Perception and neural representation of auditory restoration in the songbird

Margaret Claire Bjoring
Charlottesville, Virginia

B.A., Rice University, 2009
M.A., University of Virginia, 2018

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy
for the Degree of Doctor of Philosophy

Department of Psychology

University of Virginia
May, 2021

Thesis advisor: Professor C. Daniel Meliza Margaret Claire Bjoring

Perception and neural representation of auditory restoration in the songbird

Abstract

It is well established that the human auditory system is capable of filling in missing pieces of a speech stream through a process known as phonemic restoration, yet a neural understanding of this phenomenon has been limited by the lack of a suitable animal model. Songbirds, and especially zebra finches, have proven to be a fruitful model for speech perception and production, making them a promising model for studying restoration of complex vocalizations. In this work, I established through behavioral testing that zebra finches experience this illusion and then recorded single-unit neural data to characterize the neural representation of restored birdsong. Using a decoding method, I was able to identify neural responses consistent with restoration. These responses were widely distributed through the avian auditory cortex but especially strongly represented in the deep auditory pathway. I also found that the evidence for restoration grew stronger over the course of the missing syllable, suggesting the involvement of an internal predictive model that suppresses the incoming auditory signal.

Contents

Acknowledgments

THIS DISSERTATION has benefited greatly from the support, assistance, encouragement, and feedback of many more people than I can list here, but I would like to name a few as well as recognize the generous funding from the National Science Foundation, the National Institute for Deafness and Other Communication Disorders, the Hartwell Foundation, the Jefferson Scholars Foundation, and the University of Virginia School of Data Science which made this work possible.

First, I would like to thank my advisor, Prof. C. Daniel Meliza for his mentorship throughout my graduate studies, his guidance and expertise on the design, execution, and analysis of this research, and his feedback on this document at multiple stages of its development. I would also like to thank the other members of my committee, Profs. Alev Erisir, Jianhua "JC" Cang, and Jay Hirsh, for their support and insightful questions.

I am very thankful to the members of the Meliza Lab, past and present, who have been wonderful friends and colleagues. I especially want to acknowledge Samantha Moseley, Katherine Nimchuk, and Dr. Aysegul Gungor Aydin for their assistance in running the behavioral experiment; Ayush Sagar and Crystal Gong for their work developing and testing the behavioral apparatus used in this experiment; and Leah Kiely for her assistance with post-recording histology.

Finally, I would like to thank my parents, Robert Bjoring and Patricia Wiberg, for their unwavering support and encouragement throughout this journey, my grandfather, Kenneth Wiberg, for being an incredible scientific inspiration, and my whole family. Thank you.

1

Introduction

Speech processing

The brain faces an immense challenge in the auditory processing of speech. A typical speech rate for American English is 3.7 words per second (Goldman-Eisler, 1961), and each word may be composed of multiple phonemes which differ from each other by mere milliseconds. Voiced and voiceless consonants, such as /d/ and /t/ or /b/ and /p/, are discriminable based on a difference of 10 ms for the onset of the first formant (Liberman et al., 1958).

The solution to the problem is not so simple as tuning the auditory system to detect these differences. A speaker does not produce the same phoneme the same way every time. Within an individual, there is a large spread in phoneme production based on factors like vocal emphasis, speech speed, the position of a phoneme within a word, and the phonotactics of adjacent phonemes (Liberman et al., 1967). Between speakers, the variation is more severe. The sex and age of a speaker influences the size of the vocal tract, which alters the location of formants, and regional accents shift phoneme boundaries (Peterson and Barney, 1951). Non-native speakers may use a different inventory of phonemes than that of native speakers (Sereno et al., 2002).

It is clear that the auditory system can only partially rely on the acoustics of the auditory signal as detected by the cochlea when processing speech. Indeed, studies have shown that speech perception is strongly influenced by internal models and expectations and that manipulating expectations alters perception (McGurk and MacDonald, 1976, Kohn, 1981, Niedzielski, 1999). These internal models become even more important when the acoustic speech signal is corrupted, as in the presence of background noise (McGowan, 2015).

Much of natural speech processing occurs in noisy environments, and these environments present a particular challenge to listeners with disordered hearing or perception. School-aged children who struggle with reading also tend to have particular problems processing speech in noisy environments—like most elementary school classrooms—putting them further at risk of falling behind (Mody et al., 1997). Older adults with hearing loss also struggle with speech perception in noisy or multi-speaker environments (Dubno et al., 1984, Lunner, 2003), and this difficulty can exacerbate the isolation of this population.

The social importance of being able to perceive speech in noisy environments and the complexity of the task make it both interesting and important to understand the neural mechanisms that accomplish it. Much progress has been made in understanding the electroencephalographic (EEG) correlates of speech perception within a "cocktail party" background. Both low and high frequency oscillations preferentially track attended speech over ignored speech (Ding and Simon, 2012, Zion Golumbic et al., 2013), but the differential strengths of these entrainments are reduced in populations affected by hearing loss (Petersen et al., 2017) while EEG correlates of listening effort increase (Dimitrijevic et al., 2019). The features of the single-neuron responses that create these ensembles, however, are still inaccessible through current technologies, making animal models an important tool in this research.

Birdsong as a model for speech

Songbirds are one of only a few groups of animals, including humans, that learn aspects of their vocal communication rather than producing vocalizations innately. In zebra finches (*Taeniopygia guttata*), a popular species for studying vocalizations, juveniles exhibit a sensitive period for song learning between 25 and 90 days post-hatch (dph), during which time they must hear the song of an adult male to develop species-typical song (Ölveczky and Gardner, 2011). Juveniles then practice the learned song, producing successively more expert renditions as they progress through the vocal development stages which bear resemblance to the development of speech in human infants (Doupe and Kuhl, 1999). The fully crystallized songs are subsequently used for social behaviors such as courtship, pair-bond maintenance, and territory defense (Mooney et al., 2008).

The vocal learning that enables birds to produce complex sequences of sounds proceeds concomitantly with development of the auditory system, which is necessary for males to internalize a model of the tutor song and to compare their vocal production with that model (Mooney, 2009) and for females to tune their auditory systems toward quality male song (Hauber et al., 2016). Brainstem auditory responses mature by 20 dph (Amin et al., 2007), and this timing corresponds with the opening of the sensitive period for song learning (Doupe and Kuhl, 1999). Exposure to tutor song in both males and females during this time is critical to normal development of the auditory system, both behaviorally and at the neuron level (Woolley, 2012). In females, atypical auditory development induced by rearing in an environment deprived of a male tutor results in atypical mate preference (Chen et al., 2017), and in both sexes, perceptual discrimination ability is reduced (Sturdy et al., 2001). Within the auditory system, depriving a juvenile songbird of conspecific song results in systematic differences in the morphology and physiology of forebrain auditory neurons (Lauay et al., 2005, Chen and Meliza, 2020).

Human infants similarly experience a sensitive period for language learning that opens very early in infancy and gradually tapers off throughout late childhood and early adulthood (Newport et al., 2001). As in juvenile songbirds, perceptual development leads the development of vocal productions, although the development of perception and production is not as temporally separated in human infants as it is in juvenile songbirds. At only a few months of age, infants show a differential response to native-language and foreign-language speech (Mehler et al., 1988), and native-language specific perception develops around six months earlier than language-specific production (Kuhl et al., 1992).

The developmental parallels between human infants and juvenile songbirds are further strengthened by the acoustic similarities of birdsong and speech. Birdsong is acoustically complex with hierarchically structured sound sequences that build from single notes into syllables and then motifs, much like speech is built on phonemes that form syllables that form words (Doupe and Kuhl, 1999, Berwick et al., 2011).

Despite the behavioral similarities between songbird and human vocalizations, anatomical differences between the avian and mammalian brains have been seen as a limitation of using birdsong as a model for speech. In both groups, sound waves are detected by the cochlea and passed through a similar series of brainstem, midbrain, and thalamic nuclei (Karten, 1968, Jarvis et al., 2005, Woolley and Portfors, 2013), but the mammalian telencephalon is characterized by the six-layered cortex, with its precise stratifications and columnar processing units. The avian pallium, by contrast, is organized in a combination of lamina and nuclei (Jarvis et al., 2005, Reiner et al., 2005).

A growing body of research, however, points to a number of surprising parallels in the circuitry, molecular markers, and genetic influences in the forebrains of songbirds and humans despite their gross anatomical differences. Information processing in the avian auditory pallium is hierarchical with population coding characteristics and a canonical microcircuit that bear strong resemblance to mammalian cortex (Wang et al., 2010, Calabrese and Woolley, 2015). The extent and distribution of cortical and striatal neurons as well as their projection targets are consistent between mammalian cortex and avian pallium (Reiner et al., 2005). Neuronal markers for mammalian layer 4 input and layer 5 output neurons are present in subgroups of avian pallial neurons with the same connectivity patterns but different spatial arrangements (Dugas-Ford et al., 2012).

The true test of birdsong as a model for speech lies in its ability to answer outstanding questions about the nature of speech and speech processing. Humans will always be the best organism in which to study speech because humans are the only organism that uses it; despite its learned nature and acoustic complexity, birdsong cannot model important features of speech such as complex syntax and semantics (Berwick et al., 2011). In humans, however, we are limited to research technologies like functional magnetic resonance imaging (fMRI) or electroencephalography (EEG) which provide severely limited temporal or spatial resolution of brain activity. To understand the neuron- or circuit-level details of complex speech-like behavior, songbirds provide a compelling and productive alternative.

Reconstructing speech in noise

Returning to the challenge of perceiving speech in noisy environments, songbirds provide an ideal model for probing the neural circuitries that allow the brain to compensate for partially-degraded signals. Songbirds are social animals that often live within large colonies, and the acoustic similarities between speech and birdsong mean that both human and avian auditory systems are challenged by similar noise-induced signal degradation (Bee and Micheyl, 2008). Studying the neural mechanisms that are responsible for allowing birds to perceive conspecific song in noise will shed light on when and why certain human populations experience increased difficulty perceiving speech in noise.

Broadly, the study of speech perception in a noisy environment is known as the "cocktail party effect," a term introduced by Colin Cherry in 1953 to describe the auditory challenge of attending to a single speaker in a multi-speaker background (Cherry, 1953). This term encompasses the effects of attention on speech perception in noise, the masking effects of noise on peripheral auditory responses, and auditory streaming effects (Bronkhorst, 2015). Research in these areas can uncover not only the mechanisms of noise robustness in the auditory system but also the fundamental building blocks of speech perception through careful study of where and how speech perception breaks down when challenged by noise (e.g. Festen and Plomp, 1990, Ericson et al., 2009, Ellinger et al., 2017, Vanthornhout et al., 2019). Animal studies have the potential to add detailed neuron-level information to the behavioral and EEG research conducted with human participants as well as cross-species comparative research that may identify conserved strategies of identifying vocalizations in acoustically complex environments (Bee and Micheyl, 2008).

One auditory phenomenon that has generated interest across human and animal researchers is phonemic restoration, or auditory continuity as it is typically termed in animal research. Phonemic restoration was first described by Warren (1970) as the illusory perception of a deleted phoneme when the deletion is covered by intense, broadband noise, and the specifics of the illusion in speech were further refined by Bregman (1999). In animal models, the illusion is usually simplified to a pure tone interrupted by noise or a simple, tone-like vocalization (Miller et al., 2001, Petkov and Sutter, 2011). This type of auditory continuity is prevalent across a wide range of species, indicating a bias towards an assumption of stimulus continuity in the auditory pathway in the absence of conflicting evidence (Sugita, 1997, Braaten and Leary, 1999, Petkov et al., 2007), and the neural signature of this type of restoration is evident in the mammalian auditory pathway in primary auditory cortex (Petkov et al., 2007) and potentially as early as the auditory brainstem (Bidelman and Patro, 2016).

The limitation of auditory continuity as a model for phonemic restoration is the spectral and temporal simplicity of the stimuli compared to speech as well as the limited behavioral relevance of the stimuli. While auditory continuity can be explained through a simple feed-forward model (Husain et al., 2005, Riecke et al., 2012), phonemic restoration is influenced by the familiarity and meaningfulness of the stimuli (Ishida and Arai, 2016) as well as the context in which it is presented (Sivonen et al., 2006, Grossberg and Kazerounian, 2011). The use of more complex, speech-like stimuli, such as birdsong, in a restoration paradigm has the potential to more directly model phonemic restoration while retaining the benefits of an animal model. Seeba and Klump (2009), for example, demonstrated enhanced restoration of conspecific over heterospecific birdsong in European starlings. To date, however, there has been no investigation of the neural responses to restoration-inducing stimuli in the areas of the songbird brain responsible for high-level auditory processing.

The present work grows out of an interest in bridging this gap between the neural mechanisms and the perception of complex vocalizations as well as computational research on neural signal processing indicating that a subset of neurons in an area of the avian auditory pathway involved in the perception of conspecific vocalizations show enhanced robustness to noise in an auditory signal (Bjoring and Meliza, 2019). In the following chapters, I explore the perception and accompanying neural responses of zebra finches to partially occluded conspecific song, investigating the effects of stimulus familiarity, as well as the ways in which stimulus coding changes across the levels of the auditory forebrain at both a single-neuron and population level. Finally, I conclude with a discussion of the implications of this research on the understanding of the auditory processing of birdsong and human speech perception, as well further areas of research suggested by this work and others.

2

Perception of auditory restoration in the zebra finch

INTRODUCTION

BACKGROUND NOISE DURING SPEECH presents a serious threat to successful communication. Both environmental noises and the noise of multi-speaker backgrounds are capable of obscuring the fine acoustic features of speech and degrading intelligibility. To compensate for this, the auditory system has developed a number of mechanisms, both peripheral and central, to mitigate the effects of noise on speech intelligibility (Bronkhorst, 2015). Binaural listening with the accompanying interaural time and level differences helps to unmask speech and aid intelligibility at the peripheral level (Ellinger et al., 2017). Cues such as spectral characteristics of voices and spatial separation of sound sources promote streaming effects (Festen and Plomp, 1990, Ericson et al., 2009). Attention also affects the degree to which speech can be tracked in competing noise (Vanthornhout et al., 2019).

Beyond mechanisms for improving the detection and signal-to-noise ratio of speech, the brain also has mechanisms to fill in parts of the speech signal that were lost to noise, much like the visual system fills in the blind spot or a partially occluded object (Komatsu, 2006). Phonemic restoration is an ideal model for studying the mechanisms of auditory filling-in because it is a well-defined auditory processing paradigm that reduces many of the variables involved in natural listening. First described by Warren (1970), the phenomenon of phonemic restoration is an auditory illusion in which listeners report hearing a deleted phoneme when the deletion is replaced with noise but not when left as a silent gap. This effect is at least partially knowledge-dependent, as the illusion is stronger for real versus nonsense words and for native versus non-native speakers (Ishida and Arai, 2016).

Efforts to test phonemic restoration in animal models have focused on the auditory continuity illusion, which abstracts phonemic restoration away from a speech-specific signal. In this illusion, a simple tone is interrupted by a short burst of noise, producing the illusion that the tone has continued behind the noise. This form of auditory restoration has considerable behavioral evidence among non-human primates (Petkov et al., 2003), cats (Sugita, 1997), and rodents (Kobayasi et al., 2012). The auditory continuity illusion, however, presents several important limitations: the simple tone signals do not have behavioral relevance for the animal subjects, and the occluded section of tone can be extrapolated from the signal on either side of the occlusion, potentially eliminating the need for involvement of top-down mechanisms in the restoration.

The widespread prevalence of the auditory continuity illusion among animal models as well as behavioral studies of auditory restoration in European starlings (Braaten and Leary, 1999, Seeba and Klump, 2009) suggests that zebra finches should perceive the auditory restoration illusion. However, this assumption has not been tested. Establishing zebra finches' perception of auditory restoration would provide a model for understanding the neural basis of auditory restoration as well as the effects of development and auditory experience, including environmental and genetic manipulations, on a complex listening task. The use of behaviorally relevant stimuli, like conspecific song, also presents the opportunity to investigate the effects of knowledge and context on the perception of restoration. In humans, there exist striking differences in the perception of phonemic restoration between native and non-native speakers of a language (Samuel and Frost, 2015, Ishida and Arai, 2016) as well as between meaningful and nonsense words or phrases (Verschuure and Brocaar, 1983, Sivonen et al., 2006, Groppe et al., 2010). Seeba and Klump (2009) reported similar effects in starlings listening to conspecific and heterospecific song, but the question of whether this effect extends to social familiarity with conspecific songs has not been addressed.

Here, we adapt a paradigm established by Petkov et al. (2003) for testing the perception of a continuity illusion in macaque monkeys to assess auditory restoration in zebra finches. We further investigate whether social familiarity with the song stimuli used in the experiment increases susceptibility to the restoration illusion.

METHODS AND MATERIALS

ANIMALS

All animal use was performed in accordance with the Institutional Animal Care and Use Committee of the University of Virginia. Adult zebra finches were obtained from the University of Virginia breeding colony. Eight zebra finches were used for song familiarization and 12 were trained on the behavioral experiment.

Song recording and social familiarization

Recordings were made of the songs of eight adult male zebra finches to use as stimuli in the behavioral and electrophysiological experiments. Each singer was housed individually in a sound isolation box (Eckel Industries, Cambridge, MA) with *ad libitum* food and water on a 16:8 h light:dark schedule. A lavalier microphone (Audio-Technica Pro 70) was positioned in the box near a mirror to stimulate singing. The microphone signal was amplified and digitized with a Focusrite Scarlett 2i2 at 44.1 kHz, and recordings to disk were triggered every time the bird vocalized using Jill (`https://github.com/melizalab/jill`; version 2.1.4), a custom C++ real-time audio framework. A typical recording session lasted 1–3 days. From each bird's recorded corpus, a single representative motif was selected and high-pass filtered using a 4th-order Butterworth filter with a cutoff frequency of 500 Hz.

Subsequent to song recording, the eight males were randomly assigned to two groups of four and housed in group cages in separate rooms in the breeding colony. Experimental birds were housed in one of the group cages for at least one week to become familiar with the songs of the recorded males. Experimental birds were assigned essentially at random, but with the constraint that they had no prior social contact with the males in the group cage that they were not placed in. Thus, familiarity was counterbalanced, with half of the motifs familiar to a different half of the experimental subjects.

Behavioral experiment

Operant apparatus

Behavioral experiments were run on a single-board computer (Beaglebone Black) with a custom expansion board (`https://meliza.org/starboard`, revision A2A) that interfaced with the operant manipulanda, cue lights, house lights, and feeder. The experiments were implemented using `decide` (`https://github.com/melizalab/decide`; version 3.2.1), an event-driven framework our lab has developed for controlling behavioral experiments. Each subject was housed individually in an acoustic isolation box (Eckel Industries) with its own apparatus and single-board computer, which sent the trial data it collected to a centralized database (django-decide-host; `https://github.com/melizalab/django-decide-host`).

Acoustic stimuli were presented by the single-board computer through an Altec Lansing Orbit iML227 USB speaker. The subject interacted with the apparatus by pecking an opening in a custom printed circuit board fitted with infrared beam-break detectors and cue lights. Reinforcement was standard finch seed, delivered through a custom 3D printed inlet housing a motorized screw shaft, which was advanced by a stepper motor for 500 ms to deliver approximately 3 seeds (median; range 0–13).

Shaping Following social exposure, behavioral subjects were moved to an acoustic isolation box and allowed to acclimate for 1–3 days. Throughout the subsequent shaping, training, and testing stages of the experiment, birds were maintained in the box on an semi-open economy. They received seed from the feeder at 5–10 minute intervals throughout the day, but the feeder was shut off at least 30 minutes before beginning a training or testing session. During sessions, food was only available by completing trials. Behavior was monitored to ensure birds received adequate food, feeding intervals were adjusted to ensure the birds maintained surplus seed, and sessions were terminated if the bird went for more than 4 hours without eating.

Subjects were trained to peck the response panel using a standard autoshaping paradigm. First, a cue light located near the opening was lit just before automatic food delivery. Once the bird started pecking at the opening, automatic food delivery stopped, and reinforcement was only given after pecking. It was often helpful to suspend a small piece of string in the opening during this initial shaping stage to encourage exploration. There were three blocks of 100 trials: in the first block, the bird had to peck the lit opening once; in the second, it had to peck twice; and in the third, the bird had to peck twice, but the cue light was eliminated. One of the 12 birds was excluded during this stage because it failed to learn the pecking behavior.

Stimuli For each of the eight motifs, eight variants were constructed using a $2 \times 2 \times 2$ design (Figure 2.1). The first factor was the timing of the critical interval chosen for manipulation. Two non-overlapping intervals were chosen for each motif, avoiding both the first and last notes of the motif. The second factor was whether the motif was continuous or discontinuous. The discontinuous variants were constructed by deleting the sound in the critical interval. The third factor was the duration of white noise added to the motif. In the occluding case, the noise was only present during the critical interval; in the masked case, the noise was present throughout the motif. Following Petkov et al. (2003), edge artifacts were minimized by applying a 3 ms cosine ramp to the onsets and offsets of the noise and the gaps in the occluding case. In the masked case, noise onset and offset used a 25 ms cosine ramp.

The stimuli were amplified so that the unmodified motifs all had a RMS amplitude of 50 ± 2 dB SPL at the location where the bird interacted with the operant response panel, as measured by an NTi Audio XL2 Sound Level Meter. For each variant, the amplitude of the white noise was varied relative to the song in 5 dB increments between 20 and -15 dB signal-to-noise ratio (SNR), corresponding to SPLs between 30 and 65 dB. Thus, there were a total of 512 different stimuli. For an example of the full set of stimuli generated for one motif, see Figure 2.1.

BEHAVIORAL TASK Initial auditory shaping consisted of a single song motif. The bird pecked to initiate a trial and a song motif was presented. A second peck within 1 s of the end of the motif was rewarded. In the second stage of auditory shaping, the bird listened to two motifs and was rewarded for withholding a peck until the second motif was presented.

After shaping was completed, birds were trained on the main task, which was to detect a motif with a gap in a sequence of otherwise identical but continuous motifs presented with 200 ms inter-stimulus intervals. The position of the discontinuous motif in the sequence was random, but the first motif was always continuous. The finches had a short window from the start of the gap in the discontinuous motif to peck for a correct response, which was rewarded with seed. Pecking at any other time during the trial was a false alarm, and failure to peck when there was a discontinuous motif was a miss. False alarms and misses were punished with a 2 s “time-out” during which the house lights were extinguished and trials could not be initiated. Stimulus playback ended immediately after a peck response, whether correct or incorrect. On

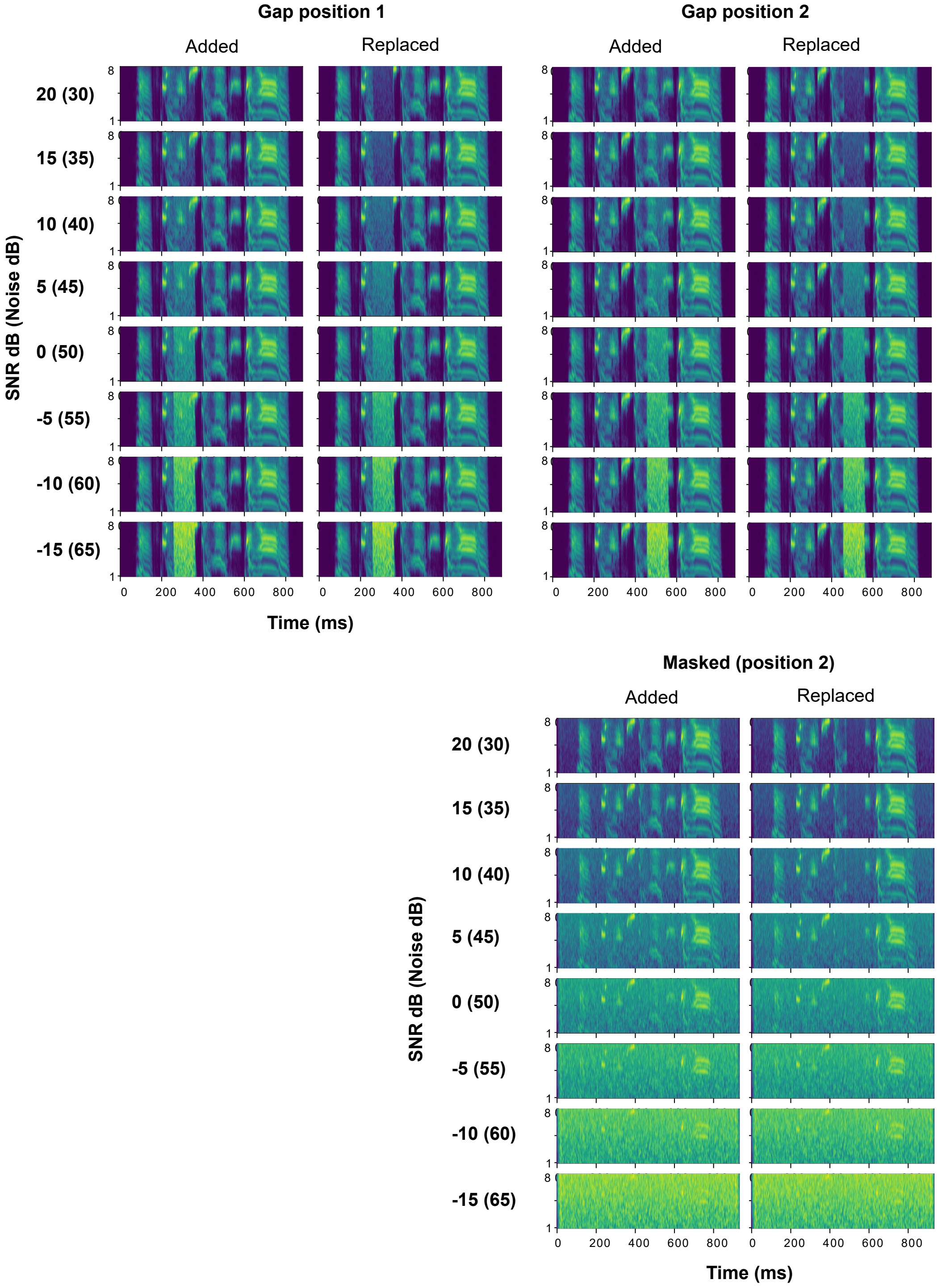


Figure 2.1: All stimulus variants for a single motif.

20% of the trials, all of the motifs were continuous, so the correct response was to withhold a peck, and a reward was delivered at the end of the motif set.

The initial training only used familiar motifs, and only the variants with occluding noise. To make the starting difficulty as low as possible while still allowing the birds to learn the task, the SNR was 20 dB, the critical interval was 200 ms, the sequence was three motifs long, and the finches had a 2 s window from the start of the gap in the discontinuous motif to peck for a correct response, which was rewarded with food. The difficulty of the task was progressively increased by incrementally shortening the response window from 2 s to 1 s, then reducing the critical interval from 200 ms to 150 ms to 100 ms, then increasing the number of motifs from 3 to 4 to 5. For the purpose of tracking performance, the log-odds ratio (LOR) was calculated empirically from sliding blocks of 50 trials as

$$LOR = \log\left(\left(\frac{H}{M} \times \frac{CR}{FA}\right)^{\frac{1}{2}}\right).$$

where H is the proportion of trials with hits, M is the proportion with misses, CR is the proportion with correct rejections, and FA is the proportion with false alarms. To be included in the study, birds had to achieve and maintain performance above an LOR of 1. Birds that failed to show a systematic increase in performance over a two week period were excluded. Five of the 12 birds were excluded during task-specific shaping. The remaining six subjects learned the task within 15700 ± 7800 trials and achieved a final LOR of 1.15 ± 0.09 (see Table 2.1 for full breakdown of trials).

Five of the 12 birds had previously been trained on a similar same-different task where they were asked to detect the different song motif in a set. Of those birds, three successfully learned the new task and were included in this study, and two were excluded during task-specific shaping.

Testing auditory restoration Auditory restoration is characterized by an illusion of continuity when a gap in a stimulus is occluded by noise of similar amplitude. Thus, as the

Table 2.1: Trial numbers per block during task-specific training.

Subject	3 Motifs	4 Motifs	5 Motifs	Ending LOR	Sex	Group
O186†	3268	3044	839	1.15 ± 0.19	M	A
P17	2703	8193	12340	1.24 ± 0.19	F	A
P24	2832	10195	13344	1.02 ± 0.26	M	A
P29	2807	13371	12588*		F	B
P30	1360	6104*			M	B
P35	2155	7671	3923*		M	A
P42	1876	6772	10203	1.20 ± 0.22	F	A
P49†	10375	2237	9327	1.20 ± 0.17	F	B
P52†	191	687‡			F	B
P76†	3731	2063	1729	1.06 ± 0.19	M	B
P8†	4980	3920*			F	B

*bird was excluded during this stage
†bird was previously trained on a similar same-different task
‡bird was excluded for low trial initiation

noise amplitude increases, the odds of responding during the correct interval is expected to decline while the false alarm rate remains constant. We tested this prediction by systematically varying the amplitude of the occluding noise from 20 to −15 dB SNR. The noise level was the same for all motifs within a trial. Trials with louder noise were introduced in stepwise blocks. That is, we started with all the trials at 20 dB SNR, then switched to a new block with 30% trials at 15 dB SNR. In the next block, there were 22% trials at 15 dB, 22% trials at 10 dB, and so forth. We stopped on the block where the lowest SNR was 0 dB (50 dB SPL) or where the performance on the lowest-SNR stimuli dropped below chance, whichever came later. This procedure ensured that a large proportion of the trials in each block were relatively easy for the bird, which helped to maintain baseline performance and reduced the likelihood that birds would become frustrated and switch to a guessing-based strategy.

During testing sessions, the unfamiliar motifs were included for the first time, and trials with occluded variants were randomly interleaved with trials with masked variants. The noise intensity distribution in the masked trials was matched to that of the occluded trials. Masking noise is an important control because it is more difficult for the birds to detect gaps in the motif but does not induce an illusion of continuity, so the decline in performance with noise amplitude is expected to be shallower. Moreover, without an illusory percept to fool the bird into thinking the stimulus is continuous, performance is only expected to decline to chance.

Data analysis The data from the behavioral experiment comprised a total of 63,134 trials from the six birds included in the study. After an initial exploratory analysis, we excluded one bird because its false alarm rate on the catch trials (with no gap) was greater than its hit rate. This left 50,809 trials (range 5085–13,611 per bird). Each trial was split into 1–5 intervals, one for each motif the bird heard (recall that trials terminated immediately after the bird responded). Each interval was coded with a single binary dependent variable, *peck*, which was 1 if the bird pecked during the interval and 0 otherwise, and with seven independent variables: *dB*, the amplitude of the noise (coded as a factor with 8 levels); *gap*, the presence of a gap in the motif; *position*, the position of the interval in the sequence (coded as a factor); *familiarity*, whether the subject had been exposed to the motif in social housing; *condition*, whether the noise was occluding or masking; *bird*, the identity of the subject (coded as a factor); and *song*, the identity of the motif variant (two gap positions per motif, coded as a factor).

A generalized linear mixed-effects model (GLMM) was used to infer the effects of noise intensity and familiarity on the subjects' ability to detect the motif with a gap in it. *Peck* was modeled as a binomial random variable with log odds that depended on a linear function of *dB*, *gap*, *position*, *familiarity*, and their interactions. Two of the higher-order interactions (*dB*position*familiarity* and *db*gap*position*familiarity*) had to be removed for the parameter estimation to converge. Random effects were included to account for variations in psychophysical curves associated with subject and motif identity (Figure 2.2 and Figure 2.3), with motif identity nested in subjects. The parameters were estimated in R using lme4 (version 1.1-23). The model specification was as follows:

```
peck ~ 0 + dB*gap*familiarity + dB*gap*position + (0 + ndB |
    subject/song) + (0 + ndB:gap | subject/song)
```

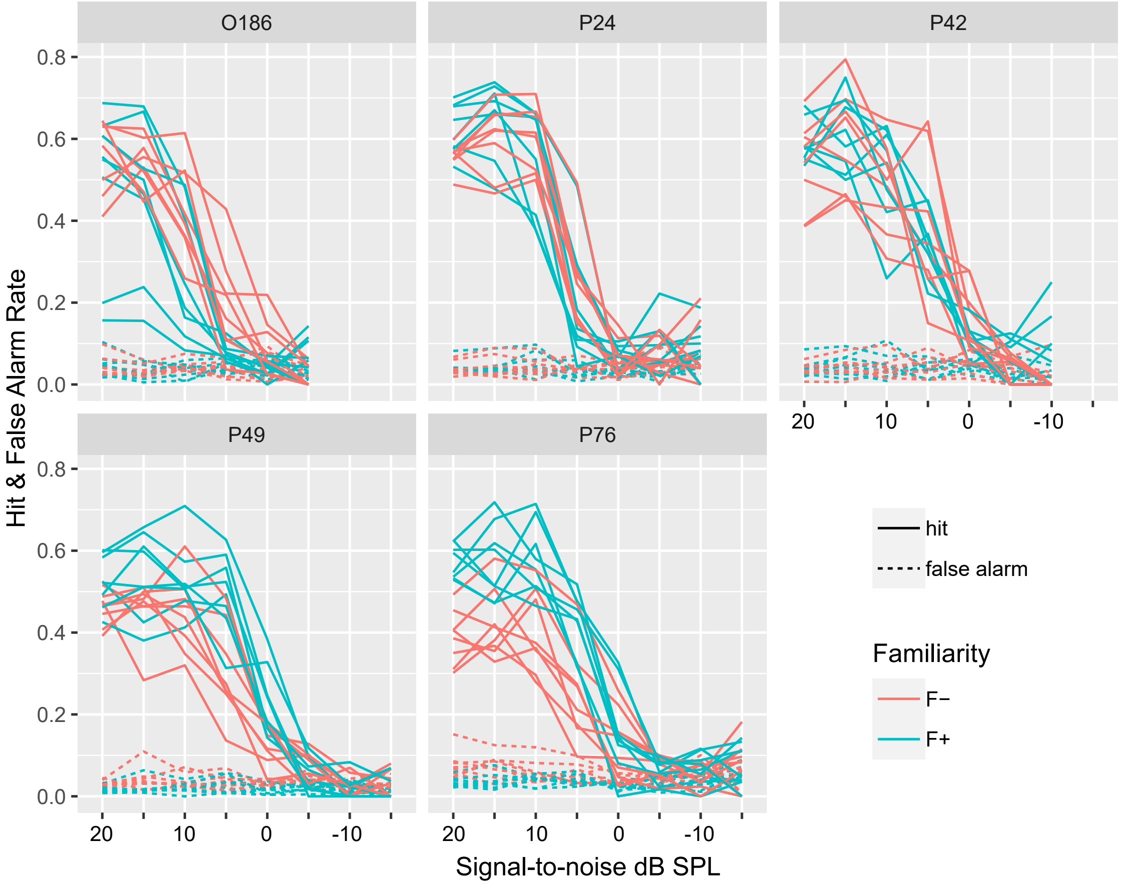


Figure 2.2: Hit and false alarm rates for each behavioral subject. For most subjects, initial performance was highest for the familiar motifs (blue). Some subjects (e.g. P76) showed different thresholds for different motifs and higher performance variability overall, while others (e.g. P24) performed very consistently across all motifs.

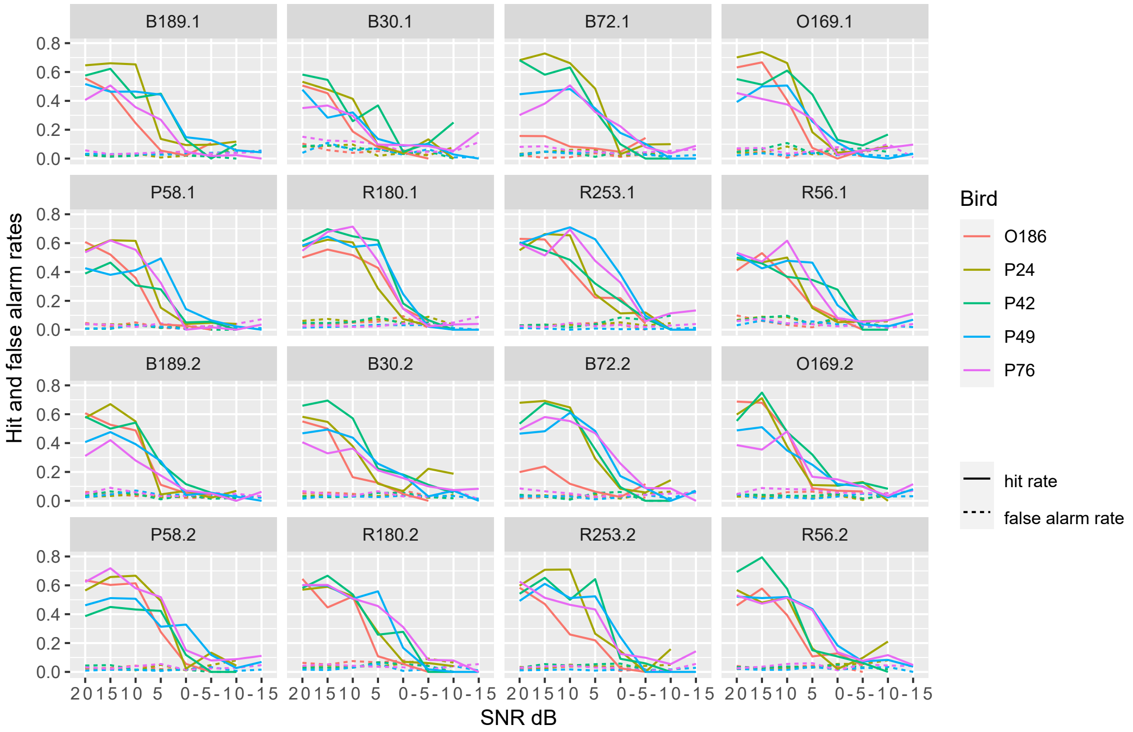


Figure 2.3: Hit and false alarm rates for each motif variant. Performance on different motifs followed a similar pattern, but detection was more difficult for some variants (e.g. B30.1), and some showed greater inter-subject variability in performance (e.g. B72.1)

In this formulation, the parameters have the following interpretations: *dB* gives the false alarm odds for unfamiliar stimuli at each of the eight noise levels; *dB:gap* gives the log odds ratio of pecking when a gap is actually present, again for unfamiliar stimuli and at each of the noise levels; *dB:familiarity* gives the effect of familiarity on the log odds of false alarms; and *dB:gap:familiarity* gives the effect of familiarity on the log odds ratio for detecting a gap. The *dB:position* and *dB:gap:position* parameters correspond to the effect of position within the sequence on false alarm rate and performance respectively. These are essentially nuisance parameters, but they need to be included because the intervals are otherwise not independent – the probability of pecking in later intervals depends on the bird not pecking earlier in the trial – and the distribution of responses was not uniform across the intervals, especially at low SNR (Figure 2.4). Because the model is nonlinear and has many interactions, effects and confidence intervals are reported as estimated marginal means, calculated using the `emmeans` R package (version 1.5.4).

The data for trials with masking noise were analyzed using the same model, but separately. This was to avoid introducing another interaction into an already complex model.

RESULTS

We tested the perception of auditory restoration in zebra finches using an oddball-detection paradigm (Figure 2.5a). Birds were trained to detect a brief discontinuity in a 50 dB conspecific song motif. To create these discontinuities, we selected a 100 ms critical interval in the motif and elided the song note within that interval, replacing the note with white noise (*Replaced* condition). The remaining motifs in the set were continuous versions of the same motif with white noise added on top of the critical interval (*Added* condition). Eight song motifs were used as stimuli in this task with two different critical intervals for each for a total of 16 stimuli. Songs were recorded from males in our colony, and the experimental birds were socialized with four of those males prior to training (Figure 2.5b) so that half of the stimuli were familiar (F+)

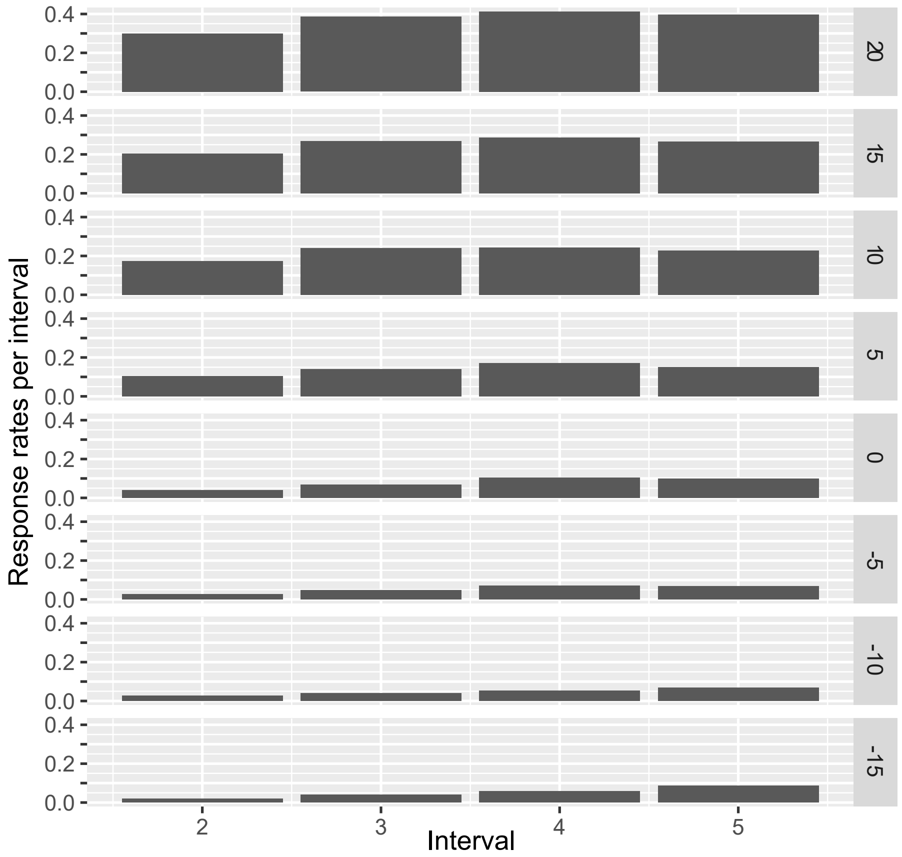


Figure 2.4: As SNR decreases, birds respond later in the motif sequence. At −10 and −15 dB SNR, there is a clear tendency of birds to delay responses until the last motif presentation.

to them and half were unfamiliar (F–). Familiarity was cross-balanced within our experimental birds so that the same stimuli were F+ for some birds and F– for other birds.

At high signal-to-noise ratios (SNR), trained birds ($N = 5$) showed proficient performance, despite the difficulty of learning the task (7 of 12 birds were excluded for failing to reach inclusion criteria). At 20 dB SNR, the mean hit rate was 0.50 ± 0.07 with a false alarm rate of 0.06 ± 0.03 (Figure 2.5c). As SNR within the critical interval decreased, oddball detection became more challenging, and hit rate decreased. The slope of the drop-off in hit rate was similar across all subjects, but subjects showed individual perceptual thresholds. Individual false-alarm rate stayed constant across noise levels and showed low between-subject variability (Figure 2.5c).

As SNR decreases and the task becomes more challenging, we expect performance to go to chance ($\text{LOR} = 0$) unless the auditory restoration illusion comes into play at low SNR. In that case, we expect performance to drop below chance: the bird is not guessing when it hears the *Replaced* stimulus because it is convinced that it heard the *Added* stimulus. To estimate this effect, given the within- and between-subject variability, we fit the data with a generalized linear mixed-effects model (GLMM) and assessed group performance with the marginal parameter estimates (Figure 2.5d). Because the behavioral task allows the *Replaced* stimulus to occur in different positions throughout the motif set, we included the oddball position in the the GLMM. The results presented here and in Figure 2.5d are taken from the estimates of performance when the *Replaced* stimulus occurs in the second position; detection in the second position is the most challenging task (equivalent to a same-different task) and will be least affected by perceptual anchoring effects (Banai and Ahissar, 2010), as we did see evidence of this effect in our data (Figure 2.6).

Task performance provided strong evidence that zebra finches are susceptible to the auditory restoration illusion. Despite the difficulty of the task, birds were able to perform at a high level at positive SNR dB with LOR well above chance (Figure 2.5d). As expected if subjects experienced the illusion, subject performance dropped below chance to $\text{LOR} = -1.19$ at –10 dB

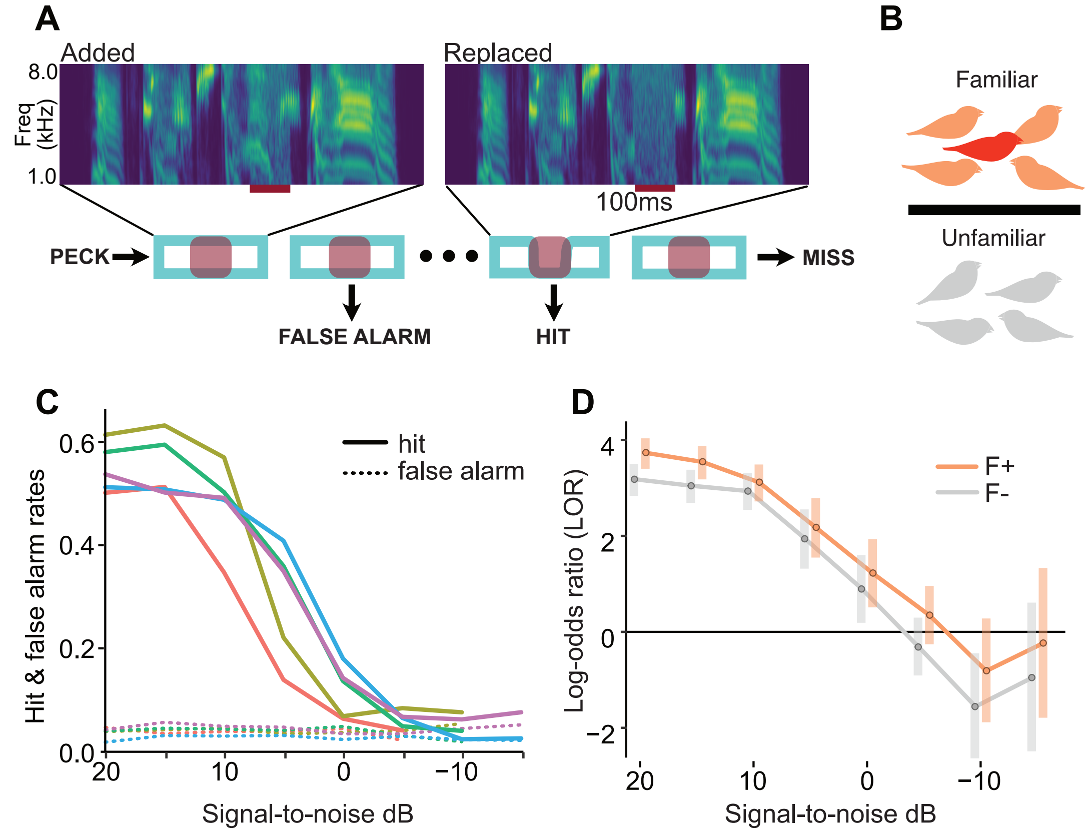


Figure 2.5: Illusory perception of continuity during oddball detection task. A. Subjects performed an oddball detection task to identify a discontinuous song in a series of continuous songs. Signal-to-noise ratios varied between 20 dB and −15 dB. The red line under the song indicates the critical interval for this stimulus. B. Prior to the behavioral task, the subject (red) was familiarized with four of the males whose songs were used in the task (orange). The other four song motifs were unfamiliar to the subject (gray). C. Perceptual thresholds differed between subjects with the hit rate dropping at lower noise levels for some birds than others (solid lines). False-alarm rate stayed constant across noise levels and showed lower between-subject variability (dotted lines). D. GLMM estimates of log-odds ratio performance across noise levels, accounting for within- and between-subject variability. Task performance dips below chance at −10 dB (LOR $= -1.19, z = -2.20, p = 0.028$, 90% CI $= \{-2.24, -0.13\}$) across all stimuli, indicating an illusory percept. Contrary to expectation, the effect was stronger for unfamiliar (F−, gray) stimuli (LOR $= -1.56, z = -2.77, p = 0.006$, 90% CI $= \{-2.66, -0.46\}$) than familiar (F+, orange) stimuli (LOR $= -0.81, z = -1.47, p = 0.14$, 90% CI $= \{-1.90, 0.27\}$), and the difference between familiarity was significant at −10 dB (LOR $= 0.74, z = 2.49, p = 0.013$, 90% CI $= \{0.25, 1.23\}$). Error bars show 90% confidence interval. Data shown are the LOR performance estimates for the first interval in which the oddball could appear. For estimates for all intervals, see Figure 2.6

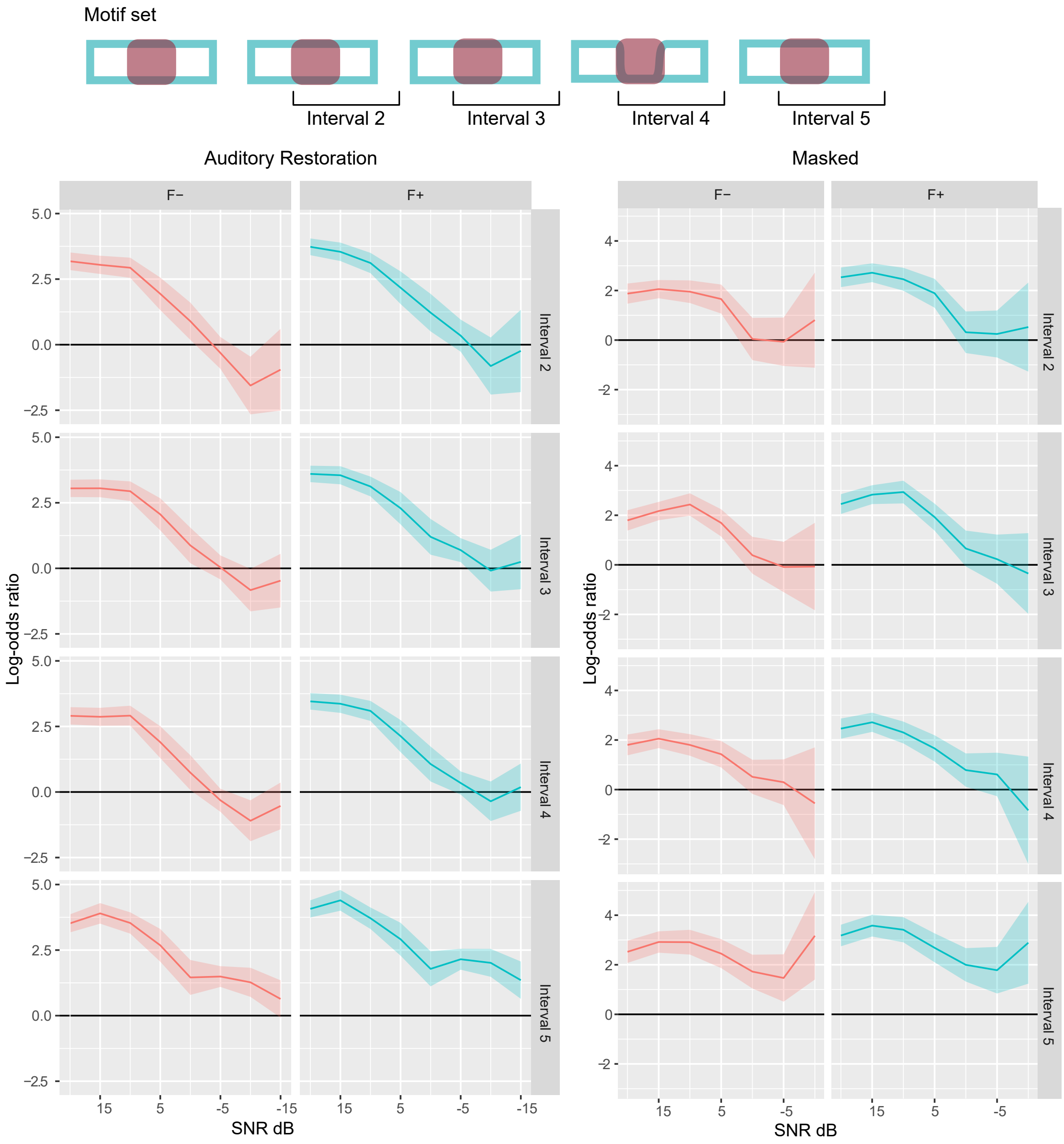


Figure 2.6: GLMM estimates for F+ and F- for all intervals. Top: An example motif set illustrating different intervals. Left: GLMM estimates and 90% confidence intervals for the auditory restoration task faceted by interval and familiarity. The increase in performance in interval 5 likely indicates both a perceptual anchoring effect and a learned guessing strategy for the last motif in the set. Right: GLMM estimates and 90% confidence intervals for the masked task.

SNR.

Interestingly, while we did see an effect of familiarity on the strength of the illusion (Figure 2.5d), it was in the opposite direction that we expected. Unfamiliar stimuli showed a large dip in performance below zero at −10 dB, indicating a strong illusory percept in these stimuli at high levels of noise. The trend to below-chance performance for familiar stimuli was more modest, indicating a weaker or absent illusion. It is important to note, however, that because our study was designed to detect an effect in the opposite direction, the lack of evidence for a strong illusion in the familiar stimuli could be due to multiple factors, including a training effect.

As a control, birds were also presented with a secondary set of stimuli in which the noise in the *Added* and *Replaced* conditions extended the entire length of the motif (masked condition, Figure 2.7a). These stimuli serve as a good control for the auditory restoration task because the intensity of the motif and the noise matches the auditory restoration task, but because the noise mask is not restricted to the critical interval, the masked condition should not induce an illusory percept. If the below-chance performance seen in the auditory restoration task was due to some factor other than the induction of an illusion, the performance on the masked condition should also reflect this dip; if performance simply goes to chance, this provides additional evidence that the behavior of the birds in the auditory restoration task was due to the illusion. Consistent with expectation, performance on the masked task went to chance but not below (Figure 2.7c). The GLMM was unable to converge to an estimate for performance at -15 SNR because both the hit and false alarm rates dropped essentially to zero (Figure 2.7b).

Overall, we found that by manipulating zebra finch song following the principles of phonemic restoration, we could detect behavioral evidence of an illusory percept of continuity at high noise levels. The performance of zebra finches trained to detect discontinuities in conspecific song dropped significantly below chance as the signal-to-noise ratio of the stimulus decreased, while their performance on a control task designed not to induce an illusory percept dropped to chance but not below. These findings make zebra finches an excellent model for exploring

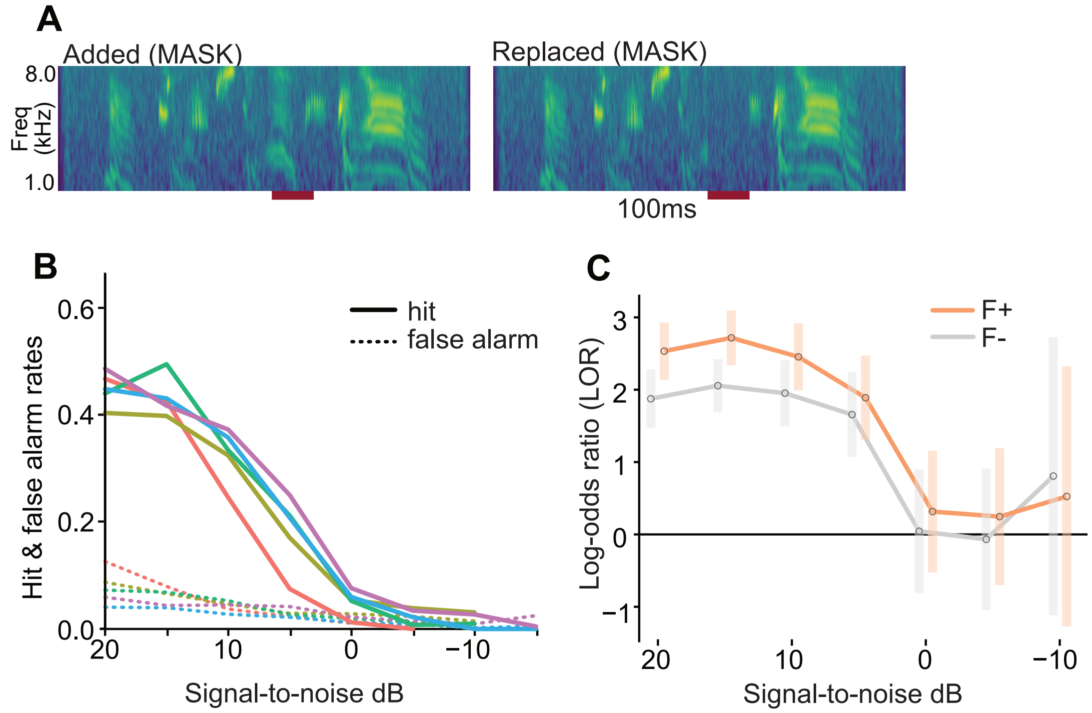


Figure 2.7: Performance on control masked condition goes to chance. A. In the masking condition, noise was added on top of the entire motif instead of the critical interval to eliminate the restoration illusion. Signal-to-noise ratios varied between 20 dB and –15 dB. The red line under the song indicates the critical interval for this stimulus. B. Hit rate was lower than for the auditory restoration task even at the highest signal-to-noise ratio (solid lines). The false-alarm rate decreased as SNR decreased, indicating a different response strategy than for the auditory restoration task (dotted lines). Both the hit and false-alarm rates decreased essentially to zero at –15 dB, which made GLMM estimation at that noise level impossible. C. GLMM estimates of log-odds ratio performance. Task performance drops to chance at 0 dB (LOR $= 0.18$) for both F+ and F- stimuli and remains at chance levels. Data for –15 dB not shown because the model could not converge, given the extremely low hit and false-alarm rates for that noise level. Data shown are the LOR performance estimates for the first interval in which the oddball could appear, for consistency with Figure 2.5. Masking trials also showed a potential perceptual anchoring effect (Figure 2.6).

the neural basis of auditory restoration of complex vocalizations.

DISCUSSION

In this study, we found that zebra finches exhibit behavioral responses consistent with perceiving illusory continuity in an occluded segment of birdsong. When presented with conspecific song interrupted by sufficiently loud white noise, birds' ability to detect the discontinuity in the songs dropped significantly below chance as the illusion convinced them that they had heard an intact song. This finding from the auditory restoration task was further strengthened by performance on a control task which dropped to chance but not below as noise levels increased.

These findings extend our understanding of the perception of illusory continuity of auditory signals briefly interrupted by noise and provide an important bridge between the auditory continuity illusion reported in many animal species and phonemic restoration of speech. While phonemic restoration requires the restoration of a speech phoneme that cannot be simply extrapolated from the speech sounds present on either side of the occluding noise, the auditory continuity illusion typically employs stationary tones (Petkov et al., 2003), simple vocalizations (Miller et al., 2001), or predictable tone progressions such as sweeps (Sugita, 1997). Because of the acoustic complexity of birdsong and its behavioral relevance to the animals, we were able to demonstrate the presence of phonemic restoration-like illusory continuity in an animal model. Braaten and Leary (1999) and Seeba and Klump (2009) demonstrated evidence suggestive of auditory restoration in European starlings listening to modified conspecific versus heterospecific song but did not assess the psychophysical tuning curves that can more conclusively establish the presence of an illusory percept, as in Sugita (1997) and Petkov et al. (2003). This is also the first demonstration of any type of auditory continuity illusion in zebra finches.

One unexpected result from this study was the effect of familiarity that we observed. Given previous literature on this subject in humans (Ishida and Arai, 2016) and the effect of conspecific versus heterospecific stimuli in starlings (Seeba and Klump, 2009), we had hypothesized stronger restoration in conspecific songs that were familiar to the subjects than in unfamiliar conspecific songs. Our results, however, only showed strong support for restoration in the unfamiliar songs (Figure 2.5). We suspect that the lack of restoration of the familiar songs could be due to a technical rather than theoretical reason: birds were trained on the task with the familiar songs both to retain familiarity with them over the length training period and to minimize exposure to the unfamiliar songs, which may have resulted in overtraining on the familiar songs. Hearing the discontinuous songs at low levels of noise so often during training may have blunted the perception of any illusion that might have been elicited by them. Investigation of the neural correlates of auditory restoration in subsequent chapters will shed additional light on whether this effect of familiarity observed in the behavioral data is a true effect or one of overtraining.

However, this result points to an interesting and important feature of auditory restoration: that familiarity with the stimuli does not appear to be necessary to induce restoration. Ishida and Arai (2016) found that the quality of restoration differed between native and non-native speakers of a language, but that both groups restored the missing phoneme. Similarly, in this study, we see evidence that unfamiliar or less familiar stimuli are capable of inducing a clear restoration response. There are two interesting possibilities that emerge from this finding. First, zebra finches may be able to use contextual cues and knowledge of conspecific syllable and spectral patterns to generate an illusory percept of continuity without knowledge of the precise note that was missing, much like the non-native speakers perceived restoration in Ishida and Arai (2016). Second, zebra finches may have mechanisms to familiarize themselves with new conspecific songs extremely rapidly, such that within a few repetitions, the unfamiliar songs become familiar. Research from Yu et al. (2020) indicates that this could be a possibility, adding further support to the idea that the lack of illusion in the "familiar" stimuli could be due to overtraining.

In this study, we used masked stimuli as a control for the auditory restoration task, which presents some advantages and drawbacks. In theory, it is a very strong control because within the critical region, the stimuli are identical, and the only response difference should be the presence or absence of the illusion (Petkov et al., 2003). When implementing this with zebra finches, however, we ran into several challenges. Their performance even at the lowest SNR was worse than for the auditory restoration condition as estimated by our GLMM, and this resulted from both lower hit rates and higher false alarm rates on this task (Figure 2.5b,c versus Figure 2.7b,c. The false alarm rate also decreased as SNR decreased, in contrast to the auditory restoration task where the false alarm rate remained stable. Both these factors seem to indicate that the zebra finches did not see these two tasks as identical and may have adopted different strategies that weaken the ability of the masked condition to control for the presence of the illusion. It was also difficult to obtain stable estimates for the performance at very low SNR because both the hit and false alarm rates were essentially zero. This is likely due to zebra finches' known aversion to white noise (Tumer and Brainard, 2007); in such an extended dose, it may have prompted the birds to simply abandon the task and wait for a more pleasant stimulus set.

An interesting finding of this study unrelated to the question of auditory restoration was the evidence of complex statistical learning in our subjects. It is known that in an oddball detection task, the oddball becomes easier to detect the later it occurs in the stimulus set, an effect known as perceptual anchoring (Banai and Ahissar, 2010). We were struck by the fact that the zebra finches appeared to take advantage of this as the SNR became more challenging by delaying their response until the last stimulus of the set to make a decision (Figure 2.4, which allowed them to continue to perform slightly above chance on the last interval across all tested SNR levels (Figure 2.6).

3

Population-level stimulus reconstruction of occluded birdsong

INTRODUCTION

THE PREVIOUS CHAPTER explored the perception-level illusion of auditory restoration in zebra finches and demonstrated strong evidence for its existence. In human speech, the perceptual phenomenon of restoration has a corresponding neural hypothesis: that there should be auditory neurons that have a response to the illusion-inducing stimulus that is indistinguishable from the continuous stimulus and shows no evidence of discontinuity (Bregman, 1999).

Petkov et al. (2007) demonstrated neural responses in macaques consistent with the neural hypothesis posited by Bregman (1999) using the paradigm of auditory continuity. Given the limitations of the auditory continuity illusion in terms of acoustic complexity and behavioral relevance, however, an analysis of auditory restoration of birdsong has the potential to provide insights into the restoration of complex, speech-like vocalizations.

The potential complexity of the neural responses, however, represents an analytical challenge. To quantify a response to an illusory stimulus, we must assess its similarity to the missing song syllable while accounting for its response to the syllable's absence and its response to the occluding noise. Furthermore, we expect large variability in the responses of single units, both because of the high degree of song selectivity in high-level auditory areas (Calabrese and Woolley, 2015), which means that a neuron that restores one song well may fail to respond at all to the other stimuli, and because some neurons inherently encode more information than others based on their connectivity (Meliza and Margoliash, 2012, Yanagihara and Yazaki-Sugiyama, 2016). This is the type of computational problem that is well suited to machine-learning approaches.

The reconstruction of a stimulus given a set of neural responses is a machine-learning technique that has been used only sparingly in sensory neuroscience in contrast to the related technique of estimating spectrotemporal (or spatiotemporal) receptive fields (STRFs) and predicting neural responses to stimuli (Sharpee, 2013). Stimulus reconstruction provides several advantages over STRFs, including the ability to use multiple neurons or recording channels in the analysis as well as ease of interpretation. Particularly useful for the study of illusions is the fact that it allows for quantification of discrepancies between the presented stimulus and the stimulus as perceived by the brain. For example, Leonard et al. (2016) used this to show bistable perceptions of an ambiguous word in a phonemic restoration experiment.

Applications of stimulus reconstruction have predominantly appeared in the human literature (Pasley et al., 2012, Schäfer et al., 2018, Alickovic et al., 2019, Broderick et al., 2019) where it is seen as a promising approach toward a speech neuroprosthesis (Akbari et al., 2019). It possesses equally strong potential, however, in animal research of sensory perception where simultaneous neural recording and behavioral reporting of perception is impractical or impossible. Mesgarani et al. (2009), for example, use stimulus reconstruction to demonstrate the effects of active and passive listening conditions.

Here, we use stimulus reconstruction to estimate zebra finch perception of birdsong that has been manipulated to induce auditory restoration. We examine the effect of familiarity at a population level and estimate the strength of restoration at different levels of the auditory processing hierarchy.

Methods and Materials

Animals

All animal use was performed in accordance with the Institutional Animal Care and Use Committee of the University of Virginia. Adult zebra finches were obtained from the University of Virginia breeding colony. Fourteen zebra finches were used for extracellular recording.

Extracellular recordings

Extracellular stimuli The stimuli for the extracellular recordings had the same basic structure as Chapter 2, but we allowed the duration of the critical interval to vary and included some additional variants. Critical intervals were selected to overlap completely with a single note in the motif but were never longer than 100ms. The design was 2×4: two critical intervals per motif, and four variants (Figure 3.1a). The variants comprised the continuous stimulus (CS), which was unaltered; the discontinuous stimulus (DS), which replaced the song note within the critical interval with silence; the noise-only stimulus (NS), which was a segment of white noise spanning the critical interval; and the replaced stimulus (RS), which replaced the note within the critical interval with white noise to produce the illusory perception of the song

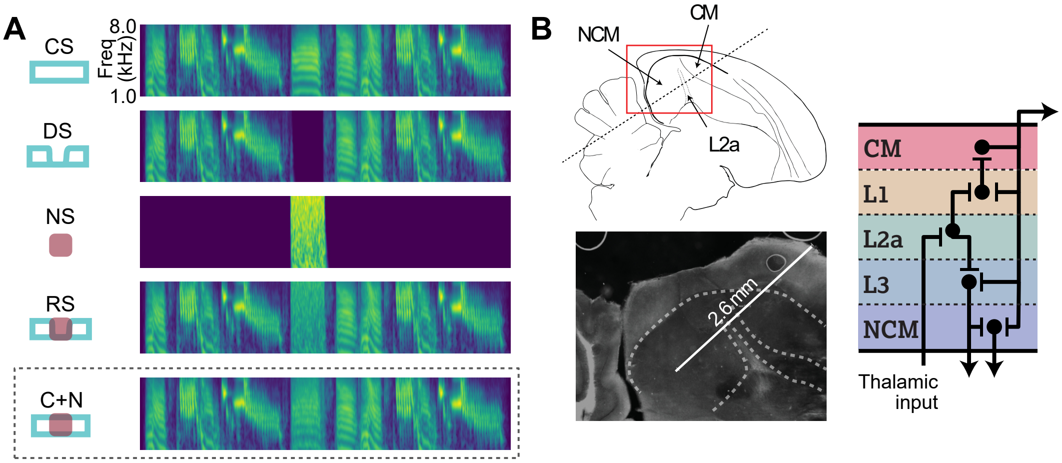


Figure 3.1: Experiment design for extracellular recording. A. Subjects were presented with four variations of each song stimulus: Continuous stimulus (CS) is an unmanipulated song motif, Discontinuous stimulus (DS) has a silent gap during the critical interval, Noise-only stimulus (NS) is a burst of white noise the same duration as the critical interval, and Replaced stimulus (RS) replaces the song note in the critical interval with noise. The fifth variant, C+N (dotted box) was not presented to the subjects but was used in subsequent analyses as an approximation of the expected illusory percept. B. The extracellular probe was inserted at an angle to record from the entire avian auditory cortex. DiI was used to mark the probe track for histological verification. The right panel shows the microcircuit of the avian auditory cortex (adapted from Calabrese and Woolley, 2015).

continuing behind the noise. As with the behavioral stimuli, 3 ms ramps were applied to the edges of the noise and gaps. The noise amplitude was +15 dB relative to the motif amplitude and was not varied.

Surgery Birds were anesthetized with isoflurane inhalation (1–3% in O2) and placed in a stereotaxic apparatus (Kopf Instruments). An incision was made in the scalp, and the skin was retracted from the skull. The recording site was identified using stereotaxic coordinates relative to the Y-sinus. A metal pin was affixed to the skull rostral to the recording site with dental cement, and the skull over the recording site was shaved down but not completely removed. The bird was allowed to recover completely for several days prior to recording.

On the day of recording, the bird was anesthetized with three intramuscular injections of 20% urethane spaced half an hour apart. The bird was placed in a 50 mL conical tube as described in Tremere et al. (2010), and the head pin was attached to a stand in the recording chamber. The thin layer of skull remaining over the recording site was removed along with the dura, and a well was formed around the recording site and filled with phosphate-buffered saline.

STIMULUS PRESENTATION Stimuli were presented with the `sounddevice` python library (version 0.3.10) through a Samson Servo 120a amplifier to a Behringer Monitor Speaker 1C. The RMS amplitude of the unmodified motifs was 70 dB SPL. Stimuli were presented in a pseudorandom order to minimize stimulus adaptation, with 1 s between each song. Each stimulus was presented 10 times.

DATA ACQUISITION Neural recordings were made using a NeuroNexus 32-channel probe in a four-shank, linear configuration (A4x8-5mm-100-400-177-A32) connected to an Intan RHD2132 Amplifier Board. Data were collected by the Open Ephys Acquisition Board and sent to a computer running Open Ephys GUI software (version 0.4.6).

The recording electrode was coated with DiI (Invitrogen). The electrode was inserted at a dorso-rostral to ventro-caudal angle that allowed for recording of all auditory forebrain regions with a single penetration (Figure 3.1b). The probe was lowered into the brain until the local field potentials (LFPs) across channels and shanks showed coordinated responses to birdsong, and the probe was allowed to rest in place for half an hour to ensure a stable recording. Recordings of responses were made across all 32 channels. After the recording, the probe was moved to successively deeper regions of the auditory pathway and additional recordings were made.

HISTOLOGY After recording, birds were administered a lethal intramuscular injection of Euthasol and perfused transcardially with a 10 U/mL solution of sodium heparin in PBS (in mM: 10 Na2HPO4, 154 NaCl, pH 7.4) followed by 4% formaldehyde (in PBS). Brains were immediately removed from the skull, postfixed overnight in 4% formaldehyde at 4 °C, cryoprotected in 30% sucrose (in 100 mM Na2HPO4, pH 7.4), blocked saggitally into hemispheres or on a modified coronal plane (Chen and Meliza, 2018), embedded in OCT, and stored at −80 °C. 60 μm sections were cut on a cryostat and mounted on slides. After drying overnight, the sections were rehydrated in PBS and coverslipped with Prolong Gold with DAPI (ThermoFisher, catalog P36934; RRID:SCR_015961). Sections were imaged using epifluorescence with DAPI and Texas Red filter cubes to locate DiI-labeled penetrations. Images of the electrode tracks were used to identify the locations of recorded units (Figure 3.1b).

Spike sorting Spikes were sorted offline using MountainSort from MountainLab (`https://github.com/flatironinstitute/mountainlab-js`), which provides automated selection of high-quality units with high isolation and low noise (Chung et al., 2017). Single units were additionally filtered by visual inspection for spheroid PCA cluster shape, very low refractory period violations in the autocorrelogram, and stability of the unit throughout the recording. These high-quality single units were included in the dataset if they showed a clear, phase-locked auditory response to at least one stimulus.

Response rates We calculated the average response rates per unit to CS, DS, NS, and RS within the critical intervals and log-transformed the rates to approximate a normal distribution, adding one spike to each unit to set a non-zero baseline. We performed regressions predicting RS from each of CS, DS, and NS with `lme4`, and we used the R^2 of each model to generate confidence intervals for the amount of variance explained using the `psychometic` package (version 2.2) in R.

Stimulus reconstruction

A linear stimulus decoding model was used to predict stimulus spectrograms from recorded neural responses (Mesgarani et al., 2009, Crosse et al., 2016, Leonard et al., 2016). The model is similar to the spectrotemporal receptive field (STRF), in which the expected firing rate of a single neuron at a given time point t is modeled as a linear function of the stimulus spectrogram immediately prior to t. In the linear decoding model, the relationship is reversed, and the expected stimulus at time t is modeled as a linear function of the response that follows. Using a discrete time notation where s_t is the stimulus in the time bin around t, and r_t is the response of a single neuron in the same time bin, then

$$E(s_t) = r_t g_0 + r_{t+1} g_1 + \cdots + r_{t+k} g_k,$$

where k is the number of time bins one looks into the future, and $\mathbf{g} = (g_0, g_1, \ldots, g_k)$ are the linear coefficients of the model. If the errors are independent and normally distributed around the expectation with constant variance σ^2, then this is a ordinary linear model. If there are n time bins in the stimulus, then the stimulus is a vector $\mathbf{s} = (s_0, \ldots, s_n)$ drawn from a multivariate normal distribution. In vector notation,

$$\mathbf{s}|\mathbf{g}, \sigma^2, \mathbf{R} \sim N(\mathbf{Rg}, I\sigma^2),$$

where $\mathbf{R}$ is the $n \times k$ Hankel matrix of the response. Without any loss of generality, the model can be expanded to include the responses of multiple neurons. If there are p neurons, then r_t becomes a p-element vector $(r_{1,t}, \ldots, r_{p,t})$, $\mathbf{R}$ becomes a $n \times pk$ matrix formed by concatenating the Hankel matrices for each of the neurons, and $\mathbf{g}$ becomes a pk-element vector.

Because the model is simply linear regression, standard tools can be used to estimate the parameters $\mathbf{g}$ and σ^2. Using a ridge penalty for regularization, the maximum likelihood estimate of $\mathbf{g}$ is

$$\hat{\mathbf{g}} = \left(\mathbf{R}^\top \mathbf{R} + \lambda \mathbf{I}\right)^{-1} \mathbf{R}^\top \mathbf{s},$$

where $\lambda\mathbf{I}$ is the identity matrix multiplied by the shrinkage penalty for the ridge regression.

Additional regularization can be achieved by projecting the response matrix $\mathbf{R}$ into an alternative basis set, such as a non-linearly spaced series of raised cosines (Pillow et al., 2005). The width of each basis function increases with lag, which gives the model high temporal resolution at short lags and lower resolution at longer lags. This allows the inclusion of longer lags without exploding the number of parameters.

In this study, the response matrix was constructed from the peristimulus time histograms (PSTHs) averaged over 10 trials for each of the 407 auditory single units from all 14 birds, using a bin size of 1 ms. Stimuli were converted to time-frequency representations using a gammatone filter bank (Slaney, 1998) implemented in the Python package `gammatone` (version 1.0) with 50 frequency bands from 1–8 kHz, a window size of 2.5 ms, and a step size of 1 ms. Power was log-transformed with a constant offset of 1, giving the transformed signal a lower bound of 0 dB.

The parameters were estimated from the Continuous (CS), Discontinuous (DS), and Noise-only (NS) stimuli, leaving out the Replaced (RS) responses for prediction, using the Python machine-learning library Scikit-learn (version 0.23.0). We combined all 50 spectral bands into a single multivariate multiple regression, and then used 4-fold cross-validation to determine the best values for the ridge penalty λ, the number of time lags n, the number of basis functions, and the linearity of the basis set. Basis functions were defined as in Pillow et al. (2005). Because the models differed in the number of parameters, Aikake information criteron (AIC) was used for scoring. For the reconstructions presented here, λ was 8.59, n was 300, the number of basis functions was 30, and the linearity factor was 30.

After fitting the model, we used the parameter estimates to decode the stimulus from the responses to the RS stimuli. Using $\tilde{\mathbf{R}}$ to denote these responses, the predicted stimulus is calculated as

$$E(\tilde{\mathbf{S}}) = \tilde{\mathbf{R}}\hat{\mathbf{g}} \tag{3.1}$$

To quantify how similar the decoded stimulus was to the actual stimulus and the other variants, we calculated the correlation coefficient between $\tilde{\mathbf{S}}$ and $\mathbf{S}_{CS}$, $\mathbf{S}_{RS}$, and $\mathbf{S}_{DS}$ within the critical interval for each of the motifs. Additionally, we calculated the correlation coefficient with $\mathbf{S}_{CS} + \mathbf{S}_{RS}$ to simulate the auditory percept of the missing syllable co-occuring with noise (C+N). As a baseline for how good the reconstruction could be, we calculated the correlation between $\tilde{\mathbf{S}}$ and $\mathbf{S}_{RS}$ in an interval outside the critical interval (which was the same for all variants).

Area-level analysis

To understand how different auditory areas in the zebra finch auditory cortex influence the overall reconstruction, we performed a leave-one-area-out analysis. Using the same cross-validated parameters as the full reconstruction, we fit the model to the entire dataset minus the units from one area. This technique allowed us to identify unique contributions made by each of the auditory areas because if the responses of one area were redundant with those of another, the model would shift its weights but the estimated reconstruction would remain unchanged. Changes in the reconstruction seen with this method mean that the model lost irreplaceable evidence when the units from one area were omitted.

To quantify the changes in reconstructions, we calculated correlation coefficients between the reconstruction and CS, DS, RS, and C+N as discussed above, and subtracted the correlation coefficient of the reconstruction without one area from the correlation coefficient of the full reconstruction:

$$\text{Influence} = CC_{\text{full}} - CC_{\text{leave-one-out}}$$

Here, the influence of a given area is positive if leaving that area out of the reconstruction made the correlation coefficient between the reconstruction and a stimulus spectrogram worse; that is, the data from that area generally tend to improve the reconstruction's correlation with the stimulus. We fit the influences of each area with a linear model using `lme4` in R with an interaction between area and stimulus condtiion and used `emmeans` to calculate the effects and confidence intervals of the interactions.

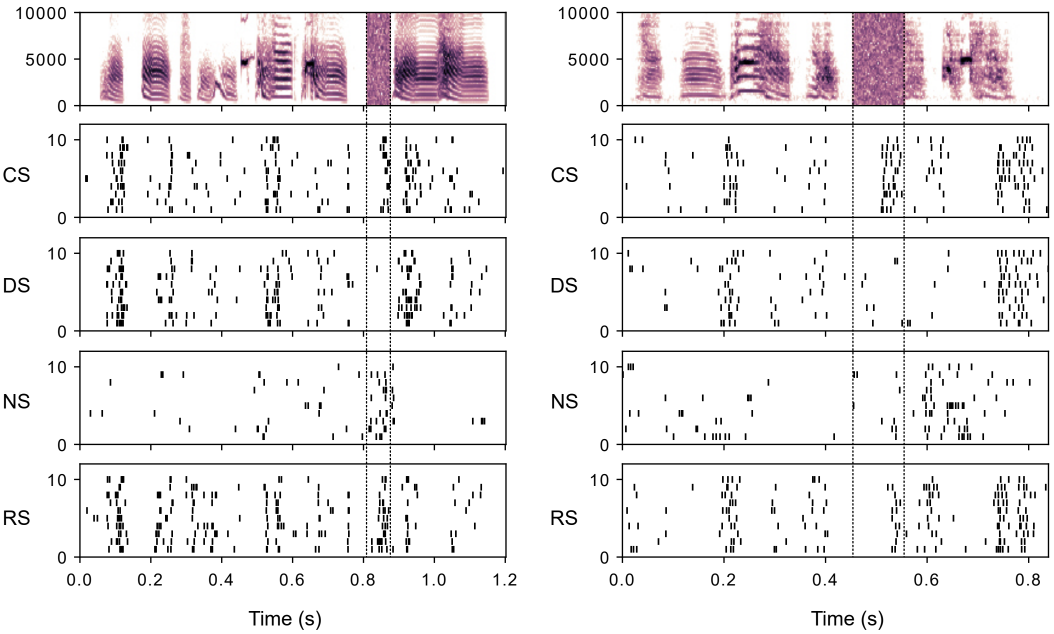


Figure 3.2: Single unit responses show auditory restoration. Two example single unit rasters of 10 trials of the RS stimuli (top) showing evidence of auditory restoration during the critical interval (dotted lines).

RESULTS

To test whether auditory neurons responded to the illusory song segment, we made extracellular recordings of 407 single units ($N = 14$ birds) across the avian auditory cortex, including the caudal mesopallium (CM, $N = 56$ units), field L subunits L1 ($N = 25$ units), L2a ($N = 33$ units), and L3 ($N = 59$ units), and the caudomedial nidopallium (NCM, $N = 90$ units) (Figure 3.1b). We presented anesthetized zebra finches with the illusion-inducing *Replaced* stimulus (RS) as in the behavioral experiment as well as three control conditions: the Continuous unmodified stimulus (CS); a Discontinuous stimulus (DS) where the note in the critical interval was elided but not replaced by any noise; and a Noise-only stimulus (NS) (Figure 3.1a).

We observed a wide variety of response patterns to these stimuli, including many responses that appeared to show auditory restoration (Figure 3.2). Across all single units, the average response rate to RS within the critical interval was most tightly correlated to the response rate to CS during the same interval, providing support for the hypothesis that these units are responding as though the song is present (Figure 3.3).

To draw conclusions from the complex responses represented in our data, we employed a machine-learning approach known as stimulus reconstruction (Crosse et al., 2016) to estimate the stimulus encoded by the entire population of recorded units. Stimulus reconstruction uses ridge regression to map between the neural responses of many units and a single stimulus spectrogram that best explains all of the observed responses. We fit our model with the three control conditions (CS, DS, and NS) using cross validation and AIC model selection to avoid overfitting and estimated the stimulus that elicited the RS responses. Examples of the stimulus reconstructions are shown in the bottom row of Figure 3.4a.

The stimulus reconstruction technique allowed us to test directly the hypothesis that the neural responses to RS encoded the CS syllable that had been replaced by noise. If restoration failed, we would expect reconstructions that look like white noise during the critical interval and a high similarity to RS. If, on the other hand, neurons encoded the illusory syllable, we would expect reconstructions with a high similarity to CS. In Figure 3.4a, we see both of these possible outcomes represented: the first two reconstructions appear similar to the CS syllable in the critical interval, while the last appears to show white noise.

For each critical interval, we calculated the similarity (correlation coefficient) between the reconstructed spectrogram and CS, DS, and RS spectrograms in the critical interval (Figure 3.4b). We also included an additional comparison (C+N) which adds transparent noise to the CS spectrogram, simulating the expected illusory percept of the noise and the restored syllable. The C+N and CS spectrograms showed the highest correlations with the reconstruction, providing strong evidence that the neural population represented the critical interval as though the elided syllable was actually present along with the noise. Correlation of the reconstruction with RS, which was the stimulus that actually elicited the responses used in the reconstruction, was significantly lower than C+N. The correlation with DS was significantly lower than the C+N, CS, and RS correlations, indicating that these neurons do not show evidence of discontinuity in their response, consistent with the hypothesis. An unmanipulated syllable was selected from

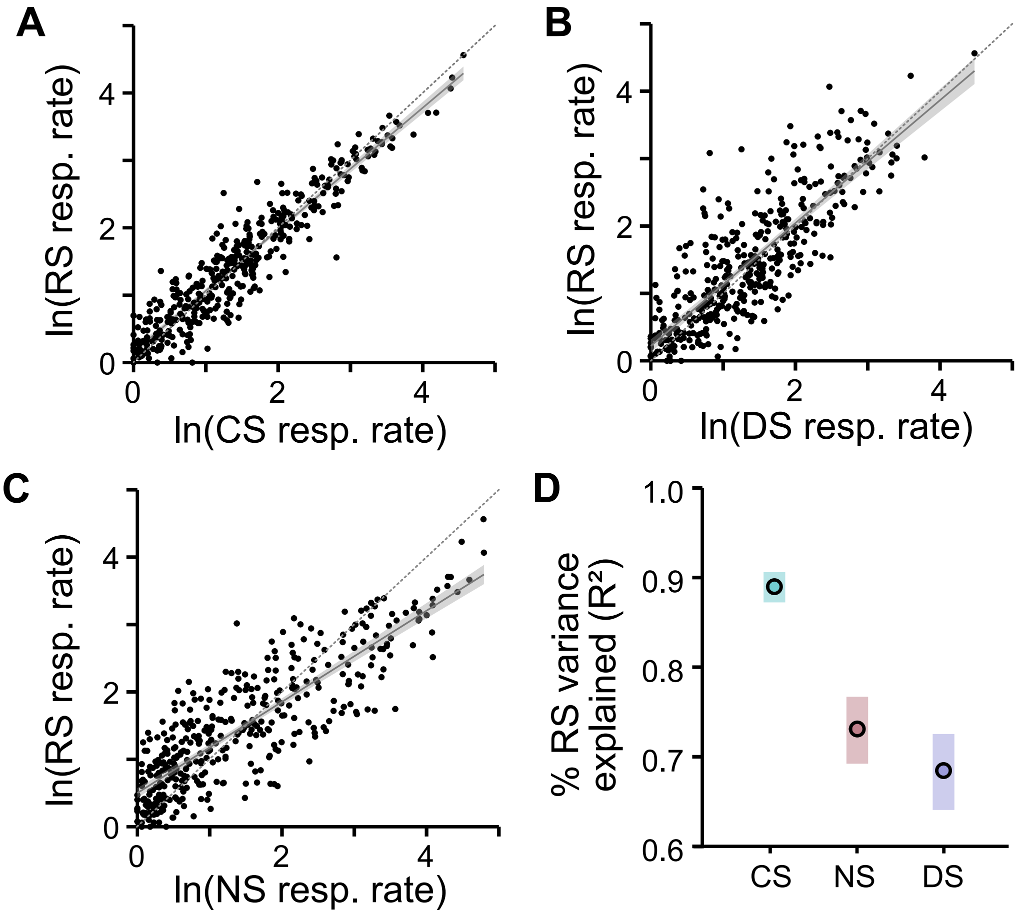


Figure 3.3: RS response rates are most consistent with CS. A-C. Log-transformed response rates to the critical interval of RS for all units ($N = 407$) compared to response rates to CS, DS, and NS stimuli. The shaded line shows the regression of the two response rates, and the dotted line indicates the line of equality. CS response rates are highly predictive of RS response rates, while the rates for DS and NS show more scatter. D. R^2 values and 90% confidence intervals for the regressions. CS rates explain close to 90% of the variance of RS rates, suggesting that neurons were responding as though the missing song syllable was present.

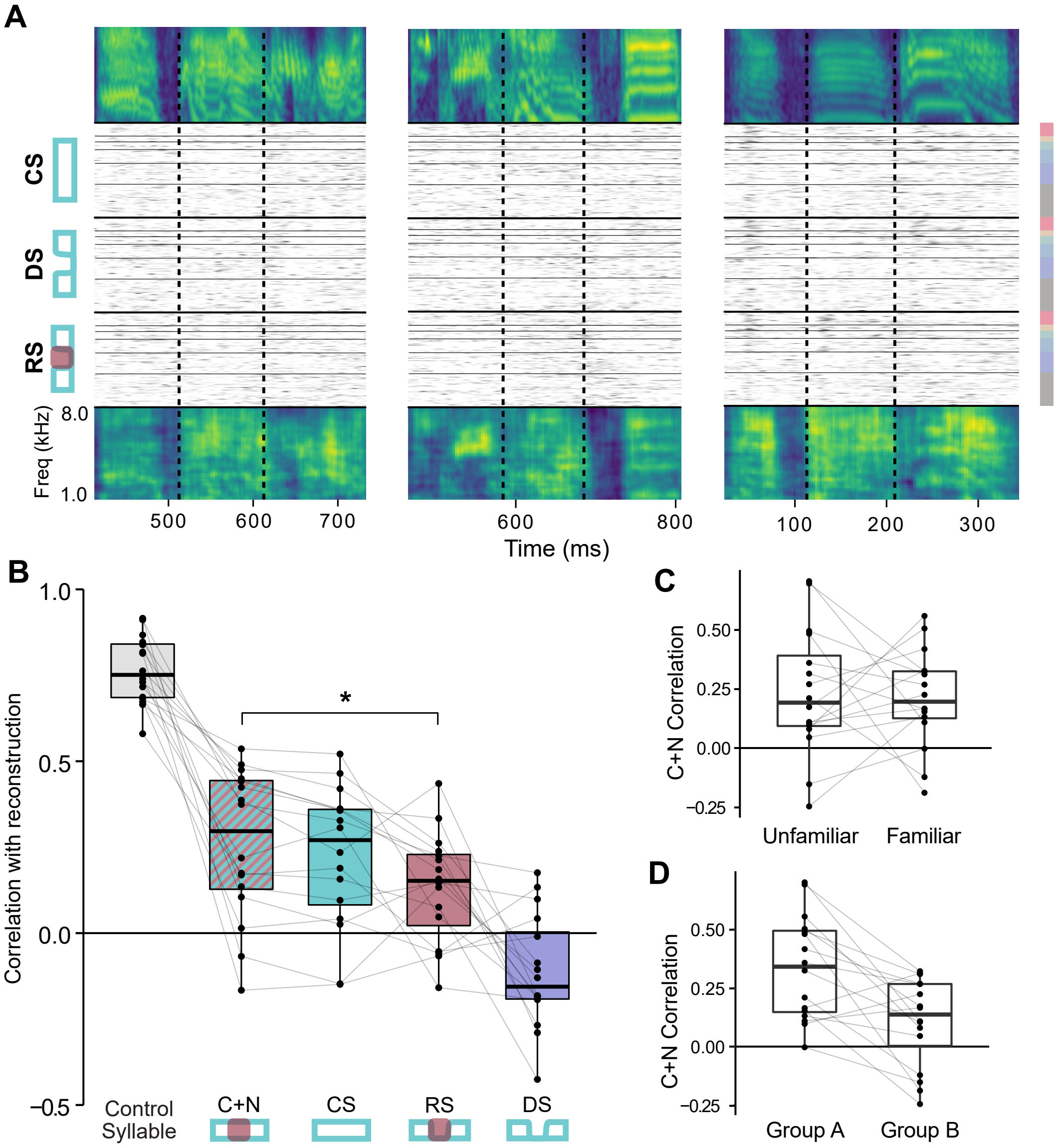


Figure 3.4: Auditory neurons encode illusory syllables. A. Example neurograms (middle) and reconstructions (bottom) of three motifs (top). Reconstructions were predicted from neural responses to RS. The dotted lines indicate the critical interval. The units in the neurograms are sorted by area with the color-coding on the right side corresponding to the microcircuit in Figure 3.1b. Gray indicates unclassified units. The first two examples show evidence of auditory restoration based on the reconstruction, while the third shows a reconstruction that appears more like white noise, as is also apparent from the RS neurogram. B. Similarity (correlation coefficient) was calculated between the critical interval of the reconstruction and the same interval of CS, DS, and RS spectrograms as well as a spectrogram constructed with transparent noise on top of the CS motif (C+N) which approximates the expected illusory percept. The reconstructions showed the highest similarity to C+N and CS spectrograms (0.26 ± 0.22 and 0.22 ± 0.20, respectively), indicating that the recorded neurons as a whole tend to encode the illusory song note during auditory restoration. C+N was significantly more similar than RS ($p = 0.03$, Kenward-Roger), the white noise stimulus that was actually present. Correlation with DS was significantly lower than C+N, CS, and RS correlations ($p < 0.0001$ for each). Control syllables from the reconstruction were correlated with the CS stimulus outside of the critical interval. C. Data were split by stimulus familiarity, and the models were fit separately to assess whether the similarity of the reconstruction to C+N varied by stimulus familiarity. Lines connect the correlations of the same stimulus variation. We detected no significant effect of familiarity on C+N correlation ($p = 0.77$, paired t-test). D. C+N correlation differed significantly across social groups ($p = 0.003$, paired t-test) despite balancing the groups for sex and age.

each stimulus as a control, and the correlation between the reconstruction of that syllable and the original spectrogram (0.76 ± 0.10) shows that all stimuli were well fit by the model.

Because of the effect of familiarity in the behavioral study, we tested whether our neural data showed a similar effect. We fit the reconstruction model separately with data from birds familiar with a given stimulus and data from birds unfamiliar with that stimulus and observed no difference in familiarity on the strength of restoration (Figure 3.4c). However, we did observe a significant difference between the two socialization groups (Figure 3.4d). The birds in these groups had similar distributions of sex (Group A: 3 males, 4 females; Group B: 4 males, 3 females) and ages (Group A: median 222 days, range 128 – 1126 days; Group B: median 163 days, range 122 – 599 days), but they were housed in separate rooms which may have provided different auditory environments (Group A: large colony; Group B: breeding pairs).

Contributions of auditory areas to restoration

Next, we investigated whether the evidence of auditory restoration seen in the stimulus reconstructions was driven by a particular auditory area or whether there was distributed support for auditory restoration throughout avian auditory cortex. To test this, we fit multiple reconstruction models, leaving out the neural responses from one of the auditory areas each time. The advantage of this approach was two-fold: first, it allowed for similar sample sizes between areas, producing reconstructions of comparable quality (Figure 3.5a), and second, it allowed us to identify contributions unique to an auditory area. If the responses of one unit are redundant with another, taking the former out of the model will result in a shift of weight to the latter and no change in the model estimates. If we detect a change in the reconstruction after removing data from one area, we know that area is producing a non-redundant effect on the model.

Figure 3.5b shows the unique influence of CM, L1, L2a, L3, and NCM on the correlation between the reconstruction and C+N, CS, DS, and RS. Across all stimulus variants, only the reconstruction of DS was significantly different between areas. The thalamorecipient area L2a significantly decreased the similarity of the reconstruction to DS, providing early bottom-up

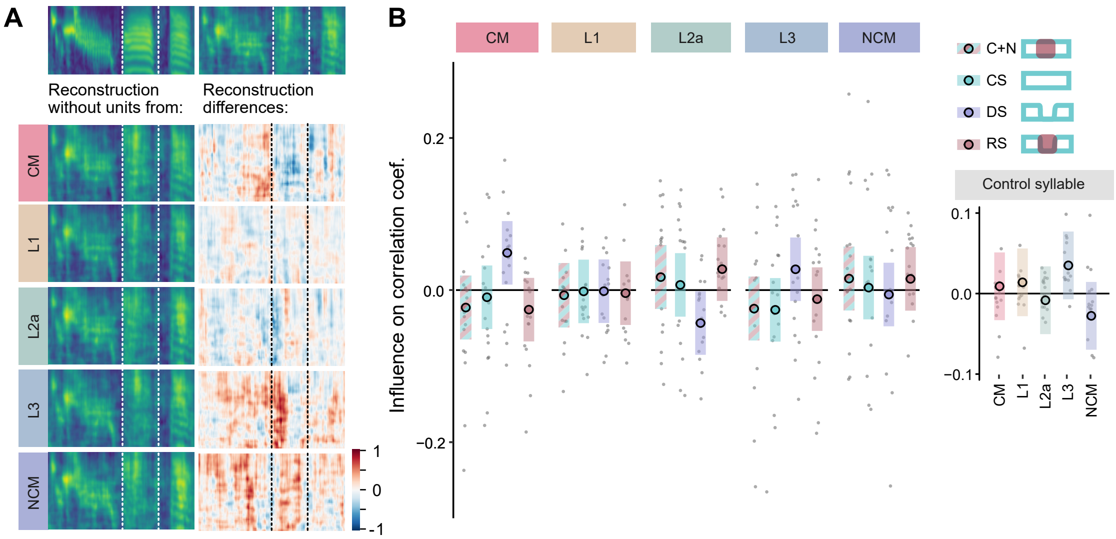


Figure 3.5: Distributed support for restoration across the auditory cortex. A. Reconstruction models were fit with a leave-one-area-out approach to isolate the contribution of cortical-level auditory areas to auditory restoration. Examples show the stimulus reconstructions made without the units from a given area (left) and the differences between the full reconstruction and the reconstruction without one area (right). Red indicates areas where the amplitude of the full reconstruction was greater than the reconstruction without the given area, and blue indicates areas where the amplitude of the full reconstruction was lower. B. Overall influence of each area on the correlation coefficients shown in Figure 3.4b. L2a decreases the similarity of the reconstruction to DS (-0.043, 90% $CI = \{-0.085, -0.001\}, t_{360} = -2.68, p = 0.039$), while CM increases the similarity of the reconstruction to DS (0.049, 90% $CI = \{0.007, 0.091\}, t_{360} = 3.04, p = 0.013$). No other effects were significant. Gray points show the difference between the leave-one-out correlation coefficient (CC) and the CC for the full data set for each stimulus. Colored points and bars show the effect estimates and confidence intervals of a regression model with an interaction between area and stimulus type. Reconstruction of the control syllables did not differ between areas.

Table 3.1: Model estimates of auditory area influences on reconstruction.

Condition	Location	estimate	df	t.ratio	p.value	LCL	UCL
Continuous	CM	-0.0092442	360	-0.5734429	0.5956873	-0.0509890	0.0325006
Cont+Noise	CM	-0.0227901	360	-1.4137397	0.2638375	-0.0645349	0.0189547
Control	CM	0.0085613	360	0.5310851	0.5956873	-0.0331835	0.0503061
Discontinuous	CM	0.0490287	360	3.0413948	0.0126374	0.0072839	0.0907735
Replaced	CM	-0.0255557	360	-1.5852973	0.2638375	-0.0673005	0.0161891
Continuous	L1	-0.0015809	360	-0.0980692	0.9359384	-0.0433257	0.0401639
Cont+Noise	L1	-0.0066314	360	-0.4113682	0.9359384	-0.0483762	0.0351134
Control	L1	0.0134334	360	0.8333122	0.9359384	-0.0283114	0.0551782
Discontinuous	L1	-0.0012966	360	-0.0804319	0.9359384	-0.0430414	0.0404482
Replaced	L1	-0.0039244	360	-0.2434429	0.9359384	-0.0456692	0.0378204
Continuous	L2a	0.0069468	360	0.4309295	0.6667775	-0.0347980	0.0486916
Cont+Noise	L2a	0.0171784	360	1.0656292	0.4788425	-0.0245664	0.0589232
Control	L2a	-0.0086476	360	-0.5364395	0.6667775	-0.0503924	0.0330971
Discontinuous	L2a	-0.0431451	360	-2.6764184	0.0389082	-0.0848899	-0.0014003
Replaced	L2a	0.0276675	360	1.7162993	0.2174188	-0.0140773	0.0694123
Continuous	L3	-0.0258075	360	-1.6009136	0.1673253	-0.0675522	0.0159373
Cont+Noise	L3	-0.0242200	360	-1.5024398	0.1673253	-0.0659648	0.0175248
Control	L3	0.0342946	360	2.1273986	0.1673253	-0.0074502	0.0760394
Discontinuous	L3	0.0274592	360	1.7033767	0.1673253	-0.0142856	0.0692040
Replaced	L3	-0.0117264	360	-0.7274218	0.4674404	-0.0534712	0.0300184
Continuous	NCM	0.0034538	360	0.2142506	0.8304729	-0.0382910	0.0451986
Cont+Noise	NCM	0.0152172	360	0.9439655	0.5955617	-0.0265276	0.0569619
Control	NCM	-0.0278333	360	-1.7265803	0.4255047	-0.0695781	0.0139115
Discontinuous	NCM	-0.0056949	360	-0.3532693	0.8304729	-0.0474397	0.0360499
Replaced	NCM	0.0148571	360	0.9216334	0.5955617	-0.0268876	0.0566019

support for auditory restoration by providing no evidence of discontinuity in the stimulus. Interestingly, CM shows the opposite pattern, increasing the similarity of the reconstruction to DS. Neither effect, however, is large compared to the correlations of the full reconstruction (Figure 3.4b), indicating that although these results point to some select differences in the processing of ambiguous stimuli across the auditory cortex, they generally show a distributed response that supports illusory continuity. For all area-level estimates, see Table 3.1

DISCUSSION

In this study, we found that the responses of neurons in the avian auditory cortex to stimuli designed to induce a perception of illusory continuity support the neural hypothesis put forth by Bregman (1999). Neural response rates during the critical interval to RS were nearly identical to CS response rates, in contrast to DS and NS response rates which were less well correlated (Figure 3.3). Furthermore, stimulus reconstructions from the RS neural responses showed the highest correlation to a spectrogram designed to approximate the illusory percept by combining both the missing syllable and noise (Figure 3.4). The reconstructions also showed no support for a discontinuous perception of the stimulus, which is one of the key requirements for inducing auditory restoration (Bregman, 1999). Taken together, these findings demonstrate neural responses in the avian auditory cortex that provide population-level support for the perception of auditory restoration of zebra finch song.

There are two ways this illusion-encoding neural response could occur: auditory restoration could be a distributed process that is supported by activity throughout the auditory cortex, or auditory restoration could be driven by the responses of a specific area or group of areas. Area-specific stimulus reconstructions provide evidence toward the more distributed view of auditory restoration, as there was no one auditory area that made a large, unique contribution toward this illusion (Figure 3.5). The response of L2a does indicate early evidence against discontinuity in the stimulus, which may contribute to further downstream refinement of the illusory percept, while the responses of CM and L3 as the superficial and deep pathways, respectively, may represent a suppression of the noise response, but a more sensitive method would be needed to fully explore the differential effects of the auditory areas. The present model is a linear decoder, and it is possible that a more complex, non-linear decoder might be provide more insight, especially with regards to secondary areas like NCM that are known to produce sparse, non-linear responses to song (Calabrese and Woolley, 2015).

Evidence of restoration in this study did not differ significantly by stimulus familiarity (Figure 3.4c). This finding adds further support for the interpretation that the behavioral familiarity difference detected in the previous chapter may have been due more to specifics of the training paradigm than a true difference in perception. The birds used for electrophysiology were familiarized with the song stimuli in the same manner as that described in Chapter 2, but they were naive to the behavioral task. This study provides evidence that at least at a neural level, stimulus familiarity does not play a strong role in auditory restoration. If zebra finches do have a strong top-down, knowledge dependent influence on their perception of these stimuli, it either occurs outside of the auditory cortex, perhaps in an association area, or our analysis was not sensitive enough to detect such a difference.

Interestingly, we did detect an effect of social group on auditory restoration (Figure 3.4d). This was quite unexpected, as the two groups were balanced in both age and sex. The most salient difference between the two groups was the rooms in which they were housed; they were housed separately to avoid experimental birds hearing the unfamiliar stimuli before the recording. Group A was housed in a room with several large flight aviaries, although the socialization group itself was housed in a smaller flight cage. Group B was housed in the same type of flight cage in a room primarily comprised of breeding pairs in family cages. It is possible that the acoustic environments of the two rooms affected these responses; strong evidence from our lab has highlighted the importance of acoustic environment on intrinsic neuronal response properties (Chen and Meliza, 2020). It is also possible that these groups differed with regard to some other unidentified factor that affected these responses.

The stimulus reconstruction technique used in this chapter has great potential for studies of neural coding in animal models. One of the challenges of sensory neuroscience in animal models is lack of access to the animal's perception of the stimulus it is experiencing without using technically challenging and time-consuming procedures of behavioral training concurrent with chronic recordings. As shown here and in Mesgarani et al. (2009), the reconstruction technique, when fit with appropriate training data, can provide a powerful estimation of the consensus population neural response represented in the stimulus space. This technique is already established in studies of human perception (Crosse et al., 2016, Leonard et al., 2016, Schäfer et al., 2018, Akbari et al., 2019, Alickovic et al., 2019, Broderick et al., 2019) but has not been widely adopted in animal research. Given the increasing availability of high-density neural probes (e.g. Jun et al., 2017, Yang et al., 2020), a technique that can integrate the neural responses of large datasets in an interpretable way is clearly welcome, and this becomes even more important as neuroscience moves toward more naturalistic stimuli (Sonkusare et al., 2019, Nastase et al., 2020). This mapping of neural responses to a stimulus becomes even more important when investigating illusions where the presented stimulus and the perception of that stimulus are misaligned; Leonard et al. (2016) used this technique with ambiguous stimuli to demonstrate multiple perceptions of the same stimuli and their supporting neural responses.

4

Illusory likelihood in zebra finch neural responses

Introduction

The results of the previous chapter show strong neural evidence of auditory restoration. Using a correlation analysis of stimulus reconstructions, we were able to demonstrate population-level support for auditory restoration, and we identified a spectrogram that combined the continuous syllable with noise (C+N) as a promising approximation of the stimulus the neural responses are coding for.

In this chapter, we extend the previous analysis with a predictive likelihood approach. Using the reconstruction model from the previous chapter, we can further assess the validity of using C+N as an approximation of auditory restoration using log likelihood ratios to predict how likely it is that the C+N stimulus elicited the observed responses rather than RS, DS, or CS alone. The likelihood approach can provide greater sensitivity than the correlation analysis as it takes into account the uncertainty in the calculation of the probability distributions, and it can also provide temporal resolution of likelihood over the course of the critical interval, revealing fine-scale changes in stimulus coding. We use this approach to perform a time-course analysis of the changes in log likelihood ratio between C+N and RS during the critical intervals of the stimuli, and we also use the greater sensitivity of the predictive likelihood analysis to estimate the contributions of auditory areas to the population-level restoration.

METHODS AND MATERIALS

POSTERIOR PREDICTIVE LIKELIHOOD

Using a Bayesian approach to estimating posterior probabilities from regression models as described by Gelman et al. (2020), we quantified the posterior probability that the predicted stimulus $E(\tilde{\mathbf{S}})$ was either the actual stimulus physically presented to the animal $\mathbf{S}_{RS}$ or the illusory percept $\mathbf{S}_{C+N}$.

For a linear regression model, the posterior predictive distribution conditional on the observations used to fit the model is a multivariate t distribution with a mean given by

$$E(\tilde{\mathbf{S}}) = \tilde{\mathbf{R}}\hat{\mathbf{g}},$$

$n - k$ degrees of freedom, and scale matrix $s^2(\mathbf{I} + \tilde{\mathbf{R}}\mathbf{V}_{\mathrm{g}}\tilde{\mathbf{R}}^\top)$, where s^2 is the sample variance of the residuals,

$$s^2 = \frac{1}{n-k}(\mathbf{S} - \mathbf{R}\hat{\beta})^\top(\mathbf{S} - \mathbf{R}\hat{\beta}),$$

and $\mathbf{V}_\mathrm{g}$ is the posterior variance of the parameter estimates,

$$\mathbf{V}_\mathrm{g} = (\mathbf{R}^\top\mathbf{R})^{-1}.$$

Thus, the posterior predictive uncertainty reflects both the unexplained variance in the model (s^2) and the posterior uncertainty in the parameter estimates ($\mathbf{V}_\mathrm{g}$). Because $n - k$ was very large (27,725), we approximated this distribution with a multivariate normal. We then calculated the log likelihood ratio ℓ between $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{C+N}|\mathbf{S})$ and $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_x|\mathbf{S})$; in other words, given the response, how much more likely was it that the stimulus was the combined continuous and noise stimulus (C+N) versus CS, RS, or DS. For Figure 4.1, we calculated the joint distribution of $p(\tilde{\mathbf{S}}|\mathbf{S})$ over the entire critical interval. For Figure 4.2, this quantity was evaluated using the marginal distributions of $p(\tilde{\mathbf{S}}|\mathbf{S})$ for each time point and spectral channel to produce a spectrotemporal likelihood ratio plot. Outside the critical interval, $\mathbf{S}_{RS}$ and $\mathbf{S}_{C+N}$ were by design equal.

To calculate the contribution of each area to the reconstruction, we recalculated the log likelihood ratio between $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{C+N}|\mathbf{S})$ and $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{RS}|\mathbf{S})$, using the area specific reconstructions from the previous chapter. Here, we used the joint distribution and compared this value $\ell^{(i)}$ to the log likelihood ratio ℓ for the full set of neurons. Negative values meant that the given area caused $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{C+N}|\mathbf{S})$ to decrease relative to $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{RS}|\mathbf{S})$, implying that the neurons in that area were encoding the noise that was actually present during the critical interval. Positive values implied that the neurons of the given area were encoding the internal expectations of what was behind the noise.

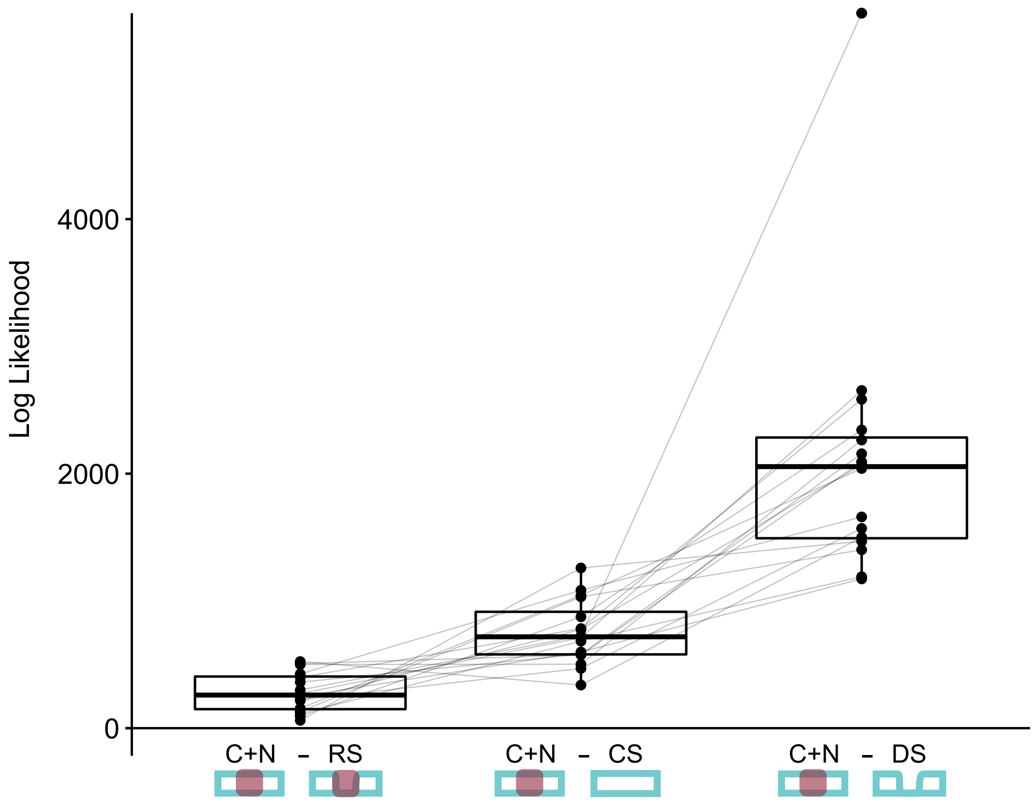


Figure 4.1: Predictive likelihood shows encoding of illusory syllable. The log likelihood ratio of C+N compared to RS, CS, and DS was strictly positive for all stimuli, providing strong evidence that neurons are coding for the illusory syllable approximated by C+N. Boxes show median and interquartile ranges. Lines connect the same stimulus variant.

RESULTS

Because the C+N approximation of the illusory percept showed high similarity with the reconstructions, we used the reconstruction model to calculate the posterior predictive likelihood of the C+N stimulus having elicited the recorded neural responses and compared it to RS, CS, and DS with a log likelihood ratio. The results of this approach showed strong evidence for the missing syllable in the neural responses (Fig. 4.1) and suggested that the log likelihood ratio could provide more sensitivity in measuring the strength of auditory restoration than the spectrogram correlations. C+N showed greater likelihood than RS, CS, or DS for every motif. For subsequent analyses, we focused on the log likelihood ratio of C+N and RS to characterize evidence for or against the missing syllable over noise only.

Time course of auditory restoration

We hypothesized that the illusory percept of the missing syllable is generated by a predictive internal model and that the neural activity associated with this internal model is in conflict with ascending sensory input. The strength of the illusion could therefore depend on how effectively the response to noise is suppressed, and suppressing this activity could take time. To test this idea, we used the decoding model to calculate the amount of evidence for the missing syllable over noise. We quantified this as the log likelihood ratio of the posterior predictive likelihood of C+N and RS, for each time point and spectral channel during the critical interval (Figure 4.2b).

Across the frequency spectrum, we saw differences in the log likelihood ratio (Figure 4.2b). The high frequencies were more likely to show large differences in log likelihood; given the power spectrum of white noise relative to zebra finch song, this result is expected, as the largest difference in the power spectra occur at high frequencies. Specific bands of lower frequency ranges also showed larger differences which may correspond to zebra finch song spectral structure.

Additionally, we saw a strong change in the log likelihood ratio over the course of the critical interval. Initial responses showed evidence for noise, as expected when a new, salient signal appears. By 20 ms, the evidence had shifted toward the missing syllable and continued to strengthen over the course of the critical interval (Figure 4.2b,c), supporting the view of a predictive internal model driving the neural response after the initial onset of noise. In the few cases where there was not strong evidence for the missing syllable (e.g., Figure 4.2c, 5th row), the likelihood still became stronger with time but in the negative direction, emphasizing that the neural coding tends to become more certain over time, either toward or away from an illusory percept.

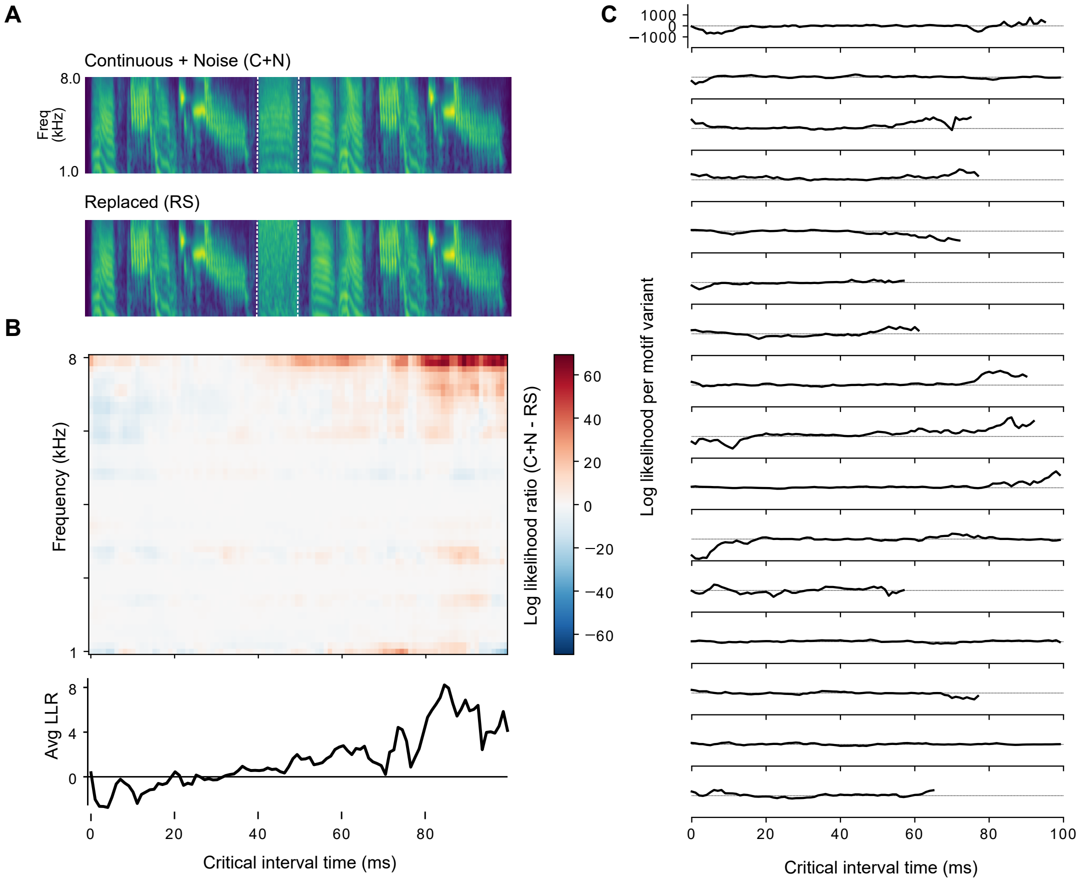


Figure 4.2: Evidence of restoration increases during the critical interval. A. C+N and RS spectrograms showing the difference in the two stimuli during the critical interval (white dotted lines). B. Average log likelihood ratio for all stimuli during the critical intervals calculated for each frequency band and time point. Red shows the log likelihood of C+N $>$ RS, and blue shows RS $>$ C+N. While neurons tend to respond to the onset of the noise, their responses ramp towards C+N within a few tens of milliseconds. Bottom trace shows the log likelihood ratio averaged across frequency bands. C. Time course of log likelihood ratio for the critical interval of each stimulus, summed across frequencies.

Influence of auditory areas and single units

We tested how removing each of the five areas from the decoder impacted the evidence for or against the missing syllable (log likelihood ratio of C+N versus RS) using the same area-specific models from Chapter 3. Because the predictive likelihood analysis showed more sensitivity to detect differences in the full model, it may also be able to detect subtle area specific differences in stimulus coding as well. As in Chapter 3, if the likelihood of C+N decreases relative to RS after removing data from one area, we know that area is making a non-redundant contribution to the illusory response.

Across all the motifs tested, we found that L3 and NCM both non-redundantly increased the likelihood of C+N relative to RS (Figure 4.3). In contrast, L1 biased the reconstruction more toward the noise. CM and L2a supported the illusory percept in that they did not decrease evidence for the missing syllable, but they did not make strong, unique contributions. These results point to both distributed support for the missing syllable as well as specific differences in the processing of ambiguous stimuli between the superficial and deep pathways of avian auditory cortex.

Discussion

This chapter introduced a more sensitive analysis of the reconstruction model and provided some key insights into the neural basis of auditory restoration. In addition to strengthening the findings from the previous chapter that auditory neural responses show evidence of the missing syllable (Figure 4.1), the predictive likelihood approach facilitated a detailed view of restoration over frequency and time.

From a mechanistic point of view, one of the most important findings of this chapter was that the encoding of the ambiguous stimulus segment can change dynamically throughout that interval and that the strength of restoration tends to increase (Figure 4.2). It is expected that there would be an initial response to the onset of the noise, as the illusory percept does not

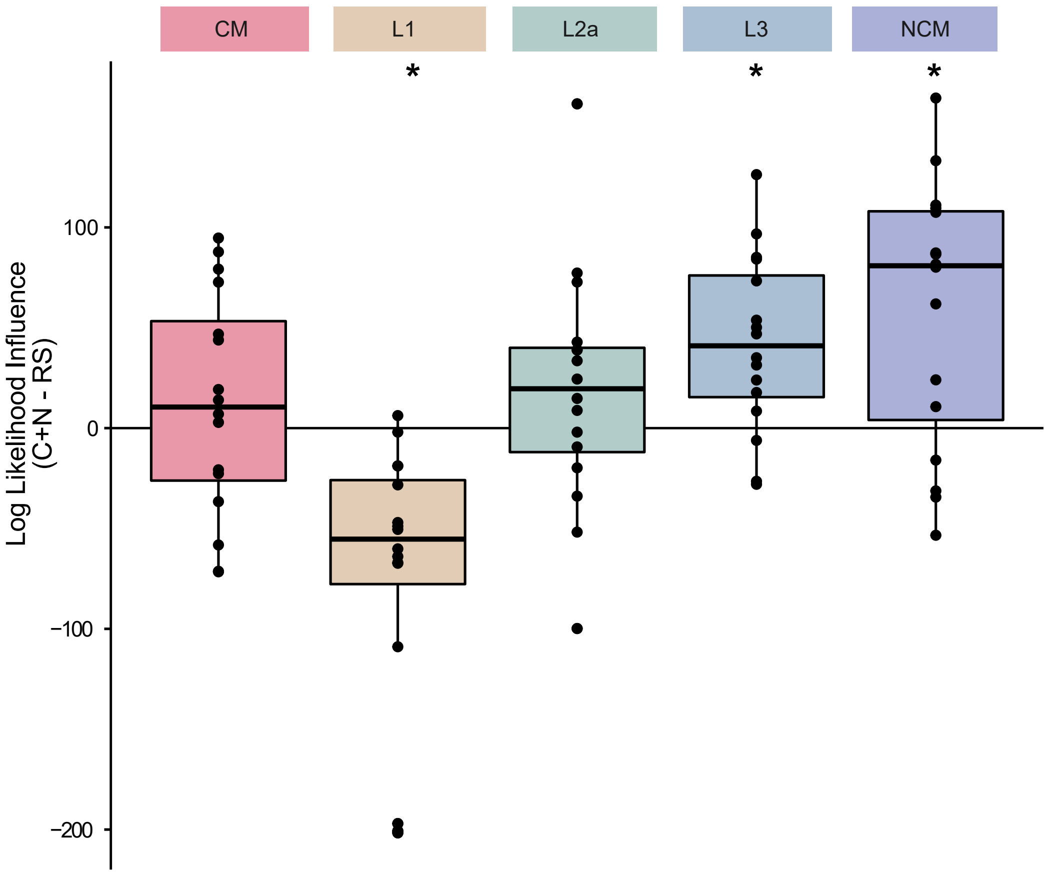


Figure 4.3: The deep auditory pathway assists with restoration. Influence of the log likelihood ratio of C+N – RS using the reconstruction models fit for Figure 3.5 for each of five auditory areas. Boxes show median and interquartile range, and points show individual stimuli. L1 reduced the likelihood relative to the full model ($p < 0.001, t_{60} = -6.55$), and L3 and NCM increased the evidence for restoration ($p = 0.034, t_{60} = 2.38$ and $p = 0.002, t_{60} = 3.59$, respectively).

abolish awareness of the noise but adds the missing syllable to it. The rapid shift from evidence for noise toward evidence for the missing syllable over the course of the critical interval supports the hypothesis of an internal predictive model gaining response weight over the ascending auditory signal. This aligns well with the model put forth by Bregman (1999) which relies on a lack of evidence contradicting continuity of the signal. An interesting extension of this would be recording responses to increasingly long critical intervals. Neither the continuity illusion nor phonemic restoration can be sustained indefinitely – in humans, the limit of illusory continuity in speech is approximately 300 ms (Bashford and Warren, 1987) – and this method could be used to characterize the full time-course of the illusion from onset to offset.

Log likelihood ratios of area-specific stimulus reconstructions provided evidence toward a more distributed view of auditory restoration, with most areas supporting perception of the missing syllable and no one area making a large, unique contribution toward this illusion (Figure 4.3). However, this analysis does highlight the importance of the deep auditory pathway of L2a to L3 to NCM, as this pathway showed increasing evidence for the missing syllable. Previous research has identified NCM as an important secondary auditory area with high selectivity for conspecific song (Meliza and Margoliash, 2012, Calabrese and Woolley, 2015) and which contains a subset of noise invariant neurons that respond selectively to conspecific song even in noisy listening conditions (Schneider and Woolley, 2013). These song-selective, noise-invariant neurons could play a prominent role in the perception of auditory restoration.

In contrast, L1 tended to increase evidence for noise, which could point toward stream segregation of song and noise within the avian auditory cortex. It has been hypothesized that stream segregation of the ambiguous speech segment during phonemic restoration is an important aspect of the illusion (Shinn-Cunningham and Wang, 2008), and the present finding of increasing evidence for the missing syllable along the deep auditory pathway and decreased evidence in a portion of the superficial pathway would support this view.

5

Conclusions

The studies described in the previous chapters establish zebra finches as an excellent model for studying the neural basis of phonemic restoration and, more broadly, mechanisms employed by the brain to facilitate auditory processing of behaviorally important signals in complex acoustic environments. Songbird models of speech have the potential to provide key new insight to longstanding questions in the field of speech perception, and such studies have the potential to uncover principles of neural coding in the avian auditory cortex as well.

Specifically, the present work demonstrated for the first time behavioral evidence for auditory restoration in zebra finches. It extended that finding with single-unit neural recordings from zebra finch auditory cortex as birds were presented with the illusion-inducing stimuli. We found neural responses consistent with auditory restoration of song syllables at both the single-unit and population level. These results bolster the neural hypothesis of phonemic restoration put forth by Bregman (1999) and previous work by Petkov et al. (2007) showing similar results in primates experiencing the tonal continuity illusion. Using a linear decoder with a predictive likelihood approach, we were able to identify the deep auditory pathway from L2a to NCM as a potential circuit involved in the emergence of the illusory percept of the missing syllable and show the dynamic increase of neural evidence for that missing syllable over the course of the occluded interval.

More broadly, this work adds to a growing understanding of auditory processing of birdsong co-occurring with noise – a type of stimuli that is undoubtedly a closer approximation to natural auditory processing than the high signal-to-noise stimuli more traditionally used for sensory processing research. Studies of avian auditory cortex by Narayan et al. (2007), Moore et al. (2013), and Schneider and Woolley (2013) have identified populations of neurons in multiple areas that show invariance in their responses to background noise, even when that background consists of songbird colony noise, which, according to the principles of the cocktail party problem, should cause the greatest interference due to the overlap in spectral structure between the song signal and the colony background noise (Bee and Micheyl, 2008). At the same time, we are gaining an increased understanding of how the complexity of acoustic environments can cause profound changes in the avian auditory system (Chen and Meliza, 2020). These lines of research present a complex picture of how the auditory system structures and tunes itself during development to facilitate auditory perception of vocalizations in noisy social settings.

Additionally, the study of auditory illusions or other high-level auditory processes like sound embedded in noise may be a promising avenue toward a more complete understanding of functional differences in the regions of avian auditory cortex. Auditory tasks that present relatively little challenge for the auditory system, such as discrimination of high signal-to-noise song stimuli, may not reveal many meaningful differences in high-level auditory areas, whereas tasks that deliberately lead to breakdowns in auditory processing, such as illusions, or tasks that present extreme challenges like songbird cocktail parties may point to functional differences in areas and pathways that will lead to a more complete understanding of auditory processing in the songbird brain. It also highlights the importance of assessing the responses of as many of the high-level auditory areas as possible to understand whether the response features of one area are unique or part of a broader auditory processing strategy.

The work presented here also highlights the potential of human speech perception research to inform and inspire research on songbird auditory perception. Both the research question – a songbird correlate of phonemic restoration (Warren, 1970) – and the core analysis of stimulus reconstruction (Crosse et al., 2016, Leonard et al., 2016) were drawn from work by human researchers. Adapting paradigms created for understanding human speech perception to songbird vocalizations has the potential to advance our understanding both of songbird auditory circuits and of fundamental principles of auditory perception across species. Bee and Micheyl (2008) and Petkov and Sutter (2011) both argue persuasively for this approach to research on animal communication, and the work presented here shows the benefits and challenges to this approach. Adapting human perception paradigms, especially behaviorally, is often not straightforward, but it can lead to a deeper understanding of conserved mechanisms of auditory processing and draw direct parallels to human perception.

References

Akbari H, Khalighinejad B, Herrero JL, Mehta AD, Mesgarani N. Towards reconstructing intelligible speech from the human auditory cortex. Sci Rep 2019 Jan;9(1):874. doi: 10.1038/s41598-018-37359-z.

Alickovic E, Lunner T, Gustafsson F, Ljung L. A Tutorial on Auditory Attention Identification Methods. Front Neurosci 2019;13:153. doi: 10.3389/fnins.2019.00153.

Amin N, Doupe A, Theunissen FE. Development of selectivity for natural sounds in the songbird auditory forebrain. J Neurophysiol 2007 May;97(5):3517–3531. doi: 10.1152/jn.01066.2006.

Banai K, Ahissar M. On the importance of anchoring and the consequences of its impairment in dyslexia. Dyslexia 2010 Aug;16(3):240–257. doi: 10.1002/dys.407.

Bashford JA, Warren RM. Multiple phonemic restorations follow the rules for auditory induction. Percept Psychophys 1987 Mar;42(2):114–121. doi: 10.3758/BF03210499.

Bee MA, Micheyl C. The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? J Comp Psych 2008 Aug;122(3):235–251. doi: 10.1037/0735-7036.122.3.235.

Berwick RC, Okanoya K, Beckers GJL, Bolhuis JJ. Songs to syntax: the linguistics of birdsong. Trends Cogn Sci (Regul Ed) 2011 Mar;15(3):113–121. doi: 10.1016/j.tics.2011.01.002.

Bidelman GM, Patro C. Auditory perceptual restoration and illusory continuity correlates in the human brainstem. Brain Res 2016 Sep;1646:84–90. doi: 10.1016/j.brainres.2016.05.050.

Bjoring MC, Meliza CD. A low-threshold potassium current enhances sparseness and reliability in a model of avian auditory cortex. PLoS Comput Biol 2019 Jan;15(1):e1006723. doi: 10.1371/journal.pcbi.1006723.

Braaten RF, Leary JC. Temporal Induction of Missing Birdsong Segments in European Starlings. Psychological Science 1999 May;10(2):162–166. doi: 10.1111/1467-9280.00125.

Bregman AS. Auditory scene analysis. 2nd edition ed. The perceptual organization of sound, MIT Press; 1999.

Broderick MP, Anderson AJ, Lalor EC. Semantic Context Enhances the Early Auditory Encoding of Natural Speech. J Neurosci 2019 Sep;39(38):7564–7575. doi: 10.1523/JNEUROSCI.0584-19.2019.

Bronkhorst AW. The cocktail-party problem revisited: early processing and selection of multi-talker speech. Atten Percept Psychophys 2015 Jul;77(5):1465–1487. doi: 10.3758/s13414-015-0882-9.

Calabrese A, Woolley SMN. Coding principles of the canonical cortical microcircuit in the avian brain. PNAS 2015 Mar;112(11):3517–3522. doi: 10.1073/pnas.1408545112.

Chen AN, Meliza CD. Experience- and sex-dependent intrinsic plasticity in the zebra finch auditory cortex during song memorization. J Neurosci 2020 Jan;p. 2137–19. doi: 10.1523/JNEUROSCI.2137-19.2019.

Chen AN, Meliza D. Phasic and Tonic Cell Types in the Zebra Finch Auditory Caudal Mesopallium. Journal of Neurophysiology 2018 Mar;119(3):1127–1139. doi: 10.1152/jn.00694.2017.

Chen Y, Clark O, Woolley SC. Courtship song preferences in female zebra finches are shaped by developmental auditory experience. Proc Biol Sci 2017 May;284(1855):20170054. doi: 10.1098/rspb.2017.0054.

Cherry EC. Some Experiments on the Recognition of Speech, with One and with Two Ears. The Journal of the Acoustical Society of America 1953 Sep;25(5):975–979. doi: 10.1121/1.1907229.

Chung JE, Magland JF, Barnett AH, Tolosa VM, Tooker AC, Lee KY, et al. A Fully Automated Approach to Spike Sorting. Neuron 2017 Sep;95(6):1381–1394.e6. doi: 10.1016/j.neuron.2017.08.030.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front Hum Neurosci 2016;10:604. doi: 10.3389/fnhum.2016.00604.

Dimitrijevic A, Smith ML, Kadis DS, Moore DR. Neural indices of listening effort in noisy environments. Sci Rep 2019 Aug;9(1):11278–10. doi: 10.1038/s41598-019-47643-1.

Ding N, Simon JZ. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. Journal of Neurophysiology 2012 Jan;107(1):78–89. doi: 10.1152/jn.00297.2011.

Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci 1999;22:567–631. doi: 10.1146/annurev.neuro.22.1.567.

Dubno JR, Dirks DD, Morgan DE. Effects of age and mild hearing loss on speech recognition in noise. J Acoust Soc Am 1984 Jul;76(1):87–96. doi: 10.1121/1.391011.

Dugas-Ford J, Rowell JJ, Ragsdale CW. Cell-type homologies and the origins of the neocortex. Proc Natl Acad Sci USA 2012 Oct;109(42):16974–16979. doi: 10.1073/pnas.1204773109.

Ellinger RL, Jakien KM, Gallun FJ. The role of interaural differences on speech intelligibility in complex multi-talker environments. The Journal of the Acoustical Society of America 2017 Feb;141(2):EL170–EL176. doi: 10.1121/1.4976113.

Ericson MA, Brungart DS, Simpson BD. Factors That Influence Intelligibility in Multitalker Speech Displays. The International Journal of Aviation Psychology 2009 Nov;14(3):313–334. doi: 10.1207/s15327108ijap1403_6.

Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. The Journal of the Acoustical Society of America 1990 Oct;88(4):1725–1736. doi: 10.1121/1.400247.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Introduction to regression models. In: Bayesian Data Analysis; 2020.p. 353–378.

Goldman-Eisler F. The Significance of Changes in the Rate of Articulation:. Language and Speech 1961 Jul;4(3):171–174. doi: 10.1177/002383096100400305.

Groppe DM, Choi M, Huang T, Schilz J, Topkins B, Urbach TP, et al. The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. Brain Res 2010 Nov;1361:54–66. doi: 10.1016/j.brainres.2010.09.003.

Grossberg S, Kazerounian S. Laminar cortical dynamics of conscious speech perception: neural model of phonemic restoration using subsequent context in noise. The Journal of the Acoustical Society of America 2011 Jul;130(1):440–460. doi: 10.1121/1.3589258.

Hauber ME, Campbell DLM, Woolley SMN. The functional role and female perception of male song in Zebra Finches. Emu - Austral Ornithology 2016 Dec;110(3):209–218. doi: 10.1071/MU10003.

Husain FT, Lozito TP, Ulloa A, Horwitz B. Investigating the neural basis of the auditory continuity illusion. J Cogn Neurosci 2005 Aug;17(8):1275–1292. doi: 10.1162/0898929055002472.

Ishida M, Arai T. Missing phonemes are perceptually restored but differently by native and non-native listeners. Springerplus 2016;5(1):713–10. doi: 10.1186/s40064-016-2479-8.

Jarvis ED, Güntürkün O, Bruce L, Csillag A, Karten H, Kuenzel W, et al. Avian brains and a new understanding of vertebrate brain evolution. Nat Rev Neurosci 2005 Feb;6(2):151–159. doi: 10.1038/nrn1606.

Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, et al. Fully integrated silicon probes for high-density recording of neural activity. Nature 2017 Nov;551(7679):232–236. doi: 10.1038/nature24636.

Karten HJ. The ascending auditory pathway in the pigeon (Columba livia). II. Telencephalic projections of the nucleus ovoidalis thalami. Brain Res 1968 Oct;11(1):134–153. doi: 10.1016/0006-8993(68)90078-4.

Kobayasi KI, Usami A, Riquimaroux H. Behavioral evidence for auditory induction in a species of rodent: Mongolian gerbil (Meriones unguiculatus). The Journal of the Acoustical Society of America 2012 Dec;132(6):4063–4068. doi: 10.1121/1.4763546.

Kohn K. Die Rolle der Sprachlauterwartung in der kategorialen Wahrnehmung. Phonetica 1981;38(5-6):309–319. doi: 10.1159/000260035.

Komatsu H. The neural mechanisms of perceptual filling-in. Nat Rev Neurosci 2006 Mar;7(3):220–231. doi: 10.1038/nrn1869.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 1992 Jan;255(5044):606–608. doi: 10.1126/science.1736364.

Lauay C, Komorowski RW, Beaudin AE, DeVoogd TJ. Adult female and male zebra finches show distinct patterns of spine deficits in an auditory area and in the song system when reared without exposure to normal adult song. J Comp Neurol 2005 Jun;487(2):119–126. doi: 10.1002/cne.20591.

Leonard MK, Baud MO, Sjerps MJ, Chang EF. Perceptual restoration of masked speech in human cortex. Nat Commun 2016 Dec;7(1):13619–9. doi: 10.1038/ncomms13619.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychological Review 1967 Nov;74(6):431–461. doi: 10.1037/h0020279.

Liberman AM, Delattre PC, Cooper FS. Some Cues for the Distinction Between Voiced and Voiceless Stops in Initial Position:. Language and Speech 1958 Aug;1(3):153–167. doi: 10.1177/002383095800100301.

Lunner T. Cognitive function in relation to hearing aid use. Int J Audiol 2003 Jul;42 Suppl 1(sup1):S49–58. doi: 10.3109/14992020309074624.

McGowan KB. Social Expectation Improves Speech Perception in Noise. Language and Speech 2015 Dec;58(Pt 4):502–521. doi: 10.1177/0023830914565191.

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976 Dec;264(5588):746–748. doi: 10.1038/264746a0.

Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C. A precursor of language acquisition in young infants. Cognition 1988 Jul;29(2):143–178. doi: 10.1016/0010-0277(88)90035-2.

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

Mesgarani N, David SV, Fritz JB, Shamma SA. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. Journal of Neurophysiology 2009 Dec;102(6):3329–3339. doi: 10.1152/jn.91128.2008.

Miller CT, Dibble E, Hauser MD. Amodal completion of acoustic signals by a nonhuman primate. Nat Neurosci 2001 Aug;4(8):783–784. doi: 10.1038/90481.

Mody M, Studdert-Kennedy M, Brady S. Speech perception deficits in poor readers: auditory processing or phonological coding? J Exp Child Psychol 1997 Feb;64(2):199–231. doi: 10.1006/jecp.1996.2343.

Mooney R. Neural mechanisms for learned birdsong. Learn Mem 2009 Nov;16(11):655–669. doi: 10.1101/lm.1065209.

Mooney R, Prather J, Roberts TF. Neurophysiology of birdsong learning. In: Eichenbaum H, editor. Learning and memory: a comprehensive reference; 2008.p. 441–474.

Moore RC, Lee T, Theunissen FE. Noise-invariant Neurons in the Avian Auditory Cortex: Hearing the Song in Noise. PLoS Comput Biol 2013 Mar;9(3):e1002942. doi: 10.1371/journal.pcbi.1002942.

Narayan R, Best V, Ozmeral E, McClaine E, Dent M, Shinn-Cunningham B, et al. Cortical interference effects in the cocktail party problem. Nat Neurosci 2007 Dec;10(12):1601–1607. doi: 10.1038/nn2009.

Nastase SA, Goldstein A, Hasson U. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. Neuroimage 2020 Nov;222:117254. doi: 10.1016/j.neuroimage.2020.117254.

Newport EL, Bavelier D, Neville HJ. Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In: Dupoux E, editor. Language, Brain, and Cognitive Development Citeseer; 2001.p. 482–502.

Niedzielski N. The Effect of Social Information on the Perception of Sociolinguistic Variables:. Journal of Language and Social Psychology 1999 Jul;18(1):62–85. doi: 10.1177/0261927X99018001005.

Ölveczky BP, Gardner TJ. A bird's eye view of neural circuit formation. Current Opinion in Neurobiology 2011 Feb;21(1):124–131. doi: 10.1016/j.conb.2010.08.001.

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, et al. Reconstructing speech from human auditory cortex. PLoS Biol 2012 Jan;10(1):e1001251. doi: 10.1371/journal.pbio.1001251.

Petersen EB, Wöstmann M, Obleser J, Lunner T. Neural tracking of attended versus ignored speech is differentially affected by hearing loss. Journal of Neurophysiology 2017 Jan;117(1):18–27. doi: 10.1152/jn.00527.2016.

Peterson GE, Barney HL. Control Methods Used in a Study of the Vowels. The Journal of the Acoustical Society of America 1951 Jun;23(1):148–148. doi: 10.1121/1.1917300.

Petkov CI, O'Connor KN, Sutter ML. Illusory sound perception in macaque monkeys. J Neurosci 2003 Oct;23(27):9155–9161. doi: 10.1523/JNEUROSCI.23-27-09155.2003.

Petkov CI, O'Connor KN, Sutter ML. Encoding of illusory continuity in primary auditory cortex. Neuron 2007 Apr;54(1):153–165. doi: 10.1016/j.neuron.2007.02.031.

Petkov CI, Sutter ML. Evolutionary conservation and neuronal mechanisms of auditory perceptual restoration. Hear Res 2011 Jan;271(1-2):54–65. doi: 10.1016/j.heares.2010.05.011.

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ. Prediction and Decoding of Retinal Ganglion Cell Responses with a Probabilistic Spiking Model. J Neurosci 2005 Nov;25(47):11003–11013. doi: 10.1523/JNEUROSCI.3305-05.2005.

Reiner A, Yamamoto K, Karten HJ. Organization and evolution of the avian forebrain. Anat Rec A Discov Mol Cell Evol Biol 2005 Nov;287(1):1080–1102. doi: 10.1002/ar.a.20253.

Riecke L, Micheyl C, Oxenham AJ. Global not local masker features govern the auditory continuity illusion. J Neurosci 2012 Mar;32(13):4660–4664. doi: 10.1523/JNEUROSCI.6261-11.2012.

Samuel AG, Frost R. Lexical support for phonetic perception during nonnative spoken word recognition. Psychon Bull Rev 2015 Dec;22(6):1746–1752. doi: 10.3758/s13423-015-0847-y.

Schäfer PJ, Corona-Strauss FI, Hannemann R, Hillyard SA, Strauss DJ. Testing the Limits of the Stimulus Reconstruction Approach: Auditory Attention Decoding in a Four-Speaker Free Field Environment:. Trends in Hearing 2018 Dec;22(35):233121651881660. doi: 10.1177/2331216518816600.

Schneider DM, Woolley SMN. Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes. Neuron 2013 Jul;79(1):141–152. doi: 10.1016/j.neuron.2013.04.038.

Seeba F, Klump GM. Stimulus familiarity affects perceptual restoration in the European starling (Sturnus vulgaris). PLoS ONE 2009 Jun;4(6):e5974. doi: 10.1371/journal.pone.0005974.

Sereno J, McCall J, Jongman A, Dijkstra T, van Heuven W. On the role of phonetic inventory in the perception of foreign-accented speech. The Journal of the Acoustical Society of America 2002 May;111(5):2363. doi: 10.1121/1.4777969.

Sharpee TO. Computational identification of receptive fields. Annu Rev Neurosci 2013 Jul;36(1):103–120. doi: 10.1146/annurev-neuro-062012-170253.

Shinn-Cunningham BG, Wang D. Influences of auditory object formation on phonemic restorationa). The Journal of the Acoustical Society of America 2008 Jan;123(1):295. doi: 10.1121/1.2804701.

Sivonen P, Maess B, Lattner S, Friederici AD. Phonemic restoration in a sentence context: evidence from early and late ERP effects. Brain Res 2006 Nov;1121(1):177–189. doi: 10.1016/j.brainres.2006.08.123.

Slaney M. Auditory Toolbox Version 2; 1998.

Sonkusare S, Breakspear M, Guo C. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. Trends Cogn Sci (Regul Ed) 2019 Aug;23(8):699–714. doi: 10.1016/j.tics.2019.05.004.

Sturdy CB, Phillmore LS, Sartor JJ, Weisman RG. Reduced social contact causes auditory perceptual deficits in zebra finches, Taeniopygia guttata. Anim Behav 2001 Dec;62(6):1207–1218. doi: 10.1006/anbe.2001.1864.

Sugita Y. Neuronal correlates of auditory induction in the cat cortex. Neuroreport 1997 Mar;8(5):1155–1159. doi: 10.1097/00001756-199703240-00019.

Tremere LA, Terleph TA, Jeong JK, Pinaud R. Bilateral multielectrode neurophysiological recordings coupled to local pharmacology in awake songbirds. Nat Protoc 2010 Feb;5(2):191–200. doi: 10.1038/nprot.2009.224.

Tumer EC, Brainard MS. Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature 2007 Dec;450(7173):1240–1244. doi: 10.1038/nature06390.

Vanthornhout J, Decruy L, Francart T. Effect of Task and Attention on Neural Tracking of Speech. Front Neurosci 2019;13:977. doi: 10.3389/fnins.2019.00977.

Verschuure J, Brocaar MP. Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. Percept Psychophys 1983 Mar;33(3):232–240. doi: 10.3758/bf03202859.

Wang Y, Brzozowska-Prechtl A, Karten HJ. Laminar and columnar auditory cortex in avian brain. Proc Natl Acad Sci USA 2010 Jul;107(28):12676–12681. doi: 10.1073/pnas.1006645107.

Warren RM. Perceptual restoration of missing speech sounds. Science 1970 Jan;167(3917):392–393. doi: 10.1126/science.167.3917.392.

Woolley SMN. Early experience shapes vocal neural coding and perception in songbirds. Dev Psychobiol 2012 Sep;54(6):612–631. doi: 10.1002/dev.21014.

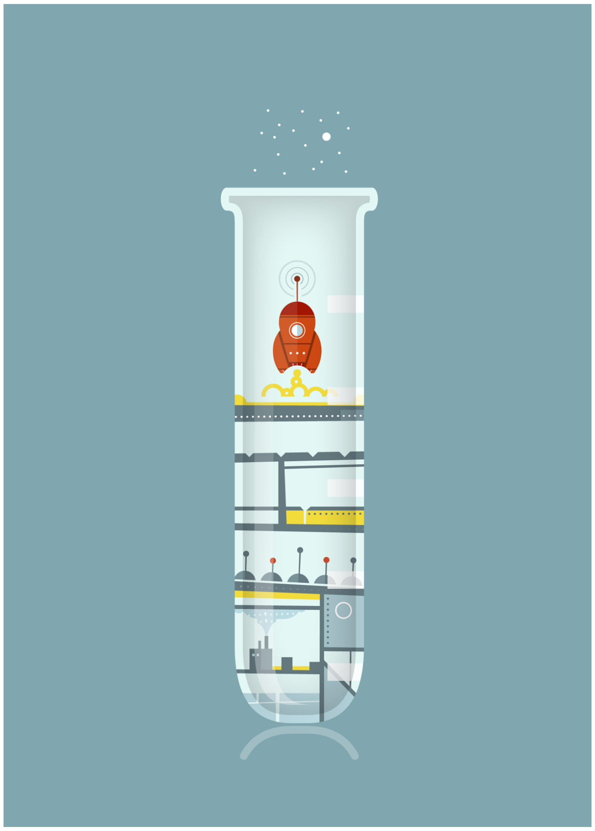
Woolley SMN, Portfors CV. Conserved mechanisms of vocalization coding in mammalian and songbird auditory midbrain. Hear Res 2013 Nov;305:45–56. doi: 10.1016/j.heares.2013.05.005.

Yanagihara S, Yazaki-Sugiyama Y. Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. Nat Commun 2016 Jun;7(1):1–11. doi: 10.1038/ncomms11946.

Yang L, Lee K, Villagracia J, Masmanidis SC. Open source silicon microprobes for high throughput neural recording. J Neural Eng 2020 Jan;17(1):016036. doi: 10.1088/1741-2552/ab581a.

Yu K, Wood WE, Theunissen FE. High-capacity auditory memory for vocal communication in a social songbird. Science Advances 2020 Nov;6(46):eabe0440. doi: 10.1126/sciadv.abe0440.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, et al. Mechanisms underlying selective neuronal tracking of attended speech at a ”cocktail party”. Neuron 2013 Mar;77(5):980–991. doi: 10.1016/j.neuron.2012.12.037.



THIS THESIS WAS TYPESET using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, "Science Experiment 02", was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.