Title
An Auditory Memory System for Individual Vocal Recognition in the Zebra Finch

Author
Yu, Kevin

Publication Date
2021

Peer reviewed|Thesis/dissertation

An Auditory Memory System for Individual Vocal Recognition in the Zebra Finch

by

Kevin Yu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Frederic Theunissen, Chair
Professor Bruno Olshausen
Professor David Foster
Professor Michael Yartsev

Fall 2021

An Auditory Memory System for Individual Vocal Recognition in the Zebra Finch

Abstract

An Auditory Memory System for Individual Vocal Recognition in the Zebra Finch

by

Kevin Yu

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Frederic Theunissen, Chair

The zebra finch is a social songbird that lives in large groups and produces vocal communication calls to facilitate social interactions. A subset of these calls can be used for individual recognition, via distinct acoustic features that are stereotyped within a bird but individualized across birds. Given their large natural group sizes and the ethological importance of individual recognition, one might expect that zebra finches would have the capacity to recognize the calls of a large number of conspecifics. In this thesis, I describe a set of neuroethological experiments to test the memory of zebra finches for individual conspecifics by their vocalizations. We hypothesized that the caudal nidopallium (NCM), a higher-order auditory region of the avian brain analogous to mammalian auditory association cortex, is involved in the learning and retention of these auditory memories. Using an operant task in which birds were trained to associate the calls of some individuals with food reward, zebra finches were found to have a large capacity for recognizing individuals by their calls and song, and that those associations could be learned with just a few training examples and persist for at least a month without reinforcement. Furthermore, lesions to NCM eliminated previously formed associations but did not prevent re-learning or learning of novel stimuli, in contrast with lesions to vocal pre-motor pathways which had no effect on the recognition ability of the birds. Finally, using the spiking activity from single neurons across the cortical-like auditory regions of the brain, we found that familiar and task-relevant vocalizations elicited more reliable neural responses, with higher information capacity, than in response to unfamiliar and less behaviorally relevant calls.

To Rocco

Contents

List of Figures

Acknowledgments

Many friends and colleagues have helped and supported me throughout the last several years to make this work possible.

I would first like to thank Frederic Theunissen, for his unwavering support, optimism, and encouragement over the years; for his uncanny ability to convince me that I knew what I was doing, even when I did not know what I was doing; and for creating a lab environment that cherishes and respects animal behavior and the natural world.

I would also like to thank Bill Wood, who took me under his wing when I first started in the lab; who worked together with me on performing all the experiments presented here; and whose ideas and expertise in experimental techniques made these projects possible.

I am also grateful for the friends, colleagues, and collaborators who have contributed to my scientific development over the years, of whom there are several. Thank you to Julie Elie, whose work on the zebra finch vocal repertoire is extensively referenced herein; who designed the original operant task upon which the behavioral experiments are based; and who has provided valuable scientific advice and feedback. I would also like to thank my labmates Leah Johnston, Pepe Alcami, Logan Thomas, Lily Gong, Izabela Rice, Apurva Prasad, Coral Chen, and Raelyn Vu for their contributions to data collection, ideas, and/or analysis in these and other projects that I have spent my time on.

I would also like to thank Will Liberti, for wonderful but occasionally endless conversations about science and life, and Zuzanna Balewski, for providing emotional support, companionship, desk space, and occasional advice on statistics.

Finally, I would like to thank my thesis committee members, Bruno Olshausen, Michael Yartsev, and David Foster, for their scientific advice and guidance; and the HWNI faculty and administrative staff—in particular, Candace Groskreutz, who was a beacon of light for graduate students like myself lost in the darkness of bureaucracy.

Go Bears.

Chapter 1

Introduction

Many animals, including humans and songbirds, share the natural ability to recognize the other members of one's species by their vocalizations. This general ability is known as individual vocal recognition (Carlson et al., 2020; Tibbetts & Dale, 2007). The neural basis for this skill is an auditory memory system capable of learning and recognizing the acoustic features that distinguish one individual from the next. Individual recognition can be accomplished through vocal communication calls that convey identifying attributes of the caller, such as sex (Blumstein & Munos, 2005; D'Amelio, Klumb, et al., 2017), age (Akçay et al., 2016; Blumstein & Munos, 2005), or size (Davies & Halliday, 1978; Favaro et al., 2017), or the specific identity of the vocalizer via distinct acoustic features (Blumstein & Munos, 2005; Favaro et al., 2017; Hare, 1998).

The skill of individual vocal recognition is particularly important in social animals, for whom the vocal recognition of a conspecific may be necessary for maintaining dominance hierarchies, familial relationships, and mate preferences (Vignal et al., 2008). Maintaining these social relationships require these memories to persist over weeks, months, and years. Long term vocal recognition memory has been observed in migratory birds (Godard, 1991), ravens (Boeckle & Bugnyar, 2012), mammals (Insley, 2000; McComb et al., 2000), and humans (Aglieri et al., 2017). In the brain, hierarchical circuits map natural sounds, decomposed into acoustic feature spaces described by modulations in time and frequency (Theunissen et al., 2000; Woolley et al., 2005) to more abstract categories (Russ et al., 2008) and their associated meaning. To facilitate learning and storage of auditory memories, these circuits are shaped by the sensory and cognitive experience of the animal. Presented here is a set of experiments in a model songbird, the zebra finch, aimed at linking auditory memory behavior to the animal's neurobiology. In these experiments, we demonstrate the zebra finch's impressive memory capacity for conspecific vocalizations and identify neural correlates of its auditory memory system.

The songbird as a model for the neural basis of individual vocal recognition

The songbird has been the subject of scientific study for several decades due to its amazing vocal learning capabilities. As a result, the behavior and neurophysiology of the songbird has been well described and songbird species have become excellent models for understanding the neural basis of vocal communication, from the perspective of both motor production (Hahnloser et al., 2002; Nottebohm, 2005) and perception (Gentner, 2004; Woolley et al., 2005).

The zebra finch is a social species of songbird known to live in colonies of over 100 individuals (Zann, 1996), travel and forage with smaller groups of other individuals (McCowan et al., 2015), and form mating pairs that can last for life (Adkins-Regan, 2002). Recognition of others is particularly important in these social networks; a fledgling may need to recognize its parents (Jouventin et al., 1999), parents need to recognize their offspring and mate, and unpaired birds may need to recognize other local singles in their area. Zebra finches communicate with a diverse repertoire of at least 12 distinct call types each with specific behavioral contexts in which they are used (Zann, 1996). Among these types, the song and distance call (DC) are particularly interesting in the context of individual vocal recognition. These calls are stereotyped, individualized (Elie & Theunissen, 2018), and socially affiliative; although in some species song may be used in territorial aggression and competition over mates (Krebs et al., 1978; McGregor et al., 1993). Considering the ethological importance of recognition and the group sizes found in nature, zebra finches may be capable of recognizing a large number of individuals by their song and DC. However, the memory capacity of zebra finches for conspecific vocalizers is hitherto unknown. Quantifying this memory capacity is the subject of the experiments described in Chapter 2.

Higher-order auditory regions of the avian pallium involved in auditory memory

Where in the avian brain are these auditory memories for conspecific vocalizers? In humans, specific regions of the auditory association cortex have been shown to be involved in speaker recognition tasks and may represent voice identity (Andics et al., 2010). Analogous regions in the avian brain are thought to be the interconnected regions of the caudal nidopallium (NCM) and the caudal mesopallium (CM) (Bolhuis & Gahr, 2006; Bolhuis et al., 2010). NCM has been suggested as a potential repository for learned vocalizations, with evidence from immediate early gene expression studies (Bolhuis et al., 2001), lesions (Canopoli et al., 2014; Gobes & Bolhuis, 2007), and electrophysiology (Chew et al., 1996; Phan et al., 2006; Thompson & Gentner, 2010; Yanagihara & Yazaki-Sugiyama, 2016).

Another possibility is that long-term memory for auditory stimuli are associated with representations in vocal motor systems for production of those sounds (Massaro & Chen, 2008; Schulze et al., 2012; Williams & Nottebohm, 1985). This "motor theory of speech perception" is supported by the discovery of auditory, tutor song selective neurons in the

anterior forebrain pathway (AFP) necessary for song learning (Doupe & Konishi, 1991; Mooney, 2014). These motor neuron responsive to auditory stimuli have been hypothesized to be part of a circuit that compares incoming sounds to the stored auditory memory of the tutor template. Are motor circuits required for the storage and retrieval of auditory memories for conspecific vocalizations? In Chapter 3, we use neurotoxic lesions to test the role of NCM and HVC in birds' ability to recall previously formed auditory memories and their ability to acquire new ones.

Neural encoding of learned vocalizations

A central question in auditory neuroscience is how auditory objects are represented in neural circuits. Much is known about the hierarchical organization of the auditory system (Meliza et al., 2010; Russ et al., 2008; Woolley et al., 2005), in which lower brain regions such as Field L (analogous to primary auditory cortex in mammals) encode acoustic features such as the spectro-temporal modulations found in natural sounds (Theunissen et al., 2000; Woolley et al., 2005), while neurons in higher-order regions such as NCM and CM are better represented by complex receptive fields (Kaardal et al., 2017) and feature sparser responses (Meliza & Margoliash, 2012). Circuits in these higher-order brain regions have been shown to undergo neuroplastity during learning, with increased selectivity to learned auditory objects or object categories during both song learning (Thompson & Gentner, 2010; Yanagihara & Yazaki-Sugiyama, 2016, 2019) and in operant tasks (Gentner & Margoliash, 2003; Meliza & Margoliash, 2012). However, selectivity for specific auditory objects is only one way that circuits might change to better encode information about learned stimuli. At the single neuron level, it has been shown that the firing rates of single units can be used to extract more information about learned song motifs than novel motifs (Jeanne et al., 2011), and at the population level, modulations in noise correlation structure have been observed in response to learned versus untrained stimuli (Jeanne et al., 2013; Theilman et al., 2021). In Chapter 4, we quantify the information in the spiking activity of single units and ensembles, recorded across primary and secondary auditory brain regions of the anesthetized zebra finch, in response to playbacks of both unfamiliar and learned vocalizations.

Outline

Chapter 2, entitled *High-capacity auditory memory for vocal communication in a social songbird*, was previously published as (Yu et al., 2020) and included here with minor modifications. It describes the results of an experiment to test the memory capacity of zebra finches for several individual vocalizers. The experimental design expands on an operant task previously used in the lab to test auditory categorization (Elie & Theunissen, 2018) by gradually increasing the number of vocalizers, allowing us to test for a large number of individual vocalizers. We analyzed if birds could respond correctly to each individual vocalizer presented and how quickly they learned to do so, and found that they could learn

vocalizers quickly, retain those memories for at least a month, and recognize a large number of individuals in this way, perhaps more than the 54 vocalizers we could test.

Chapter 3, entitled *Lesions to NCM impair individual vocal recognition memory in the zebra finch*, reports the effect of bilateral neurotoxic lesions to two brain regions, NCM and HVC, on learned auditory memories of conspecific vocalizers. We found that lesions to NCM destroy learned auditory associations, those associations can quickly be re-learned, and that memory capacity may be decreased. In contrast, lesions to HVC have little to no effect on either stored auditory associations for individual vocalizers or the ability to learn to recognize new vocalizers. This provides further evidence for NCM as a site of auditory memory for conspecific vocalizations.

Chapter 4, entitled *Neural encoding of learned communication calls in the anesthetized zebra finch* presents an analysis of single neuron responses across auditory cortical-like brain areas in the anesthetized zebra finch. This section includes a description of the custom data processing pipeline we developed for extracellular, multi-electrode array recordings, including a detailed description of a semi-automated spike sorting algorithm used to identify and isolate neurons with non-stationary spike shapes in long recordings. Neurons identified this way were used to quantify the information content of the neural response. We found that the information needed to identify individual vocalizers was distributed across neurons in all auditory areas, and that past experience with familiar and task-relevant vocalizations (learned in the memory capacity experiments of Chapter 2) influenced the reliability of the neural response and thus the information capacity of the neurons.

Chapter 2

High-capacity auditory memory for vocal communication in a social songbird

Yu, Kevin, Wood, William E., & Theunissen, Frederic

2.1 Abstract

Effective vocal communication often requires the listener to recognize the identity of a vocalizer, and this recognition is dependent on the listener's ability to form auditory memories. We tested the memory capacity of a social songbird, the zebra finch, for vocalizer identities using conditioning experiments and found that male and female zebra finches can remember a large number of vocalizers (mean, 42) based solely on the individual signatures found in their songs and distance calls. These memories were formed within a few trials, were generalized to previously unheard renditions, and were maintained for up to a month. A fast and high-capacity auditory memory for vocalizer identity has not been demonstrated previously in any nonhuman animals and is an important component of vocal communication in social species.

2.2 Introduction

In species with large vocal repertoires and sophisticated social behaviors, learning to interpret vocal signals requires a large capacity memory system. For example, a high-capacity memory for defining sounds of words is needed to process human language semantics (Brysbaert et al., 2016). Similarly, humans can recognize a large number of individuals based on the sound of their voices as well as linguistic idiosyncrasies (Aglieri et al., 2017; Perrachione et al., 2011) and must therefore have formed memories for those unique acoustic features (Belin et al., 2004). Young humans form these auditory memories rapidly and retain them

for long periods in a process called fast mapping (Markson & Bloom, 1997)—the formation of these auditory memories with few exposures and their maintenance for long periods of time. While the complexity of animal vocal communication pales in comparison with human spoken language (Hauser et al., 2002), auditory memory also plays an important role in the vocal communication of nonhuman social species. In particular, songbirds demonstrate aptitude in several communicative tasks that require auditory memories for vocal signals (Elie & Theunissen, 2020). For example, young male songbirds imitate the song of a tutor that they have stored as an auditory memory (Sakata et al., 2020); some birds can learn the alarm calls from other species to avoid dangerous situations (Potvin et al., 2018) and can even mimic alarm calls of mammals for deceit purposes (Flower et al., 2014); and territorial birds learn to recognize their neighbors based on their voice, enabling them to identify and react to unfamiliar intruders at the boundaries of their local territory (Kroodsma, 1976).

Individual recognition based on voice also plays a central role for creating and maintaining bonds in social songbird species such as the zebra finch. In the wild, zebra finches are a gregarious and nomadic species, living and traveling in multifamily colonies sometimes comprising more than 100 individuals (Zann, 1996). Zebra finches also mate for life, making strong pair bonds with their partners that are maintained through vocal communication (Elie et al., 2010; Zann, 1996). Laboratory studies have shown that their songs have a strong individual signature and can be used to recognize one's mate (Miller, 1979a), father (Miller, 1979b), and peers (Honarmand et al., 2015). Individual recognition by vocalizations is not restricted to song; distance calls (DCs) (Vignal et al., 2004), begging calls (Ligout et al., 2016), and soft contact calls (D'Amelio, Klumb, et al., 2017) are also used for individual recognition in juveniles and adults. In previous work, we have shown that all the call types of the zebra finch repertoire are individualized by distinct individual acoustical cues for each call type and that zebra finches could use those cues to discriminate between two vocalizers, irrespective of the call type (Elie & Theunissen, 2018). Given that zebra finches live in large social groups and that vocal communication plays a key role in the creation and maintenance of their social networks, we hypothesized that they might have a high-capacity auditory memory for the acoustic individual signatures found in their calls. We were also interested in investigating whether zebra finches are capable of fast mapping. To answer these questions, we tested the ability of zebra finches to learn to discriminate the identities of unseen vocalizers based on either their song or DC; the song and the DC are the two loud call types in the zebra finch repertoire with strong individual signatures that birds use to recognize and localize each other often without visual contact (Elie & Theunissen, 2016, 2018).

2.3 Results

We trained male and female zebra finches to recognize several conspecifics by their songs ($n = 19$) or DC ($n = 19$) using a modified go–no go task with food reward (Figure 2.1A). To test the birds on a large number of vocalizers, we used a 5-day learning ladder procedure

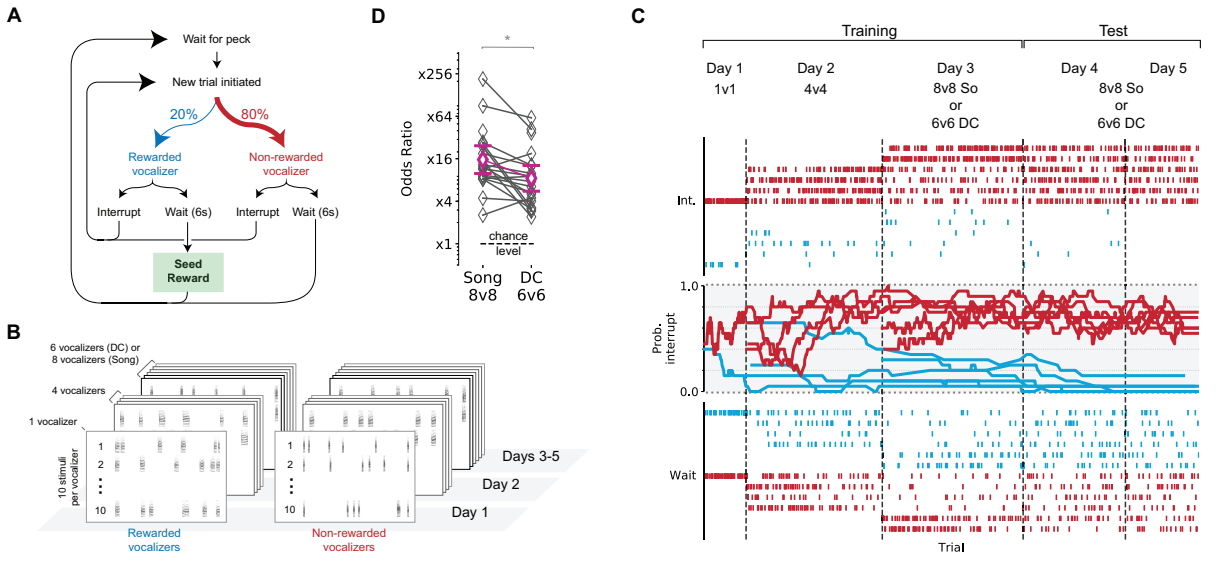


Figure 2.1: **Learning ladder for assessing auditory memory capacity.** (**A**) The structure of a single trial. Subjects initiate a trial by pecking a key. A randomly chosen 6-s stimulus file is then played (20% of trials are rewarded, and 80% of trials are nonrewarded). If the stimulus is interrupted by another peck on the same key before the 6-s playback is completed, then a new trial is immediately initiated. If the stimulus is not interrupted and the stimulus is in the rewarded group, then the subject receives 12 s of seed access from a mechanical food hopper. (**B**) The learning ladder procedure gradually introduces new rewarded and nonrewarded vocalizers to the stimulus set each day. Ten stimuli are used for each vocalizer and vocalization type. Each stimulus is, in turn, composed of random sequences of renditions of DCs or songs sampled from our repertoire library for that vocalizer (see also fig. S1 for full-size exemplar spectrograms). (**C**) The lines show the probability of stimulus interruption of individual vocalizers by a single subject in 20 trial bins (blue, rewarded; red, nonrewarded). Tick marks above the plot indicate interrupted trials, and those below the plot indicate noninterrupted trials. (**D**) Average odds ratio (OR) for song and DC assessed after training, on days 4 and 5, for all subjects ($n = 19$). Birds perform better on songs (OR, 15.5; 95% CI, 9.9 to 24.4) than on DC (OR, 8.4; 95% CI, 5.6 to 12.9) ($p = 0.004$, log-transformed paired t test). Error bars show 2 SEM.

in which subjects began by discriminating one rewarded vocalizer from one nonrewarded vocalizer, while additional vocalizers were added to the test on subsequent days (Figure 2.1, B and C). Zebra finches individualize each of their call types, and, although their song and DCs are fairly idiosyncratic and stereotyped, there is also acoustical variability across renditions produced by a single vocalizer (Elie & Theunissen, 2018). Thus, each vocalizer was represented by multiple renditions of its song or DC (Figure 2.1B).

The performance of each subject was evaluated on days 4 and 5, after they had had at least 1 day of training on each vocalizer. Overall, task performance was measured using an odds ratio (OR): the odds of interruption for nonrewarded trials (correct responses) divided by the odds of interruption on rewarded trials (incorrect responses). An OR of 1 indicates behavior at chance level, and greater than 1 indicates that the subject successfully distinguished rewarded from nonrewarded trials. Nearly all subjects had ORs significantly greater than 1, indicating that they were successful at this task, both when tested on songs (19 of 19 subjects) and on DCs (18 of 19 subjects) ($p < 0.0026$, one-sided Fisher's exact test, Bonferroni corrected; Figure 2.1D). There was no difference between males and females on this task as assessed with a mixed effects model, with subject identity as the random effect and call type (DC or song) and subject sex as the fixed effects (Figure 2.7A); the effect of subject sex on the overall log OR was not significant [$\beta = -0.163$; 95% confidence interval (CI), -1.012 to 0.687; $p = 0.707$], and neither was the interaction between subject sex and call type ($\beta = -0.449$; 95% CI, -1.315 to 0.416; $p = 0.309$).

To see whether this performance was driven by memorization of all vocalizers in the test or just recognition of a subset of them, we looked at each subject's performance in detail by evaluating their behavior per individual vocalizer (Figure 2.2). We defined the per-vocalizer OR as the ratio of the odds of interrupting a specific vocalizer by the odds of interrupting a random stimulus sampled equally from rewarded and nonrewarded trials. Using this definition, a vocalizer is memorized if the OR is significantly greater than 1 for nonrewarded vocalizers or less than 1 for rewarded vocalizers. We found that 2 of the 19 subjects were able to memorize the entire set of 16 vocalizers from their songs (12 of 19 learned at least half) and 4 of the 19 subjects were able to memorize the entire set of 12 vocalizers from DCs (15 of 19 learned at least half).

To assess the limits of the auditory memory capacity in these songbirds, for four subjects, we intermixed and doubled the size of the two stimulus sets (song and DCs) in the same session. This resulted in a set of DCs from 24 vocalizers and songs from 32 vocalizers for a total of 56 distinct vocalizers. On the first week after completing the two initial learning ladders and testing (song and DC), subjects were trained on the larger song repertoire (16v16) and DC repertoire (12v12) for 3 days each, thus doubling the total number of vocalizers in 6 days. The following week, subjects were given a single day testing session in which previously learned songs and DCs were intermixed for the first time, with only two vocalizers for each rewarding condition and call type. Under this mixed call type condition, subjects continued to self-initiate trials and interrupt the stimuli at rates seen in previous weeks. We then increased the stimulus set to all vocalizers learned thus far (32 vocalizers on song and 24 vocalizers on DC) and evaluated performance on the next 4 days. The results from these four

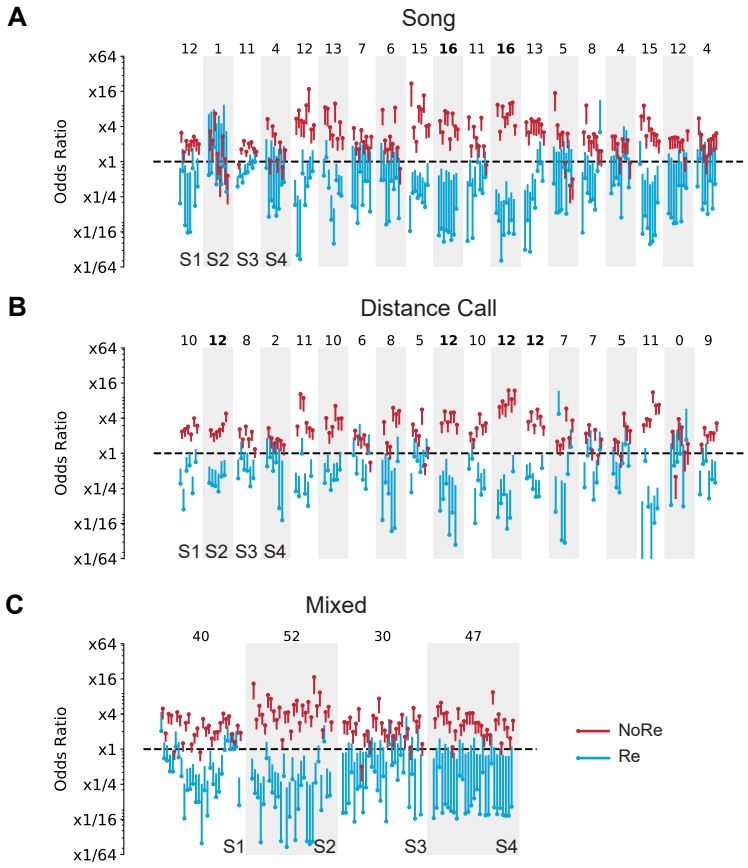


Figure 2.2: **Memory capacity for vocalizer identity over all subjects.** Discrimination performance per vocalizer and subject ($n = 19$) for songs (**A**), DCs (**B**), and both songs and DCs ($n = 4$) (**C**). The mixed condition (C) was performed by four subjects who were additionally tested with a total of 56 vocalizers: 24 vocalizers of DCs and 32 vocalizers of songs. For each subject (white/gray plot background), the dots indicate the OR of interrupting a given vocalizer. Red dots correspond to nonrewarded vocalizers (NoRe) and blue dots to rewarded vocalizers (Re). The number of vocalizers that are discriminated significantly above chance ($p < 0.05$, controlling for false discovery rate using Benjamin-Hochberg procedure) are indicated above each subject's plot (maximum number of vocalizers are 12 for DCs, 16 for songs, and 56 for the mixed condition). Note that the order of the dots on the x axis is random and that the rewarded and nonrewarded vocalizers are not paired. Error bars correspond to the one-sided 95% CI (Fisher's exact test). OR of 1 corresponds to chance. Error bars for nonrewarded stimuli are generally smaller because they are played more frequently.

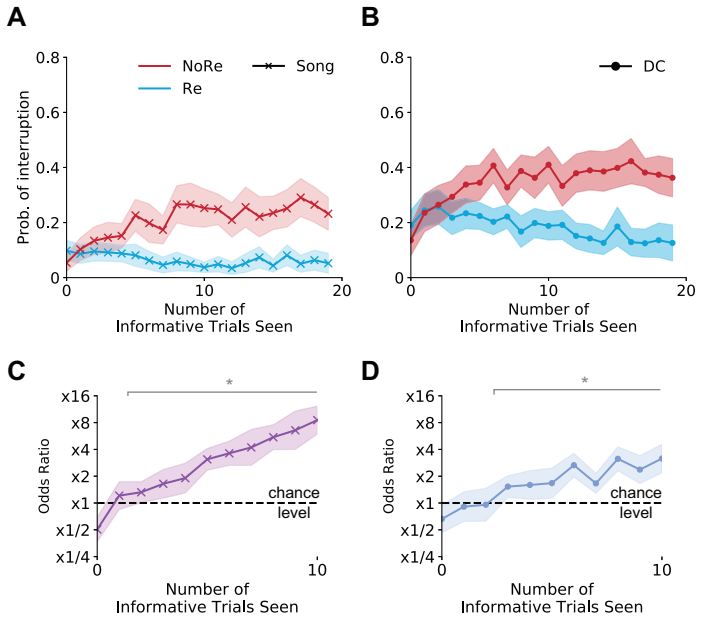


Figure 2.3: **Speed of memory acquisition.** (**A** and **B**) Learning rates are analyzed by plotting the behavioral response (probability of interruption) as a function of informative trials (see Results section) for rewarded (blue) and nonrewarded vocalizers (red). (**C** and **D**) The separation between the red and blue curves in A and B quantifies the learning and is shown in C and D as an OR of odds for nonrewarded divided by the odds of rewarded as in Fig. 2.1D [(C), song; and (D), DC]. Shaded regions show 2 SEM. Asterisks indicate region where OR was significantly greater than 0 ($n = 19$, $p < 0.05$, false discovery correction).

subjects demonstrated that 40, 52, 30, and 47 (mean, 42) vocalizers could be distinguished successfully.

To assess how quickly stimuli were learned, we generated learning curves showing the interruption probability versus the number of informative trials seen, where an "informative trial" is a trial in which the subject did not interrupt the stimulus, giving the bird an opportunity to learn the reward association (interrupted trials do not give the subject new information about whether the stimulus is rewarded or not) (Figure 2.5). For both songs and DCs, the probability of interrupting rewarded and nonrewarded stimuli is indistinguishable when no informative trials have been seen (intercepts in Figure 2.3, A and B), as one would expect. However, the interruption probabilities for rewarded and nonrewarded vocalizers begin to diverge after only a few informative trials, demonstrating very rapid learning of vocalizers' identity (Figure 2.3, A and B). There is a significant effect of call type on the rate of this divergence ($\beta = 0.155$; 95% CI, 0.086 to 0.222; $p < 0.001$, mixed effects model), suggesting that songs may be learned more quickly and with fewer examples (Figure 2.3, C and D, and Figure 2.7B). One can also notice that the default "baseline" interruption rates

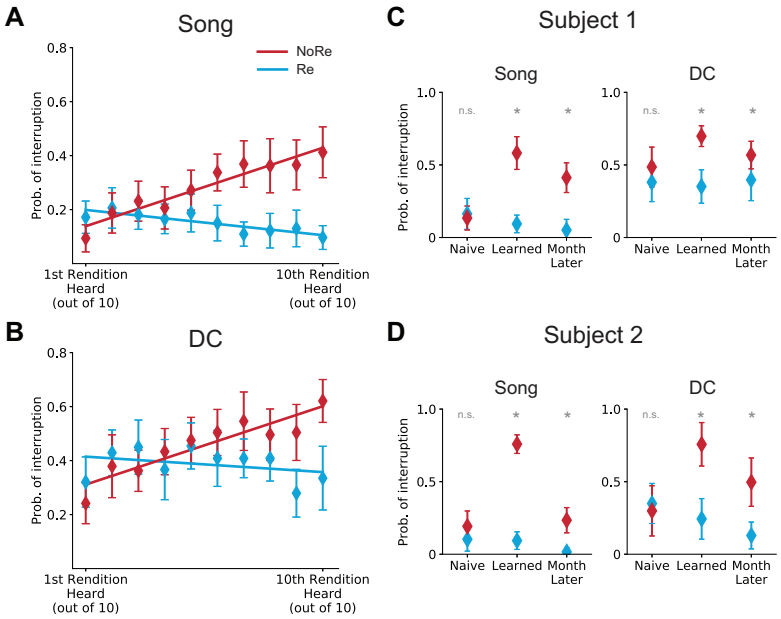


Figure 2.4: **Generalization and long-term memory.** (**A** and **B**) The plots show the average probability of interruption across all subjects ($n = 19$) for each of the 10 renditions the first time they are heard by the subject; the renditions are ordered on the x axis according to the presentation order. Error bars are 2 SEM. (**C** and **D**) Interruption rates for nonrewarded and rewarded vocalizers in two subjects (S1 and S2 of Figure 2.2) during three epochs for songs (left) and DCs (right). The three epochs shown are Naïve (initial exposure to the stimuli), Learned (last two sessions of initial learning ladders), and Month later (1 month after Learned without any reinforcement). The interruption rates to a particular vocalizer are restricted to trials before the second informative trial of that vocalizer during the relevant epoch. Asterisks indicate epochs during which nonrewarded stimuli were interrupted at a significantly higher rate than rewarded stimuli ($p < 0.05$, one-sided t test). Error bars indicate 2 SEM. *n.s.*, not significant.

differed between songs and DCs when no informative trials have been seen [song baseline, 0.08 ± 0.01 (2 SEM); DC baseline, 0.16 ± 0.02; mixed effect models, $p < 0.001$]. The difference in the baseline interruption rates or in the learning rates between male and female subjects was not significant (mixed effects models, $p = 0.563$).

As mentioned above, to encourage subjects to use the individual signature and not a particular acoustical feature present in a given rendition, a vocalizer is represented by randomly chosen call renditions. If subjects are identifying the vocalizer and not memorizing the individual recordings, then they should be able to correctly predict to which reward contingency a novel rendition belongs when they have already heard and learned some of the renditions of a vocalizer. Birds are at chance levels for the first few renditions they hear but

begin to correctly categorize previously unheard renditions after exposure to other renditions from the same vocalizer (Figure 2.4, A and B); post hoc analysis of the order in which renditions were first presented to subjects reveals that the interruption probability of unseen nonrewarded stimuli increases with the rendition presentation order ($R^2_{adj,song} = 0.90$ and $R^2_{adj,DC} = 0.81$). In the same vein, the interruption probability of rewarded stimuli decreases with the rendition presentation order for song ($R^2_{adj,song} = 0.71$), but the same decrease was not apparent for DC ($R^2_{adj,DC} = 0.00$). The slopes are steeper for the nonrewarded renditions because nonrewarded stimuli are being presented four times more frequently than rewarded stimuli; thus, they are also learned faster. Thus, birds are learning to identify the identity of the vocalizers and do not just memorize the individual sound files.

To test whether these memories are stable over longer times and without any additional reinforcement, we retested two subjects on the largest stimulus set (32 songs and 24 DCs intermixed) after a month during which they were not exposed to any of the vocalizations from the test. While their overall performance slightly decreased from optimal performance during the initial test as measured by the change in log OR [0.12 ± 0.18 (2 SEM) in subject 1 and -0.73 ± 0.23 in subject 2], the overall ORs and OR per vocalizer were still well above chance ($p < 0.001$), indicating that reward associations were retained after a month. To validate that these responses were remembered and not rapidly relearned, we examined the interruption rates for the first informative trials after 1 month and compared them to the rates found for the first informative trials during initial learning (Figure 2.4, C and D). These results indicate that these memories for rewarded and nonrewarded vocalizers are stable and can be recalled a month after learning. This is particularly remarkable given that these memories were acquired rapidly and were only reinforced for a short time.

2.4 Discussion

Zebra finches have exceptional auditory memory abilities for the individual signature found in their communication calls. We found that they are able to quickly learn to recognize the identity of up to 40 vocalizers and to maintain these auditory memories for a long period of time. The recognition of vocalizers is a nontrivial task since it requires the extraction of the individual signature present in each call while ignoring the variability across call renditions. Thus, these are not auditory memories for specific sounds but for the information bearing invariant features constituting the individual signature of the vocalizer (Elie & Theunissen, 2018). We showed that zebra finches can learn and memorize this individual signature with a very small number of exposures (less than 5), can simultaneously remember a large number of these vocalizers, and are able to use these memories to classify call renditions that they have not heard before (generalization).

The memory capacity in zebra finches for recognizing individuals from their vocalizations is large and might exceed the limits that could be tested with our experimental design. We found that 16 vocalizers based on song and 12 vocalizers based on DC could be regularly discriminated by our subjects. When subjects were tested on as many vocalizers as could

be practically tested in a single session, birds were able to discriminate up to 52 distinct vocalizers. The capacity of this auditory memory is similar to other forms of avian memory that have been well quantified, such as spatial memories in food-caching birds (Balda & Kamil, 1992) or visual memories in pigeons (Cook et al., 2005). Auditory memories for object labels have also been shown in parrots (Pepperberg, 1981) and in some mammals (Herman et al., 1984), including the exceptional example of Rico, the border collie, who could correctly fetch 200 distinct objects on vocal commands (Kaminski et al., 2004). We also found that birds make an efficient use of informative trials during their very rapid learning, as they are able to memorize the individual signature of a vocalizer after only a few examples (less than 10). This fast mapping for communicative vocal signals has only been shown in humans and dogs and is thought to be a key cognitive ability for language learning (Kaminski et al., 2004; Markson & Bloom, 1997). Last, this memory was long lasting; birds could still remember which vocalizers were assigned to reward versus nonrewarded groups after 1 month without any reinforcement. While previous experiments had shown that song exposure in zebra finches improves auditory recognition, suggestive of a capacity for long-term auditory memories for conspecific vocalizations (Braaten et al., 2007), this is the first study that quantifies the auditory memory capacity in a songbird for individual signature and demonstrates its remarkable performance. Just as in humans, we postulate that birds use an abstract neural representation of these auditory objects to facilitate both working memory manipulation and long-term memory storage (Joseph et al., 2015).

Since most songbirds are also vocal imitators, one might postulate that the memory mechanisms needed for the song imitation behavior overlap with ones that are needed for individual recognition. The auditory memories could be stored as learned motor programs (Williams & Nottebohm, 1985), and the high-level abstract representation could then be a motor code. There are many problems with such a motor theory of perception in songbirds: Individual recognition based on vocalizations is present for calls that are not learned (Elie & Theunissen, 2018); it is equally similar in male and female zebra finches, while only male zebra finches learn to sing; and male zebra finches learn a single song, but, as we have shown, they can remember the individual signature of songs and calls from a much larger number of vocalizers. Therefore, although the motor song nuclei might play a role, we and others (Gobes & Bolhuis, 2007) postulate that a separate neural mechanism representing high-level auditory features is involved in the formation and use of memories for all auditory objects that are relevant for vocal communication. The second order avian auditory pallial areas NCM (nidopallium caudal medial) and CM (caudal mesopallium) are good candidates for the locus of such an engram. NCM neurons show neural correlates of memories for the tutor song before vocal learning (Yanagihara & Yazaki-Sugiyama, 2016), and CM neurons show neural correlates for categories of natural sounds learned in operant conditioning tasks (Gentner & Margoliash, 2003; Jeanne et al., 2011). Experiments that have exploited the stimulus-specific habituation observed in NCM neurons also suggest that this auditory area can exhibit a large-capacity memory for conspecific song (Chew et al., 1996). The identity and the connectivity of neural networks involved for storing and recalling these auditory objects as well as the nature of the neural representation for vocalizations, while an active

area of research (Elie & Theunissen, 2015, 2019; Jeanne et al., 2013; Kozlov & Gentner, 2016; Moore & Woolley, 2019), remain relatively unexplored in the birdsong field (Elie & Theunissen, 2020). Just as the neural basis of the song imitation behavior has led to many insights into mechanisms of vocal production and learning (Sakata et al., 2020), we predict that future work on the neural basis of these auditory memories and their rapid formation will reveal core knowledge of the neural circuits and computations needed for recognizing learned meaning in vocal sounds, including in human speech.

The fast-learning and exceptional memory for auditory objects in songbirds is a behavioral trait that is essential for vocal communication in social species. This skill can be added to their well-studied vocal imitation behavior, their ability to learn grammar like rules (Cate & ten Cate, 2018; Gentner et al., 2006), and their capacity to combine call types to generate complex meaning (Suzuki et al., 2018). Individual recognition plays an important role for behaviors in social groups and, in particular, for fission-fusion societies such as those observed in some bird species, including the zebra finch (Silk et al., 2014), and in mammals such as in the African elephant (McComb et al., 2000). We suggest that these auditory memories for vocalizers are not only important for mate and kin recognition but also to facilitate group dynamics. Studying vocal communication in gregarious bird species should therefore include the role of higher cognitive functions, such as memory, and take into account the species social dynamics. These vocal and perceptual performances can, in turn, be added to the list of cognitive faculties that have been found in social birds, such as episodic spatial memory (Balda & Kamil, 1992; Clayton & Dickinson, 1999), social cognition (Emery et al., 2004; Vignal et al., 2004), number sense (Nieder, 2017), or puzzle solving (Heinrich & Bugnyar, 2005), and that rival the cognitive faculties found in social primates (Emery, 2006; Emery & Clayton, 2004).

2.5 Materials and Methods

Ethics statement

All animal procedures were approved by the Animal Care and Use Committee of the University of California, Berkeley (AUP-2016-09-9157) and were in accordance with the National Institutes of Health guidelines regarding the care and use of animals for experimental procedures.

Testing apparatus and software

The operant conditioning apparatus and our go–no go paradigm had been described in detail in our previous publication (Elie & Theunissen, 2018). Briefly, our operant chamber is composed of one pecking key and one food hopper (Med Associates). Subjects initiate trials by pecking the key, which triggers a 6-s auditory stimulus to be played. Sound levels are calibrated to match natural levels of intensity for each call type when vocalizations are used

as stimuli. After 6 s, a food reward is either given (if the stimulus was rewarded) or nothing happens (if the stimulus was nonrewarded). Alternatively, as the sound is played, the bird can terminate a trial and start a new one by pecking the same key. In this case, the initial trial will not result in food whether the stimulus is rewarded or not, and a new trial is immediately initiated. To maximize the rate at which reward is received in a session, the subjects learn to skip stimuli that are recognized as nonrewarded to avoid the full 6-s waiting period and move on to the next trial. Subjects are food restricted with access to water but limited seed in between test sessions to maintain motivation. Subjects were weighed before and after every test session, and seed consumed in a daily session was measured and supplemented at the end of day so that the birds maintain their weight within 10% of their starting weight. Daily handling of subjects did not seem to affect the birds' motivation or ability to do the task once they became comfortable with the experiment chamber. Once trained, birds are able to get all of their daily food allowance during the testing period.

The birds learn to use the apparatus during a shaping session that lasts approximately 1 week. During the shaping session, the bird first learns to associate pecking of the key with sounds and food reward and then learn to interrupt nonrewarded sounds. The initial shaping task involves the discrimination of two clearly distinct song stimuli. We have also performed control experiments, clearly showing that apparatus is not providing any extraneous clues that the birds could use to distinguish rewarded from nonrewarded trials (Elie & Theunissen, 2018).

The presentation of the sound stimuli, the detection of key pecks, and the operation of the food hopper were controlled by a Python program. We used a custom branch of the Python-based pyOperant software (`https://github.com/theunissenlab/pyoperant`), originally developed by J. Kiggins and M. Thielk in T. Gentner's laboratory at University of California San Diego (`https://github.com/gentnerlab/pyoperant`).

Auditory discrimination experiments

Subjects were tasked with discriminating between a set of rewarded and nonrewarded individuals based on the playback of their vocalizations. By design, 20% of trials are rewarded after the end of the stimulus playback, while 80% of trials are not rewarded so that subjects learn to peck for a new trial (interrupting the current trial) when they recognize a stimulus as nonrewarded.

For each vocalizer, we generated 10 unique stimuli that could be played on each trial so that specific extraneous acoustic features of a particular stimulus file that did not encode the vocalizer identity (e.g., length, intensity, and background noise) could not be used as a reward cue. Each song stimulus file consisted of three randomly selected song bouts of two motifs, each from the same vocalizer, separated by randomly chosen intervals such that the duration of the stimulus file would be exactly 6 s. Most introductory notes (repeated short vocalizations preceding a song bout with sometimes long internote intervals) were removed to avoid great variability in stimulus duration. Similarly, each DC stimulus file consisted of six randomly selected DC renditions from one vocalizer, separated by randomly chosen

intervals. The amplitudes of the audio files were normalized within stimuli of the same type, i.e., songs or DCs.

On the first day of the test, a subject is tasked with discriminating between one rewarded vocalizer and one nonrewarded vocalizer. Over this single session of about 8 hours, subjects learned to interrupt nonrewarded trials and to wait on rewarded trials. On subsequent days, additional vocalizers were added to the test (Figure 2.1): After the first day of 1 rewarded vocalizer versus 1 nonrewarded vocalizer (1v1), we added stimuli from three more rewarded and three more nonrewarded vocalizers, resulting in four rewarded versus four nonrewarded (4v4), again with 10 unique renditions per vocalizer. After the day of 4v4, the birds moved on to 8v8 (for songs) or 6v6 (for DCs). Because subjects do as few as 200 trials per day and we only play rewarded trials 20% of the time, a single vocalizer may be heard as few as five times per day on average once we reach 8v8. We expected that this would make learning at that stage of the ladder difficult. To aid in learning and allow the birds more opportunities to learn every stimulus, on the first day of 8v8 or 6v6, we played stimuli from the new vocalizers twice as frequently as stimuli from vocalizers previously seen on the 1v1 and 4v4 days. On the last 2 days of 6v6/8v8, the probability was set again to be equal across all vocalizers of the same reward outcome. We used these last 2 days to evaluate task performance. In a few cases, the 1v1 or 4v4 day was repeated (4 of 19 during 1v1 days, 4 of 19 during 4v4 days) because the subject failed to trigger a sufficiently large number of trials.

Vocalizers were randomly assigned to the rewarded or nonrewarded set. Moreover, we used a balanced procedure where the rewarded and nonrewarded sets were switched for each half of the birds in the experiment. Last, for DCs, male and female vocalizers were also randomly assigned to rewarded and nonrewarded sets. The zebra finch DC is sexually dimorphic (Elie & Theunissen, 2016), and by mixing male and female vocalizers in each set, we forced our subjects to use the individual signature and not the acoustic features characteristic of the sex of the vocalizer.

Subjects

Twenty adult domestic zebra finches (10 males and 10 females) were used as subjects in this study. One female subject was excluded from the song memory test analysis due to errors in stimulus selection. A different female subject was excluded from the DC memory test analysis for the same reason, resulting in n = 19 for both the song and DC analysis. Subjects were housed in a colony room (usually 10 to 30 individuals in a large flight cage) at the University of California (UC) Berkeley. Of these 20 subjects, 4 subjects were chosen (randomly) to participate in a second session with the combined and larger stimulus set, and 2 of those 4 birds were chosen in the third session to assess long-term memory.

Song vocalization recordings were from 32 male zebra finches from the Theunissen Lab at UC Berkeley, the Perkel laboratory at the University of Washington, and the Leblois laboratory, Bordeaux (France) Neurocampus. DC vocalizations came from 24 zebra finches (12 male and 12 female), all from our colony at UC Berkeley. Vocalizations used as stimuli were recorded as part of previous experiments in the laboratory, and the vocalizers were

	Interruptions	**Waits**
Nonrewarded	a	c
Rewarded	b	d

Table 2.1: Contingency matrix used to estimate the OR of interruption for nonrewarded vs rewarded vocalizer.

	Interruptions	**Waits**
Vocalizer	a	c
Random	b	d

Table 2.2: Contingency matrix used to estimate the OR of interruption for a particular vocalizer relative to a random vocalizer.

unfamiliar to the subjects in the present study. The 12 male DCs were produced by a subset of the males also used in the song stimulus set—however, reward associations were randomized (7 switched, 5 same).

Statistical analyses

Performance on the task overall was quantified as an OR obtained by dividing the odds of interrupting a nonrewarded stimulus by the odds of interrupting a rewarded stimulus. The odds of interrupting a stimulus in a given reward group was calculated by taking all trials of that reward category and computing the probability of interruption. For Fig. 1C, this was computed on the trials from the last 2 days of tests (6v6 DCs and 8v8 songs) when all vocalizers were played at equal rates. Performance on songs was compared to performance on DCs with a paired t test over subjects. All ORs and 95% CIs were computed using the Fisher's exact test using the contingency matrix shown in Table 2.1.

The odds of interruption of the nonrewarded stimulus is $O_{NoRe} = \frac{a}{c}$; similarly, the odds of interruption of the rewarded stimuli is $O_{Re} = \frac{b}{d}$. The OR is $OR = \frac{ad}{bc}$. The Fisher's exact test calculates the probability of obtaining an OR as extreme (equal or greater) by calculating the distribution of all ORs obtained for all possible contingency matrices that have the same marginals as those in the actual data. Zero values in any cell cause the OR to be undefined or go to infinity. To avoid this issue, we used the Haldane-Anscombe correction by adding 0.5 to all cells before computing the OR.

Performance per vocalizer was quantified as an OR obtained by dividing the odds of interrupting a given vocalizer by the odds of interrupting a random vocalizer during the time period of interest (Figure 2.2). The odds of interrupting a random vocalizer was computed by sampling equal numbers of rewarded and nonrewarded trials on the last 2 days of the 8v8 song and 6v6 DC ladders (Figure 2.2, A and B) or over 5 days of the 28v28 mixed set

(Figure 2.2C), using the contingency matrix shown in Table 2.2.

Learning curves (Figure 2.3) were computed as a function of informative trials, where an informative trial is defined as a trial in which the subject did not interrupt. The probability of interruption in bin k for a subject vocalizer pair is computed by pooling over all trials after the kth interruption and up to and including the $(k+1)$th noninterruption of that vocalizer. Interruption rates of 0 were adjusted by replacing them with 0.5 times the mean interruption rate across all vocalizers for the same reward contingency in that informative trial bin. Population mean and SEM were then computed across subjects. Significance in bin k was evaluated using the Bonferroni correction. Learning rate is evaluated as the rate at which the log OR between interruption rates on nonrewarded and rewarded trials increases. The effect of call type (song versus DC) on the learning rate was measured using a mixed effects model, with subject as the random effect and call type and informative trials as the fixed effects, predicting the log OR between nonrewarded and rewarded interruptions.

Acknowledgements

We thank two undergraduate research apprentices, I. Rice and A. Prasad, who helped train and test the animals. We thank L. Johnston and J. Elie for insightful comments on the manuscript and D. Perkel and A. Leblois for contributing the song stimuli. Funding: This research was funded by NIDCD R01 018321 to F.E.T. and an NSF graduate fellowship DGE 1752814 to K.Y. Author contributions: Experimental design: W.E.W. and F.E.T.; investigation: W.E.W., K.Y., and F.E.T.; data analyses and visualizations: K.Y.; writing, original draft: K.Y. and F.E.T.; writing, review and editing: W.E.W., K.Y., and F.E.T. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data (trial by trial data and stimulus audio files), code used for analysis, and documentation are available on GitHub at `https://github.com/theunissenlab/zebra-finch-memory`. Additional data related to this paper may be requested from the authors.

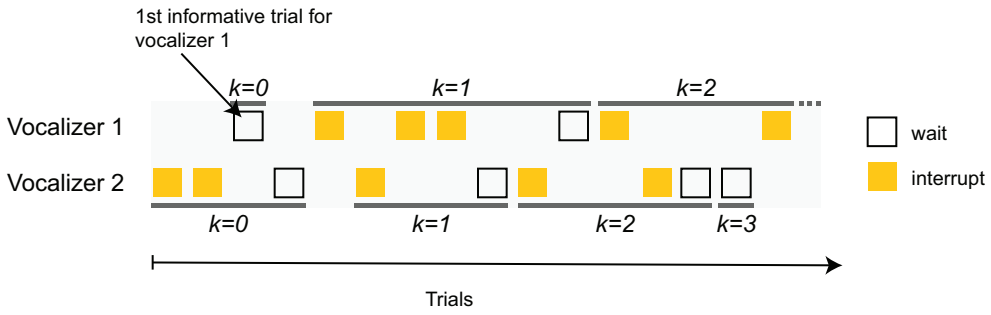


Figure 2.5: **Visual description of informative trials** used in Figures 2.3 and 2.4. Informative trials are non-interrupted trials which give the subject a chance to learn whether a vocalizer is rewarded. Shown is an example of trials from one subject in response to two vocalizers in a session. The white boxes indicate informative trials, and all trials between subsequent informative trials for one vocalizer are used to estimate the probability of interruption in that window (indicated by black lines). A trial is labeled by how many informative trials for that specific vocalizer has been seen previously (i.e. how many opportunities the subject had to learn the reward contingency for that vocalizer). In Figures 2.3A&B, each data point is the mean interruption probability across subjects for bin k, where the subject interruption probability in k is the mean interruption probability across vocalizers. In Figures 2.4C&D, probability of interruption is computed for all trials where $k < 2$.

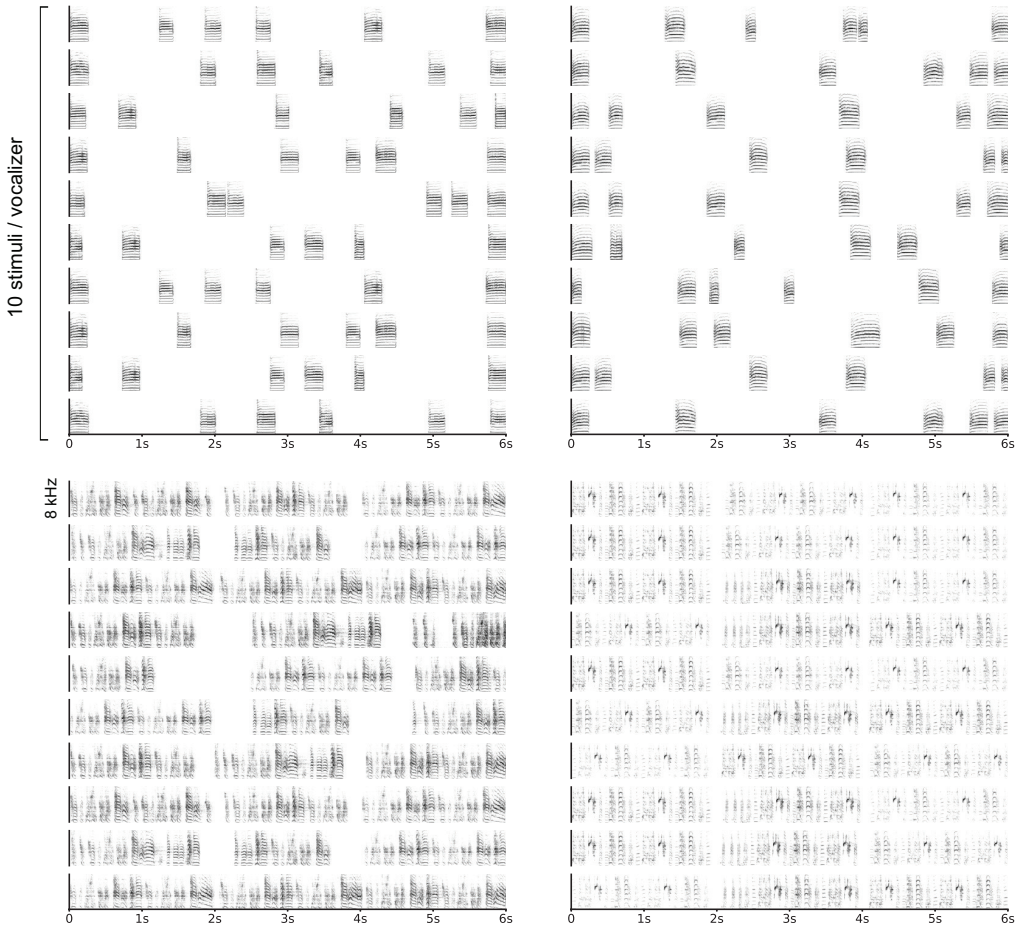
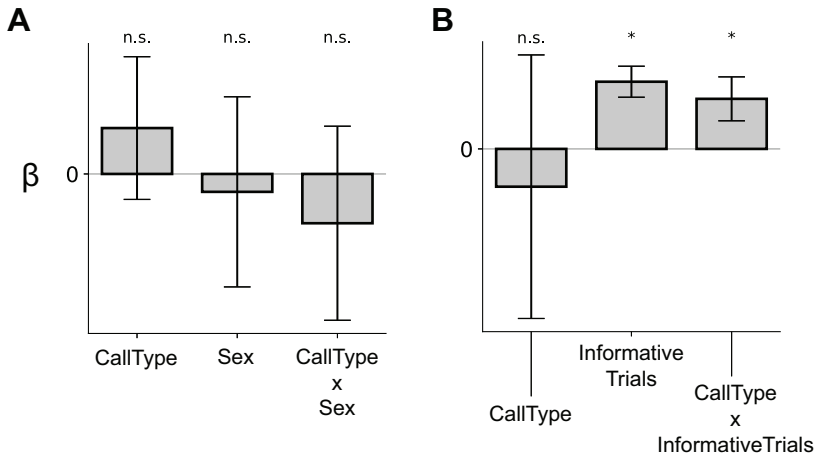


Figure 2.6: **Examples of DC and song stimuli.** Spectrograms showing examples of DC stimuli from 2 vocalizers (top quadrants) and song stimuli from 2 vocalizers (bottom quadrants). Each vocalizer is represented by 10 stimulus files containing either six DCs or three song motifs, separated by random gaps and aligned such that the resulting stimulus duration is 6 seconds long.



	β	Std. Err	z	$p > \|z\|$	[0.025	0.975]
Intercept	2.003	0.313	6.405	0.000	1.390	2.616
CallType[Song]	0.418	0.325	1.288	0.198	-0.218	1.054
Sex[M]	-0.163	0.433	-0.375	0.707	-1.012	0.687
CallType[Song] x Sex[M]	-0.449	0.442	-1.017	0.309	-1.315	0.416
Group Var	0.386	0.085				

	β	Std. Err	z	$p > \|z\|$	[0.025	0.975]
Intercept	0.322	0.517	-2.052	0.040	-0.629	-0.014
CallType[Song]	-0.117	0.204	-0.573	0.567	-0.516	0.283
InfoTrials	0.208	0.024	8.552	0.000	0.161	0.256
CallType[Song] x InfoTrials	0.155	0.034	4.495	0.000	0.087	0.222
Group Var	0.076	0.039				

Figure 2.7: **Mixed effects modeling for sex and call type.** (**A**) Mixed effect model for per-vocalizer log odds-ratios. The log odds-ratio was inverted for rewarded vocalizers ($\log_2 \frac{1}{OR}$) for a direct comparison to the non-rewarded condition ($\log_2 OR$). Call type (Song or DC) and subject sex were used as fixed effects and subject identity as the random effect. Task performance was not significantly affected by subject sex and performance was slightly higher for songs than distance calls. Error bars show 95% confidence intervals. (**B**) The log odds-ratio between interruptions of unrewarded to rewarded trials (Figure 2.3C&D) was modeled using a mixed effects model with call type (Song or DC) and number of of informative trials seen as fixed effects and subject identity as the random effect. Performance on task increases as a function of informative trials seen, indicated by significant effect of informative trials seen on the log odds-ratio. Rate of increase is greater for songs than distance calls, shown by the significant interaction term between informative trials and call type.

Chapter 3

Lesions to NCM impair individual vocal recognition memory in the zebra finch

3.1 Abstract

Social animals that use vocalizations to communicate are often able to recognize other individuals by their sounds. The neural basis for this skill is a sophisticated auditory memory system capable of mapping incoming acoustic signals to one out of many known individuals. Using the zebra finch, a social songbird that uses songs and distance calls to communicate individual identity (Elie & Theunissen, 2018), we tested the role of two higher-order brain regions in a vocal recognition task. We found that the caudomedial nidopallium (NCM), a secondary auditory region of the avian brain analogous to auditory association cortex in humans (Bolhuis & Gahr, 2006), was necessary for maintaining stored auditory memories for conspecific vocalizations, while HVC, a premotor area that gates auditory input into the vocal motor and song learning pathways (Roberts & Mooney, 2013), was not. However, neither an intact NCM nor HVC were required for acquiring new auditory memories.

3.2 Introduction

Successful vocal interactions often require individuals to recognize the identity of another vocalizer. In social species that live and move in groups, such as humans and some songbirds, this requires the brain to store memories of known individuals and to map the acoustic features of a sound to one out of potentially hundreds of known individuals, a skill known as "individual vocal recognition" (Carlson et al., 2020; Tibbetts & Dale, 2007). The zebra finch is one such social songbird that uses communication calls for speaker recognition, and its repertoire of communication calls is well documented (Elie & Theunissen, 2016; Zann, 1996). In particular, the zebra finch uses two socially affiliative vocalization types, the song and

distance call (DC), to signal vocalizer identity. These two call types have acoustic signatures that are unique to each individual, and stereotyped within an individual (D'Amelio, Klumb, et al., 2017; Elie & Theunissen, 2018). We have previously shown that zebra finches have a large capacity memory for recognizing conspecific vocalizers, and those memories are learned quickly and can persist for several weeks (See Chapter 2). How and where are these memories formed, stored, and retrieved in the brain?

Studies using lesions and fMRI have shown that the auditory association cortex in humans is broadly involved in the memory and classification of sounds, and in speaker recognition (Andics et al., 2010). One region of the avian brain thought to be analogous to auditory association cortex is NCM (caudomedial nidopallium) (Bolhuis & Gahr, 2006; Bolhuis et al., 2010). This area is at the top of the hierarchical auditory processing pathway and receives direct projections from the primary thalamo-recipient auditory region, Field L (Figure 3.1A). Substantial evidence implicates NCM as the primary site of auditory memory formation. Studies on stimulus-specific habituation properties in NCM suggest that it has a large capacity for unique vocalizations (Chew et al., 1996). It has also been observed that response strengths in NCM are lower in response to learned songs of conspecifics than novel songs (Thompson & Gentner, 2010). In song imitation learning, NCM has been implicated in memory and recognition of the tutor song in juvenile songbirds through immediate early gene studies (Bolhuis et al., 2001; Mello et al., 1995), changes in neural tuning (Phan et al., 2006; Yanagihara & Yazaki-Sugiyama, 2016), stimulus-specific habituation (Chew et al., 1995) and pharmacological inactivations (London & Clayton, 2008; Pagliaro et al., 2020). Pharmacological disruption of NCM during a learning task using pure tones reduces learning rate but not final performance, suggesting a role in association learning but not memory retrieval (Macedo-Lima & Remage-Healey, 2020). Lesions to NCM, in contrast to the manipulations cited above, have not been shown to affect song imitation learning (Canopoli et al., 2014; Canopoli et al., 2016; Canopoli et al., 2017), although lesions do impair song recognition when assayed with a tutor song preference test (Gobes & Bolhuis, 2007). Despite this body of evidence for NCM's involvement in vocal imitation learning and auditory memory, its role in recognition memory for the vocalizations of specific individuals is unknown. By pairing NCM lesions with an operant discrimination task designed to stress the memorization ability of zebra finches for conspecific calls and song, details of NCM's involvement in the storage and retrieval of these auditory memories may be revealed.

Another hypothesis, not mutually exclusive to the above, is a "motor theory of speech perception" in which auditory perception for conspecific vocalizations utilizes representations in the vocal motor system (Massaro & Chen, 2008; Schulze et al., 2012; Williams & Nottebohm, 1985). This category of idea is often proposed in studies on song learning, in which the anterior forebrain pathway (AFP) (Figure 3.1A) has been suggested as the site of "tutor song memory access" (Roberts & Mooney, 2013) through a comparison of incoming auditory feedback to a stored auditory memory of a tutor template. Song-selective "mirror" neurons found throughout the AFP could be used in a comparator circuit in which incoming songs are compared to the bird's own song as a basis for discrimination (Doupe & Konishi, 1991; Mooney, 2014). The pre-motor nucleus HVC is the only known source of auditory in-

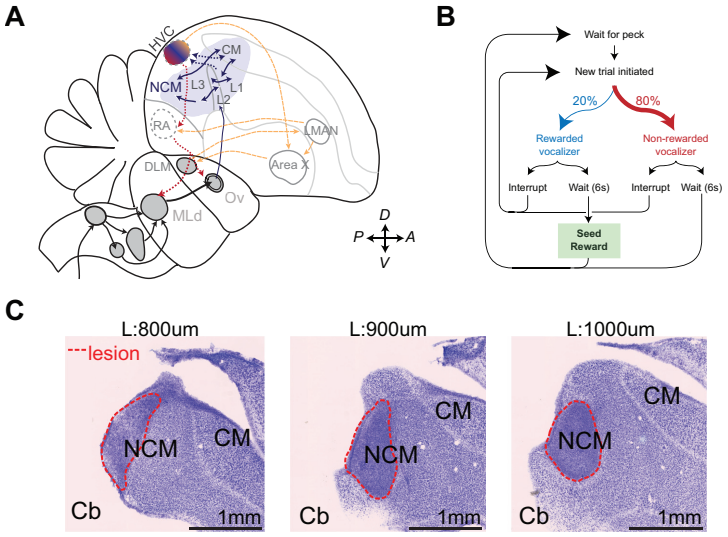


Figure 3.1: **Zebra finch auditory system; task diagram; lesion histology.** (**A**) Schematic diagram of the zebra finch auditory and vocal motor pathways. Navy: ascending auditory pathways; Red: vocal motor pathway; Orange: anterior forebrain pathway (AFP). (**B**) Task diagram for behavioral conditioning. Subjects initiate trials and hear a 6 s stimulus playback of a rewarded (Re) or non-rewarded (NoRe) vocalizer. At the end of the playback, subjects will either receive a food reward or nothing. The subject can peck again during the playback ("interrupt") to terminate the trial and begin a new trial, but will receive no reward or reward information. (**C**) Nissl stained images of NCM lesion in successive sagittal slices from the left hemisphere of one subject. Red dotted line shows approximate extent of lesion. Above each image is the lateral distance from midline of the section. Cb: Cerebellum (relative position, tissue not in image), NCM: caudomedial nidopallium, CM: caudal mesopallium.

put into the avian song system and receives auditory inputs from Av and/or NIf, which are closely connected with auditory regions CM (caudal mesopallium) and Field L, respectively (Roberts & Mooney, 2013). Furthermore, IEG expression patterns imply functional relationships between HVC and secondary auditory regions NCM and CM (Lynch et al., 2013). There is evidence that the function of auditory information in HVC is not limited to song imitation learning, a behavior that is typically restricted to males in most songbird species; for example, lesions to HVC have been shown to alter females courtship behavior and mate preference in response to male songs (Brenowitz, 1991; Del Negro et al., 1998; Perkes et al., 2019). Furthermore, lesions to HVC may alter the ability for both males and females to learn reward associations to conspecific songs, while not affecting song recognition (Gentner et al., 2000). The role of HVC is not limited to song; while HVC is not necessary for innate DC production, lesions have shown that it is necessary for production of the learned acoustic

features heard in male DCs (Simpson & Vicario, 1990). HVC has also been shown to exhibit motor and auditory responses during antiphonal social interactions that use innate calls such as the stack call (Ma et al., 2020). Thus, HVC and the greater vocal motor system could have a role in perception and recognition of the calls of conspecific calls.

In this study, we trained male and female zebra finches in an operant task used to test individual vocal recognition for several vocalizers based on songs and DCs (Chapter 2). Subjects were tested on their recognition of up to 16 vocalizers by song and 12 vocalizers by DC at a time. We then assessed if bilateral neurotoxic lesions to NCM or HVC affected each subject's ability to recall previously learned vocalizers and to learn a new set of vocalizers. By analyzing task performance during the initial exposures to vocalizers before and after lesion, we distinguished between the recall of previously learned vocalizers and the learning or re-learning of those vocalizers. We found that lesions to NCM impair the zebra finches' ability to recall previously learned vocalizers but not necessarily their ability to (re)learn and discriminate between those calls. In contrast, lesions of HVC do not seem to have any effect on recognition memory nor the learning of vocalizers in the operant task, evidence that auditory memory for individual recognition and vocal motor pathways are dissociated.

3.3 Results

Operant conditioning for conspecific vocal recognition

We trained adult zebra finches (N=21) to recognize the calls and songs of several conspecifics for food reward in a modified go/no-go paradigm as described in Chapter 2 (Figure 3.1B). In short, subjects were presented with songs or DCs from a set of rewarded vocalizers (Re) and a set of non-rewarded vocalizers (NoRe). During the playback of a stimulus, the trial could be interrupted with a key peck that immediately starts the next trial. When motivated, birds maximize the rate of food reward output by interrupting vocalizations recognized as NoRe and waiting for vocalizations recognized as Re to finish. The relative odds of interrupting NoRe trials to Re trials, quantified by an odds ratio (OR), is used as a score of the subject's performance on this memory task (Equation 3.1). An OR significantly greater than 1 indicates successful task performance, while a score of 1 indicates performance at chance level.

Subjects were trained to recognize 12 conspecific vocalizers by DCs and 16 conspecific vocalizers by songs during a pre-lesion learning phase (Figure 3.2A). These two stimulus sets were referred to as DC S1 and Song S1, respectively. Each stimulus was learned over the course of a week in a "ladder" training procedure by gradually increasing the number of vocalizers tested each day: from 1 Re and 1 NoRe on the first day (labeled 1v1), to the full set on days 4 and 5 (labeled *6v6-d2* for DCs and *8v8-d2* for songs) (Figures 3.2B&C, see Materials and Methods for details). All subjects demonstrated scores significantly greater than 1 when evaluated on all trials presented during the *6v6-d2/8v8-d2* days (Figure 3.4).

We divided subjects into three experimental groups: bilateral neurotoxic NCM lesions

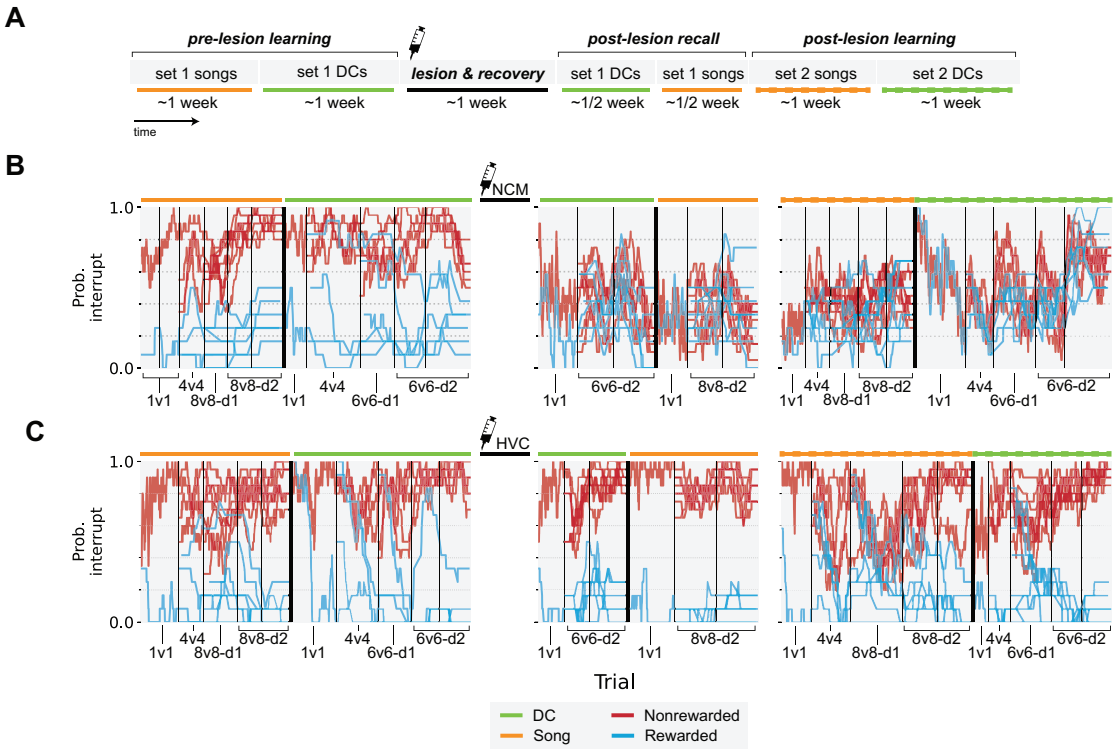


Figure 3.2: **Daily progression of memory ladder with lesions.** (**A**) Schematic of the three phases of the experiment in an example subject. First, the subject is trained on a set of songs followed by a set of DCs (pre-lesion learning of set 1). It then undergoes surgery during which NCM or HVC is lesioned (or saline/dye injected in controls) and up to one week of recovery. Next, it is tested on recall of the previously learned DCs, then songs (post-lesion recall of set 1). Finally, it is trained and tested on a new set of songs and DCs that were not introduced before lesion (post-lesion learning of set 2). The alternating pattern between song and DC sets was swapped on half of the subjects. (**B** and **C**) Example traces showing the probability of interruption per vocalizer during the three phases of the experiment in two subjects: in (B), a subject with a focal NCM lesion, and in (C), a subject with an HVC lesion. Each line shows the subject's probability of interrupting a single rewarded vocalizer, calculated in a sliding window (blue: rewarded stimuli, 12 trial bin width; red: non-rewarded stimuli, 20 trial bin width). Each day is separated by a thin vertical line. Color of the horizontal lines above the plots indicate the call type being tested (songs, orange; DCs, green). Annotations below the plot show the number of vocalizers tested each day. When the set size increases beyond 4v4, additional stimuli are introduced gradually: first with *6v6-d1* for DC or *8v8-d1* for song, during which new stimuli are presented three times more frequently than previously learned stimuli, followed by *6v6-d2* for DC and *8v8-d2* for song, during which there was no adjustment to presentation frequency.

Set #	Call type	# Vocalizers	Description
S1	Song	16	Songs first learned before lesion, re-tested after lesion
S1	DC	12	DCs first learned before lesion, re-tested after lesion
S2	Song	16	Songs first learned after lesion
S2	DC	12	DCs first learned after lesion

Table 3.1: Description of the stimulus sets used in the memory ladder.

(N=10, 5 male, 5 female), bilateral HVC lesions (N=7, 5 male, 2 female), and bilateral sham lesion controls (N=4, 3 male, 1 female). The extent and volume of lesions were validated with histology and the completeness of HVC lesions in male subjects was also validated by observing degraded song quality post-lesion. After lesion and recovery, subjects in all groups were retested on the vocalizers of S1 to test the retention of previously learned auditory memories, then finally trained and tested on two new sets of vocalizers (one with 16 songs, one with 12 DCs) collectively referred to as Set 2 (S2). The four stimulus sets used are summarized in Table 3.1, and the timeline of the experiment for an individual subject is summarized in Figure 3.2A.

Recall of learned vocalizers are affected by lesions to NCM but not HVC

To determine if lesions to HVC or NCM affected stored auditory memories of conspecific vocalizers, we compared task performance on the same set of vocalizations before and after lesion. Subjects were first trained and tested on S1 in the *pre-lesion learning* phase. Following lesion and up to one week of recovery, subjects were re-tested on S1 in the *post-lesion recall* phase.

All birds were able to learn S1 vocalizers well by the end of the initial training ladder. To quantify how quickly vocalizers were learned, we analyzed task performance as a function of informative trials seen. Here, an "informative trial" is defined as a non-interrupted trial (see Materials and Methods). The OR as a function of informative trials describes how birds improved at the task as they gained information about vocalizer-reward associations. When first exposed to the novel vocalizer stimuli of S1, healthy birds took about four informative trials to start interrupting Re from NoRe vocalizers at distinguishable rates (Figure 3.3A). We compared this initial rate of learning to task performance in sessions late in learning: the start of *6v6-d2/8v8-d2*. At the start of these sessions, subjects have had at least one prior session with all vocalizers in S1. We confirmed that NoRe trials were more likely to be interrupted than Re trials before even the first informative trial [$t(20) = 5.97$; $p < 10^{-5}$; one-sided Student's paired t test], demonstrating immediate recognition of previously learned

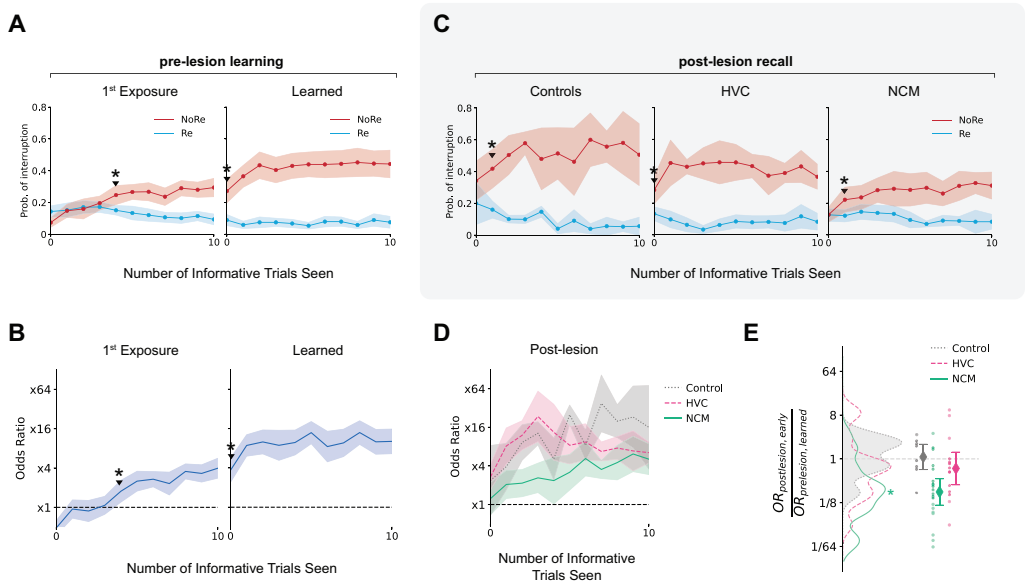


Figure 3.3: **Subjects with NCM lesions show deficits in recall of vocalizations learned before lesion, but can re-learn.** (**A**) Learning curves showing the probability of interrupting non-rewarded vocalizers (red) and rewarded vocalizers (blue) as a function of the number of informative trials seen of a given vocalizer (see Materials and Methods). Left: first 10 informative trials relative to the start of the S1 ladder before lesion. Right: first 10 informative trials relative to the start of the 4th day of the ladder, after which a subject would have had at least 1 day of experience with each vocalizer in S1. Shaded region in all figures indicate 2 SEM. The first bin where the Odds Ratio is significantly greater than 1 is labeled with an asterisk. (**B**) The same data from (A), translated into an odds ratio in each informative trial bin (see Materials and Methods). (**C**) Learning curves showing probability of interruption over the first 10 informative trials after lesion to vocalizers in S1 for the three groups: Control, HVC lesion, and NCM lesion. Probabilities in each bin are averaged over subjects and vocalizers. (**D**) As in B, the curves from (C) translated into an odds ratio for each informative trial bin. (**E**) Smoothed histograms showing the decrease in performance during the initial re-exposure to S1 vocalizers after lesion. $OR_{postlesion-early}$: OR measured before the third informative trial of each S1 vocalizer after lesion. $OR_{prelesion-learned}$: OR measured before the third informative trial of each S1 vocalizer starting from the fourth day of the ladder before lesion (i.e. *8v8-d2* or *6v6-d2* in Figure 3.2). Each dot represents the ratio $OR_{postlesion-early}/OR_{prelesion-learned}$ for one subject and vocalization type (i.e. song or DC). Diamonds show the mean for each group and error bars show 2 SE. Smoothed histogram was estimated using kernel density estimation in the log-space with a gaussian kernel and bandwidth of 0.5. Asterisk indicates distribution with mean significantly less than 1 after one-sample t test.

vocalizations and reward associations (Figure 3.3A). Qualitative inspection of interruption rates per-vocalizer over sessions and days (for example, Figures 3.2B&C) shows that interruption rates to Re and NoRe vocalizers separate quickly, often within a single session, and are stable by the end of the ladder on days *6v6-d2/8v8-d2*.

The early informative trials thus allow us to probe the recognition memory of a subject. A subject who recognizes a vocalization (and its associated reward) responds correctly starting from before the first informative trials of a session. We then compared this to the behavior of subjects immediately after lesions to NCM or HVC. OR for previously learned S1 vocalizations was measured as a function of informative trials *after* lesion in three groups: NCM, HVC, and controls. As the bird is exposed to more and more informative trials, it may have the ability to form new auditory memories in addition to recalling those acquired before lesion. Thus, we compared performance prior to the first informative trial (pure recall) and over the first 10 informative trial bins (recall with opportunity to re-learn).

Subjects with NCM lesions did not interrupt NoRe and Re vocalizers at different rates prior to the first informative trial [$t(9) = 0.68$; $p = 0.26$; one-sided Student's paired t test]. However, they managed to do so by the second informative trial [$t(9) = 2.71$; $p = 0.012$ (significant with false discovery correction)] (Fig. 3.3C). In contrast, subjects with HVC lesions did successfully distinguish NoRe and Re vocalizers before a single informative trial [$t(6) = 4.05$; $p = 0.003$], demonstrating that HVC lesions had little to no effect on the ability to store and recall the learned auditory memories needed for the task. The control subjects, who underwent surgery and the same recovery period as the NCM and HVC groups, also required one informative trial before the OR was statistically significant [$k = 0$, $t(3) = 1.19$; $p = 0.160$; $k = 1$, $t(3) = 2.80$; $p = 0.034$].

We then compared the learning curves out to $k = 10$ informative trial bins. We found that the NCM group's OR curve (Figure 3.3D) was significantly lower than both the HVC group's [$t(9) = -4.84$; $p = 7 \times 10^{-4}$; paired t test over 10 informative trial bins] and the control group's [$t(9) = -6.92$; $p = 4 \times 10^{-5}$]. There was no clear difference between the curves of the HVC and control groups [$t(9) = -1.01$; $p = 0.333$]. In the NCM group, task performance gradually recovered over several trials albeit slower than in the other two groups. This suggests that learning or re-learning was still possible despite lesion of NCM. Qualitatively, the post-lesion "learning" curve of the NCM group for S1 vocalizers resembles the initial learning curve on naive stimuli before lesion (Figure 3.3B, left), while the HVC and control group's curves resembles the performance curve for well-learned vocalizers in healthy individuals (Figure 3.3B, right), with perhaps a small deficit at the start of the post-lesion session.

The effect of lesion can be summarized with a scalar value $\delta = \frac{OR_{postlesion,early}}{OR_{prelesion,learned}}$, where $OR_{prelesion,learned}$ is the OR prior to (and including) the third informative trial during *6v6-d2/8v8-d2* before lesion, while $OR_{postlesion,early}$ is the OR measured prior to (and including) the third informative trial immediately after lesion (see Materials and Methods, Equation 3.6). We tested whether each group individually performed worse after lesion with $\delta < 1$; only the NCM group demonstrated worse performance after lesion [$t(19) = -4.91$; $p = 5 \times 10^{-5}$;

one-sample t test], while HVC ($p = 0.140$) and controls ($p = 0.611$) did not. The value of δ was also significantly different across at least two of the three groups [$F(2, 39) = 5.16$; $p = 0.010$; one-way ANOVA], which was found to be primarily driven by the difference between the control and NCM groups ($p = 0.017$; 95% CI, -4.43 to -0.37; Tukey's HSD post-hoc test), while the HVC group was indistinguishable from controls in the same window ($p = 0.635$; 95% CI, -2.94 to 1.36) (Figure 3.3E). Combined, these observations show that after NCM lesion, birds were unable to fully utilize the auditory memories they had formed before lesion, while HVC lesions had little to no effect on the birds' ability to recall those same associations.

Birds with lesions to NCM and HVC can form new auditory memories

After lesion, birds in all groups steadily increase in performance as they see more informative examples (Figure 3.3A-D). Over two full sessions of *6v6-d2/8v8-d2* after lesion, every bird in the HVC and control groups, and all but one in the NCM group, scored above chance level (Figure 3.4). Scores slightly fell compared to before lesion in the NCM group [song: $t(9) = 2.18$, $p = 0.001$; DC: $t(9) = 4.32$, $p = 0.029$; one-sided Student's paired t tests], while they did not in controls [song: $t(3) = -0.56$, $p = 0.691$; DC: $t(3) = 3.17$, $p = 0.975$]. The HVC group had mixed results, with a slight decrease in performance after lesion on song [$t(6) = 2.06$; $p = 0.043$] but not DC [$t(6) = 1.61$; $p = 0.079$]. These data are consistent with an interpretation that lesions to NCM had a destructive effect on previously learned auditory memories, but leave intact circuits and computations needed for learning or re-learning reward contingencies given renewed exposure and reinforcement. The drop in performance after lesion could also reflect a reduced capacity for auditory memory in the birds even though they are still capable of learning.

The post-lesion improvement on previously learned S1 stimuli could reflect a gradual recovery of existing memories without needing to acquire new ones. Could birds learn completely new, unfamiliar vocalizers after lesion? To answer this, we trained and tested lesioned subjects on a second stimulus set, S2, using the same training procedure and timeline as used for S1. We constructed OR learning curves for initial exposure to S2 in the same way that we did for S1 (Figure 3.5B). The learning curves for S1 and S2 can be compared by (1) how quickly OR increases above chance level, and (2) by magnitude of OR (the relative height of the curves). Before lesion, subjects distinguished Re and NoRe vocalizers in S1 by the 4th informative trial on average (Figure 3.5A). This was about the same as we observed for S2 after lesion across all groups (Figure 3.5C), suggesting that birds did not learn the new sets any faster or slower than expected. However, we found that for the same number of informative trials (out to $k = 20$), birds in the NCM group performed slightly worse than the HVC group [$t(19) = -5.05$; $p = 6 \times 10^{-5}$; paired t test over 20 informative trial bins] and controls [$t(19) = -4.69$; $p = 1.4 \times 10^{-4}$] (Figure 3.5D). There was no difference between the HVC group and controls [$t(19) = 1.01$; $p = 0.325$]. Thus, it appears that lesions to NCM

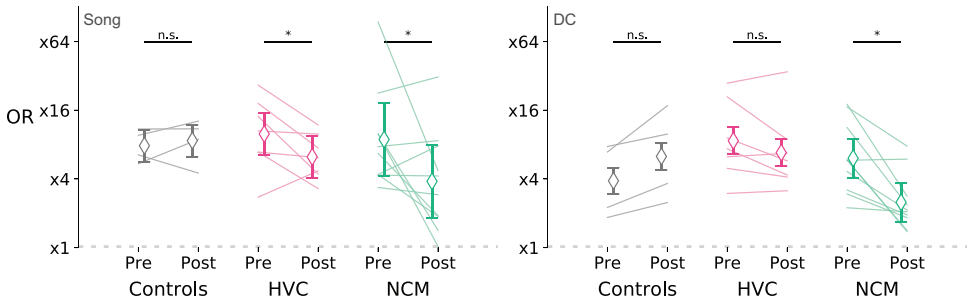


Figure 3.4: **Re-test of previously learned stimuli (S1) after lesion.** The effect of lesion on OR for songs (left) and distance calls (right) on the same set. OR was calculated on all trials during *6v6-d2/8v8-d2* with all ladder stimuli being played at equal rates. Individual lines show the change in OR for each subject, and diamonds show the mean OR pre- and post-lesion for the group. Significance was tested using a one-sided paired t test. Error bars show 2 standard errors of the sample differences from the paired t tests.

did not prevent the birds from learning to recognize new vocalizers, but the lesions may have had an effect on the speed at which they could be learned, or the total number that could be learned.

The learning curves for S2 suggest that learning may be slower or that the auditory memory capabilities are reduced in birds with NCM lesions. We next compared OR scores measured over the entire *6v6-d2/8v8-d2* sessions of S2 to see if the deficits persisted late in learning. When compared directly to the corresponding scores on S1 before lesion (Figure 3.6), control subjects had no clear change [song: $t(3) = 0.63$, $p = 0.286$, DC: $t(3) = 0.25$, $p = 0.410$; one-sided Student's paired t tests], while NCM subjects did worse on S2 than S1 overall [song: $t(9) = 1.89$, $p = 0.045$, DC: $t(9) = 4.66$, $p = 0.001$]. The HVC group did not do any worse on the new set of songs [$t(6) = -0.85$; $p = 0.572$] but did get significantly worse on DCs [$t(6) = 5.42$; $p = 0.001$]. While this could be indicative of an effect of lesion, the drop in scores could also be more simply explained if S2 distance calls were a more difficult task than the S1 distance calls (e.g. harder to tell apart acoustically). Despite the possibility that DC in S2 were more difficult, all subjects ultimately performed above chance level on the new set. Thus, fully intact HVC and NCM are not required for zebra finches to learn to recognize and discriminate between a new set of rewarded and non-rewarded vocalizers. However, there appears to be some deficit in auditory memory in NCM lesionsed subjects that prevents them from matching the same levels of performance in healthy birds and controls.

Finally, we considered if any lesion-induced difficulty in operating the task may explain the performance deficits we observed in NCM lesioned birds. Such difficulties included deafening, motor deficits that prevent the bird from operating the task apparatus, or cognitive deficits in which the bird no longer understands the task structure. We expected that such

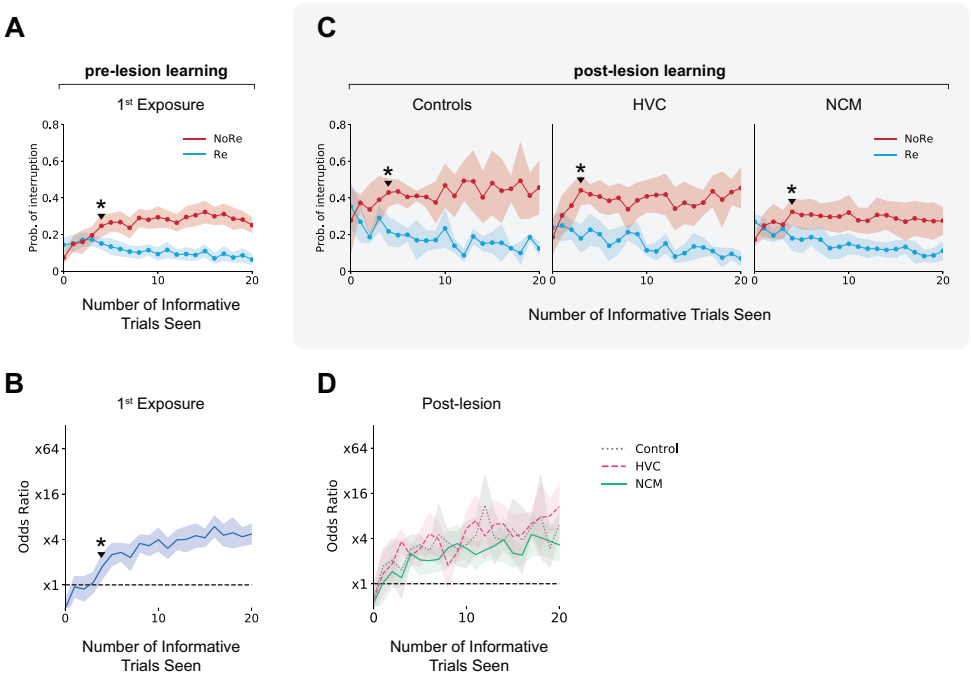


Figure 3.5: **Learning curves for a new set of vocalizers learned after lesion.** (**A**) The probability of interrupting non-rewarded (red) and rewarded (blue) vocalizers in the first 20 informative trial bins of Set 1, occurring before lesion (note that the x-axis scale has changed relative to Figure 3.3 to show the steady state of the behavior). The value in the kth bin is the probability of interruption prior to the $(k+1)$th non-interrupted (informative) trial. Shaded region in all sub-figures show 2 SEM. (**B**) The probability of interrupting non-rewarded and rewarded vocalizers in the first 20 informative trial bins of Set 1, occurring after lesion. (**C**) Data from informative trial bins shown in (A) converted to OR. (**D**) Data from (B) converted to OR and overlaid.

deficits, unrelated to the task of storing and recognizing auditory memories for individual vocalizers, would be apparent during the easiest sessions of the task after lesion: the days *1v1* with a single Re vocalizer and NoRe vocalizer. However, we found that the OR performance on *1v1* did not get worse after lesion in any group (Figure 3.7).

3.4 Discussion

Here, we establish a role of NCM in the ability to store and recall stored memories of dozens of vocalizers. We found that the ability to recall vocalizers learned before lesion was impaired in the immediate trials after lesion, lending further support to the hypothesis that NCM is a primary site of auditory memory storage. Animals rapidly recovered the ability to

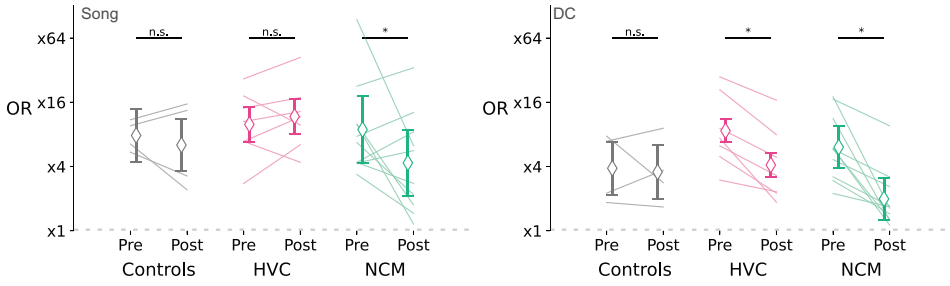


Figure 3.6: **Lesion effect on learning new vocalizers.** The effect of the lesion on OR for songs (left) and DCs (right) when learning a new set of vocalizers. OR was calculated on all trials during the two test days with all ladder stimuli being played at equal rates. Individual lines show the difference in OR for each subject, and diamonds show the mean OR pre- and post-lesion for the group. Significance was tested using a one-sided paired t test.

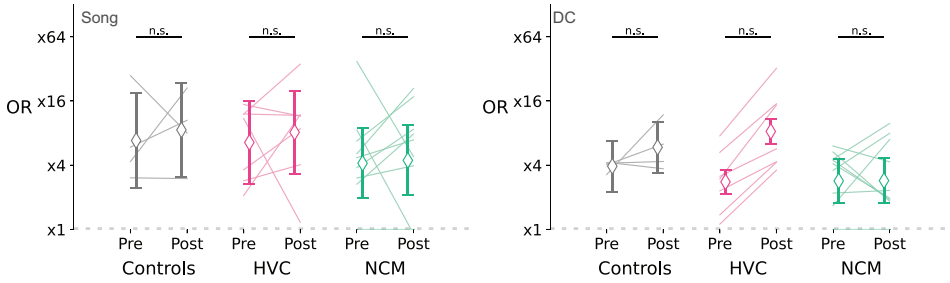


Figure 3.7: **Pre and post-lesion scores on *1v1* tests.** The effect of lesion on OR for songs (left) and distance calls (right) as measured during the *1v1* days before and after lesion. These sessions had only one rewarded vocalizer and one non-rewarded vocalizer. Significance tests using Student's paired t test.

recall memories as further informative trials were seen. In addition, birds with NCM lesions were still able to learn new vocalizer-reward associations, although slower than in healthy subjects. These results are consistent with a role for NCM in the storage and retrieval of auditory memories for vocalizations, and suggest other brain regions (CM, for instance) are also involved or can be recruited to solve the task of individual vocal recognition. In contrast, lesions to the vocal premotor region HVC did not have a discernible effect on auditory memory over the same period. This suggests that individual vocal recognition memory does not require HVC or other downstream nuclei of the vocal motor system. While this does not rule out a role for HVC in auditory perception, we could not discern any effect of HVC lesions on either the recall or formation of auditory memories for conspecific vocalizations.

Auditory memory for conspecific vocalizations in NCM

We tested zebra finches in an individual vocal recognition task, in which the DC and song of several individuals were learned and associated with reward in a multi-day song ladder training procedure. We found that subjects with NCM lesions could not recognize previously learned vocalizers as quickly, and that their maximum performance was decreased even though they were still capable of learning. Some trivial explanations for the observed deficits in task performance were ruled out. Trivial causes, such as deafening or inability to understand or operate the task apparatus, were ruled out by observing that the lesioned birds performed well in "easy" sessions when they only needed to compare the vocalizations of two individuals (Figure 3.7). We can also rule out a drop in motivation levels, as trials were self-initiated by the subjects and that lesioned subjects continued to work for a similar number of fed trials per session.

What aspect of the auditory memory behavior is affected by NCM lesion? One possibility is that one of NCM's primary roles is to recognize vocalizations as being known or unknown. Previous studies have shown the existence of changes in the activity levels in NCM based on past stimulus exposures. For example, NCM responses show stimulus-specific habituation to repeated presentations of conspecific calls and songs (Chew et al., 1996) and lower response strengths to learned vocalizations (Thompson & Gentner, 2010). Furthermore, the coding in NCM is modulated by the familiarity and behavioral salience of auditory stimuli (Chew et al., 1996; Pinaud & Terleph, 2008; Theilman et al., 2021). In such a model, NCM could be responsible for storing known acoustic signals for communications in order to recognize that an incoming auditory signal comes from a known source, while activating networks in other brain regions responsible for the semantic of behavioral representations. These other brain regions could include the interconnected auditory region CM whose activity is also modulated by task-relevant stimuli (Meliza & Margoliash, 2012; Pinaud & Terleph, 2008), more abstract representations for goal-directed action in NCL (nidopallium caudolateral) (Rinnert & Nieder, 2021), and social and reproductive behaviors associated with HVC and nearby regions (Maguire et al., 2013). The transformation from recognition of acoustic signals of a communication sound to its semantic representation could ultimately be distributed widely across the brain, as has been observed in the cortical representation of human speech (Huth et al., 2016).

Another possibility is that neurons in NCM over-represents the acoustic features that distinguish known vocalization sources, or contains neurons with selective or invariant responses to learned auditory objects In song learning, it has been observed that NCM is necessary for pitch restoration in the songs of adult males taught to shift their pitch through negative reinforcement (Canopoli et al., 2014) and that pharmacological blockade of signaling pathways in high-level auditory areas including NCM prevent song imitation quality in juvenile males (London & Clayton, 2008). In addition, single neurons in NCM have been found to be sparse, with sharper tuning for specific sound stimuli (Kozlov & Gentner, 2016; Perks & Gentner, 2015). The increased sparsity may be a consequence of complex, multi-component spectro-temporal receptive fields sensitive to specific sound combinations (Kaardal et al.,

2017; Kozlov & Gentner, 2014). These neurons may represent vocalizations as patterns in a distributed code, with populations of neurons selective for small numbers of individuals based specific acoustic features or combinations of features. In such a model, incomplete lesions may still allow the network to recover previously learned associations, but perhaps with lower fidelity or capacity. The task design using large vocalizer sets of 12-16 individuals was critical in revealing deficits in memory and recognition that smaller sets did not (e.g. Figure 3.7). Thus, using a large number of vocalizers as stimuli in tests of individual vocal recognition is important, to require the animals to engage with the task and make it difficult to solve with trivial strategies. In this experiment, the size of the stimulus set may tax the capacity of the system, or decrease the separation between acoustic and neural representations of distinct vocalization sounds. However, it is currently unknown how categorical boundaries are formed in higher-order auditory areas for the calls of different individuals. Repeating these memory and lesion experiments using synthetic stimulus sets that artificially distort or blur the acoustic boundaries between the calls of known individuals may shed light on the nature of the acoustic features of stored memories in NCM.

Implications for HVC

HVC is the critical interface between the auditory and song systems (Margoliash, 1997; Roberts & Mooney, 2013). It gates the auditory information needed to give rise to song selective neurons found throughout the vocal production pathway and the anterior forebrain pathway for song learning (Roberts & Mooney, 2013; Vicario & Yohay, 1993). We used lesions to HVC as a proxy to test the necessity of the vocal motor pathways on auditory memory for individual vocal recognition. In contrast to NCM, we saw little to no effect on task performance when HVC was lesioned. Thus, we believe that a motor representation of conspecific communication sounds is not necessary for basic perception, categorization, memory, or recognition functions for those sounds. However, this does not rule out the possibility that motor pathways could modulate or otherwise participate in these processes, or be engaged in specific behavioral contexts. For example, altered courtship behavior and mate preference caused by lesions to HVC in females (Brenowitz, 1991; Del Negro et al., 1998; Perkes et al., 2019) could be explained by changes to behavioral and sexual preferences, rather than a change in perception or ability to distinguish individuals. It is also possible that HVC will only be engaged in auditory perception in more naturalistic scenarios than the artificial setting in our memory ladder experiments, or when vocal responses are necessary such as in antiphonal calling (D'Amelio, Klumb, et al., 2017; Ma et al., 2020) and song dueling behaviors (Alcami et al., 2021).

Relation to other auditory areas involved in learning and memory

NCM is also interconnected with CM (caudal mesopallium), another secondary auditory region known to have signals correlated with task-relevance and learned sounds (Jeanne et al., 2011; Meliza et al., 2010). This area is subdivided into a lateral portion (CLM) which

receives direct projections from Field L, and a medial portion (CMM) which receives indirect projections via CLM. Neural populations in these areas show changes in response properties and network properties in response to learned auditory stimuli (Gentner, 2004; Jeanne et al., 2011; Meliza et al., 2010; Theilman et al., 2021). Furthermore, lesions in CM in some songbird species may be responsible for specificity of female song preferences, a behavioral function that is typically associated with HVC (MacDougall-Shackleton et al., 1998). In our current study, we do not address CM and instead focus our attention on NCM and HVC. However, CM also merits careful consideration as another potential site for auditory memory for conspecific vocalizers or as part of a memory network with NCM through the reciprocal connection between the two regions.

Potential limitations of interpretation

The task structure used in these lab experiments has some limitations that prevent overly broad interpretations of the results. Individual vocal recognition among conspecifics in nature is an innate behavior that occurs as a natural consequence of a bird's social interactions and relationships. Our operant task, using a learning "ladder" to progressively train zebra finches on large and larger sets of vocalizers, is meant to mimic this behavior but lacks the social elements and richness of behavioral response. First, while we limited our stimuli to song and DC, zebra finches have a larger repertoire of call types than tested here (Zann, 1996), several of which may carry individually distinctive features (Elie & Theunissen, 2016). Second, while vocalizations in our task were tied to a binary response (go/no-go) for food reward, the space of behavioral responses to the calls of conspecifics in natural scenarios is richer and may vary based on existing social relationships and context. For example, in the territorial song sparrow, playbacks of known neighbors and unknown intruders elicit different levels of aggression (Kroodsma, 1976), and zebra finches modulate their antiphonal response rates based on their relationship with the other individual (D'Amelio, Trost, et al., 2017). Finally, intensive operant conditioning on sound stimuli using playbacks may influence perception, the underlying neural code, and even use different neural pathways (Bennur et al., 2013). Live social interaction has been shown to influence the quality of song learning in birds (Chen et al., 2016; Eales, 1989; Yanagihara & Yazaki-Sugiyama, 2019), and social reinforcement in the form of video playback of other birds can be sufficient motivation for learning in an operant task (Macedo-Lima & Remage-Healey, 2020). Thus, we must consider the possibility that individual vocal recognition in natural settings with social consequences may engage different pathways and areas of the brain than those tested in our experiment.

The method of chronic neurotoxic lesions also has some intrinsic limitations. These lesions were performed by injecting a neurotoxic agent at a specific location in the brain and allowing it to diffuse into a local volume which is not necessarily confined to the target site. NCM presents a particular challenge because of its size and lack of well defined anatomical boundaries (Stripling et al., 2001). In separate subjects, we used ibotenic acid and NMA as lesioning agents as we modified our injection procedure to maximize the lesioned volume of NCM (Table 3.3). We attempted to make our lesions quite large but cannot rule out the

possibility (or likelihood) that some parts of NCM remained intact, or that off-target brain regions were partially lesioned. Adjacent regions such as primary auditory regions of Field L, the hippocampus, CM, and NCL (nidopallium caudolateral) may have been affected by these off-target lesions. In contrast, HVC is a smaller brain region and more well defined anatomically (Foster & Bottjer, 1998; Nottebohm et al., 1976) making validation of these lesions easier. Furthermore, the degradation of song and distance call in male zebra finches after HVC lesions is a well known effect (Nottebohm et al., 1976; Simpson & Vicario, 1990) and was used to validate HVC lesions.

Auditory memory and recognition is a complex process, and the chronic, irreversible lesions used in this study also cannot be used to make conclusions about some aspects of the strategies employed by the animals to solve the task. In our analysis, we used the task performance in the first informative trials after lesion to show that the NCM lesioned group could not respond correctly without at least one reward example post-lesion. This analysis is limited by how few trials there are available to analyze in this window—due to the small sample sizes, the analysis required averaging over all vocalizers and subjects in a group and the effect of lesion in a single subject for one vocalizer could not be measured. As subjects see more examples, the effect on previously learned memories and the acquisition of "new" memories become intertwined. During this time, other brain regions (e.g. other areas of the auditory system) could compensate or solve the task using different features of the stimuli. Adult neurogenesis has also been observed in NCM of the zebra finch (Pytte et al., 2010) which may also facilitate recovery of memory function in the weeks after surgery, though the effect and role of new neurons in these areas are not well understood. In the future, acute, reversible manipulations (e.g. using electrical stimulation or optogenetics) during specific phases of the learning ladder or during individual trials could be used to better distinguish the role of NCM on learning, storage, and retrieval of auditory memories and reward associations. Particularly interesting targets for manipulation are the dopaminergic innervations of NCM that have been found to correlate with learning of auditory stimuli (Chen et al., 2016; Macedo-Lima et al., 2021). Identifying and manipulating a reward signal projecting to NCM, if it exists, could help to disambiguate between memory for individual recognition, and memory for reward associations.

Familiarity versus recognition

Finally, while we interpret performance in the operant task as indicative of the individual vocal recognition capacity of the zebra finch, we also should consider if task performance could be tied to stimulus familiarity rather than individual recognition. This is particularly relevant to NCM, where previously identified neural correlates of memory are often related to repeated presentations of a stimulus, observed through decreased expression of IEGs (Bolhuis et al., 2001) and habituation (Chew et al., 1995). Given lower baseline activation levels in response to familiar stimuli, it is easy to imagine a simple circuit that uses familiarity as a reward cue without requiring the animal to interpret the stimulus sounds as belonging to a single individual. Indeed, our task design leaves open this possibility: we present

NoRe stimuli four times more frequently than Re stimuli in order to limit daily reward output and maintain subject motivation levels, as was done in (Elie & Theunissen, 2018) and the experiments of Chapter 2. A side effect of this, however, is that a subject could take advantage of this by using familiarity to a stimulus as a proxy for "non-rewardedness".

A related factor in our task structure is that incorrect responses to NoRe and Re stimuli provide asymmetric feedback to the animal: incorrect responses to NoRe (no interrupt) are promptly punished by the lack of food reward, while incorrect responses to Re (interrupt) do not give the subjects any new information, as the next trial begins immediately (Figure 3.1B). A consequence of this asymmetry is that the optimal solution to a forgotten or unknown stimulus should be to not interrupt (in order to gain information about the novel stimulus), which is identical to the correct response to a rewarded stimulus. We did not test whether birds actually adopt this "optimal" strategy on novel, oddball stimulus presentations. If they do, it leaves open the possibility that subjects may only need to recognize non-rewarded stimuli in order to prompt an interrupt response, while defaulting to non-interruption on all other stimuli, recognized or not.

Taken together, the task design of the present study (1) leaves open the possibility that a subject only needs to recognize and interrupt half the vocalizers in the set, and (2) leaves open the possibility that familiarity, and not vocal recognition, is the driver behind successful task behavior. It is some consolation, however, that recognition that a vocalization is familiar can be considered a prerequisite to recognition of an individual vocalizer itself. In any case, this ambiguity can be addressed in future studies by presenting all vocalizers with equal frequency, or by mixing the presentation frequency of rewarded and non-rewarded stimuli, while using probabilistic reward output to control motivation levels.

3.5 Materials and Methods

Operant conditioning

Animals

All animal procedures were approved by the Animal Care and Use Committee of the University of California, Berkeley (AUP-2016-09-9157) and were in accordance with the National Institutes of Health guidelines regarding the care and use of animals for experimental procedures.

Domestic zebra finches (*Taeniopygia guttata*) used in behavioral and lesion experiments were raised in our breeding colony. For these experiments, 13 adult male and 8 adult female zebra finches were chosen and divided into the following experimental groups: 10 NCM, 7 HVC, and 4 control. Subjects were housed in a colony room (usually 10 to 30 individuals in a large flight cage). During the course of the experiment (approximately 2 months, see "Ladder training procedure") a subject was housed separately from the main colony, either in an individual cage or in a shared cage with an opposite sex individual also part of the experiment.

Recordings of song and DC stimuli originated from multiple labs and originally described in (Yu et al., 2020). Song vocalization recordings were from 32 male zebra finches from the Theunissen Lab at UC Berkeley, the Perkel laboratory at the University of Washington, and the Leblois laboratory, Bordeaux (France) Neurocampus. DC vocalizations came from 24 zebra finches (12 male and 12 female), all from our colony at UC Berkeley. Vocalizations used as stimuli were recorded as part of previous experiments in the laboratory, and the vocalizers were unfamiliar to the subjects in the present study. The 12 male DCs were produced by a subset of the males also used in the song stimulus set—however, reward associations were randomized (7 switched, 5 same). Previous work has shown that it is unlikely that zebra finches can generalize vocalizer identity from one call type to another (Elie & Theunissen, 2018), so for the purposes of this study the DC and song stimuli recorded from the same bird are treated as separate individual vocalizers.

Testing apparatus and software

The behavioral task and apparatus are identical to those described in Chaper 2. Briefly, subjects were placed in an operant chamber set up with a speaker, food hopper, water bowl, and orange backlit pecking key (Med Associates). The system is operated using a custom fork of the Python-based Pyoperant software[1], originally developed[2] by J. Kiggins and M. Thielk in T. Gentner's laboratory at University of California San Diego.

Subjects were tasked with discriminating between a set of rewarded and non-rewarded individuals based on the playback of their vocalizations. Subjects initiate trials by pecking on the backlit key and a 6 second stimulus playback begins (Figure 3.1B). After 6 s, stimulus playback ends and either nothing happens (NoRe trial), or a reward is given by raising the food hopper for 12 s (Re trial). Alternatively, a subject may peck the key at any time during the 6 s playback period to terminate the trial and begin a new trial with a random stimulus. In this case, no food reward will be given regardless of if the initial trial was rewarded or non-rewarded. To maximize the rate at which reward is received in a session, subjects learn to skip stimuli that are recognized as non-rewarded to avoid the full waiting period and move on to the next trial. By design, 20% of trials are rewarded while 80% of trials are not rewarded.

Subjects are food restricted with access to water but limited seed in between test sessions to maintain motivation. Subjects were weighed before and after every test session, and seed consumed in a daily session was measured and supplemented at the end of day so that the birds maintain their weight within 10% of their starting weight. Daily handling of subjects did not seem to affect the birds' motivation or ability to do the task once they became comfortable with the experiment chamber. Once trained, birds are able to get all of their daily food allowance during the testing period.

The birds learn to use the apparatus during a shaping session that lasts approximately 1 week. During the shaping session, the bird first learns to associate pecking of the key with

[1] `https://github.com/theunissenlab/pyoperant`

[2] `https://github.com/gentnerlab/pyoperant`

sounds and food reward and then learn to interrupt non-rewarded sounds. The initial shaping task involves the discrimination of two clearly distinct song stimuli. We have also performed control experiments, clearly showing that apparatus is not providing any extraneous clues that the birds could use to distinguish rewarded from non-rewarded trials (Elie & Theunissen, 2018).

Stimulus preparation

For each individual vocalizer used as a stimulus in the task, we prepared 10 unique stimulus files composed of calls or song motif from that individual. In this paper, we use the term *vocalizer* to refer to the collection of 10 unique stimulus files of calls or song motifs from the same individual. We use this term to contrast with *rendition*, which refers to a single stimulus file out of the 10. These multiple renditions were used so that specific extraneous acoustic features of a particular stimulus file not encoding vocalizer identity (e.g. length, intensity, and background noise) could not be used as a reward cue. In Chapter 2, we showed that birds generalized their behavior over the 10 renditions from the same individual (Figure 2.4A&B), and so our analyses of task performance in this study are done at the *vocalizer* level.

Song stimuli were constructed by combining 3 example songs from one individual, while DC stimuli were constructed by combining 6 example DCs. Each song example consisted of a single song motif. Most introductory notes were removed to avoid great variability in stimulus duration. A DC example consisted of a single call, or in some cases a pair of calls if the vocalizer did not normally produce single, isolated distance calls. These examples were arranged with pseudorandom intervals such that the duration of the file would be exactly 6 seconds long. Amplitudes of the audio files were then normalized within stimuli of the same type, i.e. songs or DCs. Example spectrograms of stimulus playback files are shown in Figure 2.6.

Initial training and post-lesion tests

The full stimulus sets in these experiments included playbacks of the songs of 16 different vocalizers or the distance calls of 12 different vocalizers. Our "ladder" training procedure is designed to gradually introduce more vocalizers to a subject each day so that they are not overwhelmed by the full stimulus set right away. The five day procedure is described in Table 3.2. Each day consists of one continuous session of approximately 8 hours. Days 1-3 introduce at least 1 new Re and 1 new NoRe vocalizer each day, while Days 4 and 5, referred to collectively as *6v6-d2/8v8-d2* throughout the main text, do not introduce new vocalizers and are used for measuring overall performance.

The four stimulus sets, two sets of songs and two sets of DCs, are described in Table 3.1. In each set, half of the vocalizers were assigned to be Re and the other half assigned to be NoRe. The designation of Re or NoRe was flipped for half of the subjects (i.e. a vocalizer that was rewarded for one subject may be non-rewarded for another subject). Two of the

stimulus sets (Song S1 and DC S1) were taught to all subjects using the ladder training procedure. After lesion or sham lesion, subjects were re-tested on the stimulus sets of S1, and then trained on two new stimulus sets (Song S2 and DC S2) using the same ladder training procedure as the initial learning of S1 (Figure 3.2). The sets S2 did not feature any overlapping stimuli from the sets S1.

Lesions

Surgery

Following the behavioral tests of S1, subjects received either bilateral NCM lesions ($n = 10$), bilateral HVC lesions ($n = 7$), or sham lesions (controls, $n = 4$). Birds were food deprived for one hour prior to anesthesia and orally administered 0.5 mg/kg Meloxicam as analgesic. Birds were induced at 4% isoflurane and head-fixed into a stereotaxic apparatus (Kopf Instruments) on a stabilizing air table (Kinetic Systems). Anesthesia was maintained at 1% isoflurane. A subcutaneous injection of one drop of lidocaine (about 100 μl of 2% solution) was administered as local anesthetic. Head feathers were removed, the scalp sterilized with sterile alcohol wipes and povidone-iodine swabs, an incision was made along the midline, and the skin retracted. A craniotomy was then opened over each hemisphere around the desired injection coordinates. The mid-sagittal sinus (Y-sinus) was identified as the stereotaxic zero coordinate.

Bilateral excitotoxic lesions were made using 2% *N*-methyl-DL-aspartic acid (NMA) solution (6 NCM birds), or 0.7% ibotenic acid (IBO) (4 NCM birds, 7 HVC birds) (see Table 3.3 for protocols used). Stereotaxic coordinates for NCM and HVC were taken from existing literature and adjusted based on our own histological verification of lesioning outcomes. Each injection at one coordinate (medial/lateral, rostral/caudal) was performed at either one or two depths. For each site, a glass micropipette with tip diameter 25-50 microns was lowered using a hydraulic micro-manipulator. The pipette was lowered to the deepest injection location and solution was injected at about 2 nl/s using the Nanoject II system (Drummond Scientific Company). Before retraction to the next site or out of the brain, the pipette was left in place for 5 minutes to minimize backflow. For control subjects, the method of sham lesion varied. In two control subjects, dye was injected using the same coordinates used for the NCM group. A third control was originally in the NCM group but the lesion was found to be exceptionally small, most likely due to a clogged pipette during surgery. A fourth control was originally in the HVC group and targeted with ibotenic acid, but the lesion was found to be off target in both hemispheres; singing behavior and song quality were also not affected post-operation, further confirming the off-target lesion and validity as a control. At the conclusion of surgery, craniotomies were covered with Kwik-Cast, the skin surface sealed with Vetbond tissue adhesive, and bacitracin ophthalmic ointment was applied to prevent infection.

Subjects were then returned to their home cages with ad-lib feed. Recovery time varied by subject and ranged from 2 full days to 7 full days (mean=4.8 days). After recovery,

Day	DCs	Songs	Description
1	*1v1*	*1v1*	1 rewarded (Re) vocalizer, 1 non-rewarded (NoRe) vocalizer
2	*4v4*	*4v4*	3 Re vocalizers added (4 total) and 3 NoRe vocalizers added (4 total)
3	*6v6-d1*	*8v8-d1*	DCs: 2 Re vocalizers added (6 total) and 2 NoRe vocalizers added (6 total). Songs: 4 Re added (8 total) and 4 NoRe added (8 total). Newly added Re vocalizers are played 4 times more frequently Re vocalizers from day 2. Newly added NoRe vocalizers are played 4 times more frequently than NoRe vocalizers from day 2.
4	*6v6-d2*	*8v8-d2*	No new vocalizers added. All Re vocalizers played at the same frequency. All NoRe vocalizers played at the same frequency.
5	*6v6-d2*	*8v8-d2*	Repeat of Day 4

Table 3.2: Description of 1 week ladder training procedure referenced in the main text and illustrated in Figure 3.2.

Target	N	(L [mm], R [mm])	Depths [mm]	Injection
NCM	2	(0.4, 0.3)	1.5, 2.0	230nl 2% NMA
		(0.4, 0.8)	2.25	230nl 2% NMA
NCM	3	(0.5, 0.4)	1.5, 2.0	230nl 2% NMA
		(0.5, 0.8)	2.0	230nl 2% NMA
NCM	1	(0.5, 0.5)	1.5, 2.0	300nl 2% NMA
		(0.7, 0.8)	1.8	300nl 2% NMA
NCM	4	(0.5, 0.5)	1.0, 1.5	147nl 0.7% IBO
		(1.0, 0.5)	1.8	200nl 0.7% IBO
HVC	4	(2.2, -0.2)	0.4	200nl 0.7% IBO
		(2.2, 0.1)	1.8	200nl 0.7% IBO
HVC	3	(2.2, -0.1)	0.5	147nl 0.7% IBO
		(2.2, 0.1)	0.5	147nl 0.7% IBO
		(2.4, 0.1)	0.5	147nl 0.7% IBO

Table 3.3: Protocols for NCM and HVC lesions. All protocols were applied bilaterally as described in the main text. Coordinates are in mm relative to the Y-sinus $(0.0, 0.0)$. N: Number of subjects the protocol was applied to; L: Lateral distance from zero; R: Offset in rostral/caudal axis (positive numbers rostral); D: Depth from surface of brain; NMA: *N*-methyl-DL-aspartic acid; IBO: Ibotenic acid.

Trials	T_0	T_1	T_2
111	1	1	1
11001	1	1	3
10101	1	2	2
001011	3	2	1

Table 3.4: Examples of informative trial counts. 0 indicates an interruption, and 1 indicates a non-interruption, i.e. informative trial. T_k is the number of trials between the kth informative trial (exclusive) and the $(k+1)$th informative trial (inclusive).

subjects continued with the operant task as described in Figure 3.2 being re-tested on the stimulus set S1.

Statistical analyses

Definition of informative trials

To describe how a subject's behavior in the task changed as they gained experience with the stimuli, we defined an *informative trial* to describe a trial in which a subject had an opportunity to learn the reward contingency of a stimulus. Because of the task structure's asymmetric treatment of interrupts and non-interrupts (Figure 3.1B), subjects could only learn if a stimulus was rewarded or not if they refrained from interrupting the 6 second playback. Thus, we define an informative trial as a *non-interrupted trial*.

We used this definition to analyze task performance as a function of *number of informative trials seen* of a given vocalizer. The presentations of a vocalizer v can be divided into bins indexed by k, where k is the number of informative trials preceding a given trial. To formalize this, we define the integer value T_k^{sv} as the empirical number of trials between the kth (exclusive) and $(k+1)$th (inclusive) non-interrupted trial of a vocalizer v by subject s. Examples of how T_k is evaluated are shown in Table 3.4, and illustrated in Figure 2.5.

We allow this definition of informative trials to cross over multiple sessions, given our ladder structure in which subjects are tested on the same vocalizers over multiple days (Figure 3.2). To compare learning during different time periods (e.g. late in learning before lesion to first trials after lesion as in Figure 3.3A&B), we can use the same definition of informative trials but begin counting k from the start of the relevant period.

Task performance measured by Odds Ratio

As in Chapter 2, we quantify task performance of a single subject using the odds ratio (OR) of interrupting non-rewarded (NoRe) vocalizers to rewarded (Re) vocalizers.

$$OR = \frac{p(\text{int}|NoRe)}{1-p(\text{int}|NoRe)} \frac{1-p(\text{int}|Re)}{p(\text{int}|Re)} \tag{3.1}$$

To measure overall task performance on a given session or set of sessions, we estimate OR and 95% CIs using the Fisher's exact test. With a contingency matrix given by Table 2.1 generated from the behavioral data, Fisher's exact test estimates the odds ratio as $OR = \frac{ad}{bc}$ and evaluates significance by calculating the probability of obtaining an OR as extreme (equal or greater) by calculating the distribution OR over all possible contingency matrices with the same marginals as the observed data. Zero values in any cell would cause the OR to be undefined or unbounded. To address this, we use the Haldane-Anscombe correction by adding 0.5 to all cells before performing Fisher's exact test.

Calculation of learning curves

The learning curves in Figures 3.3 and 3.5 show the average probabilities of interrupting NoRe and Re vocalizers, and the associated OR, as a function of informative trials seen. These quantities represent averages over subjects and vocalizers. Here, we compute these curves given trial data from a set of subjects S responding to a stimulus set of vocalizers $V_{Re} \cup V_{NoRe}$, where V_{Re} is the set of rewarded vocalizers and V_{NoRe} is the set of non-rewarded vocalizers. For all analyses, distance call and song datasets were treated separately.

For each informative trial bin k, we estimate the probability that subject $s \in S$ interrupts a vocalizer $v \in V$ as a function of the size of the informative trial bin T_k^{sv}:

$$p_s(\text{int}|v,k) = \frac{T_k^{sv} - 1}{T_k^{sv}} \tag{3.2}$$

To get a subject's probability of interrupting a Re or NoRe vocalizer, we average over all vocalizers with the same reward contingency using Equation 3.3.

$$p_s(\text{int}|V,k) = \frac{1}{|V|}\sum_{v}^{V} p_s(\text{int}|v,k) \tag{3.3}$$

where $V \in \{V_{Re}, V_{NoRe}\}$ indicates reward contingency. This quantity may equal 0 if the subject s did not interrupt any vocalizer between informative trials k and $k+1$ and $T_k^{sv} = 1$ for all $v \in V$. The most likely situation in which this would occur is when a subject is performing with perfect accuracy and does not interrupt vocalizers in V_{Re} at all. This would cause numerical issues when computing odds and odds ratios (e.g. Equation 3.1). To address this exception, we replace $p_s(\text{int}|V,k)$ with $p'_s(\text{int}|V,k)$, equal to $\frac{1}{2}$ the average probability of interruption across all other subjects:

$$p'_s(\text{int}|V,k) = \frac{1}{2}\frac{1}{|S|-1}\sum_{s' \neq s}^{S} p_{s'}(\text{int}|V,k) \;\text{ if }\; \{T_k^{sv} = 1 \;\forall\; v \in V\} \;\text{ else }\; p_s(\text{int}|V,k) \tag{3.4}$$

We note that there is one more exception where p'_s can be 0: if $p_s(\text{int}|V,k) = 0$ for all subjects. In this case, we would approximate the probability of interruption as $\frac{1}{2|S||V|}$. Luckily this case did not occur during our analysis.

The learning curve logOR(k) (i.e. Figure 3.3B and Figure 3.5) is estimated by averaging the log odds ratio over subjects for each informative trial bin k:

$$\text{logOR}(k) = \frac{1}{|S|}\sum_{s}^{S}\left[\log\frac{p'_s(\text{int}|V_{NoRe},k)}{1-p'_s(\text{int}|V_{NoRe},k)} - \log\frac{p'_s(\text{int}|V_{Re},k)}{1-p'_s(\text{int}|V_{Re},k)}\right] \tag{3.5}$$

The quantity in brackets is the difference between the log odds of a subject interrupting a NoRe vocalizer $\log(\text{Odds}_{s,NoRe})$ and the log odds of a subject interrupting a Re vocalizer $\log(\text{Odds}_{s,Re})$. Each bin k is tested for significance using a paired t test over $|S|$ subjects comparing $\log(\text{Odds}_{\text{s,NoRe}})$ and $\log(\text{Odds}_{\text{s,Re}})$. To determine the first significant bin, interpreted as the fewest number of informative trials before NoRe vocalizers are interrupted more frequently than Re vocalizers, we applied this significance test to each informative trial bin k. The first significant bin is determined to be the smallest k for which the test is significant after applying the Bonferroni correction for multiple comparisons. All subsequent bins are assumed to be significant.

Analysis of initial lesion effects

In Figure 3.3E we compare the overall task performance before and after lesion using a quantity δ, representing the difference in performance before and after lesion. The goal was to compare a time period before lesion during which the stimulus set was well learned, to a period *immediately after lesion* before a subject could have had a chance to re-learn the stimulus reward associations. We previously have shown that learning can be significant within 3 informative trials upon initial exposure to a vocalizer; thus for this analysis we chose to restrict the analysis to trials prior to and including the 3rd informative trials after lesion per vocalizer. For clarity, this time period is labeled *postlesion,early*. The odds ratio computed in this window is analogous to Equation 3.5 but computed for a range of bins $k < 3$:

$$\text{logOR}_{s,k<3} = \log\frac{p'_s(\text{int}|V_{NoRe},k<3)}{1-p'_s(\text{int}|V_{NoRe},k<3)} - \log\frac{p'_s(\text{int}|V_{Re},k<3)}{1-p'_s(\text{int}|V_{Re},k<3)} \tag{3.6}$$

where the probability of a subject interrupting a vocalizer is expanded to the set of trials where the number of informative trials seen is less than 3, or $k < 3$:

$$p_s(\text{int}|v,k<3) = \frac{T^{sv}_{k<3}-3}{T^{sv}_{k<3}} \text{ where } T^{sv}_{k<3} = \sum_{k=0}^{2} T^{sv}_{k} \tag{3.7}$$

Values of Equation 3.6 evaluated on *postlesion,early* trials were then compared to a corresponding time period before lesion, which we labeled *prelesion-learned*. The trials for *prelesion,learned* were taken starting from the first *6v6-d2/8v8-d2* days of the ladder. These days were selected because they were the first days pre-lesion when all vocalizers had been presented for at least one day previously. The change in performance for a subject is measured by δ:

$$\log\delta = \log\mathrm{OR}_{postlesion-early} - \log\mathrm{OR}_{prelesion-learned} \tag{3.8}$$

Acknowledgements

The work in this chapter was made possible with the help of undergraduate research apprentices in animal training and behavioral testing: I. Rice, A. Prasad, R. Vu, C. Chen. W.E. Wood contributed equally to experimental design and investigation. Histological slice imaging and digitizing performed by L. Johnston at the CRL Molecular Imaging Center, RRID SCR_017852, supported by UC Berkeley Biological Faculty Research Fund.

Chapter 4

Neural encoding of learned communication calls in the anesthetized zebra finch

4.1 Abstract

Zebra finches are social animals who interact using vocal communication and can recognize other individuals by their calls. This requires the mapping of an incoming sound to a memory of a known individual. Secondary auditory regions in the avian brain, analogous to human association cortex, are involved in the memory of learned sounds. Previous studies have shown how prior experience with sounds can change tuning and response properties throughout the auditory system. In this study, we analyzed single unit spiking activity in cortical-like areas of the zebra finch auditory system and found evidence that the information coding capacity of single neurons in response to conspecific vocalizations was modulated by the bird's prior experience with those calls. In birds who had been trained to recognize the calls and song of several conspecific individuals in an operant task, the spiking reliability of neurons in response to task-relevant stimuli was greater than in response to non-task stimuli. We also found that, among non-task vocalizations, the information values were greater in response to familiar individuals than unfamiliar individuals. However, no explicit representations in the form of vocalizer selective units were identified. Instead, it appears that information about vocalizer identity is distributed across many neurons, and that prior experience modulates the reliability of spike timing and thus information capacity for encoding learned stimuli.

4.2 Introduction

Vocal communication requires the mapping of an incoming acoustic signal to meaningful categories, e.g. the familiarity of the vocalizer, identity of the vocalizer, or the meaning of

a sound as in human speech. In the brain, hierarchical circuits can facilitate this process by transforming the low level acoustical features of a sound (Elliott & Theunissen, 2009; Theunissen et al., 2000) into abstract, invariant representations in higher order areas (Russ et al., 2008). To facilitate learning and memory of these information bearing sounds, the circuits may be shaped by the experience of the animal. In songbird vocal communication, a particularly important skill is the ability to categorize a sound according to the identity of the vocalizer; the brain must use the information in an incoming sound to recognize each vocalizer's calls as distinct auditory object categories. Many previous studies have shown the ability of songbirds to discriminate between the sounds of other individuals (D'Amelio, Klumb, et al., 2017; Elie & Theunissen, 2018; Honarmand et al., 2015; Miller, 1979b), and in the preceding chapters we demonstrated the zebra finch's large memory capacity for individual vocal recognition and the involvement of brain regions analogous to secondary auditory cortex in that task. In this chapter, we investigate how that information is represented in neural activity of auditory regions in the zebra finch brain, and how the neural encoding changes when these vocalizations are learned and remembered.

Neurons selective for specific sounds may underlie auditory object recognition, analogous to findings in the primate visual system in which neurons selective for small numbers of objects (Ito et al., 1995) or faces (Freiwald & Tsao, 2010) may be used to represent individual identity. Electrophysiological experiments in songbirds have shown response selectivity that correlates with learning and memory for auditory objects in several areas of the avian forebrain. The song is a particularly salient and well studied example of a specific auditory memory. In song imitation learning, juvenile male zebra finches form a long lasting auditory memory for the song of a tutor. Neurons in several brain regions have shown selective tuning for the tutor song in the song learning pathway known as the anterior forebrain pathway (AFP) (Doupe & Konishi, 1991; Doupe & Solis, 1997), the premotor nuclei HVC (Nick & Konishi, 2005; Volman, 1993) and RA (Doupe & Konishi, 1991), and secondary auditory area NCM (Yanagihara & Yazaki-Sugiyama, 2016). The appearance of tutor song selective neurons in NCM (Yanagihara & Yazaki-Sugiyama, 2016) has been shown to arise as a consequence of sensory experience and correlate with the quality of sensorimotor learning later in life. Song may be a special case of an auditory memory, given that it is a sexually dimorphic behavior (in most songbird species, only males sing) and these brain circuits may be specialized for song learning and production. However, selectivity for other learned sounds have been observed in multiple species (Gentner & Margoliash, 2003; Meliza & Margoliash, 2012; Wang et al., 2020) and for other natural call categories (Elie & Theunissen, 2015). For example, in starlings neurons are more likely to exhibit high selectivity to behaviorally relevant (i.e. learned in an operant task) than unfamiliar (i.e. novel) song motifs (Gentner & Margoliash, 2003; Meliza & Margoliash, 2012).

Selectivity for one or a few stimuli cannot tell the whole story of individual recognition. In particular, it fails to capture the time varying structure of the neural response which can carry information about the type of a call or the identity of a vocalizer (Elie & Theunissen, 2019). The heterogeneity of response patterns in the auditory regions of the songbird also makes interpretation after averaging single neuron responses over the population difficult.

One complementary approach is to understand how auditory regions as a whole are modulated by learned versus unfamiliar stimuli. For example, differences in noise correlation structure between task-relevant and non-task stimuli have been observed in the caudolateral mesopallium (CLM) that can enhance the representation of behaviorally important signals (Jeanne et al., 2013). Another way that we may grapple with the heterogeneity of response patterns is by looking at sub-populations of neurons within an area that may play different computational roles. For example, inhibitory interneurons in NCM are thought to shape auditory learning and memorization (Pinaud & Terleph, 2008; Thompson et al., 2013; Yanagihara & Yazaki-Sugiyama, 2016), and may be acted on by several different neuromodulators such as estradiol (Vahaba & Remage-Healey, 2018) and dopamine (Macedo-Lima et al., 2021) to drive plasticity. These sub-populations may be distinguishable using their extracellularly recorded spike waveform shapes.

In the current chapter, we investigate the hypothesis that neural representations of conspecific vocalizations are modulated by a birds' previous experience with the stimulus. We measured stimulus response properties of neurons across primary and secondary auditory regions of zebra finches who had undergone operant training in the vocal recognition tasks described in the previous chapters. We found that at the population level, task-relevant vocalizers were not over-represented in the neural response, neither in single unit selectivity measures nor when ensemble decoding was applied. However, we did find evidence that the information capacity of the neural code, measured by the coherence, was greatest when responding to task relevant stimuli.

Data for these analyses were collected from electrophysiological recordings in the anesthetized zebra finch. To identify single neurons in our dataset, we developed a custom, semi-automated spike-sorting procedure in order to handle electrode "drift", a common issue in electrophysiological recordings (Bar-Hillel et al., 2006; Dhawale et al., 2017) when a single neuron's waveform changes shape over time during the course of a recording. The Materials and Methods section of this chapter includes a description the data pipeline software and methods used for sorting non-stationary spike shapes, which was originally developed for chronic and acute recordings where spike waveform drift is a persistent issue.

4.3 Results

Neuron classification by spike shape

We obtained recordings of single unit spiking activity in primary and secondary auditory regions of anesthetized zebra finch in response to playback stimuli of conspecific vocalizations and synthetic noise sounds. The zebra finches were previously trained in an operant task (Chapter 2) designed to test their memory for the song and distance call (DC) of several conspecifics. Vocalization stimuli were organized by their task relevance (*task-relevant* if learned in the operant task, *non-task* if not), the subject's familiarity with the source vocalizer, and the stimulus reward class (Figure 4.1, see Table 4.4 in Materials and Methods).

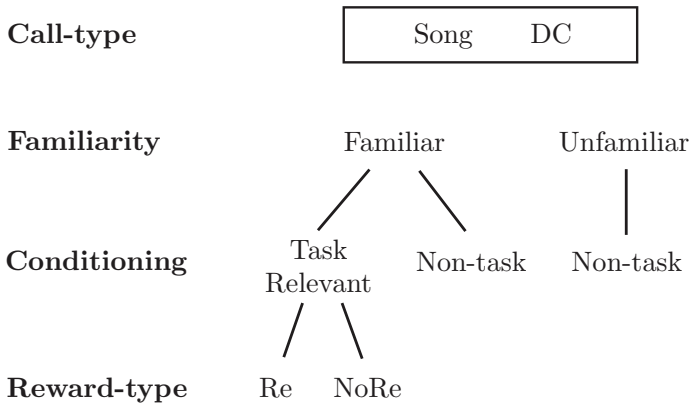


Figure 4.1: **The familiarity hierarchy of vocalization playback stimuli.** *Task Relevant*: Stimuli learned in the operant task of Chapter 2. These are further subdivided into *Re* (Rewarded) and *NoRe* (Non-rewarded). *Non-task*: Stimuli that were not used in the operant task. These include *Familiar* vocalizations that subjects were naturally exposed to in the colony room, recorded from birds either housed in the same cage as the subject or in an adjacent cage, while *Unfamiliar* vocalizations were never exposed to the subject prior to the recordings.

Each song or distance call stimulus contained three renditions.

We first categorized units by their signal quality, brain region and spike shape. Neurons were categorized as belonging to CM, NCM, or Field L (encompassing sub-regions L1, L2a, L2b, and L3), and by their spike shape as narrow-spiking (NS) or broad-spiking (BS). These classifications are thought to represent distinct cell types and functional roles: narrow-spiking units have been hypothesized to correspond to fast-spiking interneurons, while BS may represent excitatory projection neurons analogous to mammalian pyramidal cells (Mitchell et al., 2007), though these classifications have not been validated in the auditory system of the bird (Krentzel et al., 2018). Using the gap-statistic method to determine the optimal number of clusters (Tibshirani et al., 2001), we identified three clusters of neurons by spike shape (Figure 4.2A&B. One cluster of units, with peak-to-peak duration < 0.4 ms, corresponded well to previously reported narrow spiking (NS) units (Schneider & Woolley, 2013; Yanagihara & Yazaki-Sugiyama, 2016), but our method found no clear subdivision of this cluster into NS1 and NS2 clusters as reported by (Macedo-Lima et al., 2021) (Figure 4.2C). The two remaining clusters could be classified as broad-spiking (BS) neurons and corresponded well to the BS1 and BS2 clusters reported in (Macedo-Lima et al., 2021). In our analysis, we use these classifications of NS, BS1, and BS2. The distribution of these neuron types sampled in our dataset are shown in Table 4.1.

We next characterized the spiking properties of the neurons classified as NS, BS1, and BS2. We computed spontaneous firing rates in silent periods before stimulus onset, and found

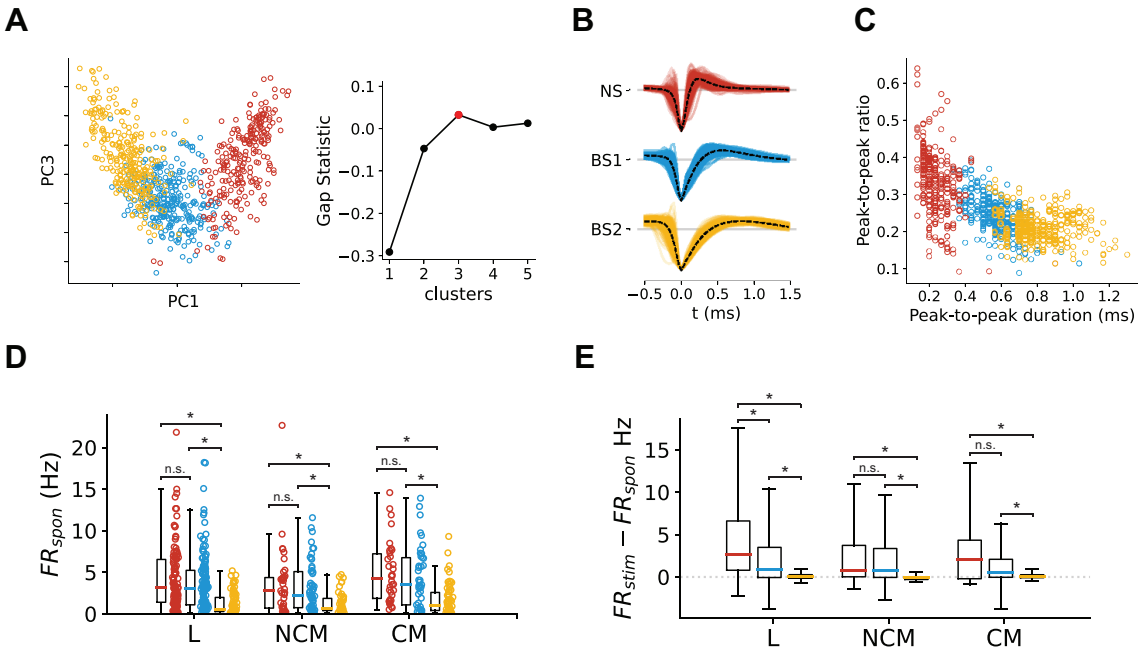


Figure 4.2: **Classification of units into broad and narrow spiking subtypes.** (**A**) *Left*: Distribution of spike times along the first and third principal components (PC1 and PC3), with colors indicating cluster from fitting a Gaussian Mixture model with 3 components. Color labels are applied to the same clusters in (B-E). *Right*: Gap statistic for Gaussian Mixture models for different cluster sizes, and the optimal value of 3 clusters shown in red. (**B**) Individual waveforms and mean waveform (dotted lines) for each of the identified clusters. (**C**) Distribution of unit peak-to-peak duration (measured from the negative peak at $t = 0$ to the time of the positive peak) against peak-to-peak ratio (the absolute value of the ratio between the magnitude of the peak at $t = 0$ and the magnitude of the positive peak). (**D**) Spontaneous firing rate FR_{spon} for the three types of units in Field L, NCM, and CM, computed as the average firing rate in 500 ms before stimulus onset. Box and whisker plot shows the median value and 1st and 3rd quartiles; whiskers extend 1.5× the inter-quartile range beyond the first and third quartiles. Significance bars show Wilcoxon signed-rank test, all tests labeled with asterisks were significant with $p < 0.002$. (**E**) Average response strength of the three neuron types in the three brain regions, measured as the difference between the mean stimulus-evoked firing rate FR_{stim} and the spontaneous rate FR_{spon}. Significance bars show Wilcoxon signed-rank test, all tests labeled with asterisks were significant with $p < 0.006$.

Region	NS	BS1	BS2	Total Units
L	113	140	76	329
NCM	39	54	43	136
CM	34	41	58	133
Other	38	56	91	185

Table 4.1: **Classification of units from each auditory brain region sampled.** L: Field L complex including L1, L2a, L2b, L3; NCM: Caudalomedial nidopallium; CM: Caudal mesopallium; Other: Undetermined or uncertain locations, including nidopallium and hippocampus, that were not clearly located in one of the other three regions. These were not used in further analyses, but most had clear auditory responses.

that NS and BS1 had similar spontaneous firing rates ($\mu_{\mathrm{NS}} = 4.2 \pm 0.6$ Hz, $\mu_{\mathrm{BS1}} = 3.8 \pm 0.4$ Hz), while BS2 units had much lower spontaneous rates ($\mu_{\mathrm{BS2}} = 1.4 \pm 0.2$ Hz) (Figure 4.2D). This relationship was consistent across both Field L, CM and NCM. NS units were also the most strongly modulated by auditory stimuli, shown by the difference between the stimulus-evoked firing rates FR_{stim} and the spontaneous rates FR_{spon} (Figure 4.2E). In contrast, BS1 units showed smaller degrees of stimulus-evoked firing rate modulation despite having similar baseline rates as NS units. BS2 units had the weakest auditory responses of the three groups when measured by the mean firing rate.

Auditory neurons were not selective for task-relevant vocalizers

Heterogeneous responses were found across the sampled population of neurons. Several neurons exhibited strong temporal alignment across repeated stimulus presentations (for example, Figure 4.3A). In some cases, the strongest auditory evoked response was an "offset" response, with peak firing rate occurring after each rendition (for example, Figure 4.3C). These diverse response patterns across different neurons and stimuli may together capture the information needed to discriminate between the vocalizations of different individuals. We wanted to test if the past experience with a stimulus in an operant memory task affected the information represented here.

We first asked if there were object selective units in our dataset that specifically encode the song or DC of one or a few individuals. We quantified the selectivity of neurons for individual vocalizer in our stimulus set using a measure defined in (Vinje & Gallant, 2000), which we refer to as the selectivity index SI (Equation 4.2). This value ranges from 0 for a broadly tuned neuron that responds equally to all stimuli, to 1 for a sharply tuned neuron that only responds to a single stimulus. The mean firing rate FR_{mean} between rendition onset and 100 ms after offset was used as the definition of response strength[1]. To validate this

[1] 100 ms after offset was included to capture offset responses of neurons with peak firing rates after the end of a stimulus; example in Figure 4.3C.

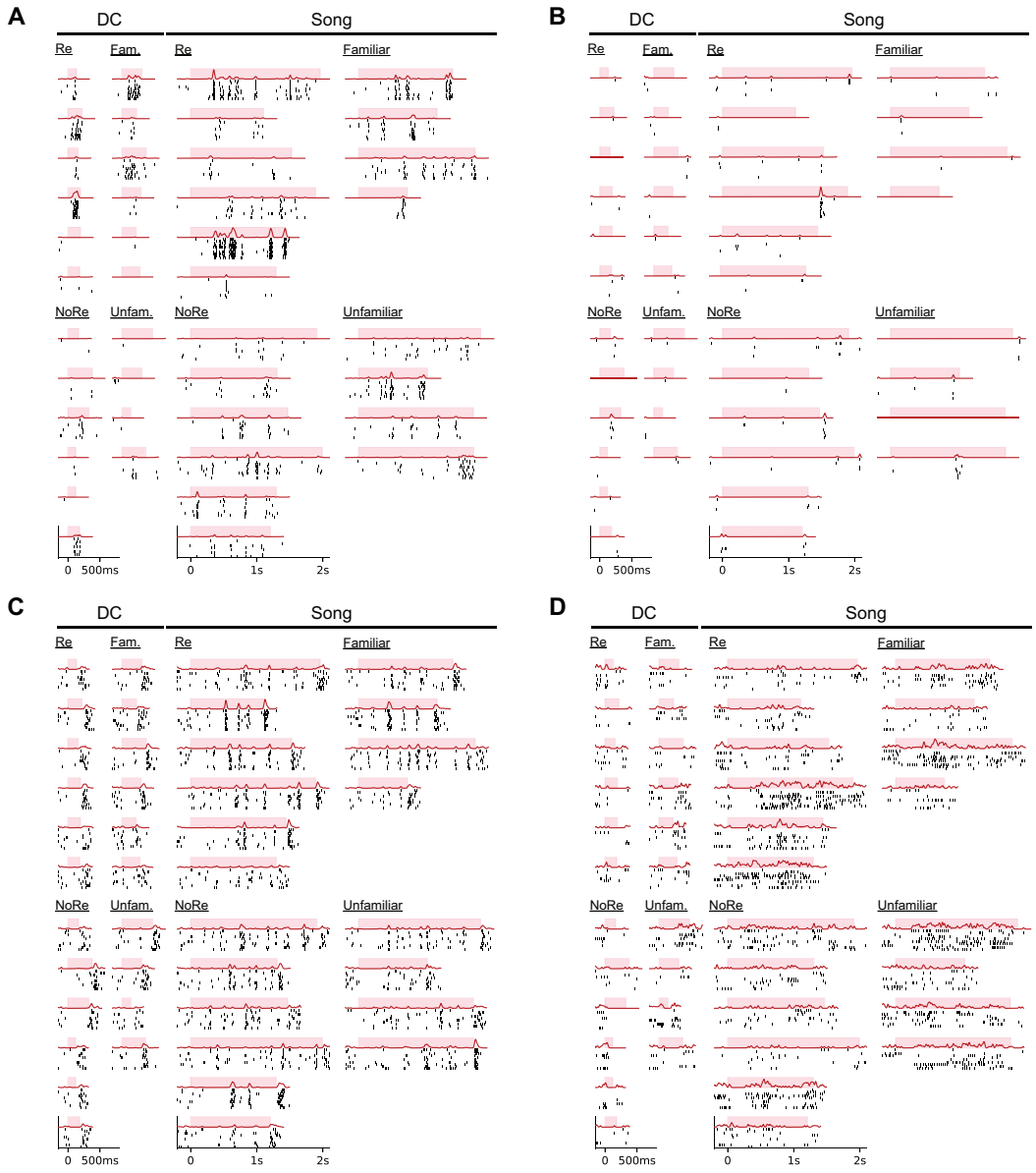


Figure 4.3: **Raster plots of spiking responses in four units.** Spiking responses of four units (A-D) to one rendition of each song and DC stimulus. Each black dot represents one spike; each row represents one stimulus presentation. Red curves above rasters show estimated PSTH computed with a Gaussian KDE. Shaded region indicates when the stimulus was on. (**A**) $BS1_L$ unit with $SI_{song} = 0.37$ and $SI_{DC} = 0.67$. (**B**) $BS1_L$ unit with $SI_{song} = 0.28$ and $SI_{DC} = 0.31$. (**C**) $BS1_L$ unit with $SI_{song} = 0.03$ and $SI_{DC} = 0.04$. (**D**) NS_{NCM} unit with $SI_{song} = 0.13$ and $SI_{DC} = 0.24$. Units (A-D) were all recorded from the same subject; (A) and (B) were recorded simultaneously on one electrode.

Region	NS	BS1	BS2	Total
L	4	7	9	20
NCM	3	0	3	6
CM	0	1	4	5
Total	7	8	16	31

Table 4.2: **Distribution of selective units.** Representative selective units were identified by taking the top five selective units within songs and DCs in four firing rate bins (0.5-1 Hz, 1-2 Hz, 2-4 Hz, 4-10 Hz). Firing rate was computed as mean stimulus evoked firing rate over all stimulus playbacks. Neurons needed at least 8 trials per playback file. Nine units met this criteria for both song and DC, resulting in 31 units total. These units are highlighted in Figure 4.4A&D and their response strengths to DC and song stimuli are mapped in Figure 4.4B&E.

metric, we estimated a null selectivity index SI_0 for each neuron by simulating a Poisson neuron with the same average firing rate but fixed to be uniform over all stimuli. The distribution of SI_0 over the population resulted in higher estimates of selectivity than from real spike times for both songs ($p < 0.001$, Wilcoxon signed rank test) and distance calls ($p < 0.001$, Wilcoxon signed rank test) (Figure 4.4B). Furthermore, S_0 has a clear inverse relationship with firing rate; this is likely due to the fact that the marginal impact of noise or variability will have a large impact when firing rates are low (see Equation 4.2). Thus, we must be careful when interpreting units with high SI but low firing rates, as even an unbiased, random Poisson unit with a low firing rate will produce high estimates of SI.

To compare the selectivity of neurons across regions, we looked separately within NS, BS1, and BS2 classes, which had very different spontaneous and stimulus evoked firing rates (Figure 4.4C&F). We found that any differences between neuron types and brain regions closely followed the null relationship: higher firing rate NS and BS1 units had distributions shifted to lower selectivity, while lower firing rate BS2 units had distributions shifted to higher selectivity.

Finding no clear evidence for over-representation of vocalizer selective units in the population, we next asked if the selective units that we did find (e.g. Figure 4.3A) were influenced by task-relevance of the stimulus in the subject's past experience. The categories of *task-relevant* and *non-task* were defined by whether or not the subject had been exposed to the stimulus during the operant memory task of Chapter 2. We imagined two ways that task relevance could potentially affect selectivity: (1) of the selective units in the dataset, more units would be selective for task-relevant stimuli than non-task stimuli, and (2) selectivity measured over task-relevant stimuli would be greater than selectivity measured within non-task stimuli.

To test (1), we selected the units with the highest selectivity and organized them by the vocalization that elicited the strongest response. We then tested if these were more likely

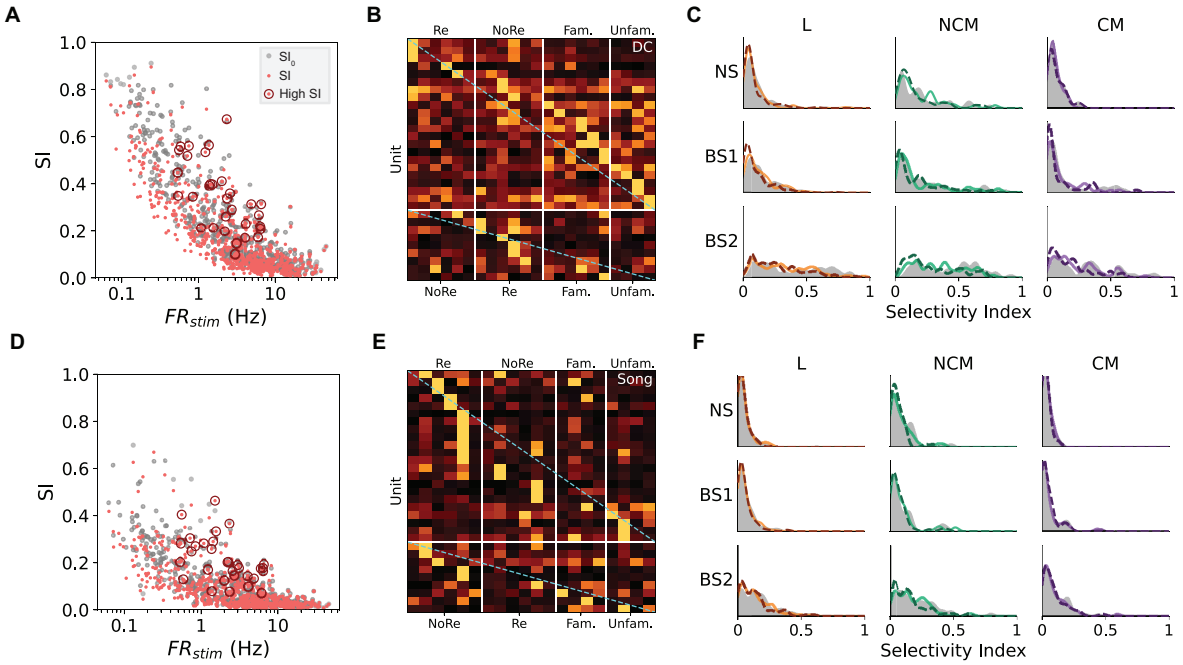


Figure 4.4: **Selectivity for vocalizer identity.** (**A**) Relationship between FR_{stim}, the mean firing rate over all playbacks, and selectivity index SI over vocalizers by DC. Pink dots represent individual neurons. Grey dots show corresponding null estimates SI_0, the selectivity of simulated neurons with Poisson spike times and a uniform firing rate over all stimuli. Thirty-one selective units are circled (criteria described in main text and in caption of Table 4.2). (**B**) Heat-map of response strengths for the thirty-one selective units circled in (A). Each row represents one neuron and each column represents one stimulus vocalizer. Columns are organized by task-relevance and familiarity as in Figure 4.1. In half the subjects, the reward contingencies were flipped; thus the heat-map is divided vertically to separate the task-relevant stimuli by reward contingency. Response strengths were normalized to the max response. Dotted lines show diagonal on which the maximum selectivity would be expected to lie if each vocalizer was equally likely to be a neuron's preferred stimulus. (**C**) Probability density (smoothed histogram) of SI for DC, organized by cell type and brain region. Shaded region denotes the distribution of SI_0. Solid line is the distribution of SI evaluated only within task-relevant stimuli, and dashed line is the distribution of SI evaluated only within non-task stimuli. Task-relevant distributions did not differ from non-task distributions (Wilcoxon rank-sum test with Bonferroni correction for 9 hypotheses). (**D-F**) Same as (A-C) but for song vocalizers.

to have their preferred stimulus belong to the task-relevant category or non-task category. We analyzed the responses to song and DC separately because of their differing SI statistics (Figure 4.4A&D). In order to avoid oversampling from spurious SI values for low firing rate units, we limited ourselves to the top 5 selective units across brain region and firing rate bins (0.5-1Hz, 1-2 Hz, 2-4 Hz, 4-10 Hz) for each call type, resulting in thirty-one selective units[2]. The response strengths of these units over the set of song stimuli and DC stimuli are shown in Figure 4.4B and 4.4E. A qualitative assessment of these data suggest that selectivity for vocalizer is distributed across the different levels of familiarity, from the task-relevant to the completely unfamiliar. There was an over-representation of one of the task-relevant songs (5th column in Figure 4.4E), with a large proportion of selective units maximally responsive to that song. This was most likely not due to the reward contingency but to the acoustics of the song, as in 2 out of the 4 subjects the reward contingency for this song motif was flipped yet still drove similarly strong responses in our sample.

We then compared the distribution of SI over task-relevant and non-task stimuli. We used repeated resampling to match the number of renditions in the two groups because SI in Equation 4.2 is sensitive to the number of stimuli tested, and there were more task-relevant stimuli in the dataset. We found no overall shift in the distribution of SI (Figure 4.4C&F) between the two groups (none significant using Wilcoxon rank-sum with Bonferroni correction over nine hypotheses). Thus, we found no evidence for modulation of selectivity based on previous task relevance or familiarity.

Ensembles in Field L and NCM have similar information about vocalizer identity

Selectivity for object identity is just one way that information about an auditory object may be represented in neural spiking patterns. In practice, the selectivity of real neurons can be described as a continuum between object selective cells (sparse representations) and non-selective cells (dense, distributed representations). To measure the information available to the network about individual vocalizer identity that might be distributed over the population, we looked at how single units and ensembles could be used to decode task relevant variables and vocalizer identities.

We applied a simple procedure to reduce the dimensionality of the neural response prior to decoding. We calculated each neuron's time-varying response to a vocalization rendition with a 500 dimensional vector representing the kernel density estimate (KDE) of spike times in the first 500 ms after stimulus onset (1 ms bins and $\sigma = 10$ ms. This smoothed representation represents spike trains with similar spike timing as similar vectors, with tolerance defined by the width of the kernel. The neural responses were then projected into the the top 10 principal components (PCs) of the neural response, explaining just over 70% of the variance. The first 2 PCs are interpretable as two primary modes of response types, with PC_1 resembling a sustained response after a short latency from stimulus onset and PC_2

[2]one of these units is the one shown in Figure 4.3A

representing an onset response followed by a sustained inversion or return to baseline. The remaining PCs appear to roughly correspond to a Fourier decomposition of the neural response in the 500 ms window. Visual inspection of the highest frequency PC suggests that the highest temporal resolution in this representation is on the order of 100 ms.

To quantify the information encoded at the ensemble level for task-relevant and non-task stimuli, we used these reduced representations of single unit responses to decode at the ensemble level. We sampled 1000 ensembles at several ensemble sizes from 1 to 40 neurons in Field L, NCM, and CM. Each ensemble was constructed from within one brain region (L, NCM, or CM), but using units across sites and subjects; as such, ensembles were not recorded simultaneously and thus noise correlations or other measures of joint activity were not be considered for analysis. An ensemble response vector was constructed by concatenating the reduced response vectors from the individual units in the ensemble, and the top 10 PCs for this ensemble response were used as input to the decoder. The ensemble neural response (Figure 4.5A) was then used to train a Gaussian Naive Bayes (GNB) classifier to predict the vocalizer identity of each rendition each vocalizer represented by three renditions and up to ten trials per rendition (see Materials and Methods). Decoders were trained separately for song and distance call stimuli.

Additionally, each classifier was separately trained and tested within either task-relevant stimuli or non-task stimuli. The classifiers were scored by their percent correct classification (PCC), the percentage of true vocalizer labels in a held out test set matching the predicted labels of the GNB classifier. The classifier was also scored by the mutual information between stimulus label and neural response from the confusion matrix generated from the posterior probability distribution over the dataset.

We found that mutual information and decoding accuracy steadily increased from single neurons up to 40 unit ensembles without maxing out (Figure 4.5B-D), showing that in this representation, discriminating vocalizer identity requires the joint activity of several units. Decoding accuracy was also approximately equal in both Field L and NCM while much lower in CM, despite stronger similarities in firing rates and SI distributions between Field L and CM than NCM (Figures 4.2 and 4.4). Furthermore, there was no difference in classifier performance when trained and tested within the task-relevant and non-task relevant groups. Thus, the information content of the ensemble neural response was not affected by previous experience.

Estimation of information in single unit spiking patterns with coherence

The decoding analyses used a reduced representation of the neural response which imposed strong assumptions about the information bearing features of the neural response. Information was discarded by (1) restricting analysis to the first 500 ms from stimulus on-

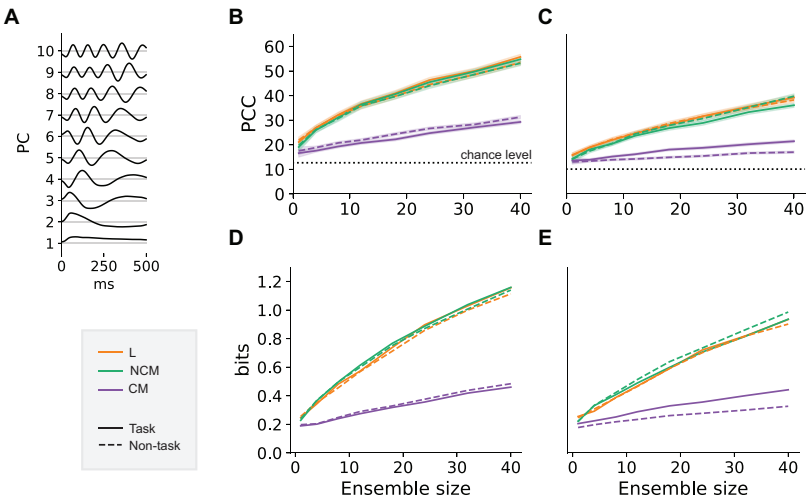


Figure 4.5: **Performance of ensemble decoders for vocalizer identity.** (**A**) First 10 principal components of the time-varying neural response across all units. (**B**) Average decoder performance (PCC: percent correct classification) over 1000 ensembles sampled from each brain region. Decoders were trained and tested on classifying the stimulus vocalizer identity. Colors indicate the brain region from which units in ensembles were sampled from. Solid curves show decoding within the task relevant stimuli, dashed curves show decoding performance within the non-task relevant stimuli. Shaded region show 2 SEM. (**C**) The same as C but for decoding within DC stimuli. (**D**) Average decoder mutual information between response and vocalizer label within song stimuli as a function of ensemble size. (**E**) Average decoder mutual information between response and vocalizer label within DC stimuli as a function of ensemble size.

set[3], (2) representing each trial with a Gaussian KDE of bandwidth 10 ms, (3) decomposition using PCA to temporal resolution on the order of 100 ms, and (4) PCA at the ensemble level. These assumptions potentially reduce the maximum performance of the classifiers because of the amount of information that is discarded in these dimensionality reduction steps.

We estimated the upper bound, or capacity, of the information in the neural response using the coherence, a measure of the reliability in the neural response from trial to trial. Less reliability in the neural response across trials can limit the capacity of a neuron to encode information about a stimulus. The neural response's signal-to-noise ratio (SNR) can be estimated by modeling the activity of the neuron on a single trial as the combination of a "true" signal (a time-varying mean firing rate) with random noise. The magnitude-squared coherence $|\gamma^2(\omega)|$ between a single trial and the "true" signal is a measure of the response SNR as a function of frequency (Hsu et al., 2004). Integrating the coherence over

[3]Offset responses as in Figure 4.3C and sparse responses as in Figure 4.3B are two clear examples where restricting analysis to an onset window loses information.

all frequencies gives an estimate of the information capacity of the neural response in bits. One advantage of this approach is that it uses the raw spike times and does not require an *a priori* assumption about the spiking timing reliability.

Of course, the "true" time-varying mean rate cannot be known in practice, but can be estimated as the mean response over repeated trials. An unbiased estimate of the coherence of each neuron's spiking response with its "true" time-varying mean rate response was made using the method described in (Hsu et al., 2004), using only a small number of trials per stimulus. The information was then estimated by integrating over frequency bins where the lower bound on the coherence estimate was non-zero (see Materials and Methods). This resulted in a sampling bias toward higher firing rate units; reliable but sparse responses from units like the one in Figure 4.3B tend to have low or zero information when measured this way due to large confidence intervals around the estimated coherence (Figure 4.9).

Stimuli were divided into *task-relevant* and *non-task* DC and song renditions, and the information of each unit was evaluated in response to all stimuli within these groups (Figure 4.6A&B). We restricted our analysis to units with non-zero information values for both categories (Table 4.3). We found that units across all regions had higher information values in response to task-relevant stimuli than non-task stimuli for both song and DC ($p < 0.001$ for both, Wilcoxon signed-rank test). Most units that had non-zero information to one or both categories of stimuli were NS_L and $BS1_L$ units. There was generally not enough data in the other brain region and neuron type groups to test this effect. However, it is interesting that Field L, the primary thalamo-recipient auditory region analogous to primary auditory cortex, also exhibited modulation in spiking reliability based on task relevance, which is a higher order feature. This modulation may be a result of network-wide plasticity, especially considering that the recordings were performed in anesthetized animals.

We next considered if the shift in to higher information in response to task relevant stimuli was a consequence of specific operant training, or general familiarity. We performed two comparisons to test these potential explanations. First, we compared the coherence in response to familiar and unfamiliar vocalizers that were not included in the operant task (Figure 4.6C&D). We found that on average, the coherence of the neural response to familiar vocalizers was higher than to unfamiliar vocalizers, suggesting that spiking reliability in the population was partially modulated by general prior exposure and not just operant conditioning.

Second, we tested whether the reward contingency (Re or NoRe) of a task-relevant stimulus affected the information of the neural code. There was no difference in the coherence between Re and NoRe responses to song ($p = 0.11$, Wilcoxon signed-rank test). Coherence in responses to Re DC was slightly higher than to NoRe DC ($p = 0.01$), but the effect was relatively small (Figure 4.6E&F). This evidence indicates that specific reward histories have a minimal influence on modulation in spiking timing information. Furthermore, it increases our confidence that the differences observed between the task-relevant and non-task categories is not an artifact of a single stimulus in the task-relevant group.

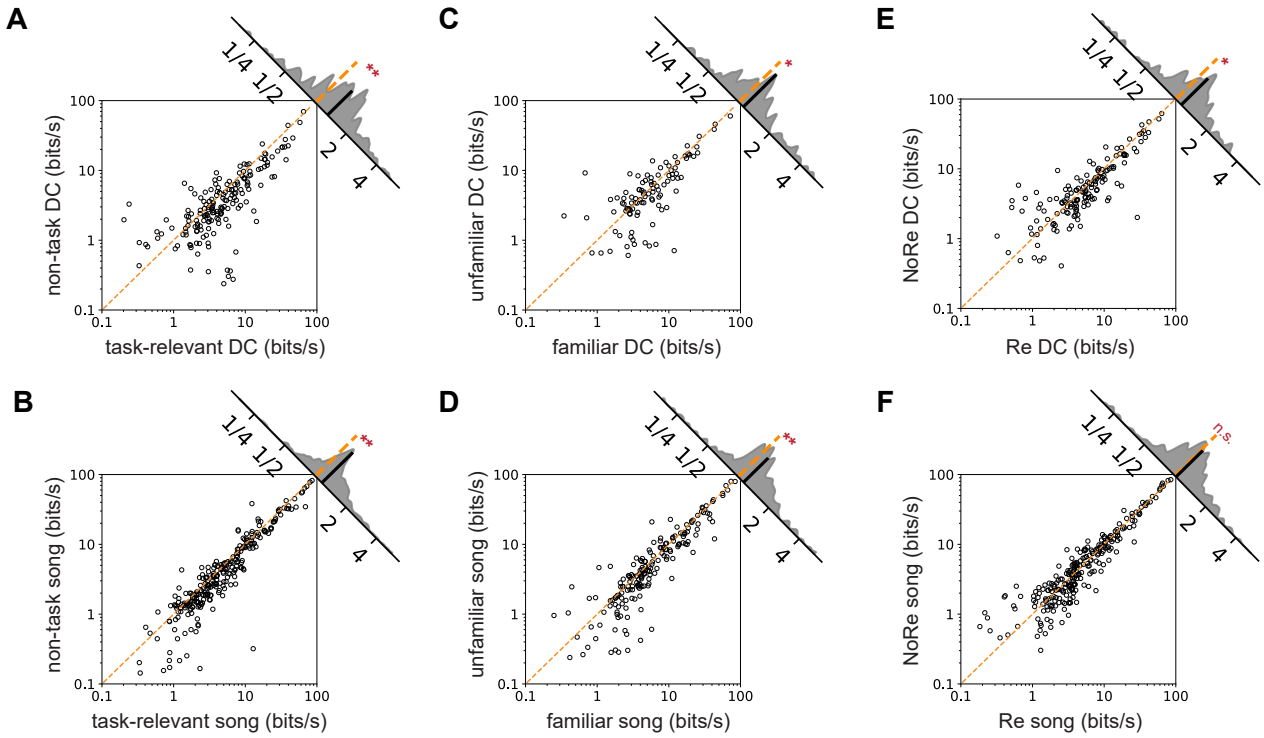


Figure 4.6: **Coherence estimates for units responding to task-relevant and non-task stimuli.** (**A** and **B**) The information (in bits/s) in response to task-relevant stimuli (x-axis) and non-task stimuli (y-axis). Top row shows data for DC, bottom row for song. Information was estimated by integrating the coherence over all frequencies where the lower bound of the coherence estimate was non-zero. Histogram shows the distribution of the ratio x/y on a log scale, i.e. a projection of the scatter data, re-scaled for visibility. Single asterisk indicates $x > y$ with $p < 0.05$ (Wilcoxon signed-rank test). Double asterisk indicates $p < 0.001$ (Wilcoxon signed-rank test). (**C** and **D**) Same as in (A, B) but comparison of information measured within familiar (non-task) stimuli and unfamiliar stimuli. (**E** and **F**) Same as in (A-D), but for Re and NoRe stimulus groups.

Song	L	NCM	CM
NS	61/81 (75%)	15/26 (58%)	9/23 (40%)
BS1	63/104 (61%)	21/32 (66%)	11/27 (41%)
BS2	13/49 (27%)	11/27 (41%)	4/30 (13%)

DC	L	NCM	CM
NS	52/81 (64%)	12/26 (46%)	8/23 (35%)
BS1	44/104 (42%)	12/32 (38%)	5/27 (19%)
BS2	7/49 (14%)	3/27 (11%)	3/30 (10%)

Table 4.3: **Neurons with non-zero information values for both task-relevant and non-task playbacks.** Each cell shows the number of neurons with significant information values for both categories, out of the total number of units in that subgroup. Only neurons with non-zero information to both categories and at least 8 trials per stimulus on average were used in the analysis of Figure 4.6.

4.4 Discussion

In the preceding chapters, we showed that the zebra finch is capable of distinguishing between several individuals using only the acoustical features in the song and distance call. In this study, we did not find explicit representations of learned vocalizer identities at the single unit or ensemble level. Instead, we found evidence that the response reliability of auditory neurons in the zebra finch is modulated by a subject's past experience with the stimulus. The information capacity of each neuron was estimated by the spiking reliability over multiple stimulus repetitions; we found information capacities that were (1) higher in response to task-relevant vocalizations than non-task vocalizations, and (2) higher to vocalizations of socially familiar birds than unfamiliar birds. These results suggest that both passive and active learning of conspecific vocalizations changes the encoding of those sounds in auditory circuits, capable of conveying more information about those sounds. Most of the neurons used in this analysis were sampled from Field L, suggesting that this experience dependent effect were not restricted to higher order areas like CM and NCM.

Selectivity for higher-order objects or object categories is a hallmark of sensory processing associated with object recognition (Russ et al., 2008). As such, we first looked for explicit representations of vocalizer identity in the form of "object selective" neurons for behaviorally relevant stimuli. We found auditory neurons with selective response for a small number of individual vocalizers (e.g. Figures 4.3A&B, 4.4B&E), but no evidence that these selective units were reflect an over-representation of behaviorally relevant vocalization categories at the population level.

On its own, selectivity can be an ambiguous or even misleading measure. For example, we simulated units with uniform firing rates over all stimuli and found that the selectivity index (Equation 4.2) is highly dependent on the window over which the firing rate is estimated

and the magnitude of the neuron's firing rate. In particular, the selectivity of low firing rate neurons will be overestimated, as random noise will have a greater marginal effect. Selective units in sensory systems are typically characterized by low firing rates to most stimuli (Willmore & Tolhurst, 2001); this makes them hard to distinguish from random low firing units with spurious high estimates of SI, and also relatively hard to trigger and detect in the first place (given a finite stimulus set).

As another example, take the sparsely firing unit shown in Figure 4.3B. From the stimulus-locked firing pattern, one may expect that it to have high selectivity, which appears inconsistent with the estimated selectivity index of $SI = 0.28$. This is because the selective response to its preferred stimulus (the fourth rewarded song motif) does not come from the raw number of spikes (firing rate), but the from its reliable spiking pattern in response to its preferred stimulus, both across trials and precision in time. This suggests that the selectivity index may be more informative when paired with other statistics that describe the temporal aspects of the neural response as well. This observation was a primary motivation to quantifying the reliability of the neural response using the coherence (Hsu et al., 2004), and may also be addressed by decoder-based selectivity indexes that can take into account the time-varying neural response (Elie & Theunissen, 2015).

Selectivity is an intuitive way to think about information encoding as it is an explicit representation of object identity (Calvo Tapia et al., 2020). A more general method to measure the information a neural code carries about a stimulus is the mutual information (Borst & Theunissen, 1999). We used a decoding approach to estimate the mutual information between ensemble responses and vocalizer identity to see if the encoded information differed between task-relevant and non-task vocalizations. The mutual information quantity measured this way can be thought of as a lower-bound on the information present in the neural code. We first found that decoder performance and mutual information steadily increased with ensemble sizes, out the largest ensembles tested (40 neurons). This suggests that the information needed to distinguish vocalizer identity is fairly distributed across the population. Ensembles sampled from Field L and NCM were better suited for distinguishing individual vocalizers than ensembles from CM. This result is slightly unexpected, as the neural response of CM units are known to be modulated by task relevance (Gentner & Margoliash, 2003; Jeanne et al., 2013; Jeanne et al., 2011; Theilman et al., 2021). Considering also that firing rates and selectivity distributions of Field L and CM were comparable (Figure 4.4C&F), and the neural population sampled was similar in CM and NCM (Table 4.1), further work will be needed to determine what qualities of the ensemble response distinguishes CM from L and NCM, and why it appears to carry less information about vocalizer identity. Using this decoding approach, we also found no clear differences in the encoding of task-relevant and non-task stimuli, although there was a possibility that task-relevant DC was easier to decode in CM (Figure 4.5E).

The decoding approach models the variability in responses within and across stimulus categories to estimate a lower bound on the information the neural response carries about a stimulus. While variability across stimulus categories increases the information capacity of the neural response, variability within the same category across trials limits how much

information a neuron can convey about the stimulus category. We estimated the reliability of the neural response using the coherence between a neuron's individual spike trains to its time-varying mean response (Hsu et al., 2004). In contrast to the selectivity and decoding approaches, the coherence makes no assumptions about how or what information is encoded, only that the information bearing feature of the response is the time-varying mean rate. We hypothesized that a subject's past experience in a vocal recognition task would affect the information capacity of neurons in the auditory system, specifically in NCM and CM, the secondary auditory areas involved in learning and memory of behaviorally relevant sounds (Gentner, 2004; Macedo-Lima & Remage-Healey, 2020; Meliza & Margoliash, 2012; Pinaud & Terleph, 2008; Thompson & Gentner, 2010). We found that the information capacities of neurons across the population were indeed modulated by past experience: the neural response was more reliable (i.e. had higher information capacity) in response to familiar stimuli and in particular the set of task-relevant vocalizations previously learned in an operant task (Figure 4.6A&B).

This finding is similar to the finding reported in (Jeanne et al., 2011), which reported that the mutual information between neuron firing rates in CLM and song motif identity were higher for task-relevant motifs. They also reported higher information values in response to rewarded stimuli than non-rewarded stimuli; we saw some weak evidence for this in the information values for rewarded versus non-rewarded distance calls (Figure 4.6C). However, our results differ in a few key ways. First, our analysis was performed on units sampled across the auditory regions of Field L, NCM, and CM which had non-zero estimates for the information capacity. The latter restriction may bias the sample towards higher firing rate units, and as such most of the units in this analysis ultimately were Field L neurons (Table 4.3). Second, the information estimated in (Jeanne et al., 2011) was computed as the mutual information between stimulus identity (i.e. song motif) and mean firing rate of the response. That analysis quantifies the *variability across stimulus categories* and was thus more directly analogous to the decoding analysis described above than the coherence based analysis.

As mentioned above, coherence as a measure of temporal spiking reliability across trials may be a useful complement to the traditional measure of selectivity to better capture our intuition of what constitutes a strong neural response. We must be careful, however, not to mistake our intuition for what a neural representation looks like to the real thing. In all the analyses presented in this chapter, we assume that the information bearing qualities of the neural response are time-locked to sensory stimulus. However, this may not be the case when complex natural behaviors such as individual vocal recognition are involved. At higher levels of cognition the relevant processes may shift from alignment to external physical stimuli to internal processes related to perceptual decision making or attention. Thus, the analytical approaches described above that represent the neural response as aligned to stimulus onset may be intrinsically biased toward stimulus-locked, lower-level responses. One solution will be to use data-driven and unsupervised methods of identifying structure in the joint neural activity of the population (e.g. Jeanne et al., 2013; Theilman et al., 2021).

One of the technical challenges addressed in this project was the spike sorting of non-stationary waveforms. The issue of waveform "drift" is well documented in extracellular

electrophysiological recordings (Bar-Hillel et al., 2006; Dhawale et al., 2017; Rey et al., 2015). The approach described in this chapter was designed for recordings in the awake, behaving zebra finch, in which the stability of the electrophysiological signal can be highly variable and contaminated with noise from motion artifacts. The processing pipeline, combining automated hierarchical clustering over time, cluster linking in a directed graph, and manual curation, was fairly successful at identifying well-isolated, individual units which often changed shape over a recording session or came in and out due to noise contamination. One successful example is the well isolated neuron of Figure 4.7G, which was tracked and distinguished from background noise for as long as possible in the recording session despite a relatively low amplitude and slow drift over the hour-long session. Two other examples are the neurons whose rasters are shown in Figure 4.3A&B. These two units were in fact recorded on the same contact and had similar spike shapes (both BS1), while one had an exceptionally low firing rate but was still able to be isolated. Anecdotally, these examples and others increased our confidence in the single unit quality and cluster isolation of single units used in our analysis.

Despite the original goal of a automated clustering algorithm with minimal human intervention, the manual curation step of the process ultimately wound up being quite time consuming for the 32-channel probes used in this project, requiring on average about a day of expert manual curation per site. The time cost suggests that the current implementation of this data processing workflow will not scale to the higher channel counts and high-density arrays that will become commonplace in the near future (Chung et al., 2019; Jun, Steinmetz, et al., 2017); however, the algorithm used here may still be useful as the basis for future implementations will rely less and less on manual curation.

One of the primary factors that limits the scope of the current project is the fact that the recordings were made in the anesthetized animal. Although the anesthetized preparation has its advantages, understanding the neural code behind a high level behavior such as individual vocal recognition will require the analysis of neural activity in the awake, behaving animal. There is evidence that anesthesia can modulate or gate sensory responses; e.g. reduced auditory selectivity in motor nuclei to the bird's own song under anesthesia (Nealen & Schmidt, 2006; Vicario & Yohay, 1993), or reduced selectivity in song selective BS neurons in NCM during sleep versus wakefulness (Yanagihara & Yazaki-Sugiyama, 2016). Thus, we cannot rule out that our reported results here would be different in the awake behaving animal. However, these results can serve as a useful baseline comparison in future experiments using awake animals.

4.5 Materials and Methods

Behavioral testing and electrophysiology

All animal procedures were approved by the Animal Care and Use Committee of the University of California, Berkeley (AUP-2016-09-9157) and were in accordance with the Na-

tional Institutes of Health guidelines regarding the care and use of animals for experimental procedures.

We performed extracellular recordings in auditory regions of four adult zebra finches (2 male, 2 female). These zebra finches were previously trained in a memory task (fully described in Chapter 2) for individual vocal recognition of songs and DCs of up to 54 conspecifics[4]. The neural recordings were performed either within the first week after the initial tests (N=2) or one month after the initial tests and immediately after a week of re-testing on the full stimulus set (N=2). The latter pair experienced a month delay without exposure to the stimuli, but the retention of learned memories for vocalizations in the stimulus sets were verified (Figure 2.4).

Subjects were anesthetized with intramuscular injection of urethane solution and head-fixed. One 32-channel multi-electrode tungsten array (MicroProbes) was mounted to a stereotaxis (Kopf Instruments) and lowered with a hydraulic micromanipulator into the right hemisphere. The signal was amplified and digitized on a digitizer chip purchased from Intan Technologies. Starting from 100 microns below the surface of the brain, a set of stimulus playback was played lasting approximately 70 minutes. After a set of playbacks, the drive was lowered 50 to 100 microns and the process repeated. The cross sectional area of each array (1mm x 2mm) spanned a large amount of the avian auditory system, the electrode trajectories intersected primary auditory areas L1, L2a, L2b, L3, as well as the secondary auditory areas of NCM, and CM (Figure 3.1). The designation of anatomical region for each unit depended on the depth of recording and histological verification in Nissl stained sagittal slices. Designations within the Field L complex (L1, L2a, L2b, L3) were considered to be the same region L. Neurons that lay outside of one of the specific auditory regions analyzed were labeled as nidopallium (Ni) and excluded. However, precise anatomical boundaries between NCM and Field L subregions are often unclear (Vates et al., 1996) and as such the designations used in this report may be considered approximations. Neural data was recorded using the Intan RHD 2000 Interface software and sampled at 30kHz.

Stimulus playbacks

During recordings, we played a stimulus set composed of conspecific vocalizations and synthetic sounds, summarized in Table 4.4. The conspecific vocalizations included distance calls and song from a subset of the rewarded (Re) and non-rewarded (NoRe) vocalizers used in the operant memory tests, collectively referred to here as *task-relevant* stimuli. Also included in the stimulus set were DC and song of birds who were housed with or next to the recorded subject (*familiar*), or call and song of birds who the subject had never been previously exposed to (*unfamiliar*). These vocalizations were not part of the operant memory tests and are collectively referred to as *non-task*. These vocalization stimuli were prepared by pseudo-randomly combining three renditions of song motifs or distance calls from the

[4]The behavioral performance of the four subjects used here correspond to the subjects S1, S2, S3, S4 in Figure 2.2

Relation	Reward?	Call Type	# Stimuli	Dur.	Description
Familiar	NoRe	Song	6m	6.0	Non-rewarded songs
Familiar	Re	Song	6m	6.0	Rewarded songs
Familiar	NoRe	DC	3m, 3f	3.0	Non-rewarded DCs
Familiar	Re	DC	3m, 3f	3.0	Rewarded DCs
Familiar	-	Song	4m	6.0	Song of males housed with or adjacent to subject for at least one month prior to recordings†
Familiar	-	DC	3m, 3f	3.0	DC of birds housed with or adjacent to subject for at least one month prior to recordings††
Unfamiliar	-	Song	4m	6.0	Song of males never before exposed to subject
Unfamiliar	-	DC	2m, 2f	3.0	DC of birds never before exposed to subject
Unfamiliar	-	Ripple	10	2.0	Modulation limited noise (ml-noise)

Table 4.4: **Playback stimuli used in anesthetized recordings.** Each stimulus file was played 10 times per site. Song and DC stimuli included 3 renditions per stimulus file separated by silent intervals. *# Stimuli*: m indicates male vocalizer, f indicates female vocalizer. *Dur.*: Stimulus duration in seconds. †For male subjects, one familiar song was the bird's own song. ††In all subjects, one familiar DC was the bird's own DC.

same vocalizer into 6-second long vocalization sequences for songs as in (Elie & Theunissen, 2016; Yu et al., 2020), or 3-second long vocalization sequences for DCs. Finally, a set of 10, 2 second long periods of modulation-limited noise, or "ripples", were included. These synthetic stimuli are white noise limited to low-frequencies of spectro-temporal modulations to match the statistics of natural sounds (Elliott & Theunissen, 2009) and have acoustic features that are known to drive primary auditory neurons. Each stimulus was repeated 10 times in a session, arranged pseudo-randomly such that the $(k+1)$th repetition of a stimulus file would not occur until the kth repetition of all stimuli. Because the stimulus set lasted for over 60 minutes and, a single neuron may not be present for all 10 repetitions (see "Spike sorting of non-stationary data").

Spike sorting of non-stationary data

The process of isolating the spikes from single neuron sources in extracellular electrode recordings is known as *spike sorting*(Rey et al., 2015, review). A challenge in spike sorting is that the signal of one neuron's action potentials on a single channel voltage trace may be contaminated by system noise, motion artifacts, and spikes from other nearby neurons. This problem is addressed by making the assumption that a single neuron's spikes appear in the voltage trace with a typical spike waveform shape which can then be used to identify its spikes from noise and other neurons.

In practice, however, this waveform may not be constant and can drift over time. This issue is exacerbated when there are large effects of motion and in long-term recordings. Open-source spike sorting solutions have been developed with some success (Chung et al., 2017; Jun, Mitelut, et al., 2017; Lee et al., 2017; Pachitariu et al., 2016), but tend to be optimized for higher density recordings and struggle with non-stationary data as described above. They also still require careful parameter tuning and manual curation, which can lead to large discrepancies in neuron identification across algorithms.

We developed a custom, open-source Semi aUtomated Spike Sorting procedure SUSS[5], which incorporates ideas from several existing approaches and was particularly inspired by the hierarchical sorting algorithm proposed in (Dhawale et al., 2017). SUSS combines an automated algorithm to group similar spike waveforms in shape and time into clusters, and a manual curation step to join those clusters over time. The general approach taken by the automated algorithm was to break up the dataset into many small, unimodal clusters of spikes localized in time, and then to link temporally adjacent clusters based on unimodality of their combined spike waveform distributions. The manual curation step following this procedure requires the researcher to manipulate and group these spike clusters rather than grouping individual spikes. Parameters for splitting up and joining clusters were hand-tuned to be specialized for the qualities of our zebra finch datasets, which were collected with multi-electrode tungsten arrays in both awake and anesthetized zebra finches.

Data preprocessing

Electrode signals were high-pass filtered with cutoff frequency at 300 Hz and the common mean across all channels subtracted from each channel. Potential spike events from the filtered voltage trace of one electrode were detected using a threshold of 4 std in 5 minute windows. The events detected this way were aligned to the peak value and cropped into snippets of size 36 samples (1.2 ms), with the peak aligned to the center sample. This produces $\mathbf{X}$, a (N, M) matrix for each channel representing N detected events and $M = 36$ samples per spike snippet. Extreme outliers were removed from $\mathbf{X}$ using unsupervised outlier detection (Breunig et al., 2000) on the first 2 PCs. This step only filtered out the most heinous instances of noise contamination, so false positives were unlikely. The dimensionality

[5]https://github.com/theunissenlab/suss-sorter

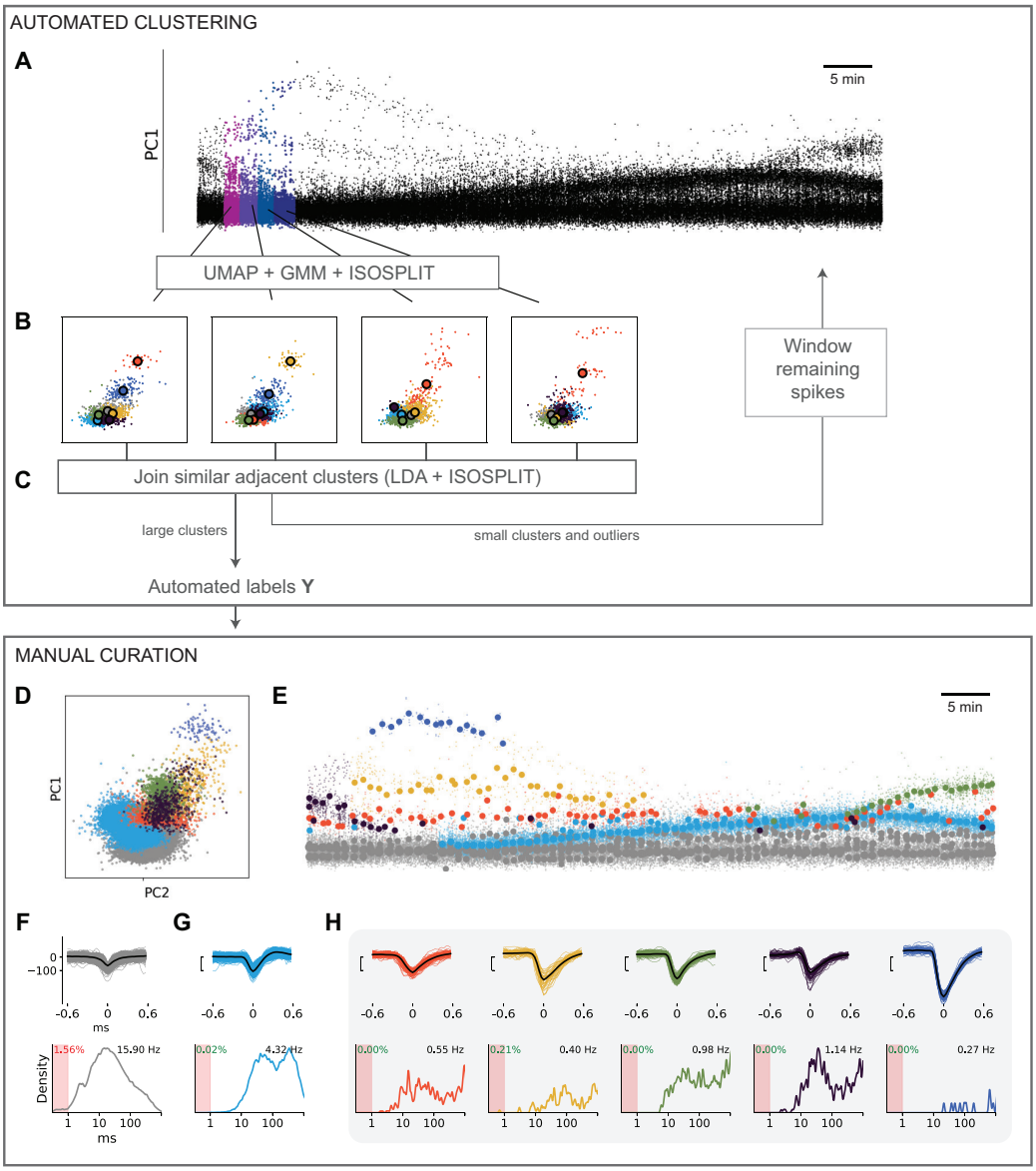


Figure 4.7: **Semi-automated spike sorting.** (**A-C**) Automated spike clustering schematic (see main text for details). (**D**) 7 clusters were identified after manual curation. (**E**) Manually identified clusters drift in shape over time. Large dots indicate cluster centroids that a user manipulates and joins during manual curation. (**F-H**) Top: Individual waveforms and mean for each cluster. Bottom: ISI (Δt) histogram. Red shaded region indicates $\Delta t < 1$ms. Percentage in top left shows % ISI violations where $\Delta t < 1$ms. Firing rate in top right calculated as $\frac{1}{\text{mean}(\Delta t | \Delta t < 10\text{s})}$. (**F**) A multiunit cluster rejected for $> 1\%$ ISI violations. (**G**) A single unit included in analysis. (**H**) 5 clusters rejected because they were not present for enough of the recording.

of $\mathbf{X}$ was reduced to $D = 3$ dimensions using the non-linear embedding UMAP (McInnes et al., 2018), producing an embedded data matrix $\mathbf{X}^{UMAP}$ of shape (N, D).

Automated spike clustering algorithm

The core algorithm is described below:

1. The first step creates a set of clusters in short time windows. Data is first chunked into windows of $N_w = 2000$ spikes, $\{\mathbf{W}_0, \cdots, \mathbf{W}_i, \cdots, \mathbf{W}_{n_{windows}}\}$, where $n_{windows} = \text{floor}(\frac{N}{N_w})$ (Figure 4.7A). The data in the ith window, $\mathbf{W}_i^{UMAP}$, is clustered using a Bayesian Gaussian mixture model (GMM) with 6 components (Figure 4.7B). This results in a label vector $\mathbf{Y}_i \in \{1, \ldots, 6\}^{N_w}$. Although we are using a GMM, the data under the UMAP projection $\mathbf{W}_i^{UMAP}$ will not generally be Gaussian. However, the purpose of this step is not to model the data distribution but simply to break up the data into several small clusters. As such, the GMM could be replaced with your favorite clustering algorithm.

2. The data in each cluster $\mathbf{W}_i[\mathbf{Y}_i = k]$ is further subdivided using the ISO-SPLIT algorithm (Magland & Barnett, 2015). The data is projected into one dimension using the first principal component of the waveform. ISO-SPLIT splits up multi-modal clusters in favor of two or more unimodal distributions. This results in an updated set of labels in each window $\mathbf{Y}_i$ with typically somewhere between 6 and 12 cluster labels per window.

3. A directed graph G is formed by first defining each spike cluster as a node $N_{i,k}$ consisting of spikes $\mathbf{W}_i[\mathbf{Y}_i = k]$. Each pair of nodes in adjacent bins $(N_{i,k}, N_{i+1,k'})$ is then compared for similarity (Figure 4.7C). To do this, spikes $\mathbf{W}_i[\mathbf{Y}_i = k]$ and $\mathbf{W}_{i+1}[\mathbf{Y}_{i+1} = k']$ are projected into 1-D using Linear Discriminant Analysis (LDA) and a two-sample Kolmogorov-Smirnov test is applied. If the K-S statistic is less than a predetermined threshold (chosen to be 0.5), we determine that the nodes are similar and an edge is added to G joining $(N_{i,k}, N_{i+1,k'})$.

4. The construction of G in Step 3 results in a directed graph, where a node in window i may have 0, 1, or 2+ incoming nodes. For a node $N_{i,k}$ with 2 or more incoming nodes, the labels $\mathbf{Y}_i[\mathbf{Y}_i = k]$ are split up even further, by reassigning labels using an LDA classifier fit to spikes of the incoming nodes' waveforms and labeled by the incoming node labels. The goal of this step is to split up clusters that have an ambiguous ancestor in a previous window and to try and create continuous paths of similar clusters over time.

5. The graph G is regenerated, following the procedure of Step 3 but using the updated labels from Step 4. This leaves us with a directed graph comprised of several weakly-connected component subgraphs whose spike waveforms are sufficiently similar in adjacent windows. Each cluster in a weakly-connected component with 2 or fewer

nodes is labeled a "leftover", while weakly-connected components with 3 or more nodes were each assigned a unique label.

6. The "leftover" clusters are designated as such because the algorithm did not find a sufficient number of similar spikes in adjacent windows. This can happen if the firing rate of the unit is slow, or if the shape of the spike changes too quickly (relative to the $N_w = 2000$ spike window).

7. Leftover spikes are then passed through the algorithm a second time (Steps 1-6, Figures 4.7A-C). During this second pass, a new "time aware" UMAP embedding is fit to the concatenation of the leftover spike waveforms and a z-scored spike timestamp T in units of hours, $[\mathbf{X}_{leftover}, \mathbf{T}_{leftover}]$. Low firing rate spikes are typically better captured on this pass, since the dense, high firing rate units would typically have been labeled and set aside after the first pass. The timestamp is included in the embedding because the spike windows of N_w may span longer timescales than recorded units can typically maintain a stable shape.

8. Finally, the result of the second pass results in a final group of leftover spikes that never found a home. At this point, windowing is no longer effective as the remaining clusters may be distributed all over the recording. Instead we simply apply UMAP to the top 20 PCs concatenated with T, and apply ISO-SPLIT to fill in the final set of cluster labels.

The above steps result in a set of cluster labels $\mathbf{Y}$ for the spike dataset $\mathbf{X}$. This data is over-clustered: there are far more data clusters than single neurons in the dataset[6]. The directed graph in Step 5 generates long paths of connected clusters through time, in which spike shapes change smoothly from window to window and the density of spikes is high enough for reliable detection. There are also several small standalone clusters consisting of noise, low firing rate units, and/or units whose shapes changed quickly during the recording. To make manual manipulation of these clusters easier, each long cluster was then broken in into several chunks with a maximum size of 2000 spikes. These smaller chunks form the basic units that the researcher manipulates in the manual curation step (see Manual curation of clusters, Figure 4.7E), and larger cluster chains are provided to the researcher as the default suggested grouping. This provided some granularity in time to strike a balance between flexibility for the researcher performing the manual curation without having too much fine grained control over individual data-points.

Manual curation of clusters

Spike clusters were manually curated using a custom GUI written in Python that allows for the joining, splitting, and deletion of clusters. Manual operations in this program were

[6]I designed the program in this way because I reasoned that, if manual curation was going to be a necessary step anyway, that it would be easier to manually join together over-clustered nodes than to split under-clustered nodes apart.

done on clusters identified in the automated procedure rather than directly on individual spike events. Clusters were joined and merged by visually observing the waveform shapes of clusters as a function of time in the dataset, primarily using a rotating projection of the top 2 PCs (y-axis) as a function of time (x-axis) (Figures 4.7D and 4.7E). By selecting one or more clusters, the researcher could visualize the distribution of spike shapes for one or more clusters and decide whether they should be joined by looking at the unimodality of the spike shape distributions, inter-spike intervals, and spike histogram aligned to stimulus onsets. The output at this stage may result in a patchy distribution of units in which a chain of clusters might abruptly stop due to changes in noise level or spike shape. To address this, we estimated a smoothed firing rate over time for each unit using a Gaussian window with bandwidth of 10 seconds, and applied a threshold of $\frac{1}{2}$ the mean firing rate. Epochs during which the smoothed rate was below the threshold were excluded under the assumption that the unit was lost or that the baseline noise level was too high for the unit to be detected. While this may artificially exclude data during which the neuron is simply quiescent, there is no way for us to verify the persistence of a cell in these conditions.

Single unit criteria

The output of the semi-automated spike sorting results in a collection of several clusters of putative units. To isolate single units, we filtered the proposed units to those with large spike shapes relative to the baseline noise ($SNR > 5.0$) and less than 1% of adjacent spikes having an inter-spike interval of less than 1 ms. SNR was defined as the peak-to-peak amplitude of the mean spike shape (measured in μV) divided by the average standard deviation of the spike shapes in the cluster measured at each time point.

An example spike sorted session and the result of manual curation is shown in Figure 4.7. During this 1 hour session, the stimulus set described in Table 4.4 was being played and 7 different clusters can be found on the single electrode after the manual curation step (Figure 4.7D-H). When projected into a 2 PCs the units are not separable, but the presence of multiple units becomes more apparent when viewed as a function of time. With our algorithm, we successfully isolated the relatively small spikes of the unit shown in Figure 4.7G from the multiunit background activity in Figure 4.7F. Unfortunately, of the 7 clusters identified, only this unit passed the criteria for analysis. The cluster in Figure 4.7F had $SNR > 5$ and had over 1% of inter-spike intervals violate the refractory period of $\Delta t < 1$ms, and so was considered to be multi-unit. On the other hand clusters in Figure 4.7H were well isolated with high SNR and low percentage of $\Delta t < 1$ms, but were not present for a long enough period of time in the recording (Figure 4.7E) to have a sufficient number of trials per stimulus for analysis (chosen to be at least 6 out of 10 trials per stimulus on average).

The single unit criteria applied here forced us to exclude many units from our analyses, such as those in Figure 4.7H. However, there are examples of successes as well; see Figure 4.3A&B for an example of two units recorded on the same electrode, with both units classified as BS1.

Classification of units

Units were classified by their spike shapes, following the broad spiking (BS) and narrow spiking (NS) classifications that have previously been observed in NCM and brain regions (Macedo-Lima et al., 2021; Meliza & Margoliash, 2012; Schneider & Woolley, 2013; Yanagihara & Yazaki-Sugiyama, 2016). Historically these classifications have been identified from the spike waveforms' peak-to-peak duration and peak-to-peak amplitude ratio, where the peaks represent the large, negative peak of the spike shape during depolarization and shallow, positive peak during hyperpolarization (Figure 4.2B).

To perform this classification, 2 ms (60 samples) spike snippets were aligned to the peak and waveforms were cut to include 0.5 ms before the peak. Each neuron was then represented by its mean spike waveform computed across all of the spike snippets assigned to it during spike sorting and normalized to the amplitude of its negative peak. Finally, we took the top 5 PCs of this distribution (98% variance explained) and applied Gaussian mixture models with a variable number of clusters. The optimal number of clusters was determined by computing the gap-statistic (Tibshirani et al., 2001) over cluster counts (Figure 4.2A), and resulting optimal cluster assignments were related to known classes or subclasses of narrow-spiking or broad-spiking units..

Analysis of firing rates

The response of each unit was characterized by its spiking responses aligned to stimulus rendition onsets. Each stimulus playback included 3 renditions of a song or DC from the same vocalizer. For each unit, a baseline spontaneous firing rate was estimated by averaging the spike rates in silent time windows before rendition onset. These time windows were taken to be as long as 500 ms, though necessarily shortened in cases when the silent period between two renditions was less than 500 ms.

Stimulus-evoked firing rates were computed by estimating the firing rate in the window from the start of a rendition to 100 ms after the end of a rendition. This short window after stimulus offset was included to capture "offset" responses of neurons such as those illustrated in Figure 4.3C where the neuron's peak firing rate occurs after the conclusion of sound. The mean stimulus-evoked firing rate $\mathrm{FR}_{\mathrm{mean}}$ to a set of stimuli is computed by taking the number of spikes evoked to each rendition divided by the duration of the rendition, and averaged over all relevant trials depending on the analysis (e.g. all trials of familiar vs unfamiliar stimuli).

To capture temporal dynamics of the neural response for decoding, we also represented the neural response as time-varying firing rate. This was measured as a peri-stimulus time histogram (PSTH) aligned to the onset of each rendition. The PSTH was computed as a kernel-density estimate of the distribution of spike times relative to stimulus onset. For a neuron's N spike arrival times relative to stimulus onset τ_i for $i = 1..N$, the KDE estimate at time t relative to stimulus onset is given by Equation 4.1.

$$\rho_{KDE}(t) = \sum_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\tau_i)^2}{2\sigma^2}\right) \tag{4.1}$$

The KDE can be estimated on a single trial to form a smoothed spike train, or over a series of stimulus repetitions to estimate the time-varying mean firing rate as in Figure 4.3.

Stimulus selectivity

We estimated the selectivity index SI of single units for one or more individual vocalizers in the dataset using a measure of sparsity defined in (Vinje & Gallant, 2000):

$$\mathrm{SI} = \frac{1 - \frac{(\sum \frac{R_s}{n})^2}{\sum \frac{R_s^2}{n}}}{1 - \frac{1}{n}} \tag{4.2}$$

Where R_s is the measured response strength (i.e. $\mathrm{FR}_{\mathrm{mean}}$) to a stimulus s and averaged over n stimuli. Values near SI = 1 are selective to a single stimulus, while values near SI = 0 are fully dense and respond to all stimuli equally. Random variance in the estimation of R_s will produce spurious estimates of the SI. For example, single spikes will have a large effect on SI computed for low firing rate units. We observed this inverse relationship between stimulus-evoked firing rates in our dataset (Figure 4.4B). To account for the natural relationship between firing rate and selectivity, we computed a null selectivity index SI_0 to pair with each estimate of SI. The null estimate was computed by taking the average firing rate of the unit over all stimuli λ and then simulating a Poisson process of a spiking unit with a fixed firing rate λ over the same set of trials and trial durations.

The selectivity of units for task-relevant stimuli and non-task stimuli (colored lines in Figure 4.4C) were computed separately for songs and distance calls. Our stimulus set included more stimulus presentations for the non-task compared to the task stimuli; SI for the non-task stimuli was estimated by random sampling so that $n = 12$ for every estimate of SI.

Coherence and information capacity

The magnitude-squared coherence, $|\gamma^2(\omega)|$, is a statistical measure that quantifies the degree of linear relationship between two signals as a function of frequency ω. When applied to a neuron's spiking activity, it quantifies the reliability of a neuron's response to repeated presentations of a stimulus, and thus its capacity to convey information about a stimulus (Hsu et al., 2004). Below, we describe the method in (Hsu et al., 2004) we used to compute unbiased estimates of the coherence between a neuron's single trial spike trains and its "true" response, the time varying mean rate. This method allows us to estimate the coherence in the 10 or fewer trials present for each stimulus in our dataset, and to estimate a lower and upper bound on the coherence. Source code of the implementation used in Python is online[7]. The

[7] `https://github.com/theunissenlab/soundsig/soundsig/coherence.py`.

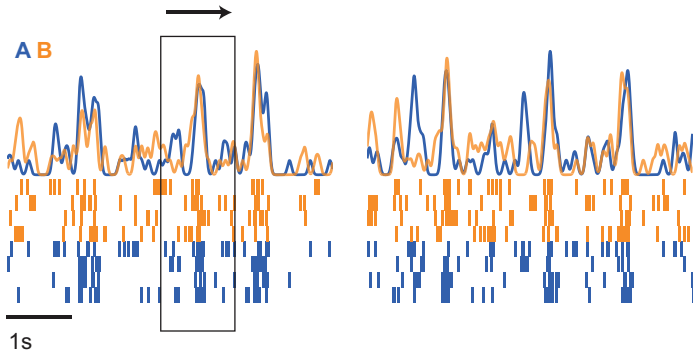


Figure 4.8: **Illustration of coherence calculation for a single unit.** Shown is a raster plot for 8 repeated presentations of two stimuli (not shown). Trials are randomly assigned to signal A (blue) or B (orange) and binned into 1 ms bins, illustrated by the smoothed PSTH above the raster. The width of bins is exaggerated in figure for illustration purposes. The cross spectral density and coherence is estimated in overlapping segments over all stimuli.

coherence between signals x and y is given by Equation 4.3, where C_{xy} is the cross spectral density of x and y and C_{xx} and C_{yy} are the auto-spectral densities of x and y.

$$\gamma^2_{xy}(\omega) = \frac{|C_{xy}|^2}{C_{xx}C_{yy}} = \frac{\langle x^*(\omega)y(\omega)\rangle\langle x(\omega)y^*(\omega)\rangle}{\langle x^*(\omega)x(\omega)\rangle\langle y^*(\omega)y(\omega)\rangle} \tag{4.3}$$

The neural response $R_i(\omega)$ on a single trial is modeled as the sum of the deterministic response of the unit $A(\omega)$ and independent noise $N_i(\omega)$. The goal is to estimate the coherence γ^2_{AR} between a single trial response R and the true response A; the problem is that in real experimental data, A is not known. Our best estimate of A is the PSTH, $\overline{R}_M = \frac{1}{M}\sum_i^M R_i$. Because all trials are included in the estimate of $\overline{R}_M$, each trial is correlated with $\overline{R}_M$ and thus the coherence estimate from this dataset $\gamma^2_{A\overline{R}_M}$ would be overestimated. The method in (Hsu et al., 2004) derives the following unbiased estimate of γ^2_{AR}:

$$\frac{1}{\gamma^2_{AR}} - 1 = \frac{M}{2}\left(-1 + \sqrt{\frac{1}{\gamma^2_{\frac{M}{2}}}}\right) \tag{4.4}$$

Here the value $\gamma^2_{\frac{M}{2}}$ is the coherence between two PSTHs computed by averaging two non-overlapping subsets of $\frac{M}{2}$ trials from the M total trials.

The practical algorithm for computing γ^2_{AR} from our data is described here:

1. We take the spiking responses of a unit to a set of stimuli. We limit our analysis to stimuli for which there are at least 8 trials. For each rendition the neural response plus

200 ms of padding before and after the rendition were included to capture the onset and offset signal.

2. Trials are randomly assigned to group α or group β. The spikes in each group are binned at 1 ms resolution (Figure 4.8 shows an example with bin width exaggerated for illustration). This forms two time-series $\alpha_{\frac{M}{2}}$ and $\beta_{\frac{M}{2}}$.

3. In order to estimate the time-averaged cross spectrum C, the signal is broken up into N_s overlapping segments of size 1024 bins with overlap of 512 bins. The signals are then tapered using 5 tapering windows of the discrete prolate spheroidal sequences (DPSS). These tapered windows reduce edge effects of spectral estimation during segmentation.

4. The cross spectral density $C_{\alpha\beta}$ and auto-spectral densities $C_{\alpha\alpha}$ and $C_{\beta\beta}$ are estimated for each taper in each segment using the Fast Fourier Transform (FFT).

5. The coherency (Equation 4.5), coherence, and error bounds on the coherence are estimated using a jackknife procedure (Thomson & Chave, 1991), averaging C over all segments.

$$\gamma_{\alpha\beta}(\omega) = \frac{\langle C_{\alpha\beta} \rangle}{\sqrt{\langle C_{\alpha\alpha} \rangle \langle C_{\beta\beta} \rangle}} \tag{4.5}$$

6. The procedure from Step 1 is repeated over multiple shuffled assignments of trials to A and B.

Given the unbiased estimate of a single unit's coherence, we can summarize it using the normal mutual information. In the discrete case, we compute it using Equation 4.6, where $\Delta\omega$ is the size of a frequency bin in the discrete FFT, here 1 kHz, and k_{max} is the index of the first frequency bin where the lower bound of the estimate of $\gamma_{A\overline{R}_M}$ is zero.

$$I = -\sum_{k=0}^{k_{max}} \log_2(1 - \gamma^2_{AR}(\omega_k))\Delta\omega \tag{4.6}$$

This information measure quantifies the noise power level and information capacity of the neuron. We use this quantity to compare the information coding capacity of the neuron under different conditions, e.g. for comparing the neuron's information capacity when task-relevant versus non-task, or for familiar versus unfamiliar stimuli.

Single unit and ensemble decoding

We used a Gaussian Naive Bayes (GNB) decoder to quantify the information in single unit or ensemble responses regarding reward history, task relevance, and call type. The input to the decoder was a dimensionality reduced representation of the neural responses of single units or ensembles.

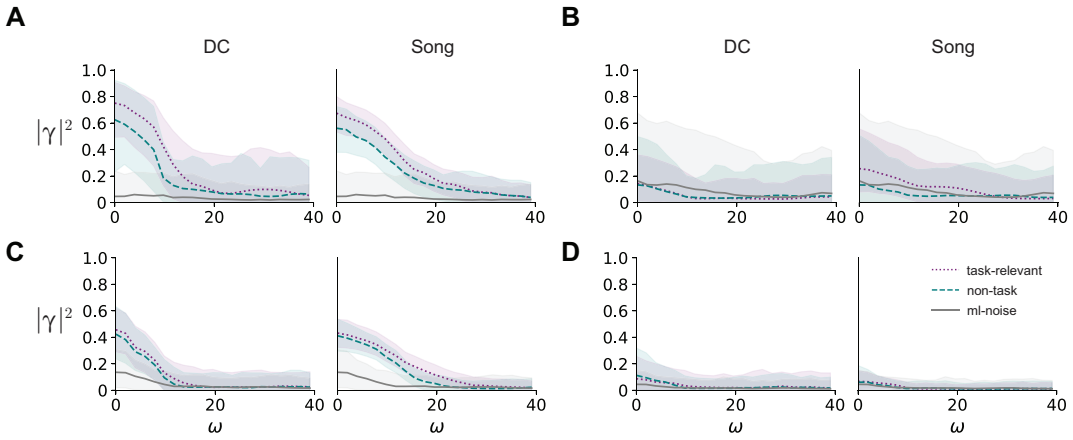


Figure 4.9: **Coherence functions of four units for task-relevant and non-task stimuli.** The coherence curves as a function of frequency for task-relevant stimuli, non-task stimuli, and modulation limited noise stimuli. Shaded regions show 95% CI of jackknife estimate for $|\gamma|^2$. (**A-D**) correspond to the units show in Figure 4.3.

As described in the Results, spikes in the first 500 ms after stimulus onset were smoothed with a Gaussian kernel (Equation 4.1) of width $\sigma = 10$ ms and sampled in 1 ms bins. Thus, each trial can be described by a vector of length 500 representing the estimated PSTH on a single trial. The dimensionality of this representation was reduced by projecting each trial into the top 10 PCs, fit to all trials for all units, resulting in each trial of a single unit (SU) being represented by a vector of PC coefficients $\mathbf{X}^{SU} = [P_1, .., P_{10}]$. The top 10 PCs explained over 70% of the variance in the dataset.

To decode within an ensembles of units, we represented the ensemble response as the concatenation of all single unit response vectors. For an ensemble of size K, we sampled units from the relevant pool of units, typically one brain region out of Field L, NCM, or CM. Units in the ensemble were sampled from any subject and recording site, and trials were aligned together randomly. In other words, these ensembles were "virtual" in the sense that units were not recorded simultaneously. To represent an ensemble response, we first concatenated the single unit response vectors (in the 10-dimensional PCs space) to form a length $10 \times K$ joint vector, then reduced to the top 10 PCs fit to only the ensemble responses: $\mathbf{X}^{ensemble} = PCA([\mathbf{X}_1^{SU}, .., \mathbf{X}_K^{SU}])$. The number of PCs was chosen somewhat arbitrarily; on average, 30 PCs explained 58% of the variance for ensembles of $K = 40$ units (max=71%) and 87% of the variance for ensembles of $K = 8$ (max=99%), while 10 PCs explained 56% for $K = 8$ (max=90%) and 31% for $K = 40$ (max=46%). Yet, we found that the decoder did not perform better for more PCs, suggesting that the relevant information was likely already lost in the initial dimensionality reduction at the single unit level.

The Gaussian Naive Bayes algorithm assumes that the response $\mathbf{X}$ is normally distributed,

conditioned on the data label l (here the vocalizer identity). Each distribution has an independent mean μ_l and variance σ_l^2, where the components of the distribution along each dimension of $\mathbf{X}$ can have independent means and variances.

$$p(\mathbf{X}|l) \sim N(\mu_l, \sigma_l^2) \tag{4.7}$$

The posterior probability over vocalizers given a response vector is

$$p(l|\mathbf{X}) \propto p(l)p(\mathbf{X}|l) \tag{4.8}$$

When appropriate, a uniform prior over label classes $p(l) = \frac{1}{|L|}$ was applied, where L is the set of vocalizers in the data. The posterior probability can be maximized to get a predicted label l, or accumulated over all trials to form a joint probability distribution over response $\mathbf{X}$ and vocalizer label l, which we treat as a confusion matrix.

During each fitting procedure, 40% of trials were held out as a test set used for evaluation of decoder accuracy. The remaining 60% of trials were used to fit the model coefficients μ_l and σ_l. We compute the accuracy of the decoder using percent correct classification (PCC), which is computed by choosing the predicted label l that maximizes $p(l|\mathbf{X})$ on each trial and determining the probability that it equals the true label. This fraction collapses the information in the likelihood distribution into a single label. To provide a more nuanced estimate of the information content in the neural response given by this decoding method, we calculate the mutual information using the joint distribution between the actual and predicted labels, $p(l, l')$. This mutual information, $MI(l; l')$ is given by Equation 4.9:

$$\mathrm{MI} = \sum_{(l,l')}^{L\times L} p(l, l') \cdot \log_2 \left(\frac{p(l, l')}{p(l)p(l')} \right) \tag{4.9}$$

Acknowledgements

Many of the algorithms used for dimensionality reduction and classification, including Local Outlier Factor outlier detection, kernel density estimates, PCA, Gaussian Mixture Model, and Gaussian Naive Bayes classifier were implemented using code from the `scikit-learn` Python library (Pedregosa et al., 2011).

Chapter 5

Conclusion

In these experiments, we make progress toward linking the vocal recognition behavior of the songbird to its neurological basis. In Chapter 2, we quantified the memory capacity of zebra finches by testing how many individuals a bird could recognize by communication calls alone. To do so, we designed an operant task which allowed us to reliably assess birds on recognition over a large set of vocalizers. We found that the memory capacity was large—subjects recognized forty unique conspecific individuals by song and distance call without signs of plateauing. This finding is roughly consistent with the natural group sizes of the zebra finch. In the wild, zebra finch mating pairs are the primary social unit (McCowan et al., 2015) but individuals are typically found foraging or traveling in small groups of up to 20 (McCowan et al., 2015); breeding colonies can consist of over 100 individuals (Zann, 1996). While previous work has shown that zebra finches are capable of recognizing their mate's vocalizations (Miller, 1979a; Vignal et al., 2008), our findings show that the zebra finch memory capacity for vocalization sounds is large enough for them to recognize individuals within their typical group sizes and likely more. Future work will be needed to test the upper limits of this capacity, especially under naturalistic conditions when birds can rely on additional social and visual cues.

In nature, zebra finches communicate with a repertoire of at least 12 ethogram-based call types (Zann, 1996), several of which carry information about the vocalizer identity (Elie & Theunissen, 2016). In contrast, much of the neuroscientific study of songbirds has typically focused on song. Restricting our attention to song may limit our understanding of individual recognition, especially considering that in most species only male birds sing and that song may be most relevant to territorial or mating related behaviors rather than more general social dynamics. In the projects described here, we included the distance call, a loud contact call produced by all members of the species and used while birds are out of visual contact. While it has been previously shown that zebra finches can use the distance call for individual recognition (D'Amelio, Klumb, et al., 2017; Elie & Theunissen, 2018; Vignal et al., 2004), we demonstrate the impressive capacity of the zebra finch for remembering reward associations for distance calls as well as song. Further work on vocal communication should include even more call types that are frequently used in social interactions, such as the stack call which is

a soft contact call elicited more frequently than both the song and distance call during close range communication (D'Amelio, Trost, et al., 2017; Ter Maat et al., 2014).

The behavioral task of Chapter 2 formed the basis of our subsequent neurophysiological experiments. We found that bilateral ablation of NCM caused deficits in memory storage and retrieval when subjects were tested on the large set of vocalizers, confirming its causal role in individual vocal recognition. These results are consistent with previous studies which showed evidence for the involvement of NCM in auditory memory behaviors (Gobes & Bolhuis, 2007; London & Clayton, 2008; Thompson & Gentner, 2010). However, lesion to NCM did not completely eliminate the ability for birds to accomplish the task, suggesting that additional brain regions may be used or recruited. Successful performance in this task may engage several areas of cognition, e.g., attention, perception, auditory memory, reward association, memory retrieval—which of these components are affected by chronic NCM lesion remains unclear. To better distinguish these possibilities, future studies can use acute manipulations of pathways involving NCM during specific phases of the task to shed light on the functional role of NCM in recognition and auditory memory; for example, through electrical stimulation, optogenetic, or pharmacological manipulations (e.g. Macedo-Lima et al., 2021).

Our analysis of single neuron response properties in NCM, CM, and Field L in the anesthetized zebra finch also revealed experience dependent changes in the encoding of vocal sounds. We found that the information capacity in response to familiar vocalizers, in particular task-relevant vocalizers, was greater across the sampled population of primarily Field L neurons. These results were consistent with (Jeanne et al., 2011), which described an increase in neural information for learned song motif identity in CLM and CM in starlings. Combined, these results suggest that the plasticity that shapes auditory circuits during learning alters spiking reliability across the entire auditory system. One particularly promising direction for future research will be to understand the connection between association learning and dopaminergic innervation of interneurons in secondary auditory regions (Macedo-Lima et al., 2021). More work will be needed to determine what specific aspects of the neural response are modulated by familiarity and the underlying mechanisms that cause the reliable responses.

We also looked for explicit representations for vocalizer identity in the form of object selective neurons. In the auditory areas we sampled, we did not find any greater selectivity to learned vocalizers than we would expect by chance. However, our intuitive understanding of neural selectivity (i.e., lifetime sparsity, Willmore & Tolhurst, 2001) was not always consistent with the selectivity index measured using mean stimulus-evoked firing rates. This was particularly true in neurons with phasic, time-locked responses, and neurons with very low firing rates. Future work should be done to better understand how variability in the neural response with a limited number of trials affects the selectivity index to avoid spurious estimates of selectivity. A complementary method for identifying representations of vocalizer identity would be through invariance; in the primate visual system, object selective neurons exhibit invariant responses in response to the same underlying stimulus (e.g. a face) over naturalistic transformations (Freiwald & Tsao, 2010; Ito et al., 1995). For vocal communication sounds, a neuron encoding individual identity could be identified by invariant responses

to all calls of one individual in the presence of background noise, overlapping calls, and even across other call types. It remains to be seen whether such neurons exist in the zebra finch auditory system, or at all.

Understanding both the behavior and neurological basis of individual vocal recognition in the songbird will require a mix of artificial and naturalistic experiments. The operant go/no-go responses used in our tasks in Chapters 2 and 3 was useful for testing birds on a large number of stimuli, but may not be an accurate representation of how a bird would process a conspecific vocalization in a natural setting. Future experiments not concerned with maximizing the memory capacity of these birds may consider measuring natural response behaviors to recognized or unrecognized sounds instead, e.g., approach/avoidance, aggression, antiphonal calling. Also, while we tested individual recognition by using the actual calls and song of many different vocalizers, future studies may consider using artificially generated calls (e.g., Sainburg et al., 2020) that interpolate between the calls of two behaviorally distinct individuals; in doing so, one could identify where birds draw categorical boundaries between different individuals in acoustic space, and then search for neural correlates of that boundary. Most importantly, electrophysiological recordings in the awake, behaving animal will be crucial in understanding how learned auditory stimuli are encoded. While the results of Chapter 4 show evidence of plasticity in the encoding of behaviorally relevant vocalizations, only in an awake, behaving animal will we be able to demonstrate how those circuits are used to map the acoustic vocal sounds into meaning and behavior.

The ability to recognize others—mates, parents, offspring, friends, rivals—forms the core of social behavior. In many species, including humans, this is accomplished through vocal communication. Through neuroethological studies like these, we are beginning to understand how different species utilize individual vocal recognition to form and maintain social relationships, and the neural computations required for transforming a sound into meaning.

Bibliography

Adkins-Regan, E. (2002). Development of sexual partner preference in the zebra finch: A socially monogamous, pair-bonding animal. *Arch. Sex. Behav.*, *31*(1), 27–33. https://doi.org/10.1023/a:1014023000117

Aglieri, V., Watson, R., Pernet, C., Latinus, M., Garrido, L., & Belin, P. (2017). The glasgow voice memory test: Assessing the ability to memorize and recognize unfamiliar voices. *Behav. Res. Methods*, *49*(1), 97–110. https://doi.org/10.3758/s13428-015-0689-6

Akçay, Ç., Arnold, J. A., Hambury, K. L., & Dickinson, J. L. (2016). Age-based discrimination of rival males in western bluebirds. *Anim. Cogn.*, *19*(5), 999–1006. https://doi.org/10.1007/s10071-016-1004-3

Alcami, P., Ma, S., & Gahr, M. (2021). *Telemetry reveals rapid duel-driven song plasticity in a naturalistic social environment*. https://doi.org/10.1101/803411

Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *Neuroimage*, *52*(4), 1528–1540. https://doi.org/10.1016/j.neuroimage.2010.05.048

Balda, R. P., & Kamil, A. C. (1992). Long-term spatial memory in clark's nutcracker, nucifraga columbiana. *Anim. Behav.*, *44*(4), 761–769. https://doi.org/10.1016/S0003-3472(05)80302-1

Bar-Hillel, A., Spiro, A., & Stark, E. (2006). Spike sorting: Bayesian clustering of non-stationary data. *J. Neurosci. Methods*, *157*(2), 303–316. https://doi.org/10.1016/j.jneumeth.2006.04.023

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends Cogn. Sci.*, *8*(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Bennur, S., Tsunada, J., Cohen, Y. E., & Liu, R. C. (2013). Understanding the neurophysiological basis of auditory abilities for social communication: A perspective on the value of ethological paradigms. *Hear. Res.*, *305*, 3–9. https://doi.org/10.1016/j.heares.2013.08.008

Blumstein, D. T., & Munos, O. (2005). Individual, age and sex-specific information is contained in yellow-bellied marmot alarm calls. *Anim. Behav.*, *69*(2), 353–361. https://doi.org/10.1016/j.anbehav.2004.10.001

Boeckle, M., & Bugnyar, T. (2012). Long-term memory for affiliates in ravens. *Curr. Biol.*, *22*(9), 801–806. https://doi.org/10.1016/j.cub.2012.03.023

Bolhuis, J. J., Hetebrij, E., Den Boer-Visser, A. M., De Groot, J. H., & Zijlstra, G. G. (2001). Localized immediate early gene expression related to the strength of song learning in socially reared zebra finches. *Eur. J. Neurosci.*, *13*(11), 2165–2170. https://doi.org/10.1046/j.0953-816x.2001.01588.x

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

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.*, *11*(11), 747–759. https://doi.org/10.1038/nrn2931

Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nat. Neurosci.*, *2*(11), 947–957. https://doi.org/10.1038/14731

Braaten, R. F., Petzoldt, M., & Cybenko, A. K. (2007). Recognition memory for conspecific and heterospecific song in juvenile zebra finches, taeniopygia guttata. *Anim. Behav.*, *73*(3), 403–413. https://doi.org/10.1016/j.anbehav.2006.08.009

Brenowitz, E. A. (1991). Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science*, *251*(4991), 303–305. https://doi.org/10.1126/science.1987645

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. *SIGMOD Rec.*, *29*(2), 93–104. https://doi.org/10.1145/335191.335388

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Front. Psychol.*, *7*, 1116. https://doi.org/10.3389/fpsyg.2016.01116

Calvo Tapia, C., Tyukin, I., & Makarov, V. A. (2020). Universal principles justify the existence of concept cells. *Sci. Rep.*, *10*(1), 7889. https://doi.org/10.1038/s41598-020-64466-7

Canopoli, A., Herbst, J. A., & Hahnloser, R. H. R. (2014). A higher sensory brain region is involved in reversing reinforcement-induced vocal changes in a songbird. *J. Neurosci.*, *34*(20), 7018–7026. https://doi.org/10.1523/JNEUROSCI.0266-14.2014

Canopoli, A., Zai, A., & Hahnloser, R. (2016). Lesions of a higher auditory brain area during a sensorimotor period do not impair birdsong learning. *Matters*.

Canopoli, A., Zai, A., & Hahnloser, R. H. R. (2017). Bilateral neurotoxic lesions in NCM before tutoring onset do not prevent successful tutor song learning. *Matters*, 12.5/20. https://doi.org/10.19185/matters.201603000018

Carlson, N. V., Kelly, E. M., & Couzin, I. (2020). Individual vocal recognition across taxa: A review of the literature and a look into the future. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *375*(1802), 20190479. https://doi.org/10.1098/rstb.2019.0479

Cate, C. T., & ten Cate, C. (2018). The comparative study of grammar learning mechanisms: Birds as models. https://doi.org/10.1016/j.cobeha.2017.11.008

Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(24), 6641–6646. https://doi.org/10.1073/pnas.1522306113

Chew, S. J., Mello, C., Nottebohm, F., Jarvis, E., & Vicario, D. S. (1995). Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proc. Natl. Acad. Sci. U. S. A.*, *92*(8), 3406–3410. https://doi.org/10.1073/pnas.92.8.3406

Chew, S. J., Vicario, D. S., & Nottebohm, F. (1996). A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc. Natl. Acad. Sci. U. S. A.*, *93*(5), 1950–1955.

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

Chung, J. E., Magland, J. F., Barnett, A. H., Tolosa, V. M., Tooker, A. C., Lee, K. Y., Shah, K. G., Felix, S. H., Frank, L. M., & Greengard, L. F. (2017). A fully automated approach to spike sorting. *Neuron*, *95*(6), 1381–1394.e6. https://doi.org/10.1016/j.neuron.2017.08.030

Clayton, N. S., & Dickinson, A. (1999). Scrub jays (aphelocoma coerulescens) remember the relative time of caching as well as the location and content of their caches. *J. Comp. Psychol.*, *113*(4), 403–416. https://doi.org/10.1037/0735-7036.113.4.403

Cook, R. G., Levison, D. G., Gillett, S. R., & Blaisdell, A. P. (2005). Capacity and limits of associative memory in pigeons. *Psychon. Bull. Rev.*, *12*(2), 350–358. https://doi.org/10.3758/bf03196384

D'Amelio, P. B., Klumb, M., Adreani, M. N., Gahr, M. L., & Ter Maat, A. (2017). Individual recognition of opposite sex vocalizations in the zebra finch. *Sci. Rep.*, *7*(1), 5579. https://doi.org/10.1038/s41598-017-05982-x

D'Amelio, P. B., Trost, L., & Ter Maat, A. (2017). Vocal exchanges during pair formation and maintenance in the zebra finch (taeniopygia guttata). *Front. Zool.*, *14*, 13. https://doi.org/10.1186/s12983-017-0197-x

Davies, N. B., & Halliday, T. R. (1978). Deep croaks and fighting assessment in toads bufo bufo. *Nature*, *274*(5672), 683–685. https://doi.org/10.1038/274683a0

Del Negro, C., Gahr, M., Leboucher, G., & Kreutzer, M. (1998). The selectivity of sexual responses to song displays: Effects of partial chemical lesion of the HVC in female canaries. *Behav. Brain Res.*, *96*(1-2), 151–159. https://doi.org/10.1016/s0166-4328(98)00009-6

Dhawale, A. K., Poddar, R., Wolff, S. B., Normand, V. A., Kopelowitz, E., & Ölveczky, B. P. (2017). Automated long-term recording and analysis of neural activity in behaving animals. *Elife*, *6*. https://doi.org/10.7554/eLife.27702

Doupe, A. J., & Konishi, M. (1991). Song-selective auditory circuits in the vocal control system of the zebra finch. *Proc. Natl. Acad. Sci. U. S. A.*, *88*(24), 11339–11343. https://doi.org/10.1073/pnas.88.24.11339

Doupe, A. J., & Solis, M. M. (1997). Song- and order-selective neurons develop in the songbird anterior forebrain during vocal learning. *J. Neurobiol.*, *33*(5), 694–709. https://doi.org/10.1002/(SICI)1097-4695(19971105)33:5⟨694::AID-NEU13⟩3.0.CO;2-9

Eales, L. A. (1989). The influences of visual and vocal interaction on song learning in zebra finches. *Anim. Behav.*, *37*(3), 507–508. https://doi.org/10.1016/0003-3472(89)90097-3

Elie, J. E., Mariette, M. M., Soula, H. A., Griffith, S. C., Mathevon, N., & Vignal, C. (2010). Vocal communication at the nest between mates in wild zebra finches: A private vocal duet? *Anim. Behav.*, *80*(4), 597–605. https://doi.org/10.1016/j.anbehav.2010.06.003

Elie, J. E., & Theunissen, F. E. (2015). Meaning in the avian auditory cortex: Neural representation of communication calls. *Eur. J. Neurosci.*, *41*(5), 546–567. https://doi.org/10.1111/ejn.12812

Elie, J. E., & Theunissen, F. E. (2016). The vocal repertoire of the domesticated zebra finch: A data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim. Cogn.*, *19*(2), 285–315. https://doi.org/10.1007/s10071-015-0933-6

Elie, J. E., & Theunissen, F. E. (2018). Zebra finches identify individuals using vocal signatures unique to each call type. *Nat. Commun.*, *9*(1), 4026. https://doi.org/10.1038/s41467-018-06394-9

Elie, J. E., & Theunissen, F. E. (2019). Invariant neural responses for sensory categories revealed by the time-varying information for communication calls. *PLoS Comput. Biol.*, *15*(9), e1006698. https://doi.org/10.1371/journal.pcbi.1006698

Elie, J. E., & Theunissen, F. E. (2020). The neuroethology of vocal communication in songbirds: Production and perception of a call repertoire. https://doi.org/10.1007/978-3-030-34683-6_7

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.*, *5*(3), e1000302. https://doi.org/10.1371/journal.pcbi.1000302

Emery, N. J. (2006). Cognitive ornithology: The evolution of avian intelligence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *361*(1465), 23–43. https://doi.org/10.1098/rstb.2005.1736

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, *306*(5703), 1903–1907. https://doi.org/10.1126/science.1098410

Emery, N. J., Dally, J. M., & Clayton, N. S. (2004). Western scrub-jays (aphelocoma californica) use cognitive strategies to protect their caches from thieving conspecifics. *Anim. Cogn.*, *7*(1), 37–43. https://doi.org/10.1007/s10071-003-0178-7

Favaro, L., Gamba, M., Gili, C., & Pessani, D. (2017). Acoustic correlates of body size and individual identity in banded penguins. *PLoS One*, *12*(2), e0170001. https://doi.org/10.1371/journal.pone.0170001

Flower, T. P., Gribble, M., & Ridley, A. R. (2014). Deception by flexible alarm mimicry in an african bird. *Science*, *344*(6183), 513–516. https://doi.org/10.1126/science.1249723

Foster, E. F., & Bottjer, S. W. (1998). Axonal connections of the high vocal center and surrounding cortical regions in juvenile and adult male zebra finches. *J. Comp. Neurol.*, *397*(1), 118–138.

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*(6005), 845–851. https://doi.org/10.1126/science.1194908

Gentner, T. Q. (2004). Neural systems for individual song recognition in adult birds. *Ann. N. Y. Acad. Sci.*, *1016*, 282–302. https://doi.org/10.1196/annals.1298.008

Gentner, T. Q., Hulse, S. H., Bentley, G. E., & Ball, G. F. (2000). Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. *J. Neurobiol.*, *42*(1), 117–133. https://doi.org/10.1002/(sici)1097-4695(200001)42:1⟨117::aid-neu11⟩3.0.co;2-m

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204–1207. https://doi.org/10.1038/nature04675

Gentner, T. Q., & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature*, *424*(6949), 669–674. https://doi.org/10.1038/nature01731

Gobes, S. M. H., & Bolhuis, J. J. (2007). Birdsong memory: A neural dissociation between song recognition and production. *Curr. Biol.*, *17*(9), 789–793. https://doi.org/10.1016/j.cub.2007.03.059

Godard, R. (1991). Long-term memory of individual neighbours in a migratory songbird. *Nature*, *350*(6315), 228–229. https://doi.org/10.1038/350228a0

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

Hare, J. F. (1998). Juvenile richardson's ground squirrels, spermophilus richardsonii, discriminate among individual alarm callers. *Anim. Behav.*, *55*(2), 451–460. https://doi.org/10.1006/anbe.1997.0613

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Heinrich, B., & Bugnyar, T. (2005). Testing problem solving in ravens: String-pulling to reach food. *Ethology*, *111*(10), 962–976. https://doi.org/10.1111/j.1439-0310.2005.01133.x

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, *16*(2), 129–219. https://doi.org/10.1016/0010-0277(84)90003-9

Honarmand, M., Riebel, K., & Naguib, M. (2015). Nutrition and peer group composition in early adolescence: Impacts on male song and female preference in zebra finches. *Anim. Behav.*, *107*, 147–158. https://doi.org/10.1016/j.anbehav.2015.06.017

Hsu, A., Borst, A., & Theunissen, F. E. (2004). Quantifying variability in neural responses and its application for the validation of model predictions. *Network*, *15*(2), 91–109.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Insley, S. J. (2000). Long-term vocal recognition in the northern fur seal. *Nature*, *406*(6794), 404–405. https://doi.org/10.1038/35019064

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.*, *73*(1), 218–226. https://doi.org/10.1152/jn.1995.73.1.218

Jeanne, J. M., Sharpee, T. O., & Gentner, T. Q. (2013). Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron*, *78*(2), 352–363. https://doi.org/10.1016/j.neuron.2013.02.023

Jeanne, J. M., Thompson, J. V., Sharpee, T. O., & Gentner, T. Q. (2011). Emergence of learned categorical representations within an auditory forebrain circuit. *J. Neurosci.*, *31*(7), 2595–2606. https://doi.org/10.1523/JNEUROSCI.3930-10.2011

Joseph, S., Kumar, S., Husain, M., & Griffiths, T. D. (2015). Auditory working memory for objects vs. features. *Front. Neurosci.*, *9*, 13. https://doi.org/10.3389/fnins.2015.00013

Jouventin, P., Aubin, T., & Lengagne, T. (1999). Finding a parent in a king penguin colony: The acoustic system of individual recognition. *Anim. Behav.*, *57*(6), 1175–1183. https://doi.org/10.1006/anbe.1999.1086

Jun, J. J., Mitelut, C., Lai, C., Gratiy, S. L., Anastassiou, C. A., & Harris, T. D. (2017). *Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction*. https://doi.org/10.1101/101030

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Häusser, M., Karsh, B., ... Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, *551*(7679), 232–236. https://doi.org/10.1038/nature24636

Kaardal, J. T., Theunissen, F. E., & Sharpee, T. O. (2017). A Low-Rank method for characterizing High-Level neural computations. *Front. Comput. Neurosci.*, *11*, 68. https://doi.org/10.3389/fncom.2017.00068

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: Evidence for" fast mapping". *Science*, *304*(5677), 1682–1683.

Kozlov, A. S., & Gentner, T. Q. (2014). Central auditory neurons display flexible feature recombination functions. *J. Neurophysiol.*, *111*(6), 1183–1189. https://doi.org/10.1152/jn.00637.2013

Kozlov, A. S., & Gentner, T. Q. (2016). Central auditory neurons have composite receptive fields. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(5), 1441–1446. https://doi.org/10.1073/pnas.1506903113

Krebs, J., Ashcroft, R., & Webber, M. (1978). Song repertoires and territory defence in the great tit. *Nature*, *271*(5645), 539–542. https://doi.org/10.1038/271539a0

Krentzel, A. A., Macedo-Lima, M., Ikeda, M. Z., & Remage-Healey, L. (2018). A membrane G-Protein-Coupled estrogen receptor is necessary but not sufficient for sex differences in zebra finch auditory coding. *Endocrinology*, *159*(3), 1360–1376. https://doi.org/10.1210/en.2017-03102

Kroodsma, D. E. (1976). The effect of large song repertoires on neighbor" recognition" in male song sparrows. *Condor*, *78*(1), 97–99.

Lee, J., Carlson, D., Shokri, H., Yao, W., Goetz, G., Hagen, E., Batty, E., Chichilnisky, E. J., Einevoll, G., & Paninski, L. (2017). *YASS: Yet another spike sorter*. https://doi.org/10.1101/151928

Ligout, S., Dentressangle, F., Mathevon, N., & Vignal, C. (2016). Not for parents only: Begging calls allow Nest-Mate discrimination in juvenile zebra finches. https://doi.org/10.1111/eth.12450

London, S. E., & Clayton, D. F. (2008). Functional identification of sensory mechanisms required for developmental song learning. *Nat. Neurosci.*, *11*(5), 579–586. https://doi.org/10.1038/nn.2103

Lynch, K. S., Kleitz-Nelson, H. K., & Ball, G. F. (2013). HVC lesions modify immediate early gene expression in auditory forebrain regions of female songbirds. *Dev. Neurobiol.*, *73*(4), 315–323. https://doi.org/10.1002/dneu.22062

Ma, S., Ter Maat, A., & Gahr, M. (2020). Neurotelemetry reveals putative predictive activity in HVC during Call-Based vocal communications in zebra finches. *J. Neurosci.*, *40*(32), 6219–6227. https://doi.org/10.1523/JNEUROSCI.2664-19.2020

MacDougall-Shackleton, S. A., Hulse, S. H., & Ball, G. F. (1998). Neural bases of song preferences in female zebra finches (taeniopygia guttata). *Neuroreport*, *9*(13), 3047–3052. https://doi.org/10.1097/00001756-199809140-00024

Macedo-Lima, M., Boyd, H. M., & Remage-Healey, L. (2021). Dopamine D1 receptor activation drives plasticity in the songbird auditory pallium. *J. Neurosci.* https://doi.org/10.1523/JNEUROSCI.2823-20.2021

Macedo-Lima, M., & Remage-Healey, L. (2020). Auditory learning in an operant task with social reinforcement is dependent on neuroestrogen synthesis in the male songbird auditory cortex. *Horm. Behav.*, *121*, 104713. https://doi.org/10.1016/j.yhbeh.2020.104713

Magland, J. F., & Barnett, A. H. (2015). Unimodal clustering using isotonic regression: ISO-SPLIT.

Maguire, S. E., Schmidt, M. F., & White, D. J. (2013). Social brains in context: Lesions targeted to the song control system in female cowbirds affect their social network. *PLoS One*, *8*(5), e63239. https://doi.org/10.1371/journal.pone.0063239

Margoliash, D. (1997). Functional organization of forebrain pathways for song production and perception. *J. Neurobiol.*, *33*(5), 671–693. https://doi.org/10.1002/(sici)1097-4695(19971105)33:5⟨671::aid-neu12⟩3.0.co;2-c

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. https://doi.org/10.1038/385813a0

Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychon. Bull. Rev.*, *15*(2), 453–7, discussion 458–62. https://doi.org/10.3758/pbr.15.2.453

McComb, K., Moss, C., Sayialel, S., & Baker, L. (2000). Unusually extensive networks of vocal recognition in african elephants. *Anim. Behav.*, *59*(6), 1103–1109. https://doi.org/10.1006/anbe.2000.1406

McCowan, L. S. C., Mariette, M. M., & Griffith, S. C. (2015). The size and composition of social groups in the wild zebra finch. *Emu*. https://doi.org/10.1071/MU14059

McGregor, P. K., Butlin, R. K., Guilford, T., & Krebs, J. R. (1993). Signalling in territorial systems: A context for individual identification, ranging and eavesdropping. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *340*(1292), 237–244. https://doi.org/10.1098/rstb.1993.0063

McInnes, L., Healy, J., Saul, N., & Grossberger, L. (2018). Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, *3*(29), 861.

Meliza, C. D., Daniel Meliza, C., Chi, Z., & Margoliash, D. (2010). Representations of conspecific song by starling secondary forebrain auditory neurons: Toward a hierarchical framework. https://doi.org/10.1152/jn.00464.2009

Meliza, C. D., & Margoliash, D. (2012). Emergence of selectivity and tolerance in the avian auditory cortex. *J. Neurosci.*, *32*(43), 15158–15168. https://doi.org/10.1523/JNEUROSCI.0845-12.2012

Mello, C., Nottebohm, F., & Clayton, D. (1995). Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *J. Neurosci.*, *15*(10), 6919–6925. https://doi.org/10.1523/JNEUROSCI.15-10-06919.1995

Miller, D. B. (1979a). The acoustic basis of mate recognition by female zebra finches (taeniopygia guttata). *Anim. Behav.*, *27*, 376–380. https://doi.org/10.1016/0003-3472(79)90172-6

Miller, D. B. (1979b). Long-term recognition of father's song by female zebra finches. *Nature*, *280*(5721), 389–391. https://doi.org/10.1038/280389a0

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron*, *55*(1), 131–141. https://doi.org/10.1016/j.neuron.2007.06.018

Mooney, R. (2014). Auditory-vocal mirroring in songbirds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *369*(1644), 20130179. https://doi.org/10.1098/rstb.2013.0179

Moore, J. M., & Woolley, S. M. N. (2019). Emergent tuning for learned vocalizations in auditory cortex. *Nat. Neurosci.*, *22*(9), 1469–1476. https://doi.org/10.1038/s41593-019-0458-4

Nealen, P. M., & Schmidt, M. F. (2006). Distributed and selective auditory representation of song repertoires in the avian song system. *J. Neurophysiol.*, *96*(6), 3433–3447. https://doi.org/10.1152/jn.01130.2005

Nick, T. A., & Konishi, M. (2005). Neural song preference during vocal learning in the zebra finch depends on age and state. *J. Neurobiol.*, *62*(2), 231–242. https://doi.org/10.1002/neu.20087

Nieder, A. (2017). Evolution of cognitive and neural solutions enabling numerosity judgements: Lessons from primates and corvids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *373*(1740). https://doi.org/10.1098/rstb.2016.0514

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, serinus canarius. *J. Comp. Neurol.*, *165*(4), 457–486. https://doi.org/10.1002/cne.901650405

Nottebohm, F. (2005). The neural basis of birdsong. *PLoS Biol.*, *3*(5), e164. https://doi.org/10.1371/journal.pbio.0030164

Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., & Harris, K. D. (2016). *Kilosort: Realtime spike-sorting for extracellular electrophysiology with hundreds of channels*. https://doi.org/10.1101/061481

Pagliaro, A. H., Arya, P., Piristine, H. C., Lord, J. S., & Gobes, S. M. H. (2020). Bilateral brain activity in auditory regions is necessary for successful vocal learning in songbirds. *Neurosci. Lett.*, *718*, 134730. https://doi.org/10.1016/j.neulet.2019.134730

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pepperberg, I. M. (1981). Functional vocalizations by an african grey parrot (psittacus erithacus). *Zeitschrift für Tierpsychologie*. https://doi.org/10.1111/j.1439-0310.1981.tb01265.x

Perkes, A., White, D., Wild, J. M., & Schmidt, M. (2019). Female songbirds: The unsung drivers of courtship behavior and its neural substrates. *Behav. Processes*, *163*, 60–70. https://doi.org/10.1016/j.beproc.2017.12.004

Perks, K. E., & Gentner, T. Q. (2015). Subthreshold membrane responses underlying sparse spiking to natural vocal signals in auditory cortex. *Eur. J. Neurosci.*, *41*(5), 725–733. https://doi.org/10.1111/ejn.12831

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, *333*(6042), 595. https://doi.org/10.1126/science.1207327

Phan, M. L., Pytte, C. L., & Vicario, D. S. (2006). Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proc. Natl. Acad. Sci. U. S. A.*, *103*(4), 1088–1093. https://doi.org/10.1073/pnas.0510136103

Pinaud, R., & Terleph, T. A. (2008). A songbird forebrain area potentially involved in auditory discrimination and memory formation. *J. Biosci.*, *33*(1), 145–155. https://doi.org/10.1007/s12038-008-0030-y

Potvin, D. A., Ratnayake, C. P., Radford, A. N., & Magrath, R. D. (2018). Birds learn socially to recognize heterospecific alarm calls by acoustic association. *Curr. Biol.*, *28*(16), 2632–2637.e4. https://doi.org/10.1016/j.cub.2018.06.013

Pytte, C. L., Parent, C., Wildstein, S., Varghese, C., & Oberlander, S. (2010). Deafening decreases neuronal incorporation in the zebra finch caudomedial nidopallium (NCM). *Behav. Brain Res.*, *211*(2), 141–147. https://doi.org/10.1016/j.bbr.2010.03.029

Rey, H. G., Pedreira, C., & Quian Quiroga, R. (2015). Past, present and future of spike sorting techniques. *Brain Res. Bull.*, *119*(Pt B), 106–117. https://doi.org/10.1016/j.brainresbull.2015.04.007

Rinnert, P., & Nieder, A. (2021). Neural code of motor planning and execution during Goal-Directed movements in crows. *J. Neurosci.*, *41*(18), 4060–4072. https://doi.org/10.1523/JNEUROSCI.0739-20.2021

Roberts, T. F., & Mooney, R. (2013). Motor circuits help encode auditory memories of vocal models used to guide vocal learning. *Hear. Res.*, *303*, 48–57. https://doi.org/10.1016/j.heares.2013.01.009

Russ, B. E., Ackelson, A. L., Baker, A. E., & Cohen, Y. E. (2008). Coding of Auditory-Stimulus identity in the auditory Non-Spatial processing stream. *Journal of Neurophysiology*, *99*(1), 87–95. https://doi.org/10.1152/jn.01069.2007

Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Comput. Biol.*, *16*(10), e1008228. https://doi.org/10.1371/journal.pcbi.1008228

Sakata, J. T., Woolley, S. C., Fay, R. R., & Popper, A. N. (2020). *The neuroethology of birdsong*. Springer Nature.

Schneider, D. M., & Woolley, S. M. N. (2013). Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron*, *79*(1), 141–152. https://doi.org/10.1016/j.neuron.2013.04.038

Schulze, K., Vargha-Khadem, F., & Mishkin, M. (2012). Test of a motor theory of long-term auditory memory. *Proc. Natl. Acad. Sci. U. S. A.*, *109*(18), 7121–7125. https://doi.org/10.1073/pnas.1204717109

Silk, M. J., Croft, D. P., Tregenza, T., & Bearhop, S. (2014). The importance of fission-fusion social group dynamics in birds. https://doi.org/10.1111/ibi.12191

Simpson, H. B., & Vicario, D. S. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci.*, *10*(5), 1541–1556. https://doi.org/10.1523/JNEUROSCI.10-05-01541.1990

Stripling, R., Kruse, A. A., & Clayton, D. F. (2001). Development of song responses in the zebra finch caudomedial neostriatum: Role of genomic and electrophysiological activities. *J. Neurobiol.*, *48*(3), 163–180. https://doi.org/10.1002/neu.1049

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2018). Call combinations in birds and the evolution of compositional syntax. *PLoS Biol.*, *16*(8), e2006532. https://doi.org/10.1371/journal.pbio.2006532

Ter Maat, A., Trost, L., Sagunsky, H., Seltmann, S., & Gahr, M. (2014). Zebra finch mates use their forebrain song system in unlearned call communication. https://doi.org/10.1371/journal.pone.0109334

Theilman, B., Perks, K., & Gentner, T. Q. (2021). Spike train coactivity encodes learned natural stimulus invariances in songbird auditory cortex. *J. Neurosci.*, *41*(1), 73–88. https://doi.org/10.1523/JNEUROSCI.0248-20.2020

Theunissen, F. E., Sen, K., & Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.*, *20*(6), 2315–2331. https://doi.org/10.1523/JNEUROSCI.20-06-02315.2000

Thompson, J. V., & Gentner, T. Q. (2010). Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *J. Neurophysiol.*, *103*(4), 1785–1797. https://doi.org/10.1152/jn.00885.2009

Thompson, J. V., Jeanne, J. M., & Gentner, T. Q. (2013). Local inhibition modulates learning-dependent song encoding in the songbird auditory cortex. *J. Neurophysiol.*, *109*(3), 721–733. https://doi.org/10.1152/jn.00262.2012

Thomson, D. J., & Chave, A. D. (1991). Jackknife error estimates for spectra, coherences, and transfer functions, advances. *Spectral Analysis and Array Processing*, 58–113.

Tibbetts, E. A., & Dale, J. (2007). Individual recognition: It is good to be different. *Trends Ecol. Evol.*, *22*(10), 529–537. https://doi.org/10.1016/j.tree.2007.09.001

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Series B Stat. Methodol.*, *63*(2), 411–423. https://doi.org/10.1111/1467-9868.00293

Vahaba, D. M., & Remage-Healey, L. (2018). Neuroestrogens rapidly shape auditory circuits to support communication learning and perception: Evidence from songbirds. *Horm. Behav.*, *104*, 77–87. https://doi.org/10.1016/j.yhbeh.2018.03.007

Vates, G. E., Broome, B. M., Mello, C. V., & Nottebohm, F. (1996). Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (taenopygia guttata). *J. Comp. Neurol.*, *366*(4), 613–642. https://doi.org/10.1002/(SICI)1096-9861(19960318)366:4⟨613::AID-CNE5⟩3.0.CO;2-7

Vicario, D. S., & Yohay, K. H. (1993). Song-selective auditory input to a forebrain vocal control nucleus in the zebra finch. *J. Neurobiol.*, *24*(4), 488–505. https://doi.org/10.1002/neu.480240407

Vignal, C., Mathevon, N., & Mottin, S. (2004). Audience drives male songbird response to partner's voice. *Nature*, *430*(6998), 448–451. https://doi.org/10.1038/nature02645

Vignal, C., Mathevon, N., & Mottin, S. (2008). Mate recognition by female zebra finch: Analysis of individuality in male call and first investigations on female decoding process. *Behav. Processes*, *77*(2), 191–198. https://doi.org/10.1016/j.beproc.2007.09.003

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*(5456), 1273–1276. https://doi.org/10.1126/science.287.5456.1273

Volman, S. F. (1993). Development of neural selectivity for birdsong during vocal learning. *J. Neurosci.*, *13*(11), 4737–4747. https://doi.org/10.1523/JNEUROSCI.13-11-04737.1993

Wang, M., Liao, X., Li, R., Liang, S., Ding, R., Li, J., Zhang, J., He, W., Liu, K., Pan, J., Zhao, Z., Li, T., Zhang, K., Li, X., Lyu, J., Zhou, Z., Varga, Z., Mi, Y., Zhou, Y.,

... Chen, X. (2020). Single-neuron representation of learned complex sounds in the auditory cortex. *Nat. Commun.*, *11*(1), 4361. https://doi.org/10.1038/s41467-020-18142-z

Williams, H., & Nottebohm, F. (1985). Auditory responses in avian vocal motor neurons: A motor theory for song perception in birds. *Science*, *229*(4710), 279–282.

Willmore, B., & Tolhurst, D. J. (2001). Characterizing the sparseness of neural codes. *Network*, *12*(3), 255–270.

Woolley, S. M. N., Fremouw, T. E., Hsu, A., & Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat. Neurosci.*, *8*(10), 1371–1379. https://doi.org/10.1038/nn1536

Yanagihara, S., & Yazaki-Sugiyama, Y. (2016). Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. *Nat. Commun.*, *7*, 11946. https://doi.org/10.1038/ncomms11946

Yanagihara, S., & Yazaki-Sugiyama, Y. (2019). Social interaction with a tutor modulates responsiveness of specific auditory neurons in juvenile zebra finches. *Behav. Processes*, *163*, 32–36. https://doi.org/10.1016/j.beproc.2018.04.003

Yu, K., Wood, W. E., & Theunissen, F. E. (2020). High-capacity auditory memory for vocal communication in a social songbird. *Sci Adv*, *6*(46). https://doi.org/10.1126/sciadv.abe0440

Zann, R. A. (1996). *The zebra finch: A synthesis of field and laboratory studies* (Vol. 5). Oxford University Press.