beyond 30 kHz or 110 kHz). Lastly, to determine whether or not a USV is present, the putative USV must pass the following two criteria: 1) The calculated spectral energy must be larger than the dynamic threshold by 50%, designated as t; and 2) the peak spectral energy must be 3.5 times larger than the mean spectral energy in the predetermined ultrasonic range, designated as f. The result of each criterion is multiplied by the spectral resolution (in kHz). **Figures 3.1B and 3.1C** demonstrate how these two criteria can help discern noise from true signals. This thresholding produces a "neighborhood" of putative USVs (**Figure 3.2A**). These threshold crossings are smoothed using a 20 ms moving average box filter (**Figure 3.2B**). The output of the smoothing procedure is used to make the final determination of the presence of a
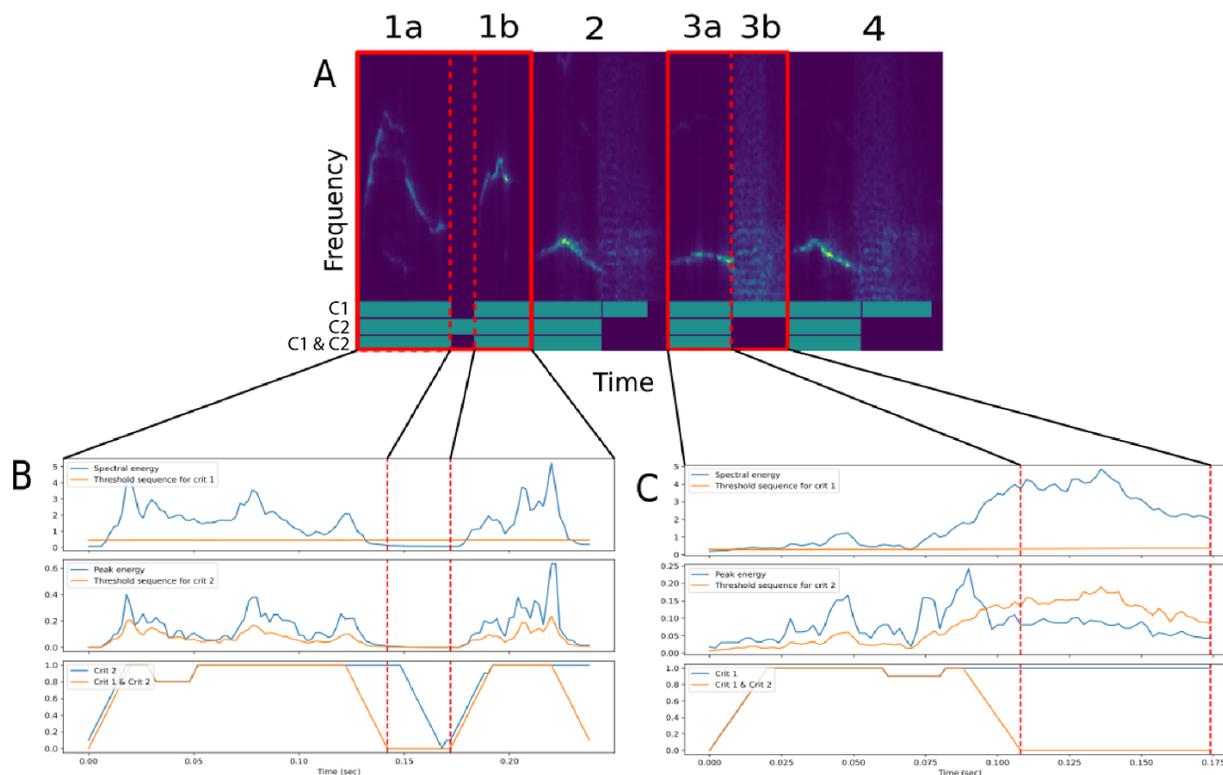
Figure 3.1: Examples of Detection Criteria.
A) Demonstration of the twofold thresholding application. The green bars of the first two lines show the detected vocalizations by each criterion. The purpose of criterion 1 is to pick our gross presence of sounds and avoid concatenation of successive USVs. The purpose of criterion 2 is to filter out non-vocal noisy components of the recordings with wide spectral energy. Segments were spliced for purposes of visualization. B) Application of the two thresholding criteria for USV 1 in (A). The first plot demonstrates the spectral energy (blue), along with the thresholding sequence of Criterion 1 (orange). The second plot presents peak energy (blue) and the thresholding sequence of Criterion 2 (orange). In the third plot, the sequence V after moving average filtering. The blue line corresponds to the case where we use only Criterion 2 for detection, while the orange line is the result of the application of both criteria. From the red dotted lines, it is clear that Criterion 2 would have resulted in the concatenation of the successive USVs 1a and 1b, which is prevented by Criterion 1. C) Application of the two thresholding criteria for USV 3 in (A). The plots display the same information as in (B). In this case Criterion 2 is necessary for filtering out noisy segments with high spectral energy like segment 3b in (A).

USV, with threshold crossings as the start and end times for each USV (**Figure 3.2C**). If any detected USVs are still separated by less than 11 ms they are then merged, as these are likely individual notes of a USV rather than separate USVs themselves (**Figure 3.2D**).
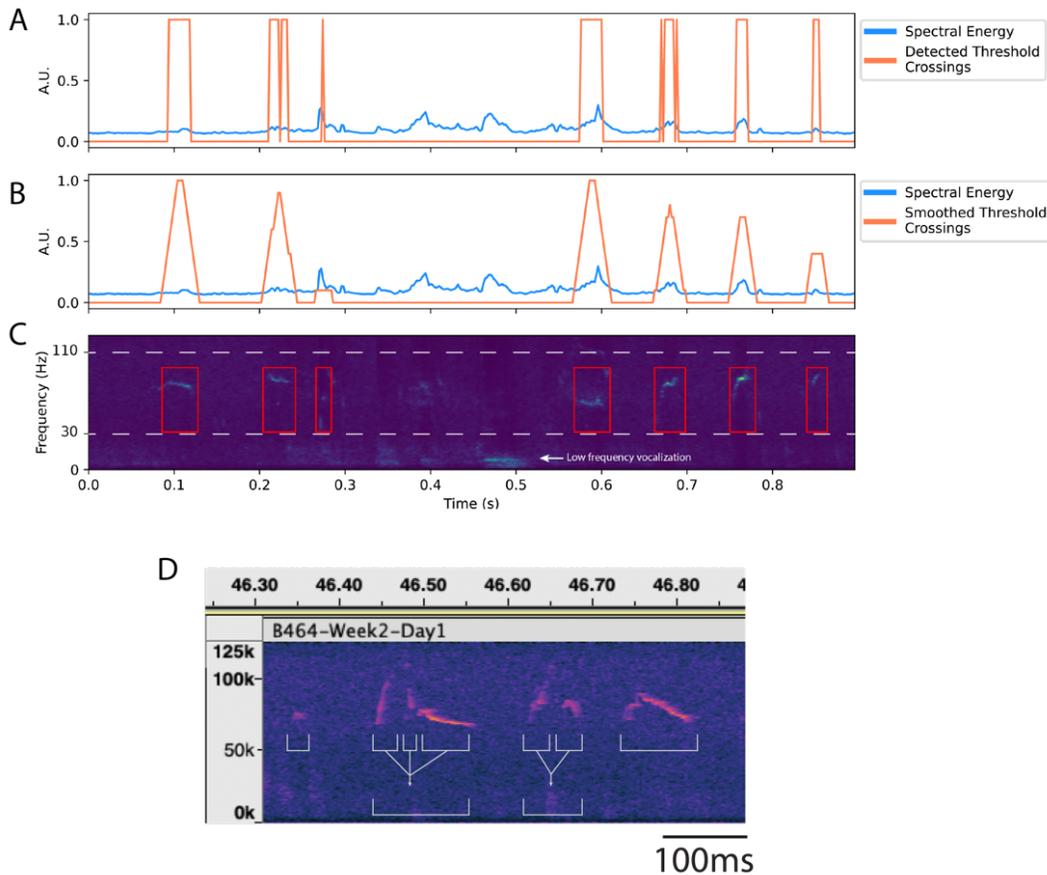
Figure 3.2: Example of Merging USVs.
A) A binary 0 or 1 is assigned for any time that the window of spectral energy surpasses threshold that is produced by the combination of the temporal and frequency threshold steps. This produces a "neighborhood" of where USVs are located. **B)** The binary position of threshold crossings is smoothed and is used as the final determination of where USV start and end times are located. **C)** Example spectrogram based on the above plots. Red boxes exhibit the final bounds determined to be where the USV is located. Note there is a portion of high energy that did not meet the threshold criteria. **D)** Diagram depicting an example of USVs composed of non-overlapping notes that would be merged after determining threshold crossings. **A-C)** From Stoumpou et al., 2022.

3.2.2 Detecting USVs Online

Detection and analysis in real time was critical for our ability to develop a behavioral training paradigm. For the online method we process audio data in chunks of 750 ms. We selected a chunk size of 750 ms after empirically determining that this was the smallest interval that approximated the accuracy of the offline method. Each subsequent chunk is overlapped with the last 13% of the previous chunk to prevent a USV that spans two chunks from being inappropriately separated. Next, we repeat a similar two-step thresholding procedure as in the offline method to find the neighborhood of USVs and apply a filter to make the final determinations. The main difference between the offline and online method is that the dynamic threshold is calculated cumulatively across all chunks preceding the current one. To compensate for the loss of information that is available in the real-time computation, we applied two weights to the convolved temporal and spectral thresholds. A weight of 0.3 is applied to the temporal threshold, and a weight of 0.7 is applied to the spectral threshold. Both thresholds were selected after empirical testing, as opposed to the 0.5 applied to each threshold in the offline detection.

3.2.3 Feature Extraction and Unsupervised Clustering

Once vocalizations are detected we use a convolutional autoencoder to inform unsupervised clustering methods to produce unbiased syllable classifications. There are many methods that can perform dimensionality reduction in large datasets. We elected to use a convolutional autoencoder for our task because they are particularly good at extracting meaningful representations of input images, and they can efficiently maintain spatial information. For our approach, we chose to use the spectrographic representation of each detected USV as an image. We needed to first create a uniform image size to perform the clustering method. We plotted a histogram of the USVs based

90

on their duration in spectrogram time frames (2ms/time frame) (**Figure 3.3**). Most USVs were below 50 time frames, or 100 ms. In selecting an image width/duration, we needed to balance the loss of temporal information from clipping long USVs and losing shape information by zero-padding short USVs. We selected a width of 64 time frames (128 ms) as it encompasses a majority of USVs and is a power of 2, which is useful for the compression we perform with the autoencoder.

First, the spectrogram of each detected USVs is saved as an image of size 64x160 pixels. If a USV is less than 64 frames in width it is zero-padded symmetrically on either side (**Figure 3.4A**). If a USV is larger than 64 frames it is clipped symmetrically at the start and end. The images are then processed through an undercomplete autoencoder (**Figure 3.4A**). Autoencoders are an ML method that takes an input and attempts to learn what latent features are most informative to reconstruct the original input. The first half of the autoencoder, called the encoder, reduces the images' information into a single vector containing what is the most informative latent features. The encoder's output is called the code. The second half of the autoencoder is the decoder. The decoder attempts to reconstruct the original input from the intermediate code representation. An
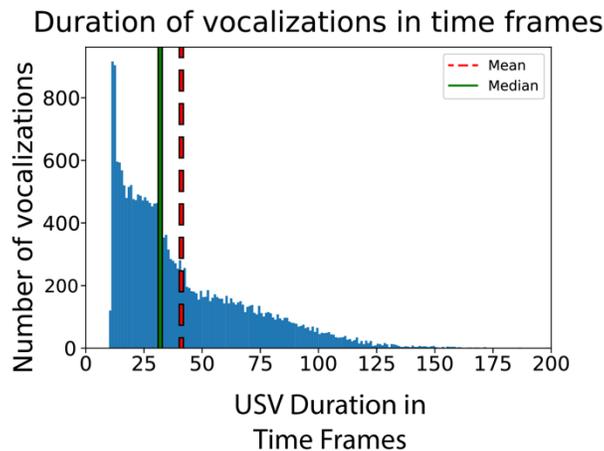
Figure 3.3: Histogram of USV Durations.
Distribution of USVs durations in Dataset D1. Solid green line indicates median duration. Dashed red line indicates mean duration. Modified from Stoumpou et al. 2022.

undercomplete autoencoder creates an informational bottleneck such that the encoder cannot simply "memorize" the inputs for reconstruction. Instead, the encoder is forced to represent the input data in a smaller informational space in the code stage, forcing the autoencoder to determine what are the most salient features that allow for accurate representation and reconstruction of the input.

We input the spectrogram images to the encoder consisting of a convolutional neural network (CNN) with 3 convolutional layers. Each convolutional layer is followed by ReLU activation function and then a max pooling layer to reduce the size of the representation between each layer. The first convolutional layer uses 64 filters, with dimensions 3x3 each. After that, a max pooling layer decreases the spatial dimensions of the activation maps by a factor of 2. This means that the output of the max pooling layer is a 32x80x64 representation. The next convolutional layer consists of 32 filters, with a max pooling layer generating an output with dimensions 16x40x32. The third and final layer includes 8 filters, and a max pooling layer, resulting in a convolutional activation map for each image with dimensions 8x20x8. We selected 8 filters in the final encoder layer as this resulted in the highest quality reconstructions without overfitting (**Figure 3.4B**). This autoencoder learns what latent features best represent a given vocalization and stores this data on a flattened vector of size 1,280. The training loss improves minimally after 2 training epochs (**Figure 3.4C**), so we selected 2 epochs in our final implementation.
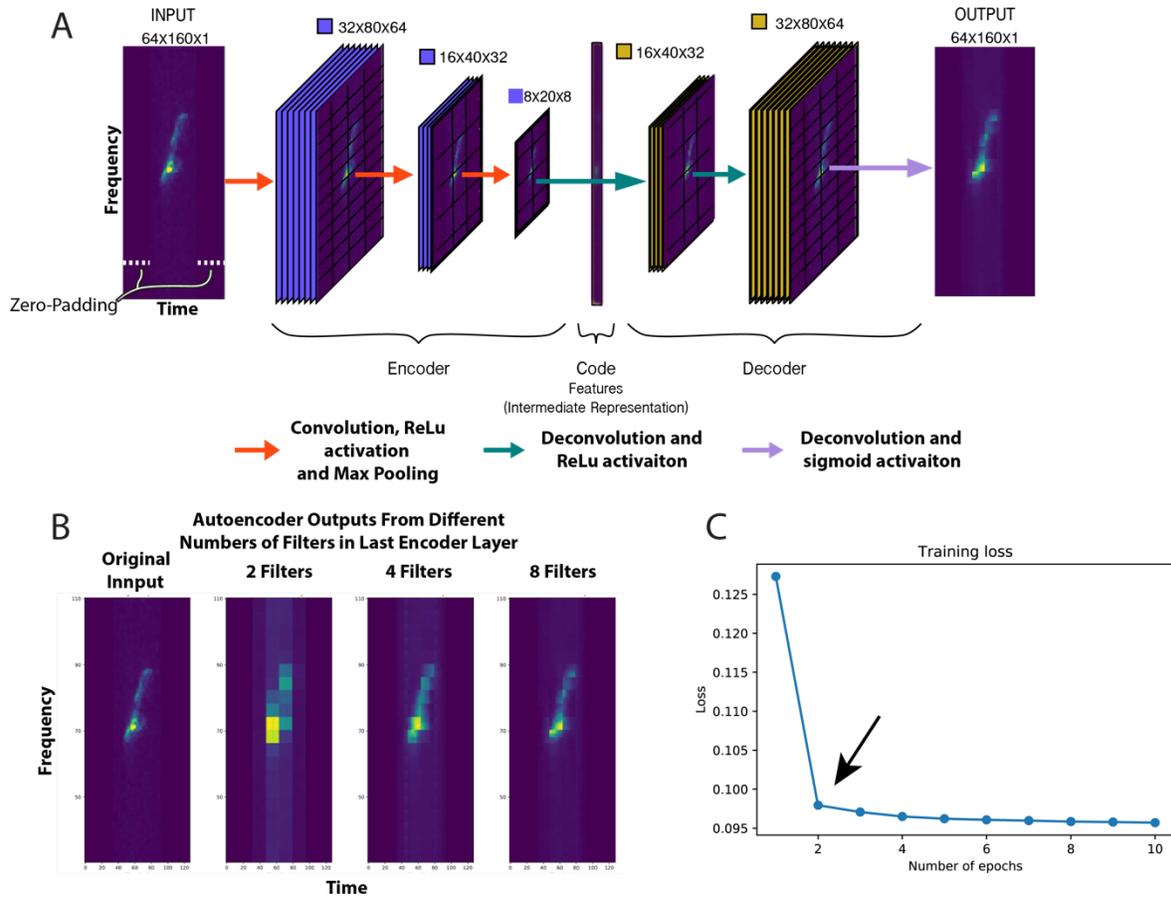
Figure 3.4: Autoencoder Architecture and Parameters.
A) The autoencoder is provided a set of input images which are processed through a set of filters that constitute the encoder. The encoder outputs a final representation that is called the code and is composed of the most informative latent features. The decoder takes the code and attempts to reconstruct the original input though a series of layers and activation functions. **B)** In the final layer of the encoder, the number of filters has a large effect on the selection of latent variables for the code. Increasing the number of layers successively improved the quality of the USV image reconstruction. **C)** Autoencoders are trained iteratively trained through multiple epochs to attempt to faithfully reconstruct the original input. We found that beyond 2 training epochs there was a marginal improvement in training loss. Modified from Stoumpou, et al. 2022.

93

To begin the clustering and classification procedures, the feature vectors of all the USVs are thresholded, normalized, and a Principal Component Analysis (PCA) is applied. A clustering method of the user's choice is then applied to this dimensionality reduced representation, including: Agglomerative, Birch, Gaussian Mixture Models, K-Means, or Mini-Batch K-Means. Users also select how many clusters they want the method to find. The final clustered representations are then visualized using tSNE. **Figure 3.5** is a schematic representation of the workflow from input to clustered output. Users can select to perform their clustering and classifications based on either "deep" features or "simple" features. The deep features as those that
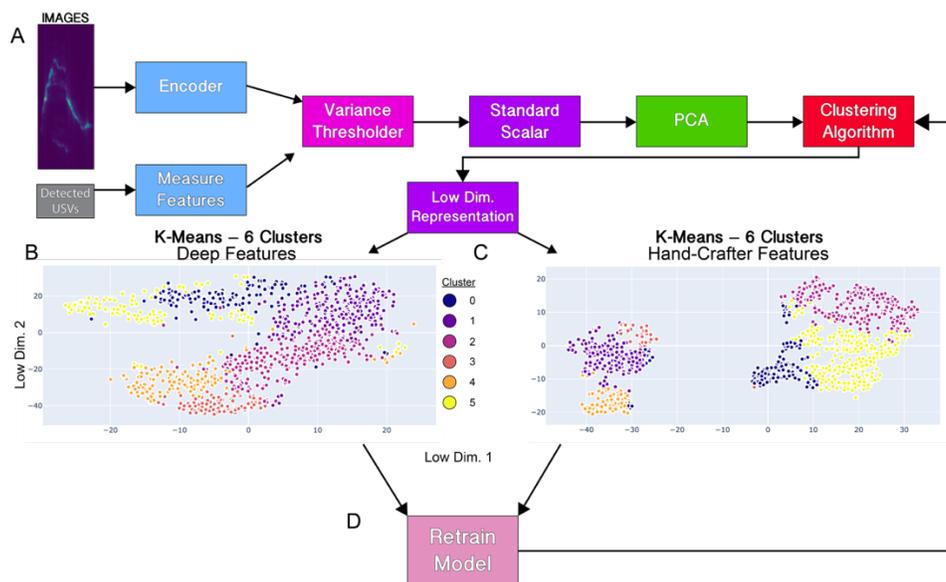
Figure 3.5: Unsupervised Clustering Workflow.
A) Flow diagram of the unsupervised clustering procedure. Output features from the autoencoder or measured simple features are passed through dimensionality reduction steps until there are clustered by a user selected method. Clusters are then visualized using tSNE. Alternatively, other visualization methods could be implemented instead of tSNE. **B)** Example of a set of clusters produced using features extracted from the autoencoder (i.e., deep features). **C)** Example of a set of clusters produced using simple acoustic features. **D)** After a manual curation, the classification model can be retrained and used to determine types of USVs as they are detected in future recordings. Figure modified from Stoumpou et al., 2022.

are derived from the latent representations of the autoencoder. The simple, or so-called "hand-crafted", features are based on the following acoustic measurements of each detected USV: bandwidth, maximum frequency, minimum frequency, mean frequency, and duration.

This process of training the autoencoder and producing clusters for classification also trains a new convolutional neural network used to classify USVs. This newly trained model can be refined through user input. In the AMVOC graphical user interface (GUI), the user can select points in the plot representing each clustered vocalization. If a user considers that a USV was mislabeled, they may reassign its label. If labels are correct, this can also be annotated by the user. These reassignments and correct annotations provide a set of scores that can then be used by the model when being retrained to further improve its labels in subsequent analyses. This trained model can then be loaded in the offline or online AMVOC modes and used for inference and classification. This semi-supervised approach for refining trained models allows the tool to be accurate enough in most use-cases, but still provides users the ability to tailor the classifications to their experiment (e.g. mouse strain) and recording equipment. These semi-supervised approaches have been used by other tools designed for the analysis and discovery of behavior and neural activity such as DeepLabCut (Mathis et al., 2018) and Suite2P (Pachitariu et al., 2016).

3.2.4 Measure of Detection Accuracy

We measured the event-level and temporal accuracy of AMVOC and of other methods (see Section 3.3.1). To calculate the accuracies of each method, we compared their results with ground truth annotations by calculating the harmonic mean (F1) of the precision (true positives relative to all detected events) and recall (true positives relative to all true events). The F1 score is multiplied by

100, leading to a possible range of scores form 0-100. A higher score indicates more similarity to the ground truth dataset, indicating a higher accuracy of detection.

3.2.5 Datasets Used

We developed and tested AMVOC using four datasets, hereafter called D1, D2, D3, and D4. Recordings for these datasets came from a collection published by Chabout et al. (2015) and Chabout et al. (2016), as well as new recordings I collected for the development of AMVOC.

D1 contains recordings from 14 B6D2F1/J and C57BL/6J male mice in both live female and female urine social contexts. We spliced 9 audio segments of 5-10 seconds each form these recordings. We created a set of ground-truth annotations for these recordings, totaling 245 syllables. Files were separated into "noisy" and "clean" based on how much non-vocal noise was present in the recording.

D2 contains 26 recordings from 9 different B6D2F1/J and C57BL/6J male mice.

D3 is composed of data from 72 recordings from B6D2F1/J male mice. There were 36 from the female context and 36 from the urine context. We randomly selected 20 second segments derived from periods of the recordings that had a vocal rate of at least 2.5 USVs/second. The selected segments were then spliced into 4 recordings: 2 composed of female-directed songs and 2 composed of urine-directed songs.

D4 is composed of data from 6 C57BL/6J mice, 5 adults and 1 P7 pups. We created a ground-truth annotation of the first 50 USVs from each recording.

3.3 Results

3.3.1 Operant Training of Volitional Vocalization – Preliminary Attempt

We first endeavored to develop an operant training paradigm using Avisoft. We used a yoked, closed-loop experimental system. The experimental and control mice were kept in separate, isolated chambers. The USVs of the experimental mouse were monitored with Avisoft, and when a sequence of USVs was produced, both mice were given a food reward using a Feeding Experiment Device (Nguyen et al., 2016) (**Figure 3.6A**). Throughout the experiment, mice were maintained on food restriction, and their weight was monitored daily so as not to go below 80% of their baseline/pre-experiment weight. Mice were placed in their testing chambers every day for one hour, any food that was not consumed during the hour was given 15-20 min after the end of the training session. The mice were kept housed together when not in their experimental chambers.

During the initial phase of the training, the mice were allowed to acclimate to the system and experimental chamber, and to receive their food ration via the FED. This acclimation period lasted 10 days. During this phase, if the experimental mouse happens to produce USVs, they are not reinforced via food reward. In the next phase, mice were presented with a social stimulus once at the beginning of the experimental session. After 17 days, we then began presenting the social stimulus every 15 min. The experiment was ended at 36 days. Even though both mice were exposed to the same social stimuli at the same time, over several weeks the experimental mouse vocalized more often than the control mice (**Figure 3.6B**). Although this was a single preliminary experiment, it gave us the optimism that this experimental approach may be possible in mice. The limitations of this system, and of Avisoft in particular, led to the development of AMVOC, which we present in more detail below.
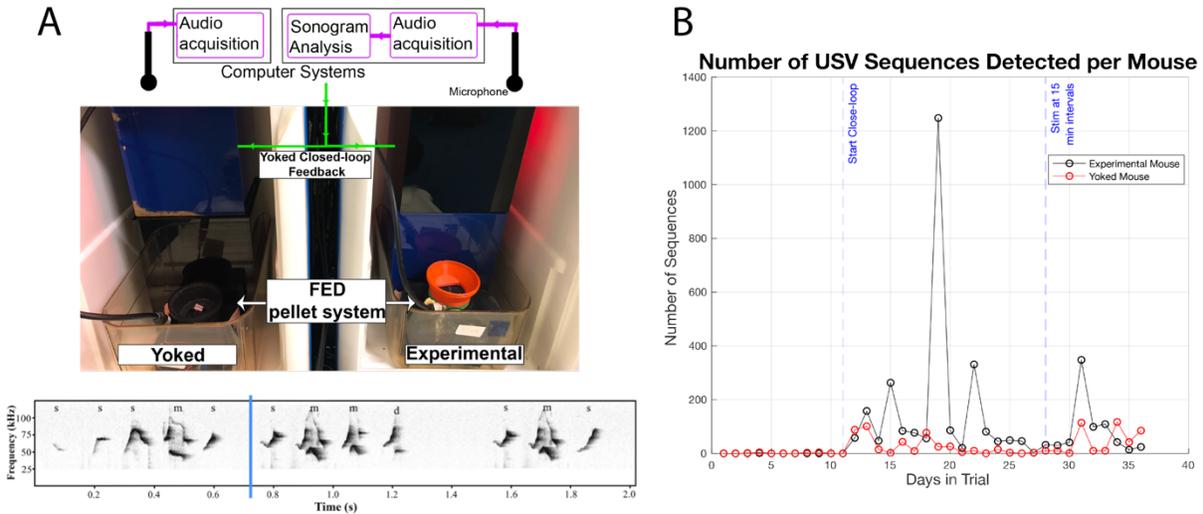
Figure 3.6: Vocal Operant Training Pilot Experiment.
A) Top: Schematic and image of experimental set up for the yoked closed-loop design where USVs are recorded from both mice, but only USVs from the experimental mouse provide a food reward to both mice. Bottom: Schematic of when a reward would be provided after a sequence of USVs (blue line). **B)** Results of this pilot experiment showing the experimental mouse (black) vocalized more than the yoked control mouse (red).

3.3.2 Selecting Parameters for Optimal Accuracy and Performance of AMVOC

In our intended use case in operant training emphasizes the detection of USVs, we term this event accuracy. We can also attempt to determine exactly when the start and end of a USV occurs, we term this the temporal accuracy. By varying t and f the event and temporal accuracy varied. We tested different combinations of these two factors on dataset D1, and determined that values of $t=0.5$ and $f=3.5$ resulted in the best performance of event detection accuracy (**Figure 3.7A**), while balancing precision and recall (**Figure 3.7B**). If we were to maximize these two values (e.g. to $t=1$ and $f=4$), the recall will decrease (**Figure 3.7B**, orange line), meaning we had higher certainty of true positives, but the number of false negatives will increase (**Figure 3.7B**, blue line). Factors t and f can be adjusted as desired by the user in the AMVOC configuration file.
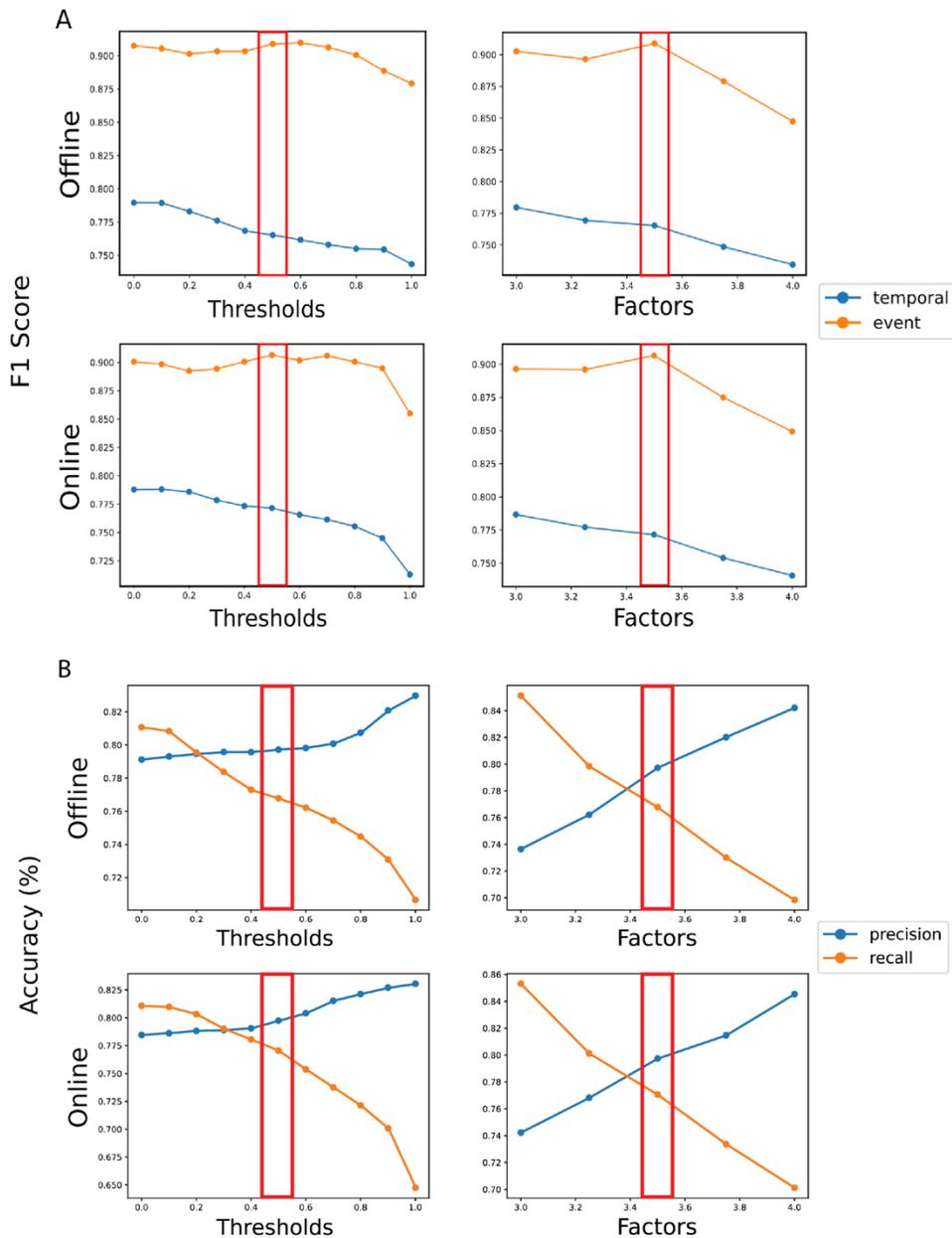
Figure 3.7: Effect of Changing *t* and *f* on Detection Accuracy.
A) Effect on the event and temporal F1 scores by varying the threshold factors *t* and *f*. **B)** Effect on the precision and recall measures of detection accuracy by varying the threshold factors *t* and *f*. Red boxes indicate selected factors of *t*=0.5 and *f*=3.5. Both metrics determined on dataset D1. **(A)** modified from Stoumpou et al. 2022.

3.3.3 Evaluating Processing Speed

We set out to develop a fast and accurate method to detect USVs. We first compared different methods' processing rates by determining the processing ratio. This ratio is a measure of how much faster the method is at processing a file compared to that file's actual duration. For example, a processing ratio of 5 means the method takes 1min to process a 5min long audio file. We compared AMVOC with the following methods: Mouse Song Analyzer 1 (MSA1) (Arriaga et al., 2012), Mouse Song Analyzer 2 (MSA2) that I helped develop (Stoumpou et al., 2022), Mouse Ultrasonic Profile Extraction (MUPET) (Van Segbroeck et al., 2017), VocalMat (Fonseca et al., 2021), and DeepSqueak (Coffey et al., 2019). We used dataset D1 for this test, and ran each file three times per method and calculated the average amount of time it took for the files to be processed. We then determined the processing ratio by dividing the mean processing time by the duration of the files. We found that AMVOC was the fourth fastest method among these methods (**Table 3.1**). The two slower methods both use machine learning (ML) image processing (Coffey et al., 2019; Fonseca et al., 2021).

Method	Processing Rate
MUPET	32.4
MSA2	29.9
MSA1	28.1
AMVOC Offline	21.2
DeepSqueak	8.2
VocalMat	4.3

Table 3.1: Comparison of Processing Rates.
Mean Processing rate determined by analyzing a set of files three times each for each method. A high processing rate indicates the method analyzes data by that factor of real-time. Table from Stoumpou et al., 2022

Because we sought out to make a USV detection method that would be fast and could be used in real-time, we needed to make sure that AMVOC's processing time was still suitable for our use. We first determined how often USVs could be found in each processing chunk. In general, most processing chunks had few, if any, USVs in the female context (**Figure 3.8A**) and the urine context (**Figure 3.8B**). Thus, it was unlikely that we would miss any significant number of USVs while chunks were processed. We next calculated processing latency from the time the spectrogram is processed until after USVs are detected in the calculated spectrogram. The median processing latency for each chunk of 750 ms is less than 39.43 ms (**Figure 3.8C**). Processing latency was not appreciably affected by the number of USVs in each chunk (**Figure 3.8D**). Lastly, we calculated the inter-syllable interval (ISI) for the combined female and urine context
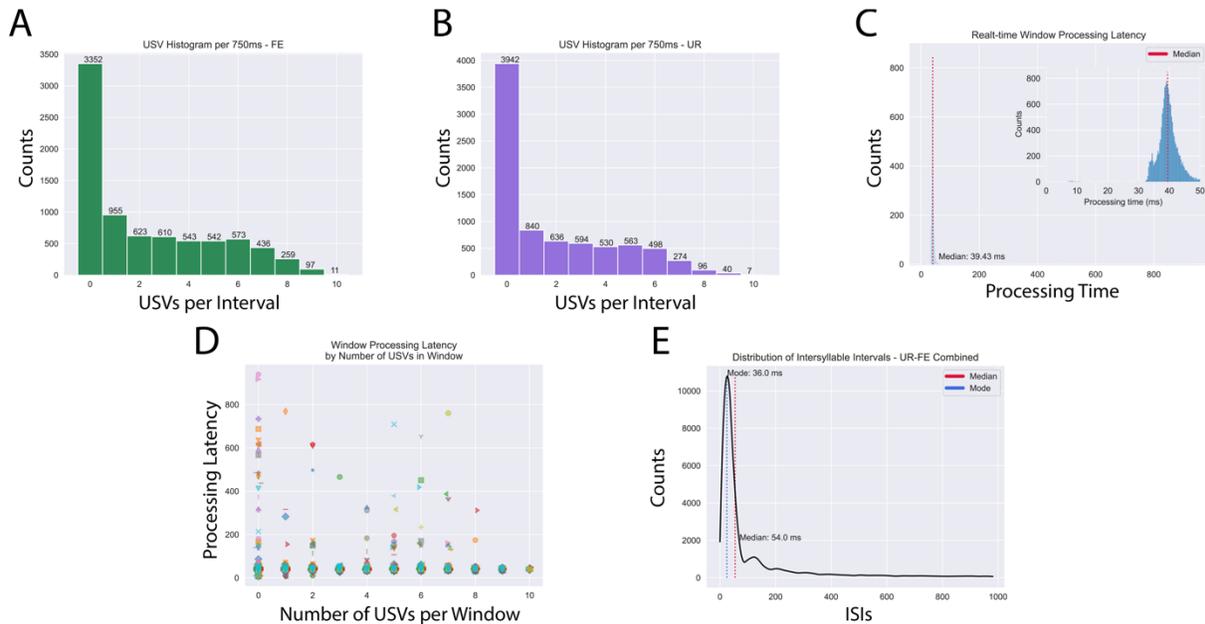
Figure 3.8: Processing Latency of AMVOC Online.
A and B) Histogram of how many USVs are detected in each chunk of online processing for recordings from male mice with a female (A) or female urine (B). **C)** Histogram of processing latencies independent of syllable content. **D)** The processing latency of each 750ms chunk based on how many USVs were detected in that chunk. **E)** Histogram of inter-syllable intervals for both female and urine condition recordings. The median ISI is 54ms and the modal ISI is 36ms. Figure modified from Stoumpou et al., 2022.

recordings. The median ISI was 54 ms and the modal ISI is 36 ms (**Figure 3.8E**). Thus, with the median ISI values being smaller than the median processing latency, we should not experience many "drops" in detection. Further, we also have a 13% overlap between segments which allows us to capture USVs that would be missed by this inter-segment processing period (see **Methods, Section 3.2.2**). The number of USVs present in a processing window did not influence the processing latency of that window. In other words, during highly vocal periods, AMVOC's online functionality is not adversely affected compared to less vocal periods. Overall, we found that AMVOC's processing speed is reasonably well suited for online processing.

3.3.4 Evaluating USV Detection Accuracy

The most important aspect for a new detection method is that it as accurate as existing methods, if not more accurate. We compared AMVOC with the same methods described in 3.3.3, using dataset D2. We divided the dataset into files that were noisy and those that were relatively free of extraneous cage noise, which we called "clean".

Because every method can be optimized with varying numbers of parameters, or trainable ML models, we used the default settings that were provided with each method in order to produce the most comparable results. DeepSqueak provided pre-trained detection models for rat and mouse USVs, so we used the pre-trained model for mice. **Table 3.2** shows the F1 scores of AMVOC and the methods we compared it with for both clean and noisy recordings. In both types of recordings, AMVOC was the most accurate method. With the exception of MSA1, the other methods performed well in the clean recordings, all scoring above 90% accuracy, but noisy files were more challenging. We also compared the temporal accuracy of each method based on the above event detection results. The temporal accuracy is a measure of how much the method agrees with the

Type	Recording Type	AMVOC-Offline	AMVOC-Online	MSA1	MSA2	MUPET	VocalMat	DeepSqueak
Event	Clean	**97**	**97**	66	94	93	90	93
	Noisy	**84**	83	33	72	57	58	81
	Mean	**90.5**	90	49.5	83	75	74	87
Temporal	Clean	84	85	46	88	85	**91**	83
	Noisy	67	68	23	71	53	58	**76**
	Mean	75.5	76.5	34.5	**79.5**	69	74.5	**79.5**

Table 3.2: Comparison of Detection Accuracies.
F1 scores for both event and temporal detection of each USV detection method tested. AMVOC had the highest scores for event detection. AMVOC did not fare as well in temporal accuracy, likely meaning that although ability to detect USVs was good, when AMVOC considers them to start or end is not as goTable 3.3: Comparison of Detection Accuracies for Dataset D4.F1 scores for both event and temporal detection of each USV detection method tested. As before, AMVOC had the highest event F1 scores. Table modified from Stoumpou et al., 2022

ground truth start and end times for each USV that was detected. AMVOC's temporal accuracy is not as good as its event accuracy, ranking 3rd and 4th in the both the clean and noisy recordings.

We next compared the processing speed with the accuracy of each method determined above. There was an overall negative correlation between processing rate and detection scores, suggesting slower methods had more accurate event detection. For example, VocalMat and DeepSqueak, which are the slowest methods (**Table 3.1**), had more accurate event detection scores than all methods except AMVOC in the clean recordings, the noisy recordings, and overall (**Figure 3.9A**). A similar trend could be seen for the temporal accuracy scores relative to processing rate in the clean recordings, noisy recordings, and overall (**Figure 3.9B**). AMVOC was more on-trend
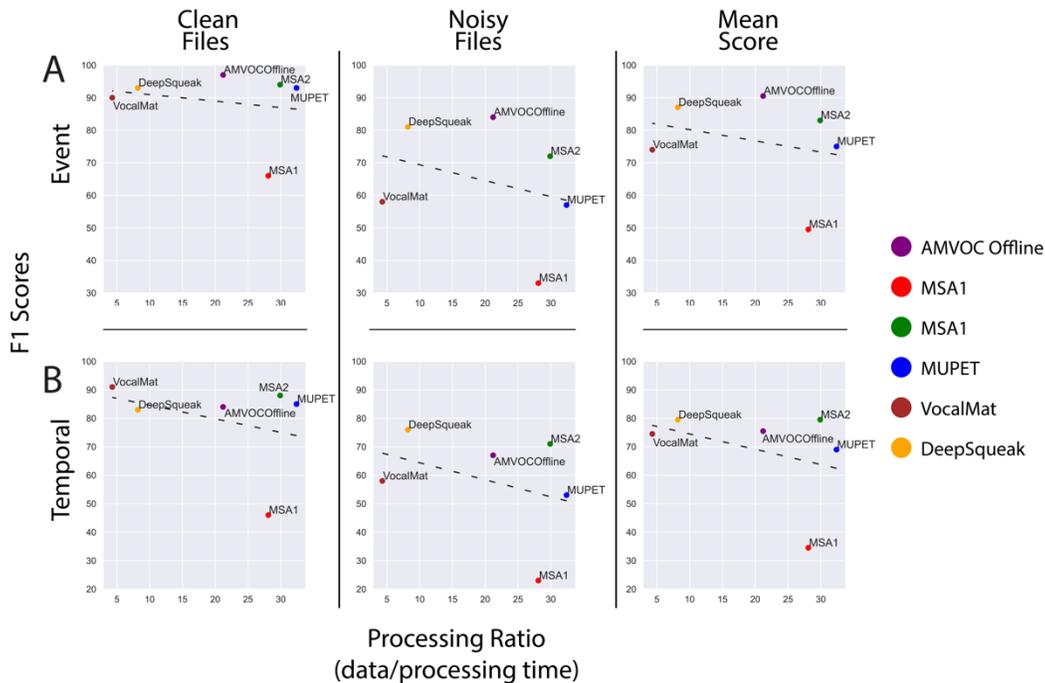
Figure 3.9: Relationship Between Processing Ratio and Detection Accuracy.
A) Event accuracy for each method compared to its processing ratio for different qualities of recording. **B)** Temporal accuracy for each method compared to its processing ratio for different qualities of recording. All comparisons had negative trends with faster methods exhibiting a tendency for lower F1 scores in both event and temporal accuracies. Figure modified from Stoumpou et al., 2022.

in terms of temporal accuracy than for event accuracy. Despite being the 4th fastest method, AMVOC consistently outperformed the other methods in event detection accuracy.

As far as the event detection results are concerned, one could argue that the recordings used to develop AMVOC were still similar to those used in our comparisons. For example, despite our test never showing VocalMat to be the most accurate method, when compared the different methods with a dataset associated with the publication of VocalMat, VocalMat was the most accurate method (data not shown). To compare methods more accurately, we annotated a new ground truth data set for recordings that had never been tested on any of the methods, dataset D4. We annotated the first 50 USVs of each recording, regardless of how long into the recording this

occurred. For each method, we evaluated their accuracy based on which vocalizations they detected up to the time at which the 50th ground-truth USV occurred. If the 50th ground-truth USV occurred at 60 seconds, and a method's 44th USV occurred at 59 seconds, while the 45th USV occurred at 62 seconds, then we only evaluated that method's accuracy up to its 44th detected syllable (**Figure 3.10**).

We evaluated each method, including AMVOC, using default parameters. Testing these methods on this previously unseen dataset, we found that AMVOC had a lower temporal F1 score than the other methods (**Table 3.3**), similar to our previous test. AMVOC was still the most accurate method for event detection in both clean and noisy recordings (**Table 3.3**). The online functionality of AMVOC retained event accuracies similar to the offline method.

3.3.5 Unsupervised Clustering and Classification

In order to perform template matching experiments, AMVOC needed to be able to perform reliable unsupervised classifications of USVs. We evaluated our unsupervised clustering by comparing the "deep" autoencoder-extracted features with the measured "simple" acoustic features (**Figure 3.11A**). Clustering was evaluated manually by scoring a random sample of USVs from each cluster
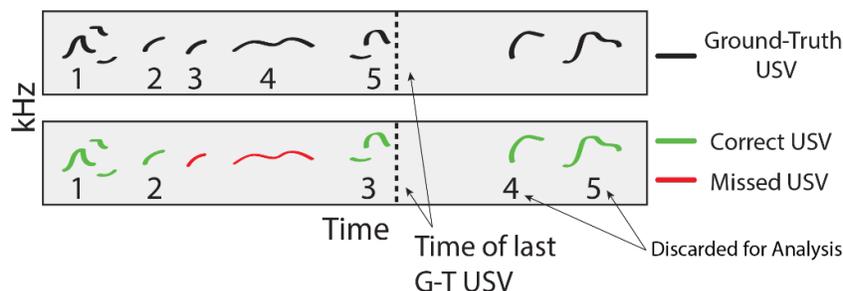
Figure 3.10: Schematic of USV Detection Method Evaluation on Dataset D4.
General schematic depicting how we compared detection results from different datasets when we generating a ground-truth annotation for only the first 50 USVs in each recording in dataset D4. The bottom diagram depicts at which USV comparison would be stopped despite more being detected passed the limit determined in the ground-truth annotations.

Type	Recording	AMVOC Offline	AMVOC Online	VocalMat	MUPET	DeepSqueak
Temporal	Rec1	64.9	64.9	63.7	**73**	63.4
	Rec2	67.6	67.6	**85.3**	61.9	79.5
	Rec3	89.1	90.9	90.7	**94.9**	89.6
	Rec4	69.9	69.6	**83.7**	68.9	61.6
	Rec5	72.1	69.1	**83**	67.4	58.6
	Rec6	41.9	43.8	34.5	27.8	77.3
	Mean	67.6 (±6.2)	67.7 (±6.1)	**73.5 (±8.7)**	65.7 (±8.9)	71.7 (±5.0)

Type	Recording	AMVOC Offline	AMVOC Online	VocalMat	MUPET	DeepSqueak
Event	Rec1	93.9	93.9	85.5	**97**	81.8
	Rec2	**99**	**99**	90.1	91.7	90.5
	Rec3	**100**	**100**	98	96.9	96.9
	Rec4	**94.1**	93.2	78.1	88.9	79.1
	Rec5	**89.5**	83.2	73.5	71.6	78.7
	Rec6	78.1	77.2	58.8	73.6	**84.8**
	Mean	**92.4 (±3.3)**	91.1 (±3.7)	80.7 (±5.6)	86.6 (±4.6)	85.3 (±2.9)

Table 3.3: Comparison of Detection Accuracies for Dataset D4.
F1 scores for both event and temporal detection of each USV detection method tested. As before, AMVOC had the highest event F1 scores. Table modified from Stoumpou et al., 2022

106

based on how well each USV in a cluster matched the other USVs in that cluster. We used dataset D3. Each evaluator assessed the clustering of each of the four recordings in three different clustering configurations: six clusters made with Agglomerative clustering, six clusters made with Gaussian Mixture clustering, and six clusters made with K-Means clustering. Each of the four recordings within each of the three clustering methods was evaluated for autoencoder-extracted features and measured acoustic features. Evaluators were blinded as to which evaluations were the deep or simple feature clustering by labeling them as either "Method 1" or "Method 2".

We found that human evaluators scored the clusters based on the autoencoder features as being significantly more accurate/consistent than clusters based on simple acoustic features (K-Means, $p=1.7 \times 10^{-4}$; Gaussian Mixture, $p=1.3 \times 10^{-4}$; Agglomerative, $p=7.2 \times 10^{-3}$, **Figure 3.11B**). Scores within cluster for each type of feature were also significantly higher for the autoencoder features (K-Means, $p=1.3 \times 10^{-9}$; Gaussian Mixture, $p=3.4 \times 10^{-9}$; Agglomerative, $p=8.3 \times 10^{-7}$, **Figure 3.11C**). Within each cluster, evaluators scored similar numbers of USVs as appropriately belonging to the group they were assigned (i.e., approved) (**Figure 3.11D**), suggesting that despite the difference in cluster score, there was within-cluster consistency for each Method. Overall, clustering with deep features outperformed clustering with simple features, indicating that the autoencoder has successfully identified latent features that are useful for the categorization of different types of USVs.
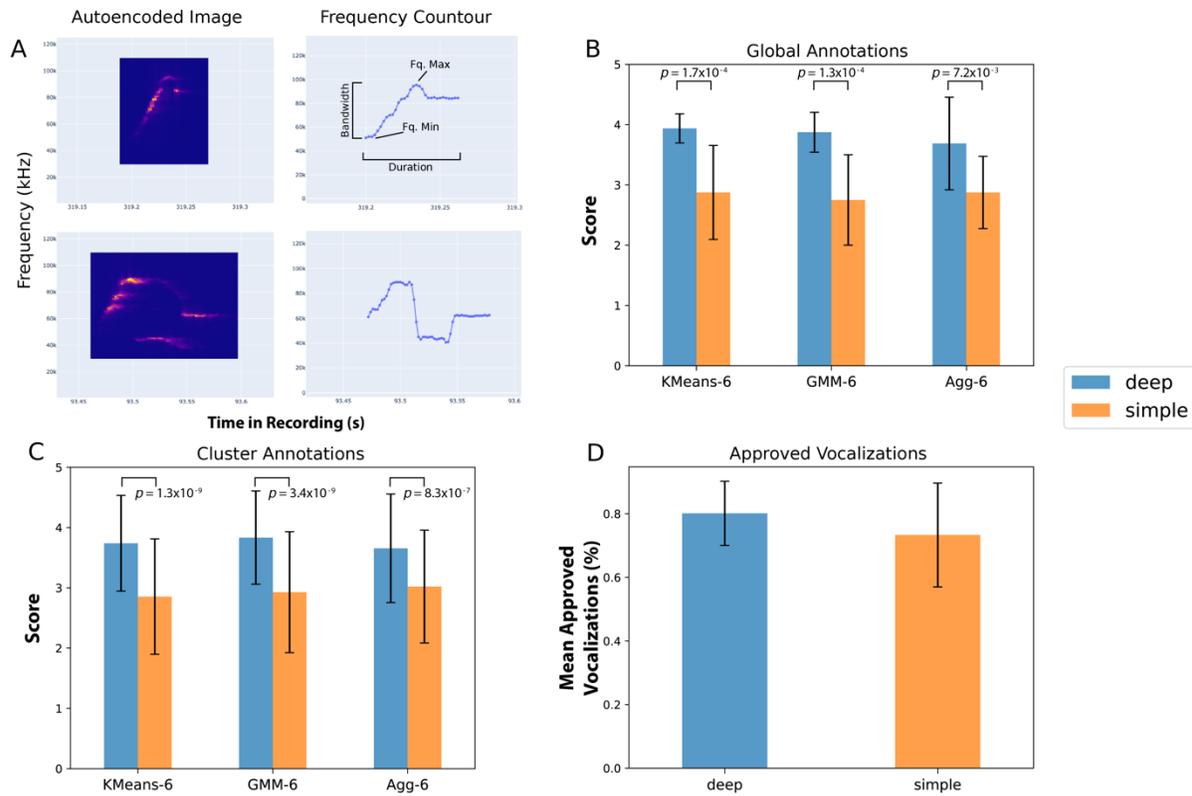
Figure 3.11: Annotation Scores of Clustering Evaluation.
A) Example of two USVs as reconstructed by the autoencoder. Right plots show the spline of the USVs based on the peak frequency. Spline is used to determine the acoustic features of bandwidth, duration, frequency minimum, and frequency max. **B)** Mean and S.D. of scores for the overall quality of the clustering across all clusters from each combination of methods evaluated. **C)** Mean and S.D. of scores for each individual cluster within each combination of methods. **D)** Percent of approved vocalizations, which is useful as a proxy for within-cluster similarity of USVs. Figure from Stoumpou et al., 2022.

3.3.6 Application of Online Detection Mechanism – Proof of Principle

3.3.6.1 Detecting USV Classes With AMVOC Online

To demonstrate the capabilities of AMVOC's ability to classify USVs in its online mode, we trained a classifier on a recording of a male mouse singing with a female present. We applied a Mini-Batch K-Means with a parameter of 5 clusters using autoencoder features and saved this classifier model. Next, we used AMVOC's online procedure to process a new audio file recorded from a different male singing to a female mouse and applied the newly trained classifier model. **Figure 3.12A** depicts the spectrogram of the second USV recording. Each detected USV was classified into a different syllable class based on one of the 5 available in the trained model. **Figure 3.12B** shows a screenshot of file as its being processed.
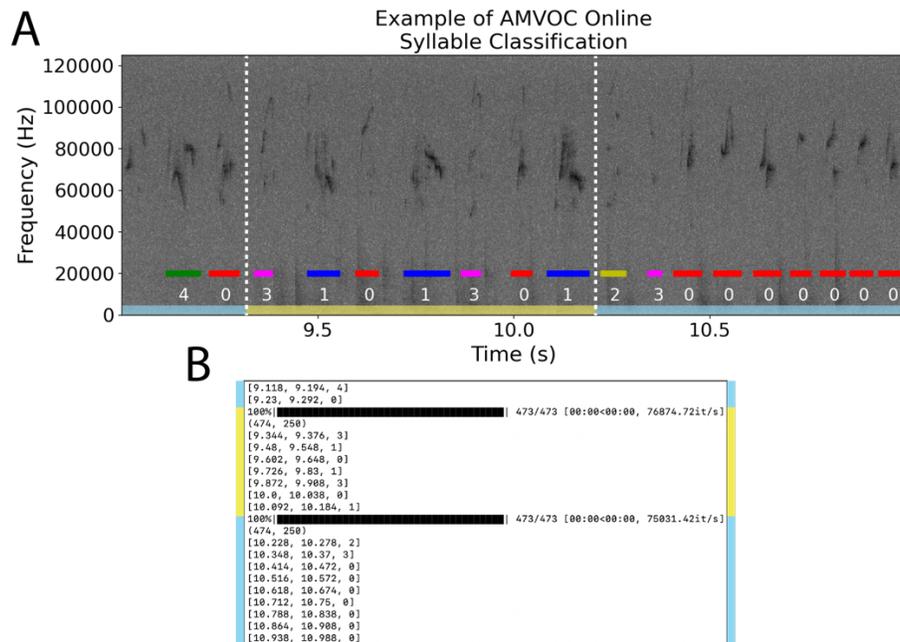
Figure 3.12: Classifying USVs with AMVOC in Real-Time.
A) Example spectrogram of USVs classified using AMVOC's real-time functionality. Colored horizontal lines mark the start and end of detected USVs. Numbers represents one of 5 possible classes from the trained model. Vertical dashes denote boundaries of processed spectrogram chunks. **B)** Screenshot of terminal window as file was processed. Each bracketed row has the start time, end time, and class of each syllable detected. Blue and yellow boxes between spectrogram and terminal window represent corresponding chunks.

3.3.6.2 Design of a Closed-Loop Operant Training Paradigm for Mice

To test the vocal capabilities of mice using AMVOC, we designed and built a low-cost, open-source operant training paradigm that can be used for our experimental goals. We designed an operant apparatus around the animals' home-cage environment. In this way, the animals self-pace their behavior and reward, thus demonstrating volitional control. This approach greatly reduces the need for human involvement (e.g., handling or moving animals between home cage and experimental chamber) which can have unmeasurable effects on the animals' training and performance.

Figure 3.13A provides a schematic of our operant training design. Each mouse is singly-housed in a cage that is place inside one of our sound isolated chambers (Igloo® coolers). Each cage has soft bedding instead of the typical bedding which can introduce a lot of acoustic noise when the mice walk around or dig. Cages our outfitted with a Feeding Experiment Device 3 (FED3) (Matikainen-Ankney et al., 2021) and a water bottle. Mice are separated from the FED3 by an acrylic barrier which only provides access to FED3's pellet well for feeding (Figure 3.13B). Two ultrasonic microphones are placed above the mouse with acoustic access through ports in the lid. One microphone is connected to the computer running AMVOC to detect vocalizations. We use the Ultramic 384K BLE (Dodotronic©) as the primary microphone connected to AMVOC. As AMVOC operates we cannot both process and save the audio without incurring a large processing cost to the latency which we have tried to diminish. Thus, the second microphone gives us a way to record ground-truth audio files of the detection events. When USVs are detected, a signal is sent to the audio recording system of the second microphone to save the ground-truth data. The ground-truth file is composed of audio during the duration of the trigger, 3 seconds prior to trigger, and
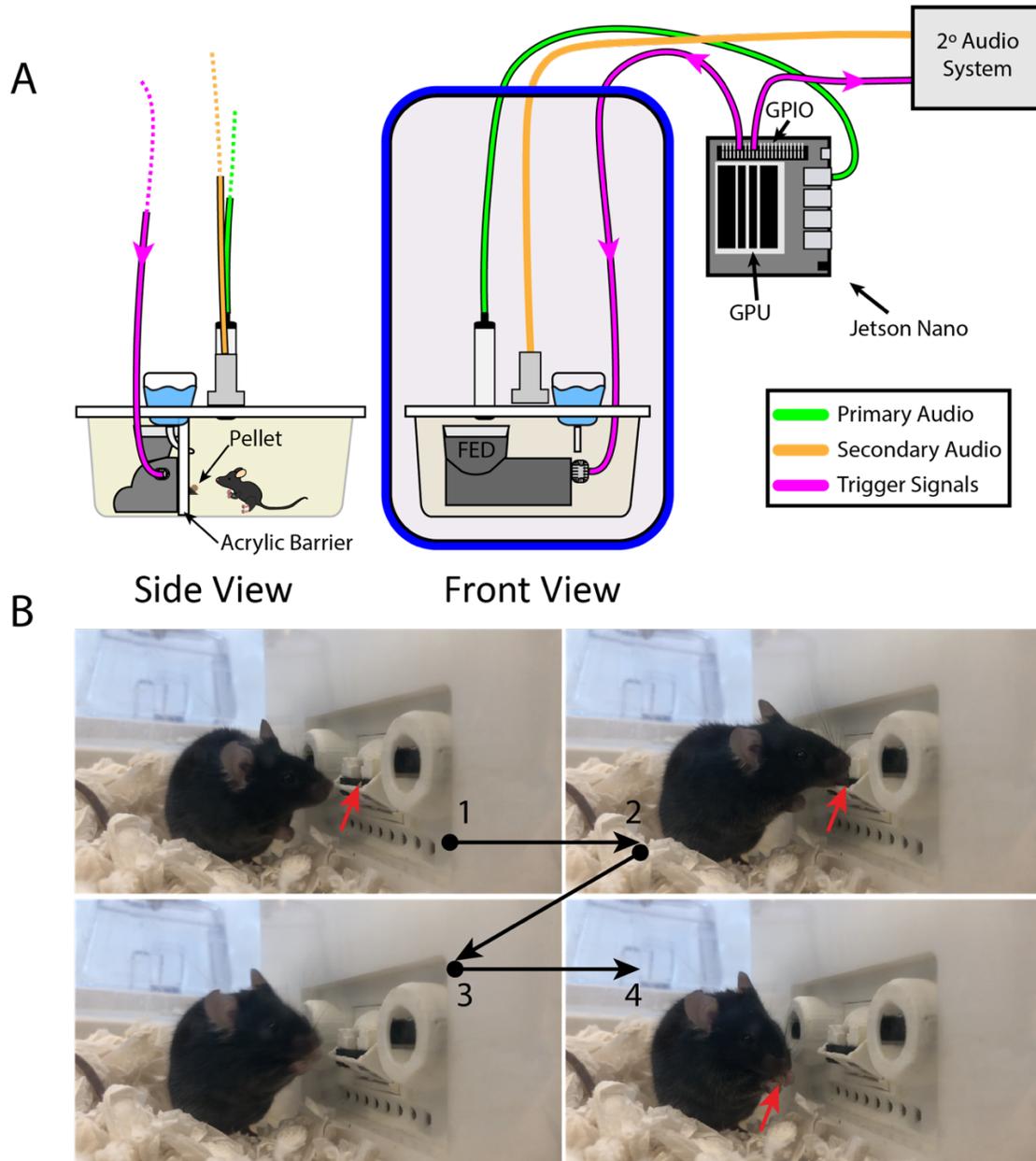
Figure 3.13: Design of Closed-Loop Operant Training Paradigm.
A) General schematic for the closed-loop operant training system. Acoustic signals detected by the primary audio recording are processed by AMVOC. When a USV is detected, the Jetson Nano sends a trigger signal to the FED3 to deliver a pellet and a second signal to the secondary audio system to capture a ground truth recording of the putative signal detected by AMVOC. **B)** Example of a mouse in its home cage eating a pellet out of the FED3. An acrylic barrier limits the mouse's access to the pellet and nose poke ports that are unused in this experimental design.

3 seconds after the trigger offset. The primary detection microphone is connected via USB to an NVIDIA Jetson Nano (a single-board computer, SBC). We selected the Jetson Nano over other SBCs currently available because it is easier to set-up to perform ML and deploy trained models for inference and classification. The Jetson Nano also comes equipped with a GPU which is preferred over central processing units (CPUs) for training ML models. If an experimenter wanted to train, and later use, a classifier on AMVOC without access to desktop equipment, the entire procedure could be run from the Jetson Nano. The general-purpose input/output (GPIO) of the Jetson Nano is configured to send a signal to the FED3 and the secondary audio recording system whenever a vocalization is detected. This signal tells the FED3 to provide the mouse a food pellet and tells the audio recording system to save a wav audio file of the segment when the USV was detected.

The FED3 is an open-source hardware system operated by an Arduino. It is designed to allow users to tailor its operation to various experimental needs. We wrote two different experimental modes that can be "flashed" onto the Arduino. The first is a food restriction program used to reduce body weight and make food rewards for motivation. A common problem with some types of food restriction is than mice will binge when given access to a full ration of food. We opted to spread out their ration throughout the day at regular intervals of smaller portions of food. As such, users can select the *n* number of pellets that are given to the mice every *x* hours. When the limit of *n* is reached before the next feeding time, no pellets are provided until the next feeding interval. If a given time interval's ration is not maxed out, then the remaining portion is rolled over into the maximum of the next feeding interval. Our second mode is strictly a closed-loop operant program. Whenever the FED3 receives a 'high' signal form its BNC port (e.g. from a detected USV), a food pellet is delivered. If a signal is received but the previous pellet has not been

removed, a new pellet is not provided. Ideally this prevents mice from singing repeatedly and amassing food in a way that does not allow them to form an association between vocalizing and food reward. In both operating modes described here, an audible tone paired with every pellet delivered. An example of the closed-loop food reward with paired tone is shown in **Figure 3.14**.
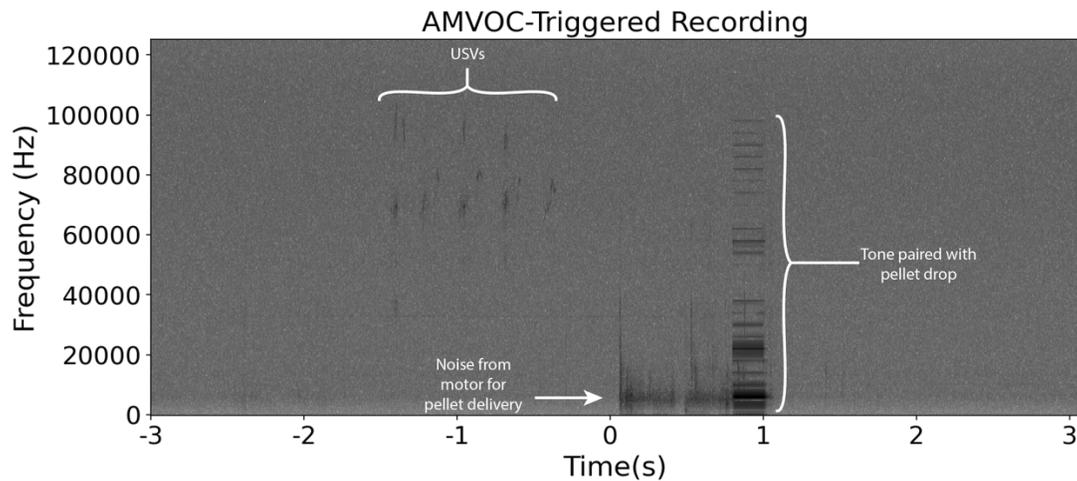
Figure 3.14: Example of Ground-Truth File Recorded in Closed-Loop.
Spectrogram of a set of USVs detected by AMVOC that lead to a pellet being delivered by the FED3. The time is centered around the time of detection. A slight delay occurs between the end of the USVs and when a pellet is delivered, due to the duration of a processing chunk. After detection, there is noise visible in the spectrogram from the FED3's motor turning to deliver a pellet. The paired tone is also visible when the pellet dropped and was detected by the FED3. Neither the motor noise and the broad spectrum of frequencies from the paired tone are detected as vocalizations by AMVOC suggesting these tools are compatible with each other.

3.3.6.3 Operant Training

Once we had established a functional method to detect USVs and provide mice with a food reward, we developed a preliminary approach for training mice in this operant paradigm (**Figure 3.15A**). We first begin by food restricting the mice for 7 days in their cage while we record their baseline level of spontaneous USVs. Following this phase, we "prime" the males with a female mouse as this increases the male's response to female urine later on (Arriaga et al., 2012; Chabout et al., 2015). A barrier is inserted into the cage, behind which a female mouse is placed with her own

source of food and water. During both of these phases the male mice are under a consistent food restriction protocol. After 3 days of priming with the female, she and the barrier are removed, and the FED3 is loaded with the closed-loop program. From here on, detected USVs will trigger the FED3 to provide a food pellet if one is currently not available. In the third phase of the training, we expose mice to female urine. During a one-hour period, 100 µl of urine is introduced every 15 min. During the last phase of the experiment, no stimuli are presented and the mouse must, on its own, produce USVs to gain access to food rewards.

USV analyses and counts were made using the ground-truth recordings form our second audio system and processed through AMVOC's offline mode. Because we switched experimental phases for each mouse at noon, we counted the number of USVs produced during the 12-hour periods representing midnight to noon (AM) and noon to midnight (PM). In our analyses, we also excluded the one-hour period during which urine was presented to the mice, as the vocal response to urine can be quite variable are not necessarily informative for our changes in vocal rate. Thus, we only counted non-elicited USVs when the closed-loop feedback was engaged.

We tested two mice with this process, Mouse 1 and Mouse 2. Mouse 1 had been used during the set-up for this experiment and was exposed to live females to test the functionality of the experimental paradigm without undergoing training. Mouse 1 had previously undergone a period of food restriction. To counter-balance this, we also performed this experiment on Mouse 2 who had not undergone any exposure to the experimental paradigm. We began the experiment on both mice at the same time, so they experienced the same days and the same periods of social stimulus exposure (female and urine).

Mouse 1 had had few USVs during the baseline period, (**Figure 3.15B**, red), and had surprisingly few during the female period (**Figure 3.15B**, orange). This mouse's USVs increased

sharply during the operant training period with urine (**Figure 3.15B**, green), and maintained this elevated level of USV production during the operant-only phase where not social stimuli were provided (**Figure 3.15B**, red). For Mouse 1 we counted 4.6x more USVs in the closed-loop only phase as compared to the baseline phase (**Figure 3.15C**).

Mouse 2 had a similar pattern of behavior to Mouse 1 during the baseline phase (**Figure 3.15D**, blue). Mouse 2 had a noticeably higher number of vocalizations when the female was present (**Figure 3.15D**, orange), as compared to what we saw from Mouse 1. In the last two phases, the USV counts dropped relative to the female phase, but on average the number of USVs was higher than in the baseline period. For Mouse 2 we counted 8.1x more USVs in the closed-loop only phase as compared to the baseline phase (**Figure 3.15E**).
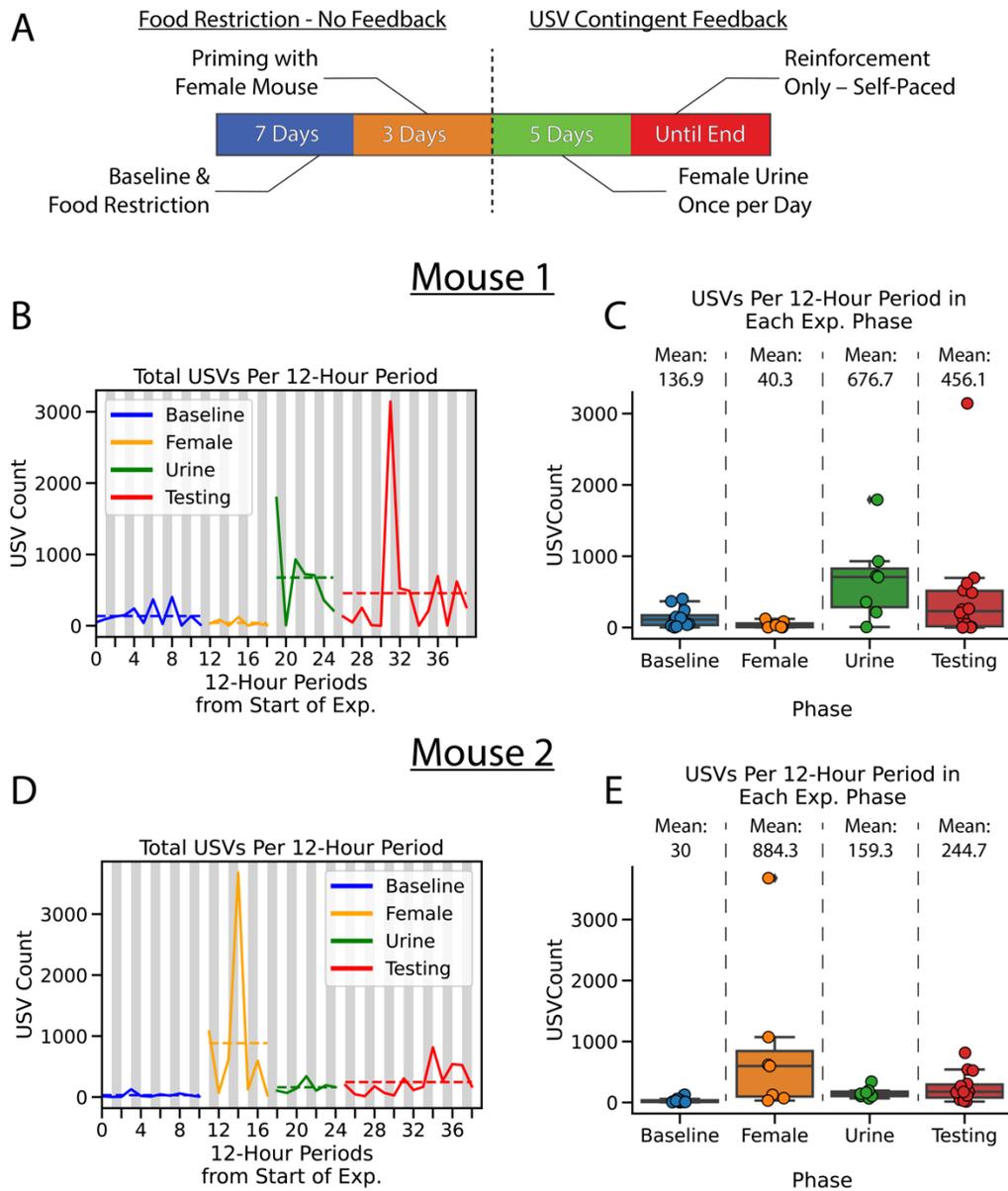
Figure 3.15: Example of Two Mice Being Trained.
A) Schematic of the different training phases, and the respective durations, that mice will go through during operant training. **B)** Plot representing the total number of USVs in a 12-hour period for Mouse 1. Colors reflect the different experimental periods. Dashed lines represent the mean number of USVs in that phase. White bars denote PM periods, and grey bars denote AM. **C)** Boxplot representation of the Data in **(B)** of the total USVs per 12-hour period with mean USV counts per period listed in the respective column. **D and E)** Plot and boxplot of total number of USVs in a 12-hour period for Mouse 2 as in **(A and B)**.

3.4 Discussion

We developed AMVOC as a new method for detecting and classifying mouse USVs. AMVOC provides high rates of accurate detection of USVs in a variety of acoustic environments. Although AMVOC is not the fastest method for analyzing and detecting USVs, it is reliably the most accurate without performing extensive parameter tuning and optimization. This makes AMVOC an ideal tool for many researchers and allows for a broader set of use cases and experimental contexts that may not always have ideal sonic environments. AMVOC also provides a GUI for users to improve the syllable classification models which allows for improving the quality of experiments that would require this classification in real-time, or simply for other offline analyses. Lastly, we demonstrate a use case for AMVOC where we integrated it into an operant training paradigm that will allow us to train different features of mouse vocal behavior.

Prior to the wide availability of automated USV detection tools, many researchers performed USV annotation and classification by hand. This allowed experimenters a great deal of flexibility in what they considered to be distinct classes of USVs based on their spectrographic shape (Scattoni et al., 2008), but this was laborious and required expert annotators. As detection methods became more automated there was a greater reliance on using empirical measures of acoustic features to determine call type, such as whether or not pitch jumps are present and in what direction (Arriaga et al., 2012; Chabout et al., 2015). The rapid changes in the accessibility of tools such as affordable graphics processing units (GPUs) have made cutting-edge ML technologies more readily usable in a standard laboratory setting. These changes in access to technology have had a large effect on studies of animal behavior (von Ziegler et al., 2021), and USVs are no exception.

Methods such as AMOVC, MUPET, and Autoencoded Vocal Analysis (AVA) (Goffinet et al., 2021) all use ML methods for the classification and analysis of the analytically detected USVs. Meanwhile DeepSqueak, VocalMat, and DeepAudioSegmenter (DAS) (Steinfath et al., 2021) are fully end-to-end dependent on ML for the detection and classification of USVs. These methods are an efficient blend of the quality of segmentation and classification that comes from hand-labeling with the promise of efficiency provided by automated methodologies; human vision is replaced by computer vision.

USV classification, either through qualitative observation or quantitative approaches, has largely been determined by human perceptual categories. Although these categories are reliably discernable by human observers, it is unclear how relevant these supervised categories are to the analysis of USVs. For example, our lab has previously used four categories of USVs based on the number and direction of pitch jumps present in a syllable (Chabout et al., 2015). Others have taken these categories and simply added additional "categories" by subdividing some of them by their duration (i.e. long and short versions of the same syllable) (Castellucci et al., 2018; Hertz et al., 2020). Others have used 10 categories based on spectrographic shape (Scattoni et al., 2008; Fonseca et al., 2021). What relevance, if any, these distinctions have on the sensory and behavioral ethology of the mouse is unclear. Further, when it comes to rare or novel types of vocalizations that occur due to a social context or genetic mutation, pre-determined categories could easily miss slight variations in USV structure. ML-based methods support a new approach that would diminish the amount of *a priori* categorization imposed on vocal classes.

The promise of unsupervised methods for classification is that they mitigate the amount of human bias insofar as the categories that are created are not based on what an experimenter believes is a relevant/important set of distinguishing factors between vocalizations. Compared to

many other species' vocalizations, mouse USVs do not form distinct clusters in low dimensional representations (Sainburg et al., 2020). USVs often form gradients in these low dimensional representations, which in a fully unsupervised classification approach can limit the types of analyses that are traditionally performed in neuroscience such as peri-event correlations with neural activity. Some degree of categorization can be useful for analyses. AMVOC, DeepSqueak, and MUPET take a semi-supervised approach where the tool is asked to discern a defined number of syllable categories. Although this means there is some degree of human influence, the actual boundaries between the categories determined by spectral structure or acoustic features are not determined by human perception. This alleviates one important distinction of a fully unsupervised method in which there are gradients of syllable differences across a low-dimensional representations of the USVs.

AMVOC was developed in parallel to many recent advances in the neurobiology of behavior. Driven by a variety of experimental questions and model organisms, there has been a boom of analysis packages that are designed to help researchers quantitatively explore and analyze animal behavior with increasing degrees of freedom. Some of these methods, like DeepLabCut (Mathis et al., 2018) have also been adapted for real-time analysis (Kane et al., 2020; Schweihoff et al., 2021). These methods have been successfully integrated with closed-loop operant methods to train mice to perform a variety of tasks in which their behavior is tracked by the behavioral models (Schweihoff et al., 2021), or even to manipulate neural activity based on behavior (Willmore et al., 2022). Ultimately, our goal is to integrate AMVOC into a similar system for vocal behavior. Our ability to fully test the vocal capabilities of mice has largely been limited by our tools (both software and hardware). By integrating AMVOC into a closed-loop operant paradigm we first want to test whether mice have volitional control of their vocalizations, which is an integral

and necessary component of vocal learning. Next, we will test whether mice can change relatively simple acoustic parameters of their vocalizations like increasing or decreasing the fundamental frequency (Xiao et al., 2018; Lattenkamp et al., 2020). Ultimately, we are interested in whether mice have any imitative abilities. For this experiment, we can implement a design similar to what has been used in budgerigars where they were trained to match a template vocalization to receive a food reward (Manabe and Dooling, 1997) — a feat only now testable in mice with the development of AMVOC. We can further integrate these experiments with electrophysiological recordings to better understand the role of M1 and the direct projection in the preparation and execution of volitional vocalizations (Hage and Nieder, 2013; Brecht et al., 2023).

With the new tools described here, we can now go beyond passive recordings of natural behavior in mice, but we can now challenge their behavioral repertoire and gain insights into cognitive and behavioral deficits or augmentations induced by targeted genetic mutations. Mice lacking *PlxnA1* have been shown to have increased manual dexterity due to more numerous direct corticospinal projections (Gu et al., 2017). Will mice exhibit changes in the vocal capacity of mice if they are challenged vocally the way the manual dexterity of mice was challenged by Gu et al. (2017)? In preliminary work in our lab, we have demonstrated that *PlxnA1$^{fl/fl}$;Rbp4-Cre* mice, which have *PlxnA1* knocked-out in layer 5 of cortex, have increased terminal arborization of LMC neurons in the reticular formation around the Amb, as well as increased innervation of the Amb. (Lomax Boyd, personal communication). Although these mice appear to have changes in their vocal repertoire, it is difficult to determine the difference between "change in vocal behavior/repertoire" and "augmented vocal skill". This neuroengineering of circuits is an exciting direction and application for the tools described here. We have not previously had adequate ways to test for "improvements" in vocal behavior in mice beyond just measuring changes in repertoire,

a feature which is known to vary by genotype and mutation, not necessarily mediated by a change in vocal "skill" (Kikusui et al., 2011; Jourjine et al., 2023). Comparative studies between human and songbirds have demonstrated gene expression patterns that appear convergently specialized for the speech and song learning circuits (Pfenning et al., 2014; Gedman et al., 2022). Introducing the sequences of the most specialized genes, or even the complement of gene expression patterns, could lead to improvements in the vocal "dexterity" of mice and make them more like vocal learners. Our operant chamber provides us the opportunity to challenge the vocal dexterity and cognitive/volitional control of vocal features, the way a *PlxnA1* mutant mouse might have increased control over its digits and wrists movements. These experiments would not only provide a better understanding of the basic neurobiology of vocal production in mammals, but we can gain experimental insight into the role that candidate genes may have in the development and maintenance of neural circuits for vocal production and vocal learning.

Data, figures, and text for the AMVOC studies of this chapter were modified from the following publication on which I was a second author:

Stoumpou, V., **Vargas, C. D. M.**, Schade, P. F., Boyd, J. L., Giannakopoulos, T., & Jarvis, E. D. (2022). Analysis of Mouse Vocal Communication (AMVOC): a deep, unsupervised method for rapid detection, analysis and classification of ultrasonic vocalisations. Bioacoustics, 1-31. https://doi.org/10.1080/09524622.2022.2099973

I contributed to the conceptualization of AMVOC, generated a number of recordings for datasets, performed analysis, and co-wrote the paper. The material seen here was used as allowed by the publication's open-source license and approval by the authors.

References

Aitken PG (1981) Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys. Brain and language 13:171-184.

Aitken PG, Wilson WA (1979) Discriminative vocal conditioning in rhesus monkeys: Evidence for volitional control? Brain and Language 8:227-240.

Ali F, Otchy Timothy M, Pehlevan C, Fantana Antoniu L, Burak Y, Ölveczky Bence P (2013) The Basal Ganglia Is Necessary for Learning Spectral, but Not Temporal, Features of Birdsong. Neuron 80:494-506.

Alstermark B, Ogawa J, Isa T (2004) Lack of Monosynaptic Corticomotoneuronal EPSPs in Rats: Disynaptic EPSPs Mediated Via Reticulospinal Neurons and Polysynaptic EPSPs Via Segmental Interneurons. Journal of Neurophysiology 91:1832-1839.

An X, Matho KS, Li Y, Mohan H, Xu XH, Whishaw I, Kepecs A, Huang ZJ (2022) A cortical circuit for orchestrating oromanual food manipulation. bioRxiv:2022.2012. 2003.518964.

Arriaga G, Jarvis ED (2013) Mouse vocal communication system: are ultrasounds learned or innate? Brain Lang 124:96-116.

Arriaga G, Zhou EP, Jarvis ED (2012) Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. PLoS One 7:e46610.

Arriaga G, Macopson JJ, Jarvis ED (2015) Transsynaptic tracing from peripheral targets with pseudorabies virus followed by cholera toxin and biotinylated dextran amines double labeling. JoVE (Journal of Visualized Experiments):e50672.

Auffret M, Ravano VL, Rossi GMC, Hankov N, Petersen MFA, Petersen CCH (2018) Optogenetic stimulation of cortex to map evoked whisker movements in awake head-restrained mice. Neuroscience 368:199-213.

Baldwin MK, Cooke DF, Krubitzer L (2017) Intracortical Microstimulation Maps of Motor, Somatosensory, and Posterior Parietal Cortex in Tree Shrews (Tupaia belangeri) Reveal Complex Movement Representations. Cereb Cortex 27:1439-1456.

Barrett JM, Martin ME, Shepherd GMG (2022) Manipulation-specific cortical activity as mice handle food. Curr Biol 32:4842-4853 e4846.

Bass AH, Gilland EH, Baker R (2008) Evolutionary Origins for Social Vocalization in a Vertebrate Hindbrain–Spinal Compartment. Science 321:417.

Belyk M, Brown S (2017) The origins of the vocal brain in humans. Neuroscience & Biobehavioral Reviews 77:177-193.

Bennett PJG, Maier E, Brecht M (2019) Involvement of rat posterior prelimbic and cingulate area 2 in vocalization control. European Journal of Neuroscience 50:3164-3180.

Berryman JC (1976) Guinea-pig vocalizations: Their structure, causation and function. Zeitschrift für Tierpsychologie 41:80-106.

Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. Nature 495:327-332.

Brecht KF, Westendorff S, Nieder A (2023) Neural correlates of cognitively controlled vocalizations in a corvid songbird. Cell Reports 42.

Broca P (1861) Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). Bulletin et Memoires de la Societe anatomique de Paris 6:330-357.

Brooks RJ, Banks EM (1973) Behavioural biology of the collared lemming [Dicrostonyx groenlandicus (Trail)]: an analysis of acoustic communication. Animal Behaviour Monographs 6:1-83.

Brudzynski SM (2007) Ultrasonic calls of rats as indicator variables of negative or positive states: acetylcholine–dopamine interaction and acoustic coding. Behavioural brain research 182:261-273.

Brudzynski SM (2013) Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. Current opinion in neurobiology 23:310-317.

Burnstein DD, Wolff PC (1967) Vocal conditioning in the guinea pig. Psychonomic Science 8:39-40.

Castellucci GA, Calbick D, McCormick D (2018) The temporal organization of mouse ultrasonic vocalizations. PLOS ONE 13:e0199929.

Cerkevich CM, Rathelot JA, Strick PL (2022) Cortical basis for skilled vocalization. Proc Natl Acad Sci U S A 119:e2122345119.

Chabout J, Sarkar A, Dunson DB, Jarvis ED (2015) Male mice song syntax depends on social contexts and influences female preferences. Front Behav Neurosci 9:76.

Chabout J, Serreau P, Ey E, Bellier L, Aubin T, Bourgeron T, Granon S (2012) Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. PLoS One 7:e29401.

Chabout J, Sarkar A, Patel SR, Radden T, Dunson DB, Fisher SE, Jarvis ED (2016) A Foxp2 Mutation Implicated in Human Speech Deficits Alters Sequencing of Ultrasonic Vocalizations in Adult Male Mice. Front Behav Neurosci 10:197.

Cheney PD, Fetz EE (1985) Comparable patterns of muscle facilitation evoked by individual corticomotoneuronal (CM) cells and by single intracortical microstimuli in primates: evidence for functional groups of CM cells. Journal of Neurophysiology 53:786-804.

Coffey KR, Marx RG, Neumaier JF (2019) DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacology 44:859-868.

Coudé G, Ferrari PF, Roda F, Maranesi M, Borelli E, Veroni V, Monti F, Rozzi S, Fogassi L (2011) Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. PLoS One 6:e26822.

Dave AS, Margoliash D (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. Science 290:812-816.

Deletis V, Fernandez-Conejero I, Ulkatan S, Costantino P (2009) Methodology for intraoperatively eliciting motor evoked potentials in the vocal muscles by electrical stimulation of the corticobulbar tract. Clin Neurophysiol 120:336-341.

Dichter BK, Breshears JD, Leonard MK, Chang EF (2018) The control of vocal pitch in human laryngeal motor cortex. Cell 174:21-31. e29.

Ebbesen CL, Brecht M (2017) Motor cortex - to act or not to act? Nat Rev Neurosci 18:694-705.

Egnor SR, Seagraves KM (2016) The contribution of ultrasonic vocalizations to mouse courtship. Curr Opin Neurobiol 38:1-5.

Ehret G (2005) Infant rodent ultrasounds–a gate to the understanding of sound communication. Behavior genetics 35:19-29.

Espadaler J, Rogić M, Deletis V, Leon A, Quijada C, Conesa G (2012) Representation of cricothyroid muscles at the primary motor cortex (M1) in healthy subjects, mapped by navigated transcranial magnetic stimulation (nTMS). Clin Neurophysiol 123:2205-2211.

Fee MS, Shraiman B, Pesaran B, Mitra PP (1998) The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. Nature 395:67-71.

Fischer J, Hammerschmidt K (2011) Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. Genes Brain Behav 10:17-27.

Fitch WT, Huber L, Bugnyar T (2010) Social cognition and the evolution of language: constructing cognitive phylogenies. Neuron 65:795-814.

Fonseca AHO, Santana GM, Bosque Ortiz GM, Bampi S, Dietrich MO (2021) Analysis of ultrasonic vocalizations from mice using computer vision and machine learning. eLife 10:e59161.

Fritsch GT, Hitzig E (1870) Uber die elektrische Erregbarkeit des Grosshirns. Arch Anat Physiol 37:300-332.

Gan-Or B, London M (2022) Cortical circuits modulate mouse social vocalizations. bioRxiv:2022.2005. 2020.492817.

Gao S-C, Wei Y-C, Wang S-R, Xu X-H (2019) Medial preoptic area modulates courtship ultrasonic vocalization in adult male mice. Neuroscience bulletin 35:697-708.

Gedman GL, Biegler MT, Haase B, Wirthlin ME, Fedrigo O, Pfenning AR, Jarvis ED (2022) Convergent gene expression highlights shared vocal motor microcircuitry in songbirds and humans. bioRxiv:2022.2007. 2001.498177.

Gillam E (2011) An introduction to animal communication. Nature Education 3.

Gioanni Y, Lamarche M (1985) A reappraisal of rat motor cortex organization by intracortical microstimulation. Brain Res 344:49-61.

Goffinet J, Brudner S, Mooney R, Pearson J (2021) Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. eLife 10:e67855.

Gooler DM, O'Neill WE (1987) Topographic representation of vocal frequency demonstrated by microstimulation of anterior cingulate cortex in the echolocating bat, Pteronotus parnelli parnelli. Journal of Comparative Physiology A 161:283-294.

Gordon EM, Chauvin RJ, Van AN, Rajesh A, Nielsen A, Newbold DJ, Lynch CJ, Seider NA, Krimmel SR, Scheidter KM (2023) A somato-cognitive action network alternates with effector regions in motor cortex. Nature:1-9.

Graziano MSA, Taylor CSR, Moore T (2002) Complex Movements Evoked by Microstimulation of Precentral Cortex. Neuron 34:841-851.

Green DB, Shackleton TM, Grimsley JM, Zobay O, Palmer AR, Wallace MN (2018) Communication calls produced by electrical stimulation of four structures in the guinea pig brain. Plos one 13:e0194091.

Grimsley JM, Monaghan JJ, Wenstrup JJ (2011) Development of social vocalizations in mice. PloS one 6:e17460.

Grimsley JM, Sheth S, Vallabh N, Grimsley CA, Bhattal J, Latsko M, Jasnow A, Wenstrup JJ (2016) Contextual Modulation of Vocal Behavior in Mouse: Newly Identified 12 kHz "Mid-Frequency" Vocalization Emitted during Restraint. Front Behav Neurosci 10:38.

Gu ZR, Serradj N, Ueno M, Liang M, Li J, Baccei ML, Martin JH, Yoshida Y (2017a) Skilled Movements Require Non-apoptotic Bax/Bak Pathway-Mediated Corticospinal Circuit Reorganization. Neuron 94:626-641 e624.

Gu ZR et al. (2017b) Control of species-dependent cortico-motoneuronal connections underlying manual dexterity. Science 357:400-404.

Haakansson J, Jiang W, Xue Q, Zheng X, Ding M, Agarwal AA, Elemans CP (2021) Aerodynamics and motor control of ultrasonic vocalizations for social communication in mice and rats. bioRxiv.

Hage SR, Nieder A (2013) Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. Nat Commun 4:2409.

Hake DF, Mabry J (1979) Operant and nonoperant vocal responding in the mynah: Complex schedule control and deprivation-induced responding. Journal of the Experimental Analysis of Behavior 32:305-321.

Halley AC, Baldwin MKL, Cooke DF, Englund M, Krubitzer L (2020) Distributed Motor Control of Limb Movements in Rat Motor and Somatosensory Cortex: The Sensorimotor Amalgam Revisited. Cereb Cortex 30:6296-6312.

Hammerschmidt K, Radyushkin K, Ehrenreich H, Fischer J (2009) Female mice respond to male ultrasonic 'songs' with approach behaviour. Biol Lett 5:589-592.

Hammerschmidt K, Whelan G, Eichele G, Fischer J (2015) Mice lacking the cerebral cortex develop normal song: insights into the foundations of vocal learning. Sci Rep 5:8808.

Harrison TC, Ayling OGS, Murphy TH (2012) Distinct Cortical Circuit Mechanisms for Complex Forelimb Movement and Motor Map Topography. Neuron 74:397-409.

Hast MH, Fischer JM, Wetzel AB, Thompson VE (1974) Cortical motor representation of the laryngeal muscles in Macaca mulatta. Brain Res 73:229-240.

Herculano-Houzel S, Kaas JH, de Oliveira-Souza R (2016) Corticalization of motor control in humans is a consequence of brain scaling in primate evolution. J Comp Neurol 524:448-455.

Hertz S, Weiner B, Perets N, London M (2020) Temporal structure of mouse courtship vocalizations facilitates syllable labeling. Communications Biology 3:333.

Hira R, Terada S-I, Kondo M, Matsuzaki M (2015) Distinct Functional Modules for Discrete and Rhythmic Forelimb Movements in the Mouse Motor Cortex. The Journal of Neuroscience 35:13311.

Holy Timothy E, Guo ZV (2005) Ultrasonic songs of male mice. PLoS Biol 3.

Hudson HM, Griffin DM, Belhaj-Saif A, Cheney PD (2015) Properties of primary motor cortex output to hindlimb muscles in the macaque monkey. J Neurophysiol 113:937-949.

Hunsperger RW, Bucher VM (1967) Affective Behaviour produced by Electrical Stimulation in the Forebrain and Brain Stem of the Cat. In: Progress in Brain Research (Adey WR, Tokizane T, eds), pp 103-127: Elsevier.

Isogai F, Kato T, Fujimoto M, Toi S, Oka A, Adachi T, Maeda Y, Morimoto T, Yoshida A, Masuda Y (2012) Cortical area inducing chewing-like rhythmical jaw movements and its connections with thalamic nuclei in guinea pigs. Neurosci Res 74:239-247.

Ivanenko A, Watkins P, van Gerven MAJ, Hammerschmidt K, Englitz B (2020) Classifying sex and strain from mouse ultrasonic vocalizations using deep learning. PLoS Comput Biol 16:e1007918.

Iwatsubo T, Kuzuhara S, Kanemitsu A, Shimada H, Toyokura Y (1990) Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. Neurology 40:309-309.

Jarvis ED (2004) Learned Birdsong and the Neurobiology of Human Language. Annals of the New York Academy of Scienes.

Jarvis ED (2019) Evolution of vocal learning and spoken language. Science 366:50-54.

Johnson AM, Doll EJ, Grant LM, Ringel L, Shier JN, Ciucci MR (2011) Targeted training of ultrasonic vocalizations in aged and Parkinsonian rats. JoVE (Journal of Visualized Experiments):e2835.

Jourjine N, Woolfolk ML, Sanguinetti-Scheck JI, Sabatini JE, McFadden S, Lindholm AK, Hoekstra HE (2023) Two pup vocalization types are genetically and functionally separable in deer mice. Current Biology.

Jürgens U (2002) Neural pathways underlying vocal control. Neurosci Biobehav R.

Jürgens U (2009) The neural control of vocalization in mammals: a review. J Voice 23:1-10.

Kane GA, Lopes G, Saunders JL, Mathis A, Mathis MW (2020) Real-time, low-latency closed-loop feedback using markerless posture tracking. Elife 9:e61909.

Karten HJ (2015) Vertebrate brains and evolutionary connectomics: on the origins of the mammalian 'neocortex'. Philos Trans R Soc Lond B Biol Sci 370.

Kay EH, Hoekstra HE (2008) Rodents. Current Biology 18:R406-R410.

Kikusui T, Nakanishi K, Nakagawa R, Nagasawa M, Mogi K, Okanoya K (2011) Cross Fostering Experiments Suggest That Mice Songs Are Innate. PLOS ONE 6:e17721.

Kirikae I, Hirose H, Kawamura S, Sawashima M, Kobayashi T (1962) An experimental study of central motor innervation of the laryngeal muscles in the cat. The Annals of otology, rhinology, and laryngology 71:222-241.

Kirzinger A, Jürgens U (1982) Cortical lesion effects and vocalization in the squirrel monkey. Brain Res 233:299-315.

Kobayasi KI, Riquimaroux H (2012) Classification of vocalizations in the Mongolian gerbil, Meriones unguiculatus. The Journal of the Acoustical Society of America 131:1622-1631.

Komiyama T, Sato TR, O'Connor DH, Zhang YX, Huber D, Hooks BM, Gabitto M, Svoboda K (2010) Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. Nature 464:1182-1186.

Kuypers HGJM (1958) Corticobulbar connexions to the pons and lower brain-stem in man: an anatomical study. Brain 81:364-388.

Lal H (1967) Operant control of vocal responding in rats. Psychonomic Science 8:35-36.

Lane H (1960) Control of vocal responding in chickens. Science 132:37-38.

Lattenkamp EZ, Vernes SC, Wiegrebe L (2018) Volitional control of social vocalisations and vocal usage learning in bats. The Journal of Experimental Biology.

Lattenkamp EZ, Vernes SC, Wiegrebe L (2020) Vocal production learning in the pale spear-nosed bat, Phyllostomus discolor. Biology Letters 16:20190928.

Lemon RN (2008) Descending pathways in motor control. Annu Rev Neurosci 31:195-218.

Leyton AS, Sherrington CS (1917) Observations on the excitable cortex of the chimpanzee, orang-utan, and gorilla. Quarterly Journal of Experimental Physiology: Translation and Integration 11:135-222.

Li CX, Waters RS (1991) Organization of the mouse motor cortex studied by retrograde tracing and intracortical microstimulation (ICMS) mapping. Can J Neurol Sci 18:28-38.

Liang F, Rouiller EM, Wiesendanger M (1993) Modulation of Sustained Electromyographic Activity by Single Intracortical Microstimuli: Comparison of Two Forelimb Motor Cortical Areas of the Rat. Somatosensory & Motor Research 10:51-61.

Madsen PT, Siebert U, Elemans CP (2023) Toothed whales use distinct vocal registers for echolocation and communication. Science 379:928-933.

Maggio JC, Whitney, G. (1985) Ultrasonic vocalizing by adult female mice (Mus musculus). J Comp Psychol 99:420-436.

Mahrt E, Agarwal A, Perkel D, Portfors C, Elemans CP (2016) Mice produce ultrasonic vocalizations by intra-laryngeal planar impinging jets. Curr Biol 26:R880-R881.

Manabe K, Dooling RJ (1997) *Control of vocal production in budgerigars (Melopsittacus undulatus): Selective reinforcement, call differentiation, and stimulus control*. Behav Process 41:117-132.

Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. Journal of Neuroscience 3:1039-1057.

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M (2018) DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature neuroscience 21:1281-1289.

Matikainen-Ankney BA et al. (2021) An open-source device for measuring food intake and operant behavior in rodent home-cages. eLife 10:e66173.

Mercer Lindsay N, Knutsen PM, Lozada AF, Gibbs D, Karten HJ, Kleinfeld D (2019) Orofacial Movements Involve Parallel Corticobulbar Projections from Motor Cortex to Trigeminal Premotor Nuclei. Neuron 104:765-780 e763.

Milojevic B, Hast MH (1964) CORTICAL MOTOR CENTERS OF THE LARYNGEAL MUSCLES IN THE CAT AND DOG. The Annals of otology, rhinology, and laryngology 73:979-988.

Molliver ME (1963) Operant Contorl of Vocal Behavior in the Cat. Journal of the Experimental Analysis of Behavior 6:197-202.

Murabe N, Mori T, Fukuda S, Isoo N, Ohno T, Mizukami H, Ozawa K, Yoshimura Y, Sakurai M (2018) Higher primate-like direct corticomotoneuronal connections are transiently formed in a juvenile subprimate mammal. Sci Rep 8:16536.

Neafsey EJ, Bold EL, Haas G, Hurley-Gius KM, Quirk G, Sievert CF, Terreberry RR (1986) The organization of the rat motor cortex: A microstimulation mapping study. Brain Research Reviews 11:77-96.

Neunuebel JP, Taylor AL, Arthur BJ, Egnor SE (2015) Female mice ultrasonically interact with males during courtship displays. Elife 4.

Nguyen KP, O'Neal TJ, Bolonduro OA, White E, Kravitz AV (2016) Feeding Experimentation Device (FED): A flexible open-source device for measuring feeding behavior. Journal of neuroscience methods 267:108-114.

Nieder A, Mooney R (2020) The neurobiology of innate, volitional and learned vocalizations in mammals and birds. Philosophical Transactions of the Royal Society B: Biological Sciences 375:20190054.

Niemczura AC, Grimsley JM, Kim C, Alkhawaga A, Poth A, Carvalho A, Wenstrup JJ (2020) Physiological and Behavioral Responses to Vocalization Playback in Mice. Frontiers in Behavioral Neuroscience 14.

Nottebohm F, Stokes TM, Leonard CM (1976) Central control of song in the canary, Serinus canarius. Journal of Comparative Neurology 165:457-486.

Nyby J, Wysocki CJ, Whitney G, Dizinno G, Schneider J (1979) Elicitation of male mouse (Mus musculus) ultrasonic vocalizations: I. Urinary cues. In, pp 957-975. US: American Psychological Association.

Okobi D (2016) A cortical locus modulates vocal motor sequences in Alston's singing mouse (Scotinomys teguina). In: Basic Medical Sciences, p 136. ProQuest: New York University.

Okobi DE, Banerjee A, Matheson AMM, Phelps SM, Long MA (2019) Motor cortical control of vocal interaction in neotropical singing mice. Science 363:983.

Olkowicz S, Kocourek M, Lučan RK, Porteš M, Fitch WT, Herculano-Houzel S, Němec P (2016) Birds have primate-like numbers of neurons in the forebrain. Proceedings of the National Academy of Sciences 113:7255-7260.

Pachitariu M, Stringer C, Schröder S, Dipoppa M, Rossi LF, Carandini M, Harris KD (2016) Suite2p: beyond 10,000 neurons with standard two-photon microscopy. BioRxiv:061507.

Pasch B, Bolker BM, Phelps SM (2013) Interspecific dominance via vocal interactions mediates altitudinal zonation in neotropical singing mice. The American Naturalist 182:E161-E173.

Penfield W, Rasmussen T (1949) Vocalization and arrest of speech. Archives of Neurology & Psychiatry 61:21-27.

Penfield W, Rasmussen T (1950) The cerebral cortex of man; a clinical study of localization of function.

Petkov CI, Jarvis E (2012) Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Frontiers in evolutionary neuroscience 4:12.

Petreanu L, Huber D, Sobczyk A, Svoboda K (2007) Channelrhodopsin-2–assisted circuit mapping of long-range callosal projections. Nature neuroscience 10:663-668.

Pfenning AR et al. (2014) Convergent transcriptional specializations in the brains of humans and song-learning birds. Science 346:1333-+.

Pierce Jr JD, Sawrey DK, Dewsbury DA (1989) A comparative study of rodent ultrasonic vocalizations during copulation. Behavioral and neural biology 51:211-221.

Portfors CV, Perkel DJ (2014) The role of ultrasonic vocalizations in mouse communication. Current opinion in neurobiology 28:115-120.

Rathelot JA, Strick PL (2006) Muscle representation in the macaque motor cortex: an anatomical perspective. Proc Natl Acad Sci U S A 103:8257-8262.

Reidenberg JS (2017) Terrestrial, semiaquatic, and fully aquatic mammal sound production mechanisms. Acoust Today 13:35-43.

Rendon NM, Keesom SM, Amadi C, Hurley LM, Demas GE (2015) Vocalizations convey sex, seasonal phenotype, and aggression in a seasonal mammal. Physiology & Behavior 152:143-150.

Richards DG, Wolz JP, Herman LM (1984) Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, Tursiops truncatus. Journal of Comparative Psychology 98:10.

Riede T, Borgard HL, Pasch B (2017) Laryngeal airway reconstruction indicates that rodent ultrasonic vocalizations are produced by an edge-tone mechanism. R Soc Open Sci 4:170976.

Rödel RM, Olthoff A, Tergau F, Simonyan K, Kraemer D, Markus H, Kruse E (2004) Human cortical motor representation of the larynx as assessed by transcranial magnetic stimulation (TMS). The Laryngoscope 114:918-922.

Sainburg T, Thielk M, Gentner TQ (2020) Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. PLoS computational biology 16:e1008228.

Salzinger K, Waller MB (1962) The Operant Control of Vocalization in the Dog. Journal of the Experimental Analysis of Behavior 5:383-389.

Scattoni ML, Gandhy SU, Ricceri L, Crawley JN (2008) Unusual repertoire of vocalizations in the BTBR T+ tf/J mouse model of autism. PloS one 3:e3067.

Schneeberger M, Parolari L, Das Banerjee T, Bhave V, Wang P, Patel B, Topilko T, Wu Z, Choi CHJ, Yu X, Pellegrino K, Engel EA, Cohen P, Renier N, Friedman JM, Nectow AR (2019) Regulation of Energy Expenditure by Brainstem GABA Neurons. Cell 178:672-685.e612.

Schweihoff JF, Loshakov M, Pavlova I, Kück L, Ewell LA, Schwarz MK (2021) DeepLabStream enables closed-loop behavioral experiments using deep learning-based markerless, real-time posture detection. Communications biology 4:130.

Seyfarth RM, Cheney DL, Marler P (1980) Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science 210:801-803.

Shapiro AD, Slater PJ, Janik VM (2004) Call usage learning in gray seals (Halichoerus grypus). Journal of Comparative Psychology 118:447.

Silva AB, Liu JR, Zhao L, Levy DF, Scott TL, Chang EF (2022) A Neurosurgical Functional Dissection of the Middle Precentral Gyrus during Speech Production. J Neurosci 42:8416-8426.

Simola N, Granon S (2019) Ultrasonic vocalizations as a tool in studying emotional states in rodent models of social behavior and brain disease. Neuropharmacology 159:107420.

Simonyan K (2014) The laryngeal motor cortex: its organization and connectivity. Curr Opin Neurobiol 28:15-21.

Simonyan K, Jürgens U (2002) Cortico-cortical projections of the motorcortical larynx area in the rhesus monkey. Brain Res 949:23-31.

Simonyan K, Jürgens U (2003) Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. Brain Res 974:43-59.

Simonyan K, Horwitz B (2011) Laryngeal motor cortex and control of speech in humans. Neuroscientist 17:197-208.

Sreenivasan V, Karmakar K, Rijli FM, Petersen CCH (2015) Parallel pathways from motor and somatosensory cortex for controlling whisker movements in mice. The European journal of neuroscience 41:354-367.

Steinfath E, Palacios-Muñoz A, Rottschäfer JR, Yuezak D, Clemens J (2021) Fast and accurate annotation of acoustic signals with deep neural networks. eLife 10:e68837.

Stoeger AS, Baotic A (2021) Operant control and call usage learning in African elephants. Philosophical Transactions of the Royal Society B 376:20200254.

Stoumpou V, Vargas CDM, Schade PF, Boyd JL, Giannakopoulos T, Jarvis ED (2022) Analysis of Mouse Vocal Communication (AMVOC): a deep, unsupervised method for rapid detection, analysis and classification of ultrasonic vocalisations. Bioacoustics:1-31.

Tennant KA, Adkins DL, Donlan NA, Asay AL, Thomas N, Kleim JA, Jones TA (2011) The organization of the forelimb representation of the C57BL/6 mouse motor cortex as defined by intracortical microstimulation and cytoarchitecture. Cereb Cortex 21:865-876.

Tschida K, Michael V, Han B-X, Zhao S, Sakurai K, Mooney R, Wang F (2018) Identification of midbrain neurons essential for vocal communication. bioRxiv.

Tschida K, Michael V, Takato J, Han B-X, Zhao S, Sakurai K, Mooney R, Wang F (2019) A Specialized Neural Circuit Gates Social Vocalizations in the Mouse. Neuron.

Van Segbroeck M, Knoll AT, Levitt P, Narayanan S (2017) MUPET-Mouse Ultrasonic Profile ExTraction: A Signal Processing Tool for Rapid and Unsupervised Analysis of Ultrasonic Vocalizations. Neuron 94:465-485 e465.

von Ziegler L, Sturman O, Bohacek J (2021) Big behavior: challenges and opportunities in a new era of deep behavior profiling. Neuropsychopharmacology 46:33-44.

Wang P, Loh KH, Wu M, Morgan DA, Schneeberger M, Yu X, Chi J, Kosse C, Kim D, Rahmouni K, Cohen P, Friedman J (2020) A leptin–BDNF pathway regulating sympathetic innervation of adipose tissue. Nature 583:839-844.

Wang R, Chen CC, Hara E, Rivas MV, Roulhac PL, Howard JT, Chakraborty M, Audet JN, Jarvis ED (2015) Convergent Differential Regulation of SLIT-ROBO Axon Guidance Genes in the Brains of Vocal Learners. Journal of Comparative Neurology 523:892-906.

Wang X, Liu Y, Li X, Zhang Z, Yang H, Zhang Y, Williams PR, Alwahab NSA, Kapur K, Yu B, Zhang Y, Chen M, Ding H, Gerfen CR, Wang KH, He Z (2017) Deconstruction of Corticospinal Circuits for Goal-Directed Motor Skills. Cell 171:440-455 e414.

Wernicke C (1874) Der aphasische symptomenkomplex. In: Der aphasische Symptomencomplex, pp 1-70: Springer.

Whitney G, Coble JR, Stockton MD, Tilson EF (1973) Ultrasonic emissions: do they facilitate courtship of mice. Journal of comparative and physiological psychology 84:445-452.

Wild JM (1993) Descending projections of the songbird nucleus robustus archistriatalis. Journal of Comparative Neurology 338:225-241.

Willmore L, Cameron C, Yang J, Witten IB, Falkner AL (2022) Behavioural and dopaminergic signatures of resilience. Nature:1-9.

Wirthlin ME, Schmid TA, Elie JE, Zhang X, Shvareva VA, Rakuljic A, Ji MB, Bhat NS, Kaplow IM, Schäffer DE, Lawler AJ, Annaldasula S, Lim B, Azim E, Consortium Z, Meyer WK, Yartsev MM, Pfenning AR (2022) Vocal learning-associated convergent evolution in mammalian proteins and regulatory elements. bioRxiv:2022.2012.2017.520895.

Xiao L, Chattree G, Oscos FG, Cao M, Wanat MJ, Roberts TF (2018) A Basal Ganglia Circuit Sufficient to Guide Birdsong Learning. Neuron.

Yamawaki N, Raineri Tapies MG, Stults A, Smith GA, Shepherd GM (2021) Circuit organization of the excitatory sensorimotor loop through hand/forelimb S1 and M1. Elife 10.

Yang HW, Lemon RN (2003) An electron microscopic examination of the corticospinal projection to the cervical spinal cord in the rat: lack of evidence for cortico-motoneuronal synapses. Exp Brain Res 149:458-469.

Yurlova DD, Volodin IA, Ilchenko OG, Volodina EV (2020) Rapid development of mature vocal patterns of ultrasonic calls in a fast-growing rodent, the yellow steppe lemming (Eolagurus luteus). Plos one 15:e0228892.

Zhao X, Ziobro P, Pranic NM, Chu S, Rabinovich S, Chan W, Zhao J, Kornbrek C, He Z, Tschida K (2021) Sex- and context-dependent effects of acute isolation on vocal and non-vocal social behaviors in mice. bioRxiv.

Zheng D-J, Okobi Jr DE, Shu R, Agrawal R, Smith SK, Long MA, Phelps SM (2022) Mapping the vocal circuitry of Alston's singing mouse with pseudorabies virus. Journal of Comparative Neurology 530:2075-2099.

Zingg B, Peng B, Huang J, Tao HW, Zhang LI (2020) Synaptic Specificity and Application of Anterograde Transsynaptic AAV for Probing Neural Circuitry. The Journal of Neuroscience 40:3250-3267.

Zingg B, Chou XL, Zhang ZG, Mesik L, Liang F, Tao HW, Zhang LI (2017) AAV-Mediated Anterograde Transsynaptic Tagging: Mapping Corticocollicular Input-Defined Neural Pathways for Defense Behaviors. Neuron 93:33-47.