©2019

MINGWEN DONG

ALL RIGHTS RESERVED

STATISTICAL LEARNING OF TRANSITION PATTERNS BETWEEN VARIABLE STIMULI

By

MINGWEN DONG

A dissertation submitted to the

School of Graduate Studies

Rutgers, the State University of New Jersey

In partial fulfillment of the requirement

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

David S. Vicario, PhD

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May 2019

ABSTRACT OF THE DISSERTATION

Statistical learning of transition patterns between variable stimuli

by MINGWEN DONG

Dissertation Director:

David S. Vicario, Ph.D.

Learning transition patterns between variable sounds is essential for vocal communications. For example, spoken speech usually consists of a series of words in a specific order. Having a variant-independent representation for a word and knowing the transition patterns between words are critical for speech perception. To investigate these questions at the neural level, we recorded extracellular neural activity from multiple sites bilaterally in the zebra finch auditory forebrain while presenting auditory stimuli in two separate experiments. In the first experiment, infrequent repetitions of a song syllable were presented after either an alternating or shuffled sequence of syllables. At all tested inter-stimulus intervals (1s, 3s, or jittered from 0.8 to 1.2s), neurons in the secondary auditory area (caudomedial nidopallium, NCM) were sensitive to the violation of transition patterns. In contrast, neurons were less sensitive to the violation of transition patterns in the primary auditory area (Field L2). These results suggest that neurons in NCM can learn transition patterns between sounds after passive

exposure independent of inter-syllable intervals (at least for all tested ISIs). In the second experiment, naturally-produced variants of zebra finch songs were presented in either blocked or shuffled order. The response temporal profiles for different variants of the same zebra finch song were more similar in NCM than in L2. Furthermore, in NCM but not L2, the response temporal profiles became more similar to each other after repeated passive exposure. These results suggest that variant-independent representation emerges hierarchically in the auditory system and that passive exposure may further facilitate that representation. Together, these two experiments provide insights into how the zebra finch auditory system can form variant-independent representations of complex sounds and learn the transition patterns between those sounds. Because similar neural mechanisms may serve the statistical learning and perceptual invariance capacities of the human auditory system, this approach may help us understand the neural basis of speech perception and ultimately contribute to treatments for certain auditory processing disorders.

ACKNOWLEDGEMENT

First, I would like to thank my supervisor Dr. David S. Vicario for his inspiring guidance and endless support throughout my research. I also thank my committee members Dr. John P. McGann, Dr. Kasia M. Bieszczad, and Dr. Timothy M. Otchy for their contributions. I am thankful to the members of the Vicario Lab, especially Dr. Mimi L. Phan, Efe Soyman, Brittany A. Klawiter, and Basilio Furest for their great support and friendship. Special thanks go to Mahinaz Mohsen and David Natanov, who have helped me with histological and data analyses for this thesis. I would also like to thank my family and my friends for their unconditional love and support. Finally, I am truly grateful to Qing Chang, who has been taking care of me like a big sister over these years at Rutgers.

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION ii

ACKNOWLEDGEMENT iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

GENERAL INTRODUCTION 1

EXPERIMENT 0: NEURAL CORRELATE OF TRANSITION VIOLATION IN THE SONGBIRD AUDITORY FOREBRAIN 5

EXPERIMENT 1: STATISTICAL LEARNING OF TRANSITION PATTERNS OVER MULTIPLE TEMPORAL SCALES 7

Abstract 7

Introduction 8

Methods 11

Results 19

Discussion 24

EXPERIMENT 2: VARIANT-INDEPENDENT REPRESENTATION OF ZEBRA FINCH SONGS IN THE AUDITORY FOREBRAIN 28

Abstract 28

Introduction 29

Methods 33

Results 38

Discussion 43

GENERAL DISCUSSION 47

REFERENCES 53

FIGURES 60

LIST OF FIGURES

Figure 1 Thresholded multi-unit, histology, and syllable stimuli. 60

Figure 2 Alternating oddball paradigm and the stimulus sequences used in the control, alternating, and oddball conditions. 61

Figure 3 Schematic illustration of neural oscillation under conditions where two stimuli are alternating. 62

Figure 4 Surprise index in the oddball condition based on MUA 63

Figure 5 SI in the control and alternating conditions in NCM based on MUA 64

Figure 6 SI in the control and alternating conditions in L2 based on MUA. ... 65

Figure 7 Oscillatory magnitude, peri-stimulus time histogram (PSTH), and raw waveforms from three examplar oscillatory sites. 66

Figure 8 SI in the oddball condition based on SUA. 67

Figure 9 SI in the control and alternating conditions in NCM based on SUA. 68

Figure 10 SI in the control and alternating conditions in L2 based on SUA ... 69

Figure 11 Thresholded multi-unit, histology, and variant stimuli 70

Figure 12 Blocked, shuffled, and contrast condition in the variant experiment. 71

Figure 13 Characterization of response temporal profiles and neural decoding 72

Figure 14 Neural dissimilarity in the blocked and shuffled condition based on MUA. 73

Figure 15 Decoding accuracy in the blocked and shuffled condition based on MUA 74

Figure 16 Neural dissimilarity in the blocked and shuffled condition based on SUA. 75

Figure 17 Decoding accuracy in the blocked and shuffled condition based on SUA. 76

Figure 18 Decoding accuracy based on the population responses from MUA.77

Figure 19 Neural dissimilarity for variants within a song in the contrast and shuffled condition based on MUA. 78

Figure 20 Decoding accuracy for variants within a song in the contrast and shuffled condition based on MUA. 79

Figure 21 Neural dissimilarity for variatns within a song in the contrast and shuffled condition based on SUA. 80

Figure 22 Decoding accuracy for variants within a song in the contrast and shuffled condition based on SUA. 81

Figure 23 The average adaptation rate in L2 and NCM from trial 6 to trial 25 based on MUA 82

Figure 24 The average adaptation rate in L2 and NCM from trial 6 to trial 25 based on SUA..... 83

Figure 25 Decoding accuracy for different songs based on MUA in the contrast condition 84

Figure 26 Decoding accuracy for different songs based on SUA in the contrast condition. 85

Figure 27 Change of neural dissimilarity across trials within each variant based on MUA. 86

Figure 28 Change of neural dissimilarity across trials within each variant based on SUA 87

GENERAL INTRODUCTION

Natural vocalizations often consist of several basic units like syllables following certain transition patterns. For example, in its simplest form, spoken speech is a series of syllables in a specific order, that singly or in combination, represent words and form grammatical sentences. Learning transition patterns can help predict future sounds, detect deviant sounds that may indicate danger or novelty, and facilitate auditory processing (Saffran, Johnson, Aslin, & Newport, 1999) (Cornella, Leung, Grimm, & Escera, 2012) (Mittag, Takegata, & Winkler, 2016). However, learning transition patterns is not an easy task. First, in the sound stream, there may not be any explicit cues or rewards associated with the transition probabilities. Therefore, animals may have to learn transition probabilities without external reinforcement, i.e. through statistical learning (Turk-Browne, 2012) (Cate & Okanoya, 2012). Second, natural sounds, especially those from vocal communicators, are often produced with variations in both temporal and acoustic dimensions (Mooney, 2009) (Glaze & Troyer, 2006). Even from the same animal, the same sound is rarely produced identically without variations (e.g., speech, or songs of a zebra finch). Therefore, learning transition patterns requires invariant representations of sounds that have small variations. Otherwise, the same sounds and transitions will happen so infrequently that they would be impossible to learn. On the other hand, learned transition patterns could help a listener recognize variable sounds by providing extra context information. Therefore, invariant representation of variable sounds and statistical

learning of transition patterns may not be isolated problems but closely related to each other. However, it is still not fully understood how the auditory system forms invariant representation of sounds and learns transition patterns between sounds.

Statistical learning of transition patterns has been studied in humans and animals. Most of these studies showed that either the human subjects, animals, or auditory neurons were sensitive to transition patterns between sounds (Aslin, 2017) (Saffran, Johnson, Aslin, & Newport, 1999) (Daikoku, 2018) (Turk-Browne, 2012) (Yaron, Hershenhoren, & Nelken, 2012) (Bennett, Murawski, & Bode, 2015) (Lu & Vicario, 2014) (Ono, Okanoya, & Seki, 2016). Most of the studies used short and fixed inter-stimulus intervals (ISI, < 1s) (Ulanovsky, Las, & Nelken, 2003). However, in natural vocalizations, the intervals between sounds (e.g., syllables/words in speech and zebra finch songs) are not always fixed and could be much longer than those used in the laboratory studies (Grabe & Low, 2002) (Glaze & Troyer, 2006).

Invariant representation of sounds has also been studied in humans and animals. Most of these studies used stimuli like artificially distorted short vocalizations or complex vocalizations superimposed with different levels of noises (Ohms, Gill, Van Heijningen, Beckers, & Cate, 2009) (Carruthers et al., 2015) (Sadagopan & Wang, 2008) (Blackwell et al., 2016) (Billimoria, Kraus, Narayan, Maddox, & Sen, 2008) (Blackwell et al., 2016) (Cate & Okanoya, 2012). However, the variability in vocalizations are not limited in these cases and the

variability in the inter-syllable intervals has not been much studied. Furthermore, none of the studies have looked at how passive exposure affects the neural representation of variants of a sound.

The current study tries to bridge these gaps mentioned above and investigates two questions: how the auditory system learns transition probabilities between sounds over multiple temporal scales and how the auditory system forms an invariant representation of natural variants of a sound (perceptual invariance). Because zebra finches and humans show similar behaviors in vocal learning (Bolhuis & Gahr, 2006) (Brainard & Doupe, 2013), and we know a good deal about basic auditory processing in the zebra finch brain, these experiments were conducted using zebra finch as the model animal. We recorded extracellular neural responses from the primary (Field L2) and secondary auditory areas (caudomedial nidopallium, NCM) while presenting sounds. In the first experiment, rare repetitions of one sound were presented after frequent alternations over multiple inter-stimulus intervals (ISI, 1s, 3s, or jittered from 0.8 to 1.2s). The results showed that neurons responded more strongly to the second stimulus of the repeated pair than normal in NCM at both variable and long ISIs. In contrast, neurons in Field L2 did not show this difference in responses. In the second experiment, variants of zebra finch songs were presented in both blocked and shuffled order. The results showed that temporal profiles of neural responses to variants of the same song were more similar in NCM than in Field L2. Furthermore, the response temporal profiles changed

more across repeated presentations in NCM than in L2. Together, these results show that sensitivity to transition patterns and invariance to song variants (i.e., variable inter-syllable intervals) may emerge hierarchically in the zebra finch auditory system and provide insights into the neural mechanisms of transition processing that may be important for the rapid processing and perception of speech.

EXPERIMENT 0: NEURAL CORRELATE OF TRANSITION VIOLATION IN THE SONGBIRD AUDITORY FOREBRAIN

Deviants are stimuli that violate one's prediction about the incoming stimuli. Studying deviance detection helps us understand how nervous system learns temporal patterns between stimuli and forms prediction about the future. Detecting deviant stimuli is also critical for animals' survival in the natural environment filled with complex sounds and patterns. Using natural songbird vocalizations as stimuli, we recorded multi-unit and single-unit activity from the zebra finch auditory forebrain while presenting rare repeated stimuli after regular alternating stimuli (alternating oddball experiment). The results showed that neurons were sensitive to rare repetitions in regular alternations. In the absence of expectation, repetition suppresses neural responses to the 2nd stimulus in the repetition. When repetition violates expectation, neural responses to the 2nd stimulus in the repetition were stronger than expected. These results showed that neural encoding of a stimulus depends not only on the acoustic features of the stimulus but also on the preceding stimuli and the transition patterns between them. These results also imply that the classical oddball effect may result from a combination of repetition suppression and deviance enhancement. Classification analyses showed that the difficulties in decoding the stimulus responsible for the neural responses differed for deviants in different experimental conditions. These findings suggest that learning transition patterns and detecting deviants in natural sequences may depend on a hierarchy of neural mechanisms, which may be

involved in more complex forms of auditory processing that depend on the transition patterns between stimuli, such as speech processing. However, in the speech, both the sounds (e.g., syllables and words) and their intervals in-between are variable instead of fixed. To further investigate the neural mechanisms of statistical learning of transition patterns in these more complex situations, two follow-up experiments were conducted.

See published work for more details about this experiment[1].

[1] Dong, M., & Vicario, D. S. (2018). Neural correlate of transition violation and deviance detection in the songbird auditory forebrain. Frontiers in systems neuroscience, 12, 46.

EXPERIMENT 1: STATISTICAL LEARNING OF TRANSITION PATTERNS OVER MULTIPLE TIME SCALES

Abstract

Statistical learning of transition patterns between sounds—a striking capability of the auditory system—plays an essential role in animals' survival. However, the neural mechanisms underlying this capability are still not fully understood, partially because of the lack of good animal models. We recorded multi-unit and single-unit activity in the zebra finch auditory forebrain while presenting rare repetitions of a single sound in a long sequence of sounds patterned in either an alternating or random order at different inter-stimulus intervals (ISI). Stimulus repetition was a deviant in the alternating condition, but not in the random condition. At all ISIs tested (1s, 3s, or jittered at 0.8 - 1.2s), deviant repetition enhanced neural responses in the alternating condition in the secondary auditory area (caudomedial nidopallium, NCM) but not in the primary auditory area (Field L2); in contrast, responses were reduced in the control condition in both L2 and NCM. In the alternating condition at 1s ISI, a subgroup of neurons showed oscillatory activities in phase with the two alternating stimuli. Even when a stimulus was repeated, neural activities showed evidence of continued oscillation. Together, these results demonstrate that neurons in the songbird auditory forebrain can learn transition patterns between sounds at multiple ISIs and neural oscillation may be a mechanism that encodes transition patterns

between sounds. Further studies using the current paradigm may help us understand the neural mechanisms of statistical learning and even speech comprehension.

Introduction

In the natural environment, individual sounds often occur in complex temporal orders with variable intervening intervals (e.g., words in spoken speech). Transition patterns could characterize the regularities in the sound sequences despite the variabilities in timing. Learning transition patterns is useful for predicting future stimuli, detecting deviant stimuli, and facilitating vocal communication (Nordby, Roth, & Pfefferbaum, 1988) (Cornella, Leung, Grimm, & Escera, 2012) (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015).

The auditory system can learn transition patterns between sounds without any external reinforcement. This phenomenon is called statistical learning and has been demonstrated in both humans and animals (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015) (Aslin, 2017) (Kikuchi et al., 2017). For example, after being exposed to sequences of tones with fixed transition patterns, human infants and adults showed surprise responses when they heard sequences that violated the transition patterns in the previously experienced sequences (Saffran, Johnson, Aslin, & Newport, 1999) (Saffran, Aslin, & Newport, 1996) (Aslin, 2017). Similar statistical learning phenomena have also been reported in songbirds, monkeys and other animals (Lu & Vicario, 2014)

(Kikuchi et al., 2017) (Dong & Vicario, 2018). However, it is still not fully understood how transition patterns are learned and encoded in the auditory system (Wacongne, Changeux, & Dehaene, 2012) (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015). Neural oscillation is one possible mechanism because both human and animal studies have suggested that neural oscillation and sequence learning may be related (Arnal & Giraud, 2012) (Doelling & Poeppel, 2015) (Sameiro-Barbosa & Geiser, 2016) (Kikuchi et al., 2017). For example, presenting sounds at fixed intervals can entrain neural oscillation in the auditory system and influence auditory perception (Sameiro-Barbosa & Geiser, 2016). Neural oscillation may be a mechanism for predicting future stimuli (Arnal & Giraud, 2012). Sequence learning can influence the phase-amplitude coupling of neural oscillation in the auditory cortex of humans and monkeys (Kikuchi et al., 2017). However, we still know little about how neural oscillation encodes transition patterns and predicts future stimuli. Also, it is unknown whether neurons are sensitive to transition patterns when the inter-stimulus interval (ISI) is long or variable because most laboratory studies used short and fixed ISIs (Saffran, Aslin, & Newport, 1996) (Saffran, Johnson, Aslin, & Newport, 1999) (Mauk & Buonomano, 2004) (Yaron, Hershenhoren, & Nelken, 2012) (Lu & Vicario, 2014) (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015).

The zebra finch is one of the best-developed animal models to study these questions because they produce complex vocalizations and use them for vocal communication (Elie et al., 2010) (Elie & Theunissen, 2018). In the current study,

we recorded multi-unit activity (MUA, Figure 1A) and single-unit activity (SUA) from the adult male zebra finch auditory forebrain while playing rare repetitions of one sound after presenting two sounds in either alternating or shuffled order (Nordby, Roth, & Pfefferbaum, 1988) (Cornella, Leung, Grimm, & Escera, 2012). The same paradigm was conducted using three different ISIs: fixed 1s, fixed 3s, and jittered (0.8 to 1.2s). The results showed that neurons were sensitive to transition patterns at all three ISIs in the caudomedial nidopallium (NCM, secondary auditory area) but not in Field L2 (primary auditory area). Stimulus repetition was a deviant in the alternating condition, but not in the shuffled condition. In the alternating condition, deviant repetition enhances neural responses to a stimulus; in contrast, in the control condition, repetition reduces neural responses. When two sounds were presented frequently in alternation at 1s ISI, some recording sites in NCM showed oscillatory responses. When one stimulus was repeated after frequent alternations, neural responses at those sites showed continuing oscillation to some extent as if the stimuli were alternating. In contrast, almost no sites showed oscillatory responses when ISI was 3s or jittered, or in the condition where stimuli were presented in shuffled order.

Our results show that neurons learn transition patterns between sounds and detect deviant sounds even when ISI is long (3s) or variable (jittered). This ability seems to occur hierarchically in the auditory forebrain. Our results also show that neural oscillation may be one mechanism of encoding transition

pattern when ISI is short (1s) and fixed. Together, these results provide new evidence for the predictive coding hypothesis (Friston, 2010), and further suggest that a hierarchy of neural mechanisms over different temporal scales may underlie the prediction of future events.

Methods

Subjects

This study used 16 adult (> 130 days old) male zebra finches. All birds were obtained from a local vendor and housed in a general aviary with other zebra finches at Rutgers University under a 12h : 12h light/dark cycle and provided with water and food *ad libitum*. All experimental procedures were approved by the Institutional Animal Care and Use Committee of Rutgers University.

Surgery

Birds were prepared for electrophysiological recording under isoflurane anesthesia (1-2% in oxygen). The anesthetized bird was placed in a stereotaxic device, feathers on the scalp were removed and 0.04 cc Marcaine (0.25%) was injected under the scalp. Then, a midline horizontal incision was made and enlarged to expose the skull. The outer layer of the skull was removed over the region of interest around the bifurcation of the mid-sagittal sinus. Dental cement was then used to form a small round chamber over the opening, and a metal pin was attached to the skull to keep the bird's head fixed during subsequent awake

electrophysiological recording. The bird received an injection of 0.04 cc Metacam (5 mg/mL) for post-operative analgesia and was closely monitored for recovery.

Electrophysiological recording

After two days of recovery, birds were recorded in a walk-in sound attenuation chamber (Industrial Acoustics Company, Bronx, NY). The bird was restrained in a custom tube and fixed to the stereotaxic frame by clamping the previously implanted pin. A small craniotomy exposed the dura over the recording area. Two silicon probes (NeuroNexus, Ann Arbor, MI), one for each hemisphere, were lowered into the auditory forebrain (1 mm lateral from midline, 1.5 mm rostral to Y-point for L2, 0.5 mm rostral to Y-point for NCM). Each probe had 16 recording sites (0.4-1 MΩ impedance at 1 kHz) in a 4-by-4 grid layout (Figure 1B). The probes were implanted in a para-sagittal plane such that the 4-by-4 grid layout extended in rostral-caudal and dorsal-ventral axes. Each probe was used for the right hemisphere in half of the birds and for the left hemisphere in the other half. Prior to insertion, the probes were dipped into a DiI solution (10% in ethanol; Sigma Aldrich, St. Louis, MO) and allowed to dry; this labeled probe insertion tracks for later histological analyses. Figure 1B shows the location of recording probes along with the two main structures of the auditory forebrain: Field L2 and caudomedial nidopallium (NCM). Field L2 is analogous to the primary auditory cortex in mammals. NCM is to the right of L2 and is similar to the superficial layer of the primary auditory cortex or the secondary auditory cortex in mammals

(Jarvis et al., 2005) (Wang & Karten, 2010) (Brainard & Doupe, 2013) (Calabrese & Woolley, 2015).

All stimuli were equated for RMS amplitude, and played back at a peak amplitude of 65 dB SPL (“A” scale) from a speaker placed 30cm in front of the bird aligned with the midline. White noise shaped with the amplitude envelope of zebra finch song was then used to search for responsive sites. Once the electrodes showed auditory-evoked activity characteristic of the target area, playback of experimental stimuli began. Two signal processors (Power 1401, CED, Cambridge, England) were used for stimulus presentation and neural recording. Neural activity was amplified (x 10,000), filtered (0.5 - 5 kHz bandpass), digitized (25 kHZ), and stored for further analysis.

Multi-unit activity (MUA) was obtained by thresholding the raw waveforms (3 standard deviations above the mean) for each recording site (Figure 1A). Single unit activity (SUA) was discriminated by feeding the raw waveforms into the automatic spike sorting algorithm “waveclus” (Quiroga, Nadasdy, & Ben-Shaul, 2004) (Chaure, Rey, & Quiroga, 2018). Sorted single units were included in the analysis only if the percentage of inter-spike intervals less than 2ms (contamination rate) was less than 2%. For each electrode/unit, neural response to each stimulus trial was computed by subtracting the average firing rate during the baseline period (1/4 of the inter-stimulus interval before the stimulus) from the firing rate during the stimulus period (plus 10% of stimulus duration).

Auditory stimuli

The stimuli were syllables from zebra finch songs (recorded in our laboratory) and canary songs (recorded from our laboratory and sampled from on-line resources). Zebra finch and canary syllables had different acoustic features, measured using Sound Analysis Pro (Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra, 2000) and potentially belonged to different categories. Figure 1C shows examples of zebra finch and canary syllables and their major acoustic differences (e.g., frequency modulation, pitch, entropy, ...). Within each condition (see below), a zebra finch syllable and a canary syllable of the same duration (140ms to 190ms) were paired. The particular stimuli used for the different conditions were counterbalanced across birds.

Alternating oddball paradigm

The alternating oddball paradigm included 3 conditions: alternating, control, and oddball (Figure 2). In the alternating condition, the two stimuli were first presented in an alternating order for 25 times (...ABABAB...), then rare repetitions of one of the two stimuli (AA or BB) were presented after a variable 4 to 10 common alternations (...ABABABA**BAA**BAB...). In total, there were 20 AA repetitions and 20 BB repetitions. For the repeated pairs, the 2nd stimulus was called the deviant because it violated the alternating regularities from the preceding sequence, while the 1st stimulus was called the standard. In the control condition, the stimulus sequence was generated from the alternating

condition: the deviant, standard, and the stimulus immediately before it were kept at the same position as a triplet whereas the stimulus sequence between the triplets was shuffled. Consequently, the 2nd stimulus in the repetitions was deviant in the alternating condition but not in the control condition (still called deviant for notation purposes). The typical oddball condition was also included for comparison purposes, as follows. Two stimuli (A & B) were presented in two blocks. In the 1st block, stimulus A was presented after a variable 4 to 10 repetitions of stimulus B. In the 2nd block, the roles of the two stimuli were reversed. For notation purposes, the rare stimulus was called the deviant and the stimulus immediately preceding it was called the standard.

Alternating and control conditions were conducted at 3 different inter-stimulus intervals (ISI): 1s fixed, 3s fixed, 1s jittered (randomly drawn between 0.8 and 1.2s after each trial). 3s fixed ISI tests whether neurons are sensitive to transition patterns when ISI is long. Jittered 1s ISI tests whether neurons are sensitive to transition patterns when ISI is variable. The oddball condition was conducted only at 1s jittered ISI for comparison purpose.

Criterion for Responsive Sites and Units

Recording sites (multi-unit) and units (single-unit) were included for data analysis if they responded to at least one deviant stimulus. Any given recording site or single-unit was considered to be responsive to a stimulus if these conditions were met:

1. The firing rate during the stimulus period was significantly different from that during the baseline period based on the paired Wilcoxon test ($p< 0.05$).

2. The average neural response to the stimulus was above baseline (firing rate > 1 spikes/s). Consequently, only excitatory recording sites and units were included in the analyses.

Quantify neural response differences elicited by the deviant and the standard

The surprise index (SI) quantifies differences in neural response to the deviant and standard (Ulanovsky, Las, & Nelken, 2003).

$$\mathrm{SI} = \frac{R_d - R_s}{R_d + R_s} \quad [1]$$

In the alternating oddball experiment, R_d and R_s represent the average neural responses to two stimuli when they were deviant and when they were standard. If the site/unit responded only to one of the stimuli, SI was calculated using the neural responses from the effective stimulus.

Quantify the magnitude of neural oscillation

For each recording site in each condition, we calculated the responses to each stimulus as firing rate during stimulus presentation. For each deviant, we define its baseline trials as those between the current and previous deviant (the 3 trials immediately after the previous deviant were excluded to remove potential post-

effects from the deviant). If the neural responses to the two stimuli during baseline were significantly different, we then calculated the oscillation magnitude as:

$$oscillation_x = \begin{cases} (D_x - \overline{S}_x) \div \mathrm{std}(S_x) & if \quad \overline{S}_x < \overline{S}_y \\ -(D_x - \overline{S}_x) \div \mathrm{std}(S_x) & if \quad \overline{S}_x > \overline{S}_y \end{cases} \quad [1]$$

where, $\overline{S_x}$ and $\overline{S_y}$ are the average responses when stimulus X and Y are not deviant; D_x is the response to X as deviant at current trial; $\mathrm{std}(S_x)$ is the standard deviation of neural responses to stimulus X during baseline.

If a stimulus elicited small neural responses during baseline, its oscillation magnitude is positive when it elicited larger responses when repeated than during baseline (Figure 3). If a stimulus elicited large neural responses during baseline, its oscillation magnitude is positive when it elicited smaller responses when repeated than during baseline. A recording site is oscillatory if the oscillation magnitudes of both stimuli are significantly larger than 0, because it indicates that the neural responses to the 2nd stimulus in the repetition is oscillating as if the stimuli were alternating. To remove potential effects from repetition suppression, the calculation of the average oscillation magnitude of a site only includes the stimulus that usually elicits small neural responses.

Histology

At the end of each recording experiment, the bird was sacrificed with an overdose of pentobarbital (0.15 ml of 39 mg/ml; Vortech Pharmaceutical, Dearborn, MI), and perfused with 0.9% saline and 3.3% paraformaldehyde. After several days of fixation, the brain was cut into 75 um sagittal sections using a Vibrotome and visualized with an epifluorescence microscopy. Figure 1B shows the location of recording probes along with the two main structures of the auditory forebrain: Field L2 and Caudomedial Nidopallium (NCM).

Statistical Analyses

In the analysis using surprise index (SI), each sample is one site/unit. For within-subject comparisons, we used the paired sample *t*-test; for between-subject comparisons, we used the independent sample *t*-test; for comparisons with hypothesized population means, we used the one sample *t*-test. When the normality assumption was not met, corresponding non-parametric statistical tests (Wilcoxon test and Mann Whitney *U*-test) were used. In the SI analysis, different statistical tests were used for comparisons based on MUA and SUA data. For MUA, comparisons between conditions were mostly within the same group of electrodes and thus we performed within-subject comparisons; for SUA, comparisons are across potentially different groups of units and thus we performed between-subject comparisons because spike sorting was done separately for different experimental conditions and units from different

conditions cannot be guaranteed to be the same. The significance level was set at 0.05 (Bonferroni adjusted *p*-values were reported when multiple comparisons were conducted). All analyses were conducted using customized scripts in Spike2, Matlab, and Python.

Results

Oddball effect seen in both Field L2 & NCM

In the typical oddball condition, a rare deviant was presented after repeating standard sounds and the inter-stimulus interval (ISI) was jittered (0.8 to 1.2 s). The surprise index (SI) was used to quantify the differences in neural responses to the deviant and standard (see Methods). A positive SI indicates that neural responses to the deviant are larger than to the standard, whereas a negative SI indicates that responses to the deviant are smaller than to the standard.

The SI was significantly larger than 0 in both Field L2 and NCM ($t=26.150$, $p < 0.001$, $n=116$ for NCM; $t=16.709$, $p < 0.001$, $n=150$ for L2; one-sample t-test), suggesting that a stimulus elicits larger neural responses when it is the rare deviant than when it is the standard (Figure 4). The SI was significantly larger in NCM than in Field L2 ($t=10.971$, $p < 0.001$, $n=116$; paired sample t-test), suggesting that neural responses to a sound are more sensitive to preceding stimuli in NCM than in Field L2. This result is consistent with previous findings that oddball responses are enhanced in primary auditory areas (Ulanovsky, Las, & Nelken, 2003) (Beckers & Gahr, 2012) (Hershenhoren, Taaseh, Antunes, &

Nelken, 2014). The data show an additional phenomenon: the oddball effect exists even when ISI is variable.

Single-unit activity (SUA) showed similar results (Figure 8) as seen for multi-unit activity (MUA). The SI was significantly larger than 0 in NCM ($t=7.422$, $p < 0.001$, $n=38$; one sample t-test) but not in Field L2 ($t=1.219$, $p > 0.240$, $n=18$; one sample t-test). Again, the SI was significantly larger in NCM than in Field L2 ($t=3.095$, $p=0.003$, $n_1=38$, $n_2=18$, independent sample t-test).

Neural responses are sensitive to transition patterns at multiple temporal scales in NCM

Sound stimuli were presented in the alternating and control conditions at multiple ISIs (see Methods; Figure 2). In the alternating condition, rare repetitions were presented after a sequence of alternating sounds; the 1st stimulus is standard because it follows the alternating pattern whereas the 2nd stimulus of the repetition is deviant because it violates the alternating transition patterns in the preceding sequence. In the control condition, the sound sequence was shuffled, so that a repeated stimulus was not deviant. In this case, stimuli were repeated at the same point in the overall sequence as for the alternating condition; the 1st and 2nd sound were labeled as standard and deviant, respectively. Responses to these two sequential sounds were compared by computing the SI.

In the control condition, the SI was significantly smaller than 0 at 1s ($W=1958$, $p < 0.001$, $n=167$; Wilcoxon test), 3s ($W=2933$, $p<0.001$, $n=195$), and

jittered ISI (W=2947, $p < 0.001$, n=196). These results showed that neural responses to the 2nd stimulus in the repeated pair were smaller than to the 1st stimulus, regardless of the ISI tested (Figure 5). This suppression effect lasted at least 3 seconds and occurred even when ISI was jittered (0.8 to 1.2s). Because the stimulus sequence was random and the 2nd stimulus in the repeated pair did not violate any regularities (neither expected nor unexpected), these results suggest that repetition suppresses neural responses to a stimulus in the absence of expectation.

In contrast, the SI in the alternating condition was significantly larger than in the control condition at 1s (W=4682, $p = 0.014$, n=156; Wilcoxon test), 3s (W=6097, $p<0.001$, n=190), and jittered ISI (W=5629, $p < 0.001$, n=182), even though it was still significantly smaller than 0 ($p < 0.05$ for all three ISIs). Thus, the responses to the deviant 2nd stimulus in the repeated pair were larger than usual, suggesting that these neurons detected the violation of the alternating pattern. This enhancement effect lasted at least 3 seconds and was also seen for jittered ISI (0.8 to 1.2s). Together, these results suggest that neurons in NCM are sensitive to transition patterns between sounds over multiple time scales and could detect deviants that violate transition patterns in the preceding stimulus stream.

Compared with SI calculated using MUA, single-unit data showed noisier results (Figure 9). The SI was significantly larger in the alternating condition than in the control condition at 1s ISI (U=1255.5, $p= 0.0426$, $n_1=62$, $n_2=50$; Mann–

Whitney U test), but not at 3s or jittered ISI (U=2313.5, p=0.344, n_1=66, n_2=73 for 3s; U=1911.5, p=0.190, n_1=60, n_2=70; Mann–Whitney U test). In the control condition, the SI was significantly smaller than 0 at 1s ISI (p=0.01, n=50, Wilcoxon test) but not 3s or jittered ISI (p=0.125, n=73 for 3s; p=0.267, n=48 for jittered; Wilcoxon test). In the alternating condition, the SI was not different from 0 at all three ISIs (p > 0.397 for all comparisons, n = 60, 62, and 66 for 1s, 3s and jittered, respectively).

Neural responses are not sensitive to transition patterns in Field L2

Neurons in Field L2 showed different behavior from those in NCM (Figure 6). As in NCM, the SI in the control condition was significantly smaller than 0 at all three ISIs (W > 2152, p < 0.001 for all comparisons, n=155, 120, and 163 for 1s, 3s, and jittered ISI, respectively; Wilcoxon test). However, in contrast to NCM, the SI in the alternating condition was not significantly different from that in the control condition at all tested ISIs (p > 0.05 for all conditions; n = 139, 118, and 143 for 1s, 3s, and jittered ISI, respectively; Wilcoxon test). For the jittered ISI, the SI was slightly larger in the alternating condition than in the control condition but was not statistically significant (note that the scale in y-axis is different in Figure 4).

Results from SUA in Field L2 were similar to those seen with MUA (Figure 10). The SI in the control condition was not different from 0 at all three ISIs (W<193, p > 0.281 for all comparisons, n=31, 24, 20 for 1s, 3s, and jittered ISI,

respectively; Wilcoxon test). The SI in the alternating condition was not significantly different from that in the control condition at all tested ISIs (U=266.5, p= 0.328, n_1=24, n_2=24 for 1s; U=189, p=0.117, n_1=24, n_2=20 for 3s; U=325, p=0.5, n_1=21, n_2=31 for jittered; Mann–Whitney U test).

Neural oscillation encodes transition patterns at 1s ISI

The deviance detection we observed requires that neurons learn transition patterns from experience and form predictions about future stimuli. In the alternating condition at regular ISI, neural oscillation is one possible mechanism for encoding the transition pattern and could underlie prediction. If one stimulus (a zebra finch syllable) elicits a large response whereas the other stimulus (a canary syllable) elicits a small neural response, the neural responses will oscillate between small and large as the two stimuli alternate. If neural oscillation encodes transition pattern, a repeated zebra finch syllable should elicit smaller responses than usual while a repeated canary syllable should elicit larger responses than usual (Figure 3).

In the alternating condition at 1s ISI, 13 out of 240 recording sites (5%) showed oscillatory behavior (Figure 7). In contrast, we found significantly fewer oscillatory sites when ISI was jittered ($p = 0.01$, χ^2 test) or 3s ($p < 0.01$). In the control (random order) condition, no such sites were found ($p < 0.01$). The oscillation magnitude (Method section) was also significantly larger in the alternating 1s condition than in any other conditions ($p <= 0.04$, , $n_1 = 13$, n_2=3;

Mann–Whitney U test). Together, these results suggest that neural oscillation may be encoding the alternating transition pattern for at least some neurons, and could underlie predictive neural responses at least for short, fixed ISI.

Discussion

The current study investigated the hierarchical emergence of sensitivity to violation of transition patterns in the songbird auditory forebrain. In Field L2, neural responses were sensitive to the occurring probabilities of a sound but not sensitive to transition patterns between sounds. In contrast, neurons in NCM were sensitive to both. Our results also suggest that neural oscillation may be one of the neural mechanisms for encoding alternation patterns when ISI is short and fixed.

Repetition suppresses neural responses at multiple time scales (Figure 5 & 6). In both Field L2 and NCM, the SI was significantly smaller than 0 In the control condition all tested ISIs (1s, 3s, jittered), suggesting that the suppression from repetition lasted at least 3 seconds and stayed even when ISI was jittered (0.8 to 1.2s). Because the stimulus sequence was random in the control condition and the 2nd stimulus in the repetition did not violate any regularities (neither expected nor unexpected), these results suggest that repetition suppresses neural responses to a stimulus in the absence of expectation. Because repetition suppression was observed in both Field L2 and NCM, it supports the idea that simple repetition suppression occurs relatively early in the

auditory system (Wehr & Zador, 2005) (Alves-Pinto, Baudoux, Palmer, & Sumner, 2010).

Deviance enhances neural responses to a stimulus at multiple time scales (Figure 5). In NCM, the SI was significantly larger in the alternating condition than in the control condition at all tested ISIs (1s, 3s, or jittered). Neural responses to 2nd stimulus in the deviant repetition were larger than usual, suggesting that the repetition violated the alternating pattern of the preceding stimulus sequence. This enhancement effect lasted at least 3 seconds and occurred for jittered ISI (0.8 to 1.2s). In contrast, no significant enhancement was seen in Field L2. Together, these results suggest that sensitivity to the violation of transition patterns may emerge hierarchically in the auditory system, which is consistent with previous reports (Wacongne et al., 2011) (Cornella, Leung, Grimm, & Escera, 2012) (Chennu et al., 2013) (Ono, Okanoya, & Seki, 2016).

When ISI is 1s and fixed, the learned transition patterns may be encoded via neural oscillation. This oscillation result is consistent with previous reports that statistical learning of sequence can affect neural oscillation (Arnal & Giraud, 2012) (Kikuchi et al., 2017) but more directly shows how neural oscillation may affect neural responses to a stimulus. The observed oscillation may also be explained by chained suppression-release or release-suppression. If the suppression effect of a stimulus depends on the magnitude of responses elicited by it, the strong responses to the 2nd repeated canary syllable will be a result of weak suppression from the 1st canary syllable whereas the weak responses to

the 2nd zebra finch syllable will be a result of strong suppression from the 1st zebra finch syllable. If this was the case, we should also observe oscillation in the jittered condition. However, not much oscillation was found when ISI was jittered. Another possible interpretation for the neural oscillation is sleep, which can be associated with oscillatory activity (Steriade, McCormick, & Sejnowski, 1993). If the bird had gone to sleep during testing, we would expect to observe many oscillatory sites but only a small fraction of sites showed oscillation. Also, sleep cannot easily explain why only neural responses to a non-effective stimulus increased. Together these results suggest that the observed neural oscillation probably is due to passive exposure to alternating sounds. However, it is not clear whether this oscillation is encoding the alternation pattern or the rhythm of the sound. It has been reported that neural oscillation could be entrained by the rhythm of the sound (Doelling & Poeppel, 2015). This is one limitation of the current study and requires further investigation. In the end, because neurons in NCM were also sensitive to violation of alternation patterns at 3s or jittered ISI when neural oscillation was not observed, it suggests that neural oscillation is at most one of the mechanisms to encode transition patterns. When ISI is long or variable, other neural mechanisms must be needed for encoding the transition patterns (Wacongne et al., 2011) (Cornella, Leung, Grimm, & Escera, 2012) (Chennu et al., 2013) (Lu & Vicario, 2014) (Ono, Okanoya, & Seki, 2016).

Further investigation to characterize which type of neurons show the oscillation patterns and how they contribute to deviance detection may provide a

circuit model and let us understand how the auditory system learns transition patterns when ISI is short and fixed. Another interesting question is whether the observed oscillation is an innate oscillation entrained by the stimulus or induced *de novo* by the stimulus pattern (Spaak, Lange, & Jensen, 2014) (Keitel, Quigley, & Ruhnau, 2014) (Daikoku, 2018). In the end, the mechanisms that encode learned transition patterns in neurons, especially when ISI is long and variable, will require further investigation. Our paradigm is one example of using a simple artificial grammar to study auditory pattern processing (Milne, Petkov, & Wilson, 2017) (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015). Further work with this type of approach may also deepen our understanding of how animals and even humans learn transition patterns between sounds (such as those in speech).

EXPERIMENT 2: VARIANT-INDEPENDENT REPRESENTATION OF ZEBRA FINCH SONGS IN THE AUDITORY FOREBRAIN

Abstract

Invariant representation of a given natural sound that is produced with variations is important for identifying specific signals used in vocal communication and individual recognition. For example, the same word spoken in different situations can be acoustically different. However, listeners still perceive it correctly. To investigate how variant-independent representation emerges in the auditory system, we recorded neural responses from both the primary thalamo-recipient auditory area (Field L2) and a secondary area (caudo-medial nidopallium, NCM) of zebra finches while presenting naturally–produced variants of the same zebra finch song. The response temporal profiles for different variants at individual recording sites were more similar to each other in NCM than in Field L2. Moreover, in NCM, the response temporal profiles for different variants converged with passive exposure, but changed little in Field L2. In addition, a population-level analysis of response magnitude showed that responses to different variants in NCM were more similar to each other than in Field L2. Together, these results suggest that invariant representation of zebra finch song variants emerges hierarchically in the auditory forebrain. Because variant-independent representation can be essential for word recognition, findings in the

zebra finch model may provide insights into basic neural mechanisms that serve speech perception.

Introduction

Perceptual invariance, the ability to perceive variable stimuli as the same, is a striking and important capability of vocal communicators like humans (Johnson, 2008) (Weatherholtz, & Jaeger, 2016). Natural vocalizations often vary in many dimensions (e.g., duration, loudness) while following certain patterns. Some of the variations are not informative, and the auditory system must ignore them while extracting the relevant pattern that identifies a given sound or phoneme (Sharpee, Atencio, & Schreiner, 2011). For example, the same word spoken in different situations (or by different speakers) sounds different acoustically, but the listener must recognize the word while ignoring the variations. How our auditory system forms such variant-independent representations is still not fully understood.

Several studies have investigated invariant representation of vocalizations varying in different dimensions using animal models (Ohms, Gill, Van Heijningen, Beckers, & Cate, 2009) (Carruthers et al., 2015) (Sadagopan & Wang, 2008) (Blackwell et al., 2016) (Billimoria, Kraus, Narayan, Maddox, & Sen, 2008) (Blackwell et al., 2016) (Cate & Okanoya, 2012). At the behavioral level, it has been suggested that the zebra finch may form a speaker-independent representation of human speech (Ohms, Gill, Van Heijningen, Beckers, & Cate,

2009). At the neural level, it has been suggested that variant-independent representation emerges hierarchically in the auditory system. In the primary auditory area (Field L2) of zebra finches, a subset of neurons displayed invariant responses to sound played at different intensities (Billimoria, Kraus, Narayan, Maddox, & Sen, 2008). In the secondary auditory areas (caudo-medial nidopallium, NCM), neurons (especially those with broader spike shape) responded more similarly to calls belonging to the same semantic category than in the primary auditory areas (Elie & Theunissen, 2015). Also, when zebra finch songs—consisting of multiple syllables—were presented with different levels of background noise, neural responses were less affected in the secondary auditory areas than in the primary auditory areas (Schneider & Woolley, 2013) (Moore, Lee, & Theunissen, 2013). Similarly, in rats, the population responses were more invariant to the distortions of short ultrasonic vocalizations in the secondary auditory cortex than in the primary auditory cortex (Carruthers et al., 2015). These results suggest that animals can form variant-independent representations of vocalizations and that the invariance emerges hierarchically in the auditory system. However, vocalizations can be more complex than short ultrasonic vocalizations and zebra finch calls. For complex vocalizations like zebra finch songs, the variations may not be limited to background effects but can include variations in various parameters, e.g., inter-syllable intervals. It remains unclear how neurons respond to natural variants of sound objects and how passive exposure may affect those neural responses.

The zebra finch is a powerful model for studying the variant-independent neural representation of complex sounds. Each adult male zebra finch sings a complex song that serves individual recognition and mate selection (Elie et al., 2010) (Elie & Theunissen, 2018) (Zann, 1996). Even though each bird's song has a fixed pattern after crystallization, individual examplars vary in both temporal and acoustic dimensions, such as in inter-syllable intervals (Mooney, 2009) (Glaze & Troyer, 2006). The temporal cues in the songs are important for individual recognition and perceptual discrimination of different songs (Gentner & Margoliash, 2003) (Gentner, Fenn, Margoliash & Nusbaum, 2006) (Shaevitz & Theunissen, 2007). The auditory system therefore must ignore the temporal variations of songs from one individual while detecting temporal variations of songs across different birds. Past studies identified an area, the caudomedial nidopallium (NCM), where neurons respond more strongly to conspecific songs than heterospecific songs (Chew, Mello, Nottebohm, Jarvis, & Vicario, 1995) (Phan, Pytte, & Vicario, 2006). Further, neural responses in NCM encode stimulus familiarity through a process of stimulus-specific adaptation to individual zebra finch songs, suggesting that NCM plays important roles in song recognition. The observed adaptation is not an overall reduction in neural responses (Chew, Mello, Nottebohm, Jarvis, & Vicario, 1995). For each neuron, the temporal profile of responses to different parts of the stimulus shows different degrees of change during adaptation. Some parts of the neural responses decrease whereas other parts stay stable, and may even increase. If response changes occur to parts of the stimuli that happen to be the differences between

song variants, it would mean that neural responses to variants will become more similar to each other after adaptation. In contrast, neurons in thalamo-recipient Field L2 -- one of the primary inputs to NCM -- do not show a preference for conspecific songs, encode stimulus familiarity, or display stimulus-specific adaptation (Chew, Mello, Nottebohm, Jarvis, & Vicario, 1995) (Ono, Okanoya, & Seki, 2016) (Brainard & Doupe, 2013). These results all suggest that neural responses in secondary auditory area NCM may encode song pattern memories that are more invariant to natural variants of zebra finch songs than in Field L2.

We recorded neural responses from L2 and NCM of zebra finches while presenting variants of a zebra finch song (produced by an individual), in either blocked or shuffled order. We found that temporal profiles of neural responses to natural variants of a zebra finch song were more similar to each other in NCM than in Field L2. Furthermore, with passive exposure, temporal profiles of responses to variants became more similar in NCM than in L2. Overall, population responses to song variants were more similar to each other in NCM than in Field L2. These results suggest that invariance emerges hierarchically in the zebra finch auditory forebrain and passive exposure may facilitate the formation of variant-independent representations.

Methods

Subjects

This study used 15 adult (> 130 days old) male zebra finches. All birds were obtained from a local vendor and housed in a general aviary with other zebra finches at Rutgers University under a 12h : 12h light/dark cycle and provided with water and food ad libitum. All experimental procedures were approved by the Institutional Animal Care and Use Committee of Rutgers University.

Surgery, histology, and criterion for responsive sites and units

See methods section in experiment 1.

Auditory stimuli

The stimuli were song motifs of adult male zebra finches. Each bird was isolated in a sound-proof box for one or two days to record hundreds of the song motifs. For each bird, one manually selected motif was selected and fed into an algorithm that automatically detected similar motifs from all recordings (gardner-lab/find-audio). The motifs extracted by the algorithm were then visually inspected and 8 motifs of different durations were selected as the stimuli. Figure 11C shows the spectrograms of one set of stimuli. For each bird, stimuli were presented in blocked, shuffled, and “contrast” conditions (Figure 12). The order of blocked, shuffled, and contrast condition was counterbalanced across birds. In the blocked condition, 8 variants of one zebra finch song were presented in a

blocked order. In the shuffled condition, 8 variants of a different zebra finch song were presented in a shuffled order. In the contrast condition, 4 variants from each of 2 different zebra finch songs (a total of 8 stimuli) were presented in a shuffled order. 4 variants were chosen by taking every other variant from the 8 variants of a song.

Neural dissimilarity

To calculate the dissimilarities between the temporal profiles of different neural responses (Figure 13A), the spike counts during the stimulus-evoked response period were first grouped into 10ms bins because peak mutual information estimations in NCM are seen at 5ms to 10ms temporal resolutions (Soyman, 2018). The duration of the response period for each stimulus set was equal to the maximum stimulus duration for that particular set plus 10% of the stimulus duration. To develop a dissimilarity metric that is only sensitive to the temporal profiles, but not to the total firing rates, neural responses were standardized by taking the z-score of each bin by normalizing it with the average and the standard deviation across all bins within the same trial (Figure 13A). Then, neural dissimilarity was quantified by calculating the Euclidean distance between these z-scored response profiles of the same unit to different pairs of stimulus presentations as

$$\text{Neural Dissimilarity} = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \quad [1]$$

where A_i and B_i are the binned response profiles in the two stimulus presentations and n is the number of bins in the longer stimulus. Similar Euclidean distance-based metrics have been widely used as measures of spike train dissimilarity (Rossum, 2001).

When comparing the blocked and shuffled condition, the neural dissimilarity was calculated across 8 variants for one song at each trial. In the contrast condition, because every other variant was taken out of 8 variants of a song, the average difference in song duration was larger than in the shuffled/blocked condition. To adjust for this difference, the neural dissimilarity in the shuffled condition was calculated by treating variant 1, 3, 5, 7 (number indicates relative stimulus duration) as one song and variant 2, 4, 6, 8 as another different song when comparing with the contrast condition. The neural dissimilarity was then calculated as their average.

Decoding based on response temporal profiles

Neural dissimilarity calculations described above were used to decode stimulus identities from the temporal profiles of neural responses (Figure 13B). The dissimilarities of a particular response to the responses on all presentations of each stimulus were averaged, which produced 8 average neural dissimilarities,

one for each of the 8 stimuli. The response was assigned to the stimulus with the minimal average neural dissimilarity. The decoding accuracy was calculated by counting how many of the 8 stimuli were correctly classified for a given stimulus presentation trial. The chance level for correct decoding probability was 1/8 = 0.125.

In the contrast condition, we calculated the decoding accuracy both for variants within each song and for different songs. When comparing with the decoding accuracy in the contrast condition, the decoding accuracy in the shuffled condition was calculated in the following way to adjust for the bigger acoustic and temporal differences between variants in the contrast condition. In comparison with the contrast condition, the decoding accuracy in the shuffled condition was calculated for variants within each song by treating variant 1, 3, 5, 7 (number indicates relative stimulus duration) as one song and variant 2, 4, 6, 8 as another different song. Their average was used as the decoding accuracy for the shuffled condition.

Decoding based on population responses

To explore whether the change in neural dissimilarity influences the decodability of neural responses by downstream neurons, linear discriminant analysis (LDA) was employed using the following assumptions:

- For each bird, the population neural response to a stimulus trial is defined as the responses (firing rates) from all responsive electrodes/units

(separately for L2 and NCM). $x = [x_1, x_2, \ldots, x_i, \ldots, x_n]$, n is the total number of responsive electrodes/units.

- The population neural response to a stimulus is a random variable following a multivariate normal distribution. $x \sim MVN(\mu\rightarrow, \Sigma)$.

- The population responses to the deviant and the standard have the same covariance matrix but different means.

Even though these assumptions may not be biologically plausible, they are consistent with the view that neural responses are samples from some underlying distributions (Hoyer & Hyvärinen, 2003) (Buesing, Bill, Nessler, & Maass, 2011). The LDA classifier was used because it requires a relatively small sample size to train and is equivalent to the naive Bayes classifier when shrinkage parameter is set to 1. The LDA that was used is the implementation from the sklearn package in Python (Pedregosa et al., 2011). Note that this approach to population decoding uses the response magnitudes (firing rates), not the temporal profiles in the previous analyses.

Statistical analysis

For within-subject comparisons, we used the paired sample *t*-test; for between-subject comparisons, we used the independent sample *t*-test; for comparisons with hypothesized population means, we used the one sample *t*-test. When the normality assumption was not met, corresponding non-parametric statistical tests (Wilcoxon test and Mann Whitney *U*-test) were used. For MUA, comparisons

between blocked and shuffled conditions are within-subject because they are repeated measures. Comparisons between NCM and Field L2 are between-subject because they consist of sites/units from different brain areas. For SUA, comparisons are across potentially different groups of units and thus we always performed between-subject comparisons because spike sorting was conducted separately for different experimental conditions and units from different conditions cannot be guaranteed to be the same. The significance level was set at 0.05 (Bonferroni adjusted *p*-values were reported when multiple comparisons were conducted). All analyses were conducted using customized scripts in Spike2, Matlab, and Python.

Results

Response temporal profiles to variants are more similar in NCM than in Field L2

Thresholded multi-unit responses were recorded form 160 sites in NCM and 165 sites in Field L2, based on histological reconstruction. In both the blocked and shuffled condition, the neural dissimilarity between variants was significantly smaller in NCM than in Field L2 ($t=-2.42$, $p=0.016$, $n_1=148$, $n_2=156$ for blocked condition; $t=-10.744$, $p<0.001$, $n_1=160$, $n_2=165$ for shuffled condition; independent sample t-test), suggesting that the temporal profiles of responses to different variants were more similar in NCM than in Field L2 (Figure 14). In NCM, the neural dissimilarity was significantly smaller in the shuffled condition than in

the blocked condition (t=10.218, p < 0.001, n=145; paired sample t-test), suggesting that order of variant presentation affects the response temporal profiles for variants. In Field L2, the neural dissimilarity between variants was similar in the blocked and shuffled conditions (t=1.028, p=0.305, n=154; paired sample t-test), suggesting that response temporal profiles do not depend on the order of variant presentation.

Average decoding accuracy based on the temporal profiles of responses to variants showed similar phenomena as seen for neural dissimilarity (Figure 15). In both blocked and shuffled conditions, the decoding accuracy was significantly lower in NCM than in Field L2 (t=-10.860, p<0.001, n_1=153, n_2=163 for shuffled condition; t=-7.937, p<0.001, n_1=148, n_2=153 for blocked condition; independent sample t-test). In NCM, the decoding accuracy was significantly lower in the shuffled condition than in the blocked condition (t=-6.433, p < 0.001, n=143, paired sample t-test). In Field L2, the decoding accuracy in the shuffled condition was also significantly lower than in the blocked condition (t=-2.839, p=0.005, n=149, paired sample t-test), even though the Field L2 decoding accuracy was very high in both conditions, with many sites showing accuracy at or near 1.0 (perfect decoding).

Single-unit activity (SUA) showed similar but noisier results. Both neural dissimilarity and decoding accuracy were much lower for SUA than for MUA (p < 0.001 for all comparisons). Again, the neural dissimilarity was smaller in NCM than in L2 (Figure 16), but the difference was not statistically significant (t=-1.286,

p=0.202, n_1=42, n_2=46 for NCM; t=-0.126, p=0.900, n_1=33, n_2=28 for L2; independent sample t test). The decoding accuracy was also lower in NCM than in L2 but the difference was not statistically significant (t=-1.5, p=0.131, n_1=18, n_2=21 for blocked condition; t=-0.730, p=0.468, n_1=22, n_2=36 for shuffled condition; independent sample t-test) (Figure 17).

Passive exposure affects responses to variants

For each recording site, we fitted a linear regression of trial (6 to 40) on neural dissimilarity or decoding accuracy. The slope measures how neural responses change across trials.

In NCM, the neural dissimilarity between variants decreased significantly from trial 6 to trial 40 in the blocked condition (t=-11.095, p<0.001, n=148; one sample t-test) but not in the shuffled conditions (t=-1.903, p=0.0588, n=160) (Figure 14). In L2, the neural dissimilarity decreased significantly from trial 6 to trial 40 in the blocked condition (t=-5.489, p<0.001, n=156, one sample t-test) but not in the shuffled condition (t=-1.899, p=0.059, n=165). In the blocked condition, the neural dissimilarity decreased significantly faster in NCM than in L2 (t=-5.493, p<0.001, n_1=148, n_2=156, independent sample t-test).

The decoding accuracy changed in a complex pattern (Figure 15). In NCM, the decoding accuracy increased during the first few trials but then decreased slowly in the blocked condition. In NCM, the decoding accuracy decreased significantly from trial 6 to trial 40 in the blocked condition in NCM (t=-

2.341, p=0.021, n=148; one sample t-test) but not in the shuffled condition (t=0.818, p=0.415, n=153). In L2, the decoding accuracy did not decrease significantly from trial 6 to trial 40 in either blocked (t=1.145, p=0.254, n=153; one sample t-test) or shuffled condition (t=0.311, p=0.756, n=63).

SUA showed more complex results (Figure 16 & 17). In NCM, the neural dissimilarity decreased significantly from trial 6 to trial 40 in both blocked and shuffled condition (t=-3.324, p=0.002, n=42 for the blocked condition; t=-2.08, p=0.043, n=46 for the shuffled condition; one sample t-test). In L2, the neural dissimilarity did not increase from trial 6 to trial 40 in either blocked or shuffled condition (t=-0.291, p=0.773, n=33 for the blocked condition; t=-0.481, p=0.634, n=28 for the shuffled condition; one sample t-test). The decoding accuracy did not change much from trial 6 to trial 40 in either NCM/L2 or blocked/shuffled condition ($p > 0.05$ for all tests), potentially due to small sizes.

Population responses to different variants are more similar to each other in NCM than Field L2

After pooling the results from blocked and shuffled conditions, the decoding accuracy based on population responses was significantly lower in NCM than in L2 (U=153.5, p = 0.014, n_1=20, n_2=22; Mann-Whitney u test) (Figure 18). Without pooling, the same trend appeared in each condition but was not statistically significant, potentially due to small sample size.

Contrast from a different song does not affect neural responses to song variants

We hypothesized that contrast from a different song in the contrast condition would make the response temporal profiles for variants of the same song more similar. The result is opposite of this hypothesis. In both NCM and L2, the neural dissimilarity between song variants were not smaller in the contrast condition than in the shuffled condition ($p > 0.05$ for both comparisons) (Figure 19). The decoding accuracy showed similar results (Figure 20). In both NCM and L2, the decoding accuracy was not smaller in the contrast condition than in the shuffled condition ($p > 0.05$ for both comparisons).

SUA showed similar but noisier results (Figure 21 & 22). In both NCM and L2, the neural dissimilarity was not smaller in the contrast condition than in the shuffled condition ($p > 0.05$). The decoding accuracy was not smaller in the contrast condition than in the shuffled condition, either ($p > 0.05$).

Adaptation and decoding accuracy between songs

In the end, we verified that the neurons in NCM displayed faster adaptation to zebra finch songs than neurons in L2. The adaptation rate in NCM was significantly more negative than in L2 in all three conditions ($p < 0.001$ for all three comparisons) (Figure 23). SUA showed similar results. The adaptation rate was more negative in NCM than L2 in all conditions (Figure 24). Note that, the

comparisons were not statistically significant in SUA, potentially due to the small number of single-units isolated ($p > 0.05$).

MUA showed that the decoding accuracy between songs was significantly higher in L2 than in NCM ($p < 0.001$, $n_1=142$, $n_2=150$; Mann-Whitney u test) (Figure 25). SUA did not show this difference ($p = 0.33$, $n_1=20$, $n_2 = 52$; Mann-Whitney u test) (Figure 26).

Discussion

The current experiment investigated how neurons in the songbird auditory forebrain respond to song variants produced by an individual zebra finch, when presented in either blocked or shuffled order. We found that the response temporal profiles for variants were more similar to each other in NCM than in Field L2. In addition, in NCM, the neural dissimilarity between variants changed with passive exposure

The results suggest that variant-independent representation emerges hierarchically in the zebra finch auditory forebrain at both population and single-neuron/site level. As a population, neural responses were more variant-independent in NCM than in L2. This is consistent with previous reports from rats using distortions of ultrasonic vocalizations as variants (Carruthers et al., 2015), suggesting that the avian auditory system may be hierarchically similar to that of mammals. At the single neuron/site level, we compared the response temporal profiles to variants of different duration (resulting from both syllable duration and

inter-syllable intervals). The response temporal profiles for variants were more similar to each other in NCM than in L2. One possible explanation is that neurons in NCM often fire less action potentials than in L2 and the difference in absolute responses explain our results. However, this is unlikely because the responses have been normalized to z-scores within each trial in our calculation of neural dissimilarity so differences in absolute response magnitude have been removed (see Methods, neural dissimilarity). Another possibility is that the neural dissimilarities between variants were smaller in NCM because the response temporal profiles were much more variable. Indeed, the within-variant neural dissimilarity was smaller in L2 than NCM. However, the magnitude of this difference was much smaller compared with the difference in between-variant neural dissimilarity (Figure 27 & 28).

Another important phenomenon is that the neural dissimilarity and decoding accuracy based on the response temporal profiles changed with passive exposure: between-variant dissimilarity decreased across trials. This is consistent with our hypothesis that response temporal files for song variants may be converging to a common template after passive exposure in NCM. Note that the decoding accuracy based on the response temporal profiles only decreased in the blocked condition but not in the shuffled condition. However, the decoding accuracy based on the response temporal profiles increased in the shuffled condition in NCM, opposite to the predictions of our hypothesis. The increase seems to be caused by a dramatic decrease of within-variant neural dissimilarity

with passive exposure (Figure 27 & 28). Note that, however, even after the increase, the decoding accuracy was still significantly lower in the shuffled condition than in the blocked condition.

The neural responses in NCM were more sensitive to the order of variant presentation than in Field L2, whereas the neural responses in Field L2 were more sensitive to the acoustics of sound variants. This hierarchical sensitivity to stimulus sequence is consistent with previous results but on a much longer temporal scale (Yaron, Hershenhoren, & Nelken, 2012) (Ono, Okanoya, & Seki, 2016) (Dong & Vicario, 2018). The neural dissimilarity between variants decreased significantly faster with exposure in NCM than in Field L2. The decoding accuracy was also more different between the blocked and shuffled presentation in NCM than in Field L2. One interpretation is that the sound is much more difficult to predict in the shuffled condition than in the blocked condition and the predictability influences the neural responses to a sound. Consistent with this, in the blocked condition, when a new variant was presented for the first time, it is like an oddball because both the syllables and inter-syllable intervals differed from those in the previous variant. Correspondingly, the neural responses increased significantly more in NCM than in Field L2. The sensitivity to stimulus sequence may also explain why neural response are less well explained by the spectro-temporal receptive field in higher auditory areas (e.g., NCM) than primary auditory areas (e.g., L2) (Woolley, Fremouw, Hsu, & Theunissen, 2005) (Kim & Doupe, 2011).

Using natural variants of zebra finch songs as we have done has the advantage of being ethologically relevant. However, it also has the limitation that acoustic variations were not explicitly controlled. The variants we used were compared for duration (which includes both syllable and inter-syllable durations), but subtler acoustic parameters may have varied as well. In the future, generative neural network models or biophysical models could be used to systematically change the songs in a controlled way (Boari, Perl, Amador, Margoliash, & Mindlin, 2015) (Sainburg, Thielk, Theilman, Migliori, & Gentner, 2018). This would enable study of the particular variations in zebra finch songs that are ignored by the auditory neurons in NCM and what features are used to form the variant-independent representation. Another possible direction is to train zebra finches on behavioral tasks using novel stimuli and study whether zebra finches could form new variant-independent representations. Because variant-independent representation is essential for word recognition across speakers and instances, more studies along these lines may help us understand the neural mechanisms of rapid speech processing.

GENERAL DISCUSSION

Current experiments investigated two fundamental questions in the auditory processing of complex vocalizations: 1) statistical learning of transition patterns from a sequence of sounds; 2) emergence of variant-independent representation of sounds. The results suggest some potential circuit neural mechanisms for these two striking capabilities of the auditory system.

For neural encoding of transition patterns at 1s ISI, the observation that a subset of neurons in NCM showed “oscillatory” responses suggest that neural oscillation may be one potential mechanism. The alternating transition patterns can be encoded by neural responses that oscillate between large and small. Prediction for future stimuli is simply represented in the magnitude of the oscillatory responses from a group of oscillatory neurons. Previous studies have shown that about half of neurons in NCM are inhibitory, and that inhibition can contribute to the temporal pattern of auditory responses (Pinaud et al., 2008). In particular, blocking GABAergic inhibition increases the phasic and suppresses the tonic component of responses to song and call stimuli (Pinaud et al., 2008). The neural oscillation may emerge through a rebalancing of inhibition in NCM caused by the alternating sounds, which in turn increases the neural responses to the effective stimulus (zebra finch syllables) and decreases the responses to the non-effective stimulus (canary syllables). Alternatively, the oscillatory responses may be generated by spontaneous activity of NCM neurons that may be similar to cortical layer V neurons in mammalian nervous system (Sanchez-

Vives and McCormick, 2000) (Compte, Sanchez-Vives, McCormick, Wang, 2003), that is then entrained by the repeated alternating sounds. If some neurons in NCM receive both oscillatory prediction responses and stimulus-locked L2 responses, they could detect the deviance of a sound by comparing the two different types of responses. More complex probabilistic and non-adjacent transition patterns may be encoded by a linear combination of neural oscillation at different time scales (Lu & Vicario, 2014). However, for encoding transition patterns at jittered and longer ISI, neural oscillation seems to be insufficient; more complex neural mechanisms like state-dependent computation may be required (Buonomano & Maass, 2009) (Buonomano, Bramen & Khodadadifar, 2009).

Several neural mechanisms may underlie the hierarchical emergence of variant-independent representation from L2 to NCM. One mechanism may be to convert the dense code in L2 to a sparser code in NCM. The neurons in L2 show a classic tonotopy similar to that seen in layer 4 of mammalian A1 (with “Mexican hat” receptive fields ad brisk phasic responses; Terleph, Mello & Vicario, 2006). As a result, they responded precisely to the acoustic features of zebra finch songs and were sensitive to small variations in the inter-syllable intervals. These precisely timed neural responses can be converted into a sparse population code in NCM, as suggested by other studies using sequences of click sounds (Lim, Lagoy, Shinn-Cunningham & Gardner, 2016) (Schneider & Woolley, 2013). This conversion may be achieved by non-linear transformation and combination of

responses from multiple neurons in L2, potentially via inhibition (Pinaud et al., 2008) (Bendor, 2015). The responses resulting from this combination may be less sensitive to small variations in the inter-syllable intervals. Another possible mechanism could reflect stimulus-specific adaptation. Previous studies have suggested that the noise-invariant representation of a sound is correlated with adaptation to stimulus statistics (Rabinowitz, Willmore, King & Schnupp, 2013). In the variant experiment, each variant has its own unique features that modify an imputed “prototypical” sound object. The convergence of responses seen with adaptation, may reflect the way NCM neurons decrease neural responses to the small variations unique to each variant. In this way, the response temporal profiles for different variants will become more similar to each other. Further studies using artificial stimuli with systematically controlled parameters may help us understand the neural mechanisms underlying variant-independent representation.

The results from the alternating and variant experiments are not isolated but related to each other, even though they were conducted separately. Despite the differences in ISI and complexity of stimuli (syllables in Experiment 1 versus song motifs in Experiment 2), both experiments showed that neurons were more sensitive to the order of stimulus presentation in NCM than in Field L2. The first experiment showed that neurons in NCM were sensitive to the violation of alternation patterns at 1s, 3s, or jittered ISI. The second experiment showed that response temporal profiles in NCM depends on whether variants were presented

in blocked or shuffled order at 3s ISI. In contrast, response temporal profiles in Field L2 were less sensitive to the order of stimulus presentation. Also, both results may be related processes of stimulus-specific adaptation. Learning transition patterns may be seen as a form of pattern-specific adaptation, which may be similar to the neural mechanism for the emergence of variant-independent representation. For example, NCM neurons have been shown to adapt to repeated pairs of non-contiguous sounds, independent of intervening sounds, and then show surprise effects when those non-contiguous patterns are violated (Lu & Vicario, 2014).

Both experiments showed that the auditory system could “ignore” the variabilities (in inter-syllable intervals or other parameters) and respond consistently to the “prototypical” features that identify each sound, even though the inter-syllable intervals were different. In the first experiment, the results showed neurons in NCM were sensitive to violations of alternation patterns independent of ISI (1s, 3s, or jittered 0.8 to 1.2s). In the second experiment, the inter-syllable intervals were much shorter (tens of milliseconds) and varied across variants. The results showed that the response temporal profiles for different variants were more similar to each other in NCM than in Field L2, suggesting that neural responses to a zebra finch song depended less on the syllable timing in NCM than in Field L2. This may help explain how juvenile zebra finches recognize and learn one zebra finch song despite the way its father may sing different variants of the same song (Slater, Eales & Clayton, 1988) (Bottjer,

Miesner & Arnold, 1984) (Funabiki & Konishi, 2003). Because the human auditory system must ignore the variabilities in the inter-syllable intervals to learn words and ignore the inter-word intervals to learn transition patterns between words (Saffran, Aslin, & Newport, 1996) (Aslin, 2017), further studies using the approach described here may help us understand the neural mechanisms of speech acquisition.

There are two possible follow-up experiments. One is to investigate whether neurons in NCM are sensitive to violations of transition patterns when stimuli are variable instead of fixed – this would combine the variants of Experiment 2 with the sequence manipulations of Experiment 1. If neurons in NCM can learn transition patterns when each sound is variable instead of being always identical, it will provide a more natural model for statistical learning of transition patterns during speech/language acquisition. The other possible follow-up experiment is to study the stimulus-independent encoding of deviant sounds. Most studies investigating deviance detection at the neural level have compared the neural responses when a stimulus is surprising versus when it is not. However, for neurons in the auditory system, both the identity and deviance of a sound is unknown. To detect a stimulus as deviant, neurons may need a representation of deviance that is independent of responses that are specific to individual acoustic stimuli. For example, when a neuron increases its responses to an unknown sound, it could be either due to the sound being deviant or being

played at higher intensity. These properties could be teased apart experimentally, modelled, and perhaps tested behaviorally.

Studies along these lines could provide a deeper understanding about how the preceding stimulus sequence and variations of the stimulus affect the neural responses to a sound and how the nervous system predicts the next sound (e.g., the most likely next word or syllable is based on what has just been heard). Thus, the same neural mechanisms may subserve both the perceptual invariance for acoustically variable sounds that is critical for speech acquisition (Kuhl, 1994) and the statistical learning of transition patterns that facilities rapid speech processing (Kuhl, 1994) (Aslin, 2017).

REFERENCES

Alves-Pinto, A., Baudoux, S., Palmer, A. R., & Sumner, C. J. (2010). Forward masking estimated by signal detection theory analysis of neuronal responses in primary auditory cortex. *Journal of the Association for Research in Otolaryngology*, *11*(3), 477–494.

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398.

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1–2), e1373.

Beckers, G. J. L., & Gahr, M. (2012). Large-scale synchronized activity during vocal deviance detection in the zebra finch auditory forebrain. *The Journal of Neuroscience*, *32*(31), 10594–10608.

Bendor, D. (2015). The role of inhibition in a computational model of an auditory cortical neuron during the encoding of temporal information. *PLoS computational biology*, *11*(4), e1004197.

Bennett, D., Murawski, C., & Bode, S. (2015). Single-Trial Event-Related Potential Correlates of Belief Updating. *ENeuro*, *2*(5). https://doi.org/10.1523/ENEURO.0076-15.2015

Billimoria, C. P., Kraus, B. J., Narayan, R., Maddox, R. K., & Sen, K. (2008). Invariance and sensitivity to intensity in neural discrimination of natural sounds. *Journal of Neuroscience*, *28*(25), 6304–6308.

Blackwell, J. M., Taillefumier, T. O., Natan, R. G., Carruthers, I. M., Magnasco, M. O., & Geffen, M. N. (2016). Stable encoding of sounds over a broad range of statistical parameters in the auditory cortex. *European Journal of Neuroscience*, *43*(6), 751–764.

Boari, S., Perl, Y. S., Amador, A., Margoliash, D., & Mindlin, G. B. (2015). Automatic reconstruction of physiological gestures used in a model of birdsong production. *Journal of Neurophysiology*, *114*(5), 2912–2922.

Bolhuis, J. J., & Gahr, M. (2006). Neural mechanisms of birdsong memory. *Nature Reviews Neuroscience*, *7*(5), nrn1904.

Bottjer, S. W., Miesner, E. A., & Arnold, A. P. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*, *224*(4651), 901-903.

Brainard, M. S., & Doupe, A. J. (2013). Translating birdsong: songbirds as a model for basic and applied medical research. *Annual Review of Neuroscience*, *36*, 489–517.

Buesing, L., Bill, J., Nessler, B., & Maass, W. (2011). Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS Computational Biology*, *7*(11), e1002211.

Buonomano, D. V., Bramen, J., & Khodadadifar, M. (2009). Influence of the interstimulus interval on temporal processing and learning: testing the state-dependent network model. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1865-1873.

Buonomano, D. V., & Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, *10*(2), 113.

Calabrese, A., & Woolley, S. M. (2015). Coding principles of the canonical cortical microcircuit in the avian brain. *Proceedings of the National Academy of Sciences*, *112*(11), 3517-3522.

Carruthers, I. M., Laplagne, D. A., Jaegle, A., Briguglio, J. J., Mwilambwe-Tshilobo, L., Natan, R. G., & Geffen, M. N. (2015). Emergence of invariant representation of vocalizations in the auditory cortex. *Journal of Neurophysiology*, *114*(5), 2726–2740.

Cate, C. ten, & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: Natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1598), 1984–1994.

Chaure, F., Rey, H. G., & Quiroga, R. Q. (2018). A novel and fully automatic spike sorting implementation with variable number of features. *Journal of Neurophysiology*.

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibánez, A. M., Agust\in and Owen, & Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *Journal of Neuroscience*, *33*(27), 11194–11205.

Chew, S. J., Mello, C., Nottebohm, F., Jarvis, E., & Vicario, D. S. (1995). Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proceedings of the National Academy of Sciences*, *92*(8), 3406–3410.

Compte, A., Sanchez-Vives, M. V., McCormick, D. A., & Wang, X. J. (2003). Cellular and network mechanisms of slow oscillatory activity (< 1 Hz) in a cortical network model. *Journal of neurophysiology*.

Cornella, M., Leung, S., Grimm, S., & Escera, C. (2012). Detection of simple and pattern regularity violations occurs at different levels of the auditory hierarchy. *PLoS One*, *7*(8), e43604.

Daikoku, T. (2018). Neurophysiological markers of statistical learning in music and language: Hierarchy, entropy and uncertainty. *Brain Sciences*, *8*(6), 114.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2–19.

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences*, *112*(45), E6233–E6242.

Dong, M., & Vicario, D. S. (2018). Neural correlate of transition violation and deviance detection in the songbird auditory forebrain. *Frontiers in Systems Neuroscience*, *12*, 46.

Elie, J. E., Mariette, M. M., Soula, H. A., Griffith, S. C., Mathevon, N., & Vignal, C. (2010). Vocal communication at the nest between mates in wild zebra finches: A private vocal duet? *Animal Behaviour*, *80*(4), 597–605.

Elie, J. E., & Theunissen, F. E. (2015). Meaning in the avian auditory cortex: Neural representation of communication calls. *European Journal of Neuroscience*, *41*(5), 546–567.

Elie, J. E., & Theunissen, F. E. (2018). Zebra finches identify individuals using vocal signatures unique to each call type. *Nature Communications*, *9*(1), 4026.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Funabiki, Y., & Konishi, M. (2003). Long memory in song learning by zebra finches. *Journal of Neuroscience*, *23*(17), 6928-6935.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204.

Gentner, T. Q., & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature*, *424*(6949), 669.

Glaze, C. M., & Troyer, T. W. (2006). Temporal structure in zebra finch song: Implications for motor coding. *Journal of Neuroscience*, *26*(3), 991–1005.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7*(515–546).

Hershenhoren, I., Taaseh, N., Antunes, F. M., & Nelken, I. (2014). Intracellular correlates of stimulus-specific adaptation. *The Journal of Neuroscience*, *34*(9), 3303–3319.

Hoyer, P. O., & Hyvärinen, A. (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior. In *Advances in neural information processing systems* (pp. 293–300).

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., ... & Striedter, G. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*(2), 151.

Johnson, K. (2008). Speaker Normalization in Speech Perception. *The handbook of speech perception*, 363.

Keitel, C., Quigley, C., & Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: Entrainment of intrinsic rhythms or frequency-following response? *Journal of Neuroscience*, *34*(31), 10137–10140.

Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A. E., Nourski, K. V., Gander, P. E., … others. (2017). Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biology*, *15*(4), e2000219.

Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A. E., Nourski, K. V., Gander, P. E., … Petkov, C. I. (2017). Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLOS Biology*, *15*(4), e2000219. https://doi.org/10.1371/journal.pbio.2000219

Kim, G., & Doupe, A. (2011). Organized representation of spectrotemporal features in songbird auditory forebrain. *Journal of Neuroscience*, *31*(47), 16977-16990.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current opinion in neurobiology*, *4*(6), 812-822.

Lim, Y., Lagoy, R., Shinn-Cunningham, B. G., & Gardner, T. J. (2016). Transformation of temporal sequences in the zebra finch auditory system. *Elife*, *5*, e18205.

Lu, K., & Vicario, D. S. (2014). Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proceedings of the National Academy of Sciences*, *111*(40), 14553–14558. https://doi.org/10.1073/pnas.1412109111

Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annu. Rev. Neurosci.*, *27*, 307–340.

Milne, A. E., Petkov, C. I., & Wilson, B. (2017). Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience*.

Mittag, M., Takegata, R., & Winkler, I. (2016). Transitional probabilities are prioritized over stimulus/pattern probabilities in auditory deviance detection: Memory basis for predictive sound processing. *Journal of Neuroscience*, *36*(37), 9572–9579.

Mooney, R. (2009). Neural mechanisms for learned birdsong. *Learning & Memory*, *16*(11), 655–669.

Moore, R. C., Lee, T., & Theunissen, F. E. (2013). Noise-invariant neurons in the avian auditory cortex: Hearing the song in noise. *PLoS Computational Biology*, *9*(3), e1002942.

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988). Event-Related Potentials to Breaks in Sequences of Alternating Pitches or Interstimulus Intervals. *Psychophysiology*, *25*(3), 262–268.

Ohms, V. R., Gill, A., Van Heijningen, C. A., Beckers, G. J., & Cate, C. ten. (2009). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20091788.

Ono, S., Okanoya, K., & Seki, Y. (2016). Hierarchical emergence of sequence sensitivity in the songbird auditory forebrain. *Journal of Comparative Physiology A*, *202*(3), 163–183.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Phan, M. L., Pytte, C. L., & Vicario, D. S. (2006). Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(4), 1088–1093.

Pinaud, R., Terleph, T. A., Tremere, L. A., Phan, M. L., Dagostin, A. A., Leao, R. M., ... & Vicario, D. S. (2008). Inhibitory network interactions shape the auditory processing of natural communication signals in the songbird auditory forebrain. *Journal of neurophysiology*.

Quiroga, R. Q., Nadasdy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, *16*(8), 1661–1687.

Rabinowitz, N. C., Willmore, B. D., King, A. J., & Schnupp, J. W. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS biology*, *11*(11), e1001710.

Rossum, M. van. (2001). A novel spike distance. *Neural Computation*, *13*(4), 751–763.

Sadagopan, S., & Wang, X. (2008). Level invariant representation of sounds by populations of neurons in primary auditory cortex. *Journal of Neuroscience*, *28*(13), 3415–3426.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Sainburg, T., Thielk, M., Theilman, B., Migliori, B., & Gentner, T. (2018). Generative adversarial interpolative autoencoding: Adversarial training on latent space interpolations encourage convex latent distributions. *ArXiv Preprint ArXiv:1807.06650*.

Sameiro-Barbosa, C. M., & Geiser, E. (2016). Sensory entrainment mechanisms in auditory perception: Neural synchronization cortico-striatal activation. *Frontiers in Neuroscience*, *10*, 361.

Sanchez-Vives, M. V., & McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature neuroscience*, *3*(10), 1027.

Schneider, D. M., & Woolley, S. M. N. (2013). Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron*, *79*(1), 141–152.

Shaevitz, S. S., & Theunissen, F. E. (2007). Functional connectivity between auditory areas field L and CLM and song system nucleus HVC in anesthetized zebra finches. *Journal of neurophysiology*.

Sharpee, T. O., Atencio, C. A., & Schreiner, C. E. (2011). Hierarchical representations in the auditory cortex. *Current Opinion in Neurobiology*, *21*(5), 761–767.

Slater, P. J., Eales, L. A., & Clayton, N. S. (1988). Song learning in zebra finches (Taeniopygia guttata): progress and prospects. In *Advances in the Study of Behavior* (Vol. 18, pp. 1-34). Academic Press.

Soyman, E. (2018). *The effects of passive familiarization on neural and behavioral discrimination of acoustic signals* (PhD thesis). Rutgers University-School of Graduate Studies.

Spaak, E., Lange, F. P. de, & Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *Journal of Neuroscience*, *34*(10), 3536–3544.

Steriade, M., McCormick, D. A., & Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, *262*(5134), 679–685.

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour*, *59*(6), 1167–1176.

Terleph, T. A., Mello, C. V., & Vicario, D. S. (2006). Auditory topography and temporal response dynamics of canary caudal telencephalon. *Journal of neurobiology*, *66*(3), 281-292.

Turk-Browne, N. B. (2012). Statistical learning and its consequences. In *The influence of attention, learning, and motivation on visual search* (pp. 117–146). Springer.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398. https://doi.org/10.1038/nn1032

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, *32*(11), 3665–3678.

Wacongne, C., Labyt, E., Wassenhove, V. van, Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, *108*(51), 20754–20759.

Wang, Y., Brzozowska-Prechtl, A., & Karten, H. J. (2010). Laminar and columnar auditory cortex in avian brain. *Proceedings of the National Academy of Sciences*, *107*(28), 12676-12681.

Weatherholtz, K., & Jaeger, T. F. (2016). Speech perception and generalization across talkers and accents. In *Oxford Research Encyclopedia of Linguistics*.

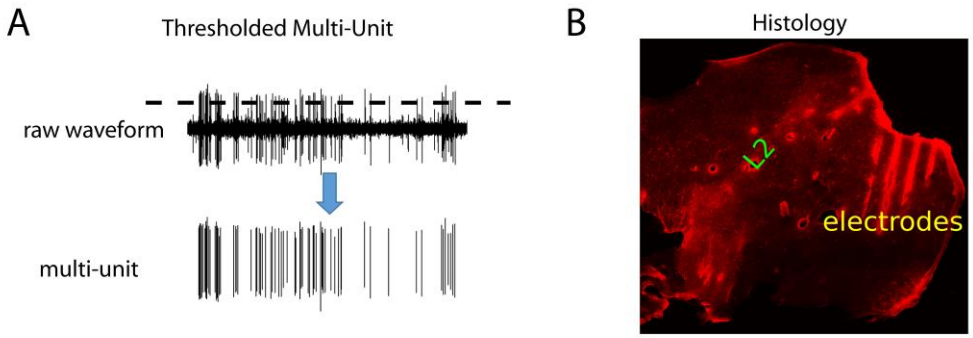
Wehr, M., & Zador, A. M. (2005). Synaptic mechanisms of forward suppression in rat auditory cortex. *Neuron*, *47*(3), 437–445.

Woolley, S. M., Fremouw, T. E., Hsu, A., & Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience*, *8*(10), 1371.

Yaron, A., Hershenhoren, I., & Nelken, I. (2012). Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron*, *76*(3), 603–615.

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

FIGURES



	pitch goodness	pitch	FM	entropy
zebra finch	300.63 ± 181.39	1.15 ± 0.97	31.70 ± 10.97	− 3.60 ± 0.68
canary	55.12 ± 30.79	3.31 ± 0.82	17.92 ± 11.30	− 5.94 ± 0.48

Figure 1 Thresholded multi-unit, histology, and syllable stimuli. (A) Multi-unit activity (MUA) was obtained by thresholding the raw waveforms (3 standard deviation above the mean). (B) Histological verification of recording sites. Roughly equal number of sites were recorded from NCM and Field L2. (C) Example spectrograms of one zebra finch syllable and one canary syllable. (D) Major acoustic differences between zebra finch and canary syllables, measured using Sound Analysis Pro(Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra, 2000).

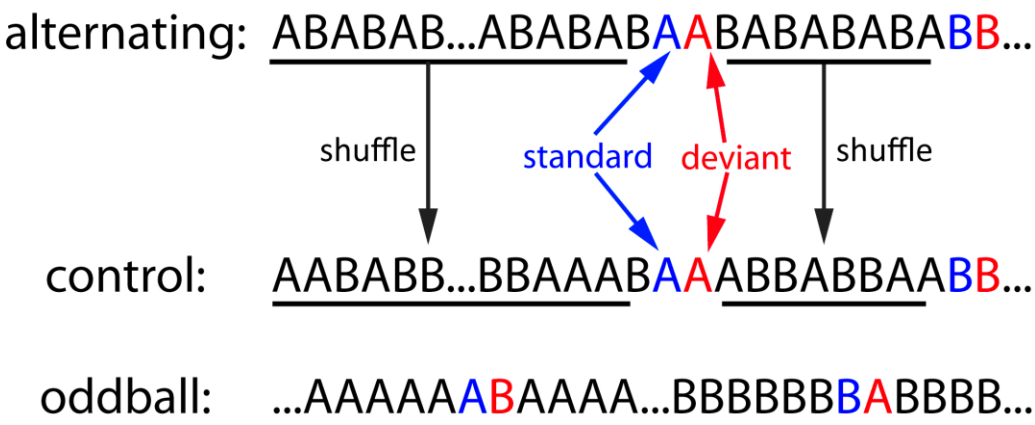


Figure 2 Alternating oddball paradigm and the stimulus sequences used in the control, alternating, and oddball conditions. In the alternating condition, two stimuli were initially presented in alternation for 25 times to familiarize the bird with the stimuli and the alternation pattern. Then, rare repetitions were presented after a variable 4–10 regular alternations. The deviant (2nd stimulus in the repetition), standard (1st stimulus in the repetition), and the stimulus immediately before them formed a “triplet.” In the control condition, the number of stimulus trials and the positions of the triplets were the same as those in the alternating condition, however, stimulus sequences between the triplets were shuffled in the control. In the oddball condition, two stimuli were presented in two blocks with different probabilities. For notation purposes, when a stimulus is presented with low probability, it is called the deviant and the stimulus immediately preceding it is called the standard. Deviant and standard are color-coded with red and blue, respectively.

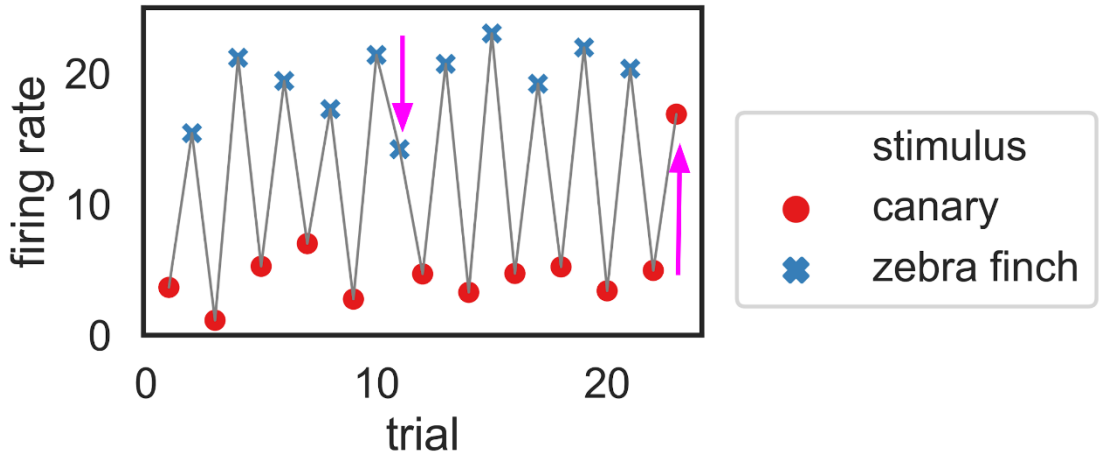


Figure 3 Schematic illustration of neural oscillation under conditions where two stimuli are alternating. A recording site is oscillating if its responses are like the illustration above, where the oscillation magnitudes of both stimuli are significantly larger than 0.

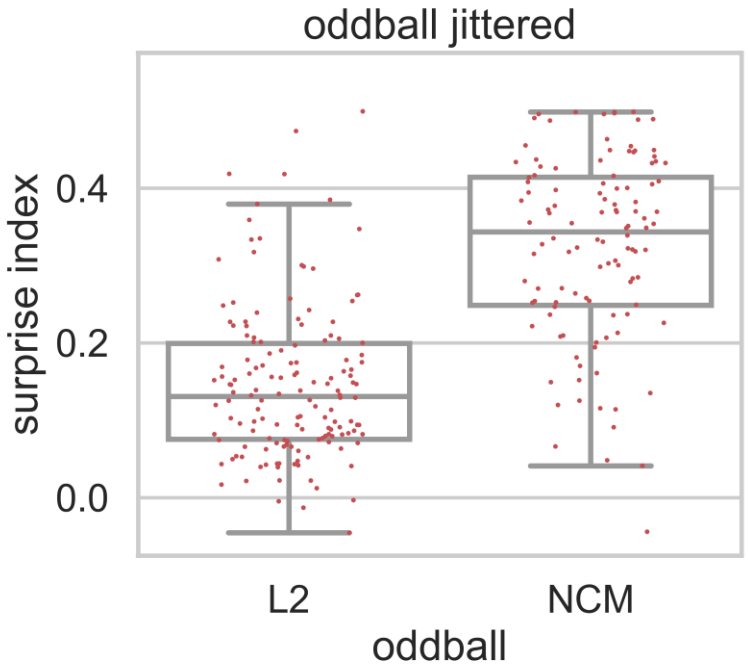


Figure 4 Surprise index (SI) in the oddball condition based on MUA. The SI was significantly larger than 0 in both L2 and NCM. The SI was significantly larger in NCM than in L2. Each dot represents SI from one recording site. The box shows the quartiles of the dataset while the whiskers extend to the rest of the distribution, except for the potential outliers. Most dots are above 0 showing that SI is significantly larger than 0.

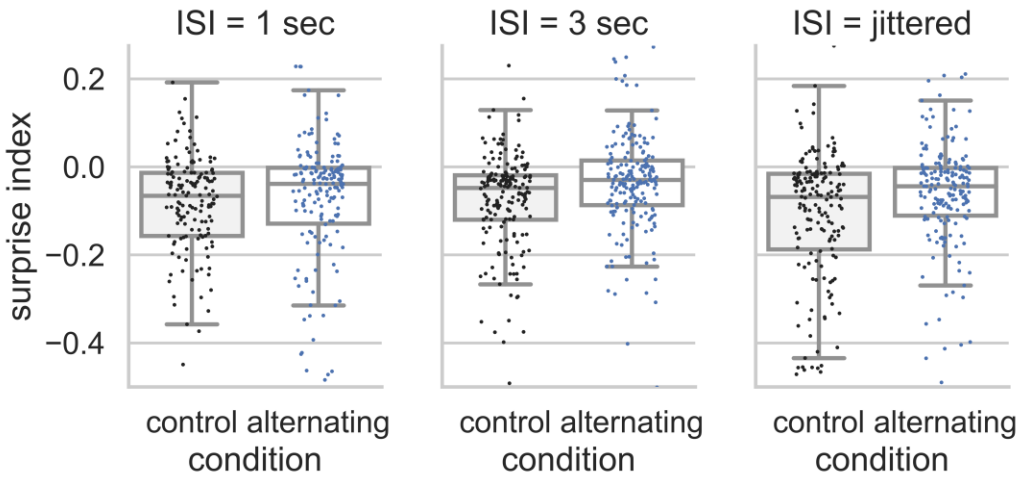


Figure 5 SI in the control and alternating conditions in NCM based on MUA. Under all ISIs, SIs in the control conditions were significantly smaller than 0, and SIs in the alternating condition were significantly larger than those in the control condition.

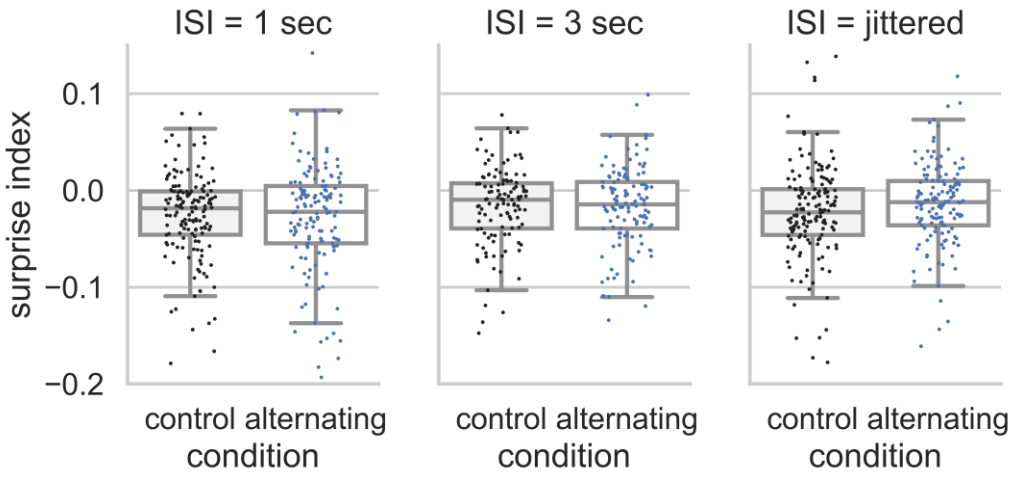


Figure 6 SI in the control and alternating conditions in L2 based on MUA. Under all tested ISIs (1s or 3s), SI in the alternating conditions were not significantly different from those in the control condition, and SI was significantly smaller than 0. Note that under jittered (0.8 - 1.2s) ISI, the results showed a trend that SI was larger in the alternating than in the control condition but it was not significant.

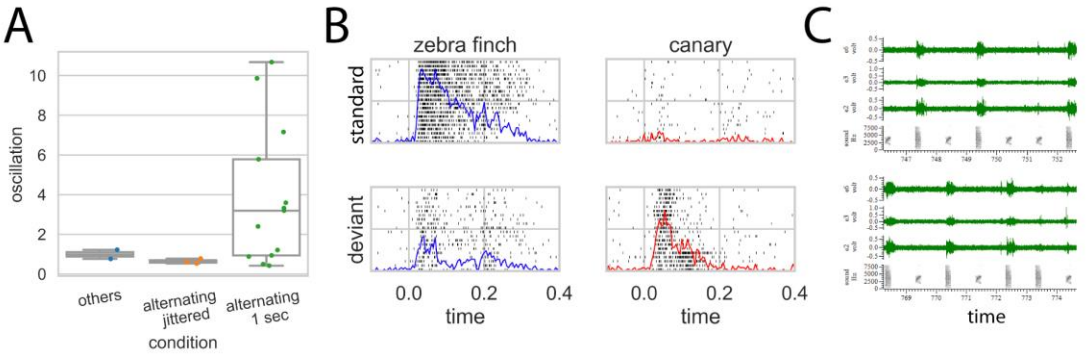


Figure 7 Oscillatory magnitude, peri-stimulus time histogram (PSTH), and raw waveforms from three examplar oscillatory sites. (A) In the alternating condition at 1s fixed ISI, 13 out of 240 (5%) recording sites showed oscillatory responses. In contrast, there were significantly fewer oscillatory sites in the alternating condition at jittered ISI or any other conditions combined. The oscillation magnitude was also significantly larger in the alternating 1s condition than in any other conditions. (B) Raster plot and PSTH from one example recording site. The top left shows the raster plot and PSTH when a zebra finch syllable is a standard; the bottom left shows the raster plot and PSTH when the same zebra finch syllable is deviant. The figures on the right show the raster plot and PSTH when a canary syllable is standard (top) and deviant (bottom). (C) Example raw recording waveforms when a zebra finch or canary syllable was repeated after alternation.

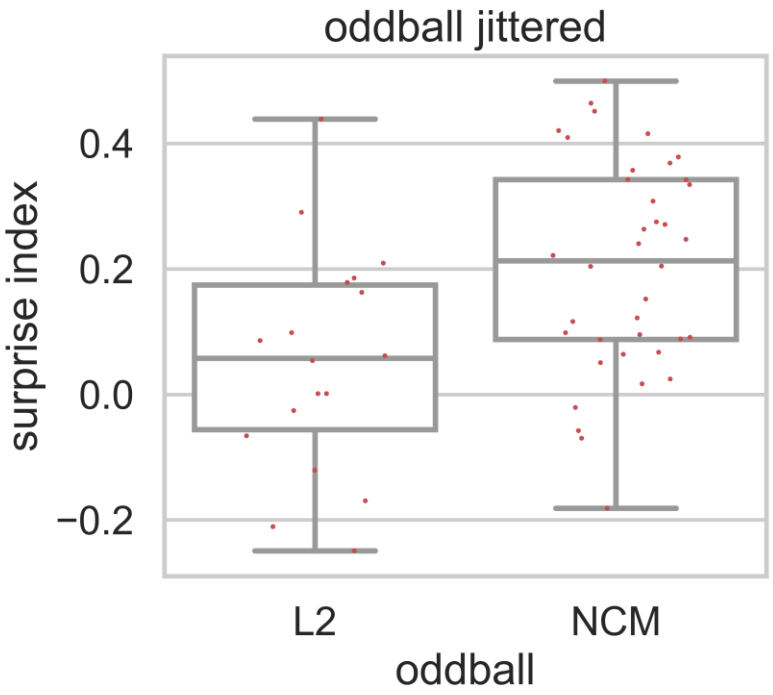


Figure 8 SI in the oddball condition based on SUA. The SI was significantly larger than 0 in NCM but not in L2. The SI was significantly larger in NCM than in L2. Most dots in NCM are above 0, showing that SI is significantly larger than 0.

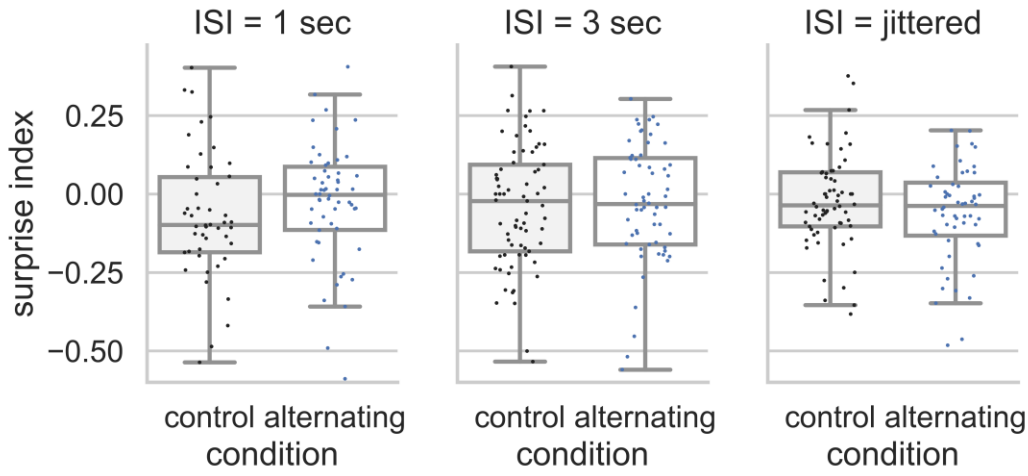


Figure 9 SI in the control and alternating conditions in NCM based on SUA. SI in the alternating condition was significantly larger than that in the control condition when ISI was 1s but not when ISI was 3s or jittered.

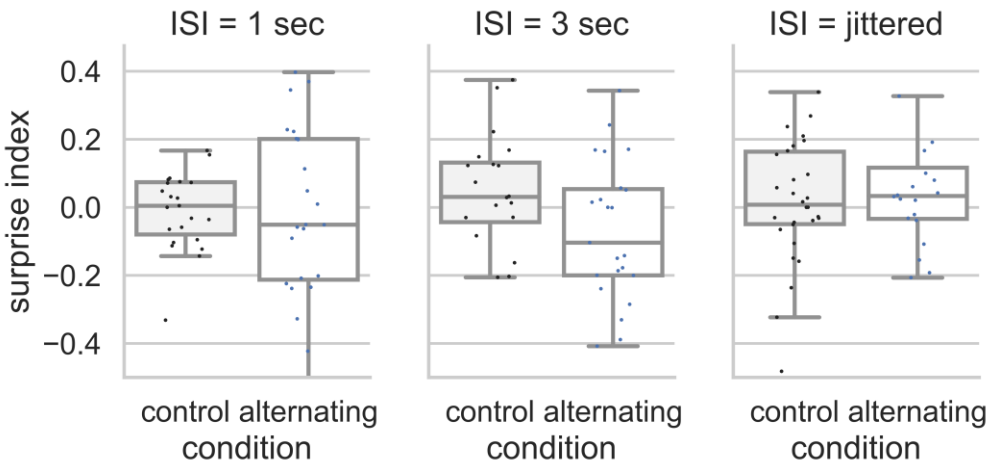


Figure 10 SI in the control and alternating conditions in L2 based on SUA. Under all tested ISIs (1s or 3s), SI in the alternating conditions were not significantly different from those in the control condition.

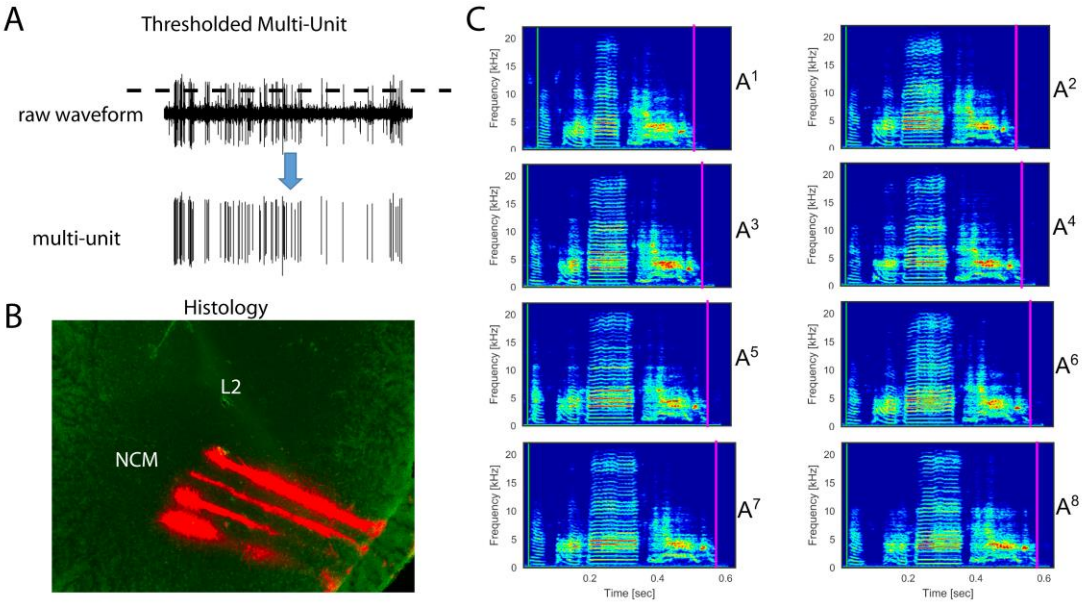


Figure 11 Thresholded multi-unit, histology, and variant stimuli. (A) Multi-unit activity (MUA) was obtained by thresholding the raw waveforms (3 standard deviation above the mean). (B) Histological verification of recording sites. Roughly equal number of sites were recorded from NCM and Field L2. (C) Example spectrograms of 8 variants of one zebra finch song motif.

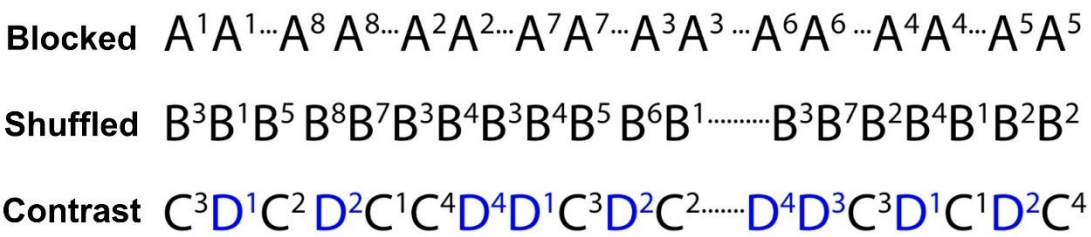


Figure 12 Blocked, shuffled, and contrast condition in the variant experiment. In the blocked condition, 8 variants of a zebra finch song (produced by one individual) were presented 40 repetitions each. In the shuffled condition, 8 variants of another zebra finch song were presented 40 repetitions each but in a shuffled order. In the contrast condition, 4 variants of each of 2 different zebra finch songs were presented in a random order.

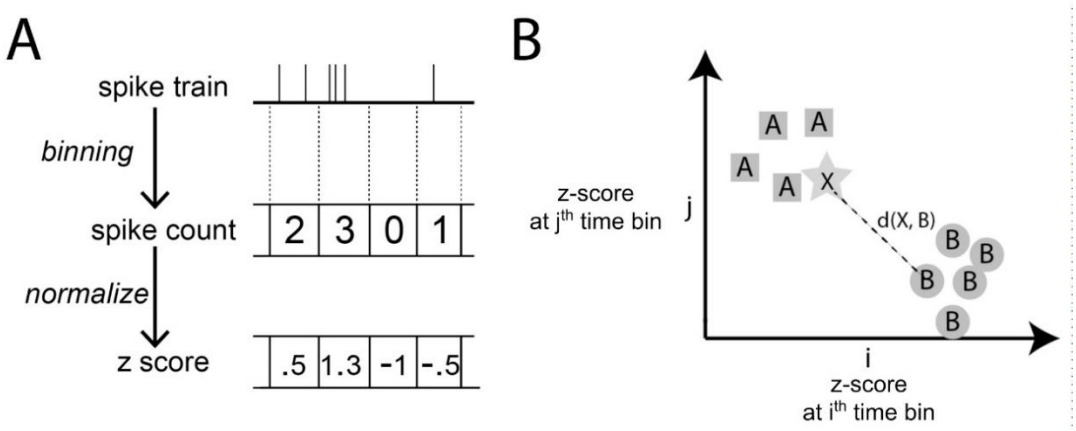


Figure 13 Characterization of response temporal profiles and neural decoding. (A) Normalization of spike trains to z-score. (B) The identity of neural responses X id predicted as stimulus whose average neural dissimilarity to X is smallest. In this illustration, neural responses X will be predicted as coming from stimulus A. Decoding accuracy is the probability that neural response X is correctly predicted.

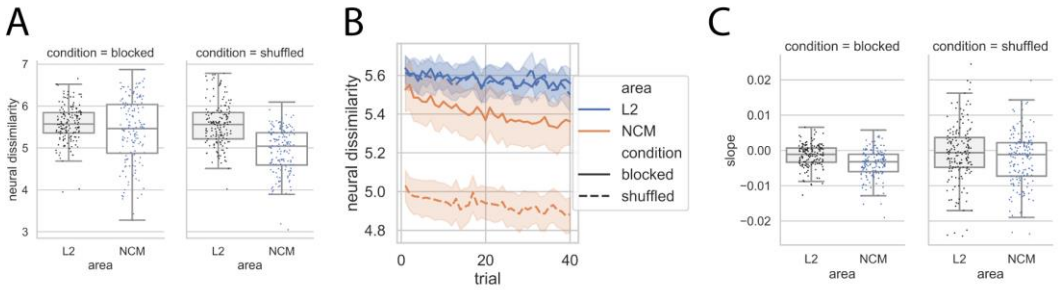


Figure 14 Neural dissimilarity in the blocked and shuffled condition based on MUA. (A) Neural dissimilarity in L2 & NCM. The neural dissimilarity depended both on brain areas and the order of stimulus presentation. In both blocked and shuffled condition, the neural dissimilarity was significantly smaller in NCM than in Field L2. In NCM, neural dissimilarity was significantly smaller in the shuffled condition than in the blocked condition. In contrast, neural dissimilarities in the blocked and shuffled condition were not different in Field L2. (B) Change of neural dissimilarity across trials. In both NCM and Field L2, the average neural dissimilarity decreased across trials. The decrease was significantly faster in NCM than in L2. The shadowed areas show the 95% confidence interval based on the standard deviation. (C) For each site, a linear regression of trial (6 to 40) on neural dissimilarity was fitted. There were more sites showing negative slopes in NCM than in L2 and the average slope was also significantly more negative in NCM than in L2. These results suggest that response temporal profiles for different variants became more similar to each other in NCM but not as much in L2.

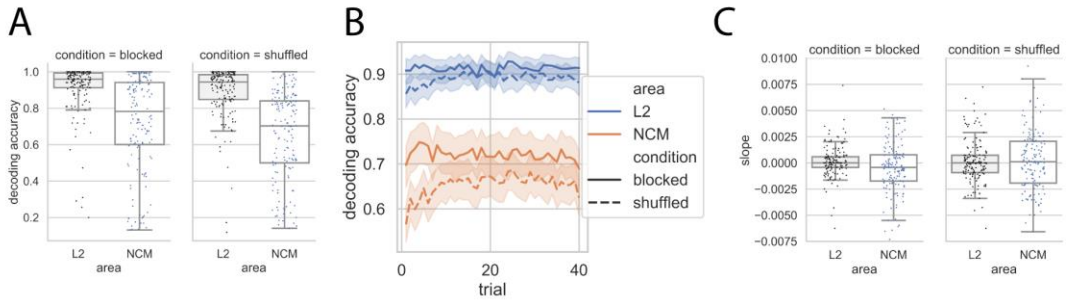


Figure 15 Decoding accuracy in the blocked and shuffled condition based on MUA. (A) Decoding accuracy based on the response temporal profile in NCM and Field L2. Overall, the decoding accuracy was significantly lower in NCM than in L2. In both NCM and L2, the decoding accuracy was lower in the shuffled condition than in the blocked condition. (B) Decoding accuracy across trials. The decoding accuracy in L2 did not change across trials. In NCM, the decoding accuracy changed differently depending on the order of variant presentation. In the shuffled condition, the decoding accuracy increased during the first 10 trials and then plateaued. In the blocked condition, the decoding accuracy increased in the first 6 trials and then slowly decreased from trial 6 to trial 40. (C) For each site, a linear regression of trial (6 to 40) on decoding accuracy was fitted. In NCM, the negative slope indicated that the decoding accuracy decreased from trial 6 to trial 40 in the blocked but not shuffled condition. In L2, the slope of decoding accuracy from trial 6 to trial 40 was not different from 0.

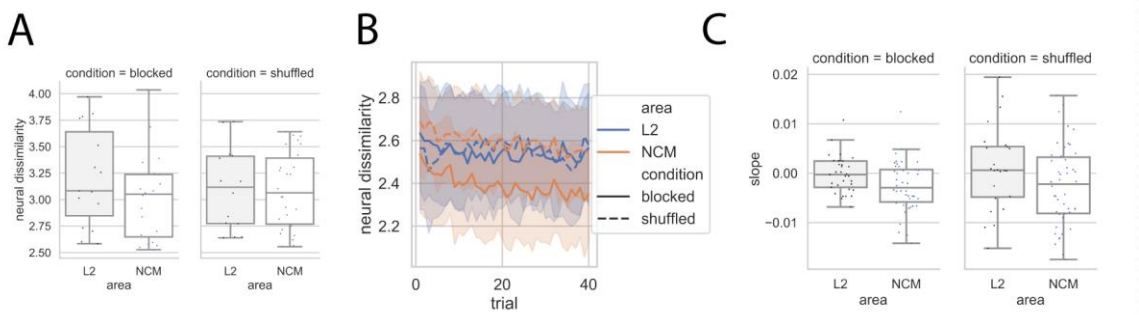


Figure 16 Neural dissimilarity in the blocked and shuffled condition based on SUA. (A) In both blocked and shuffled condition, the neural dissimilarity was smaller in NCM than in Field L2. (B) In NCM, the average neural dissimilarity decreased from across trials. In L2, the average neural dissimilarity was relatively stable across trials. (C) Regression slope of neural dissimilarity from trial 6 to trial 40. In both blocked and shuffled condition, the slope was more negative in NCM than in L2, suggesting that response temporal profiles for different variants became more similar to each other in NCM than in L2.

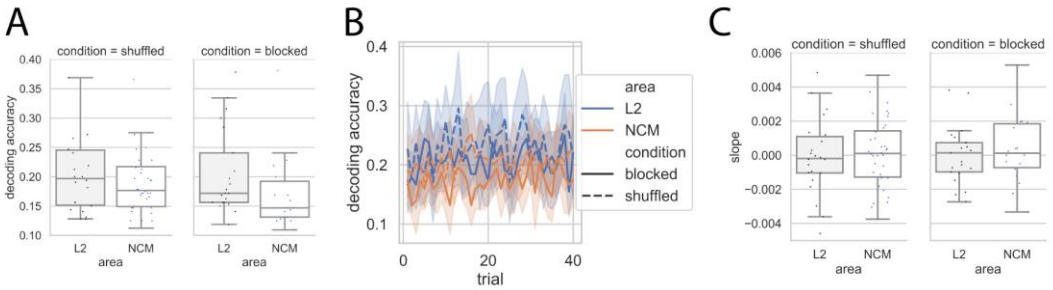


Figure 17 Decoding accuracy in the blocked and shuffled condition based on SUA. (A) Decoding accuracy based on the response temporal profile in NCM and Field L2. Overall, the decoding accuracy was lower in NCM than in L2. (B) The decoding accuracy for SUA was very variable and did not change across trials in either L2 or NCM. (C) The regression slope from trial 6 to trial 40 was not different from 0 in either NCM/L2 in or blocked/shuffled condition.

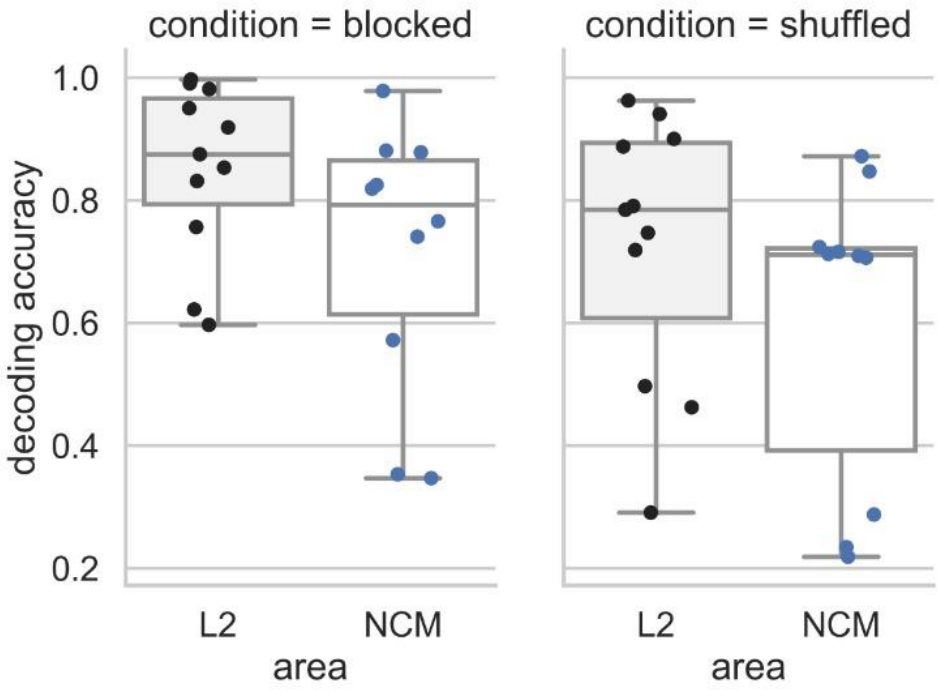


Figure 18 Decoding accuracy based on the population responses from MUA. The decoding accuracy was lower in NCM than in Field L2. Within each bird, recording sites are separated into L2 (black) and NCM (blue) sites based on histological reconstruction. The population neural response magnitude (firing rate; see methods) within each area were used to decode the identity of song variants. Lower decoding accuracy indicates that population responses to variants were more difficult to distinguish and thus more similar to each other. Each dot represents the decoding accuracy from one bird in a given area and condition.

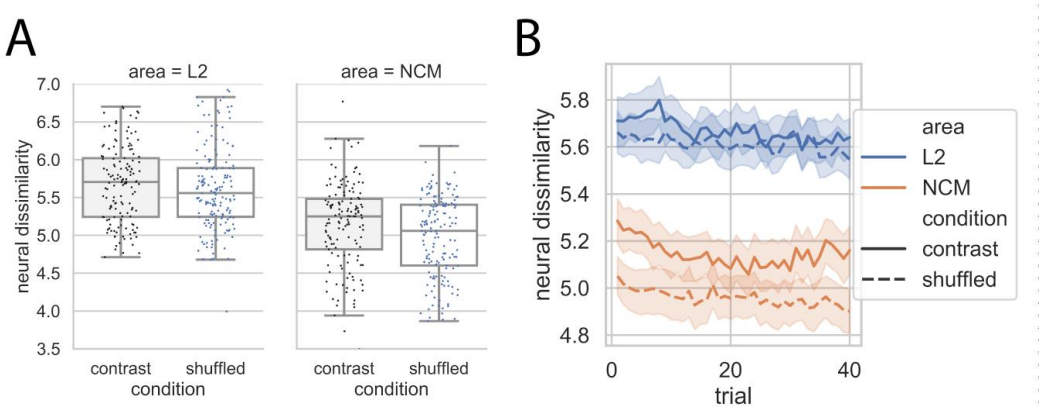


Figure 19 Neural dissimilarity for variants within a song in the contrast and shuffled condition based on MUA. (A) The neural dissimilarity for variants within a song was similar in the contrast and shuffled conditions. In both L2 and NCM, the neural dissimilarity within a song was not lower in the contrast than in the shuffled condition. However, the neural dissimilarity was lower in NCM than in L2. (B) The neural dissimilarity did not change across trials in L2. In NCM, the neural dissimilarity decreased from trial 6 to trial 40 in the contrast condition.

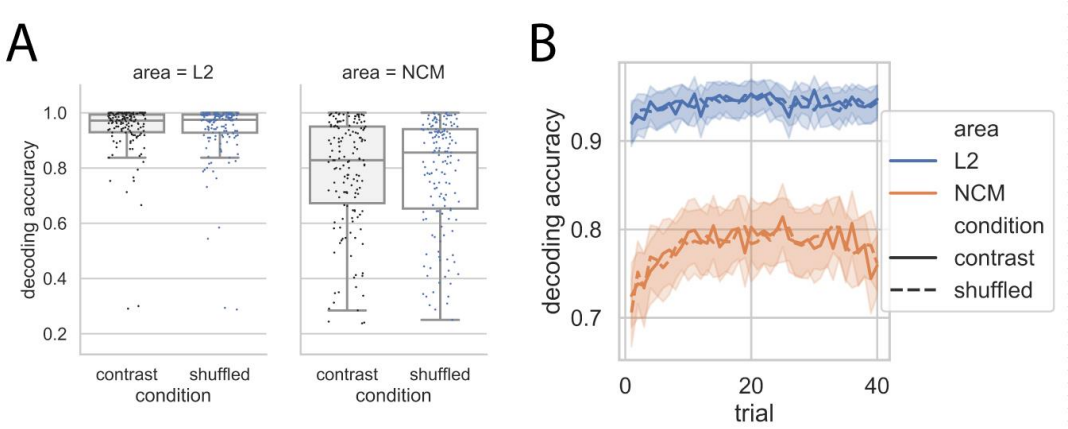


Figure 20 Decoding accuracy for variants within a song in the contrast and shuffled condition based on MUA. (A) In both L2 and NCM, the decoding accuracy for variants within a song was similar in the contrast and shuffled condition. The decoding accuracy was significantly lower in NCM than in L2. (B) The decoding accuracy changed in a similar fashion in the contrast and shuffled condition. In L2, the decoding accuracy did not change across trials. In NCM, the decoding accuracy increased during the first few trials in both conditions.

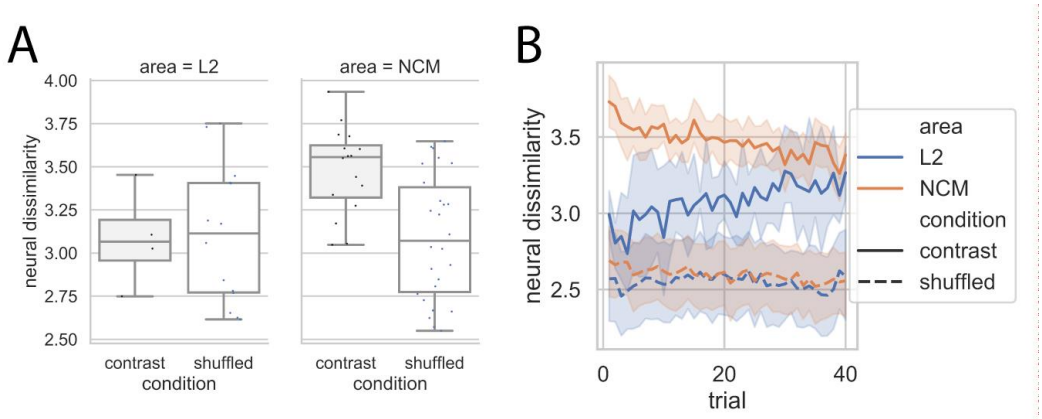


Figure 21 Neural dissimilarity for variants within a song based on SUA in the contrast and shuffled condition. (A) In L2, the neural dissimilarity was not different in the contrast and shuffled conditions. In NCM, the neural dissimilarity was not smaller in the contrast than in the shuffled condition. (B) The neural dissimilarity stayed stable across trials in the shuffled condition in both L2 and NCM. In the contrast condition, the neural dissimilarity decreased in NCM whereas it increased in L2. This unexpected pattern may be due to the small number of isolated.

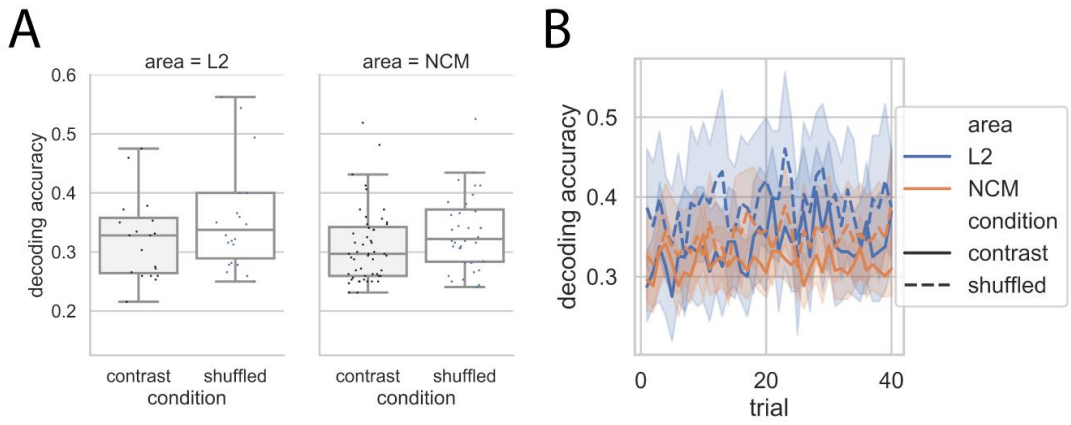


Figure 22 Decoding accuracy for variants within a song in the contrast and shuffled condition based on SUA. (A) In both L2 and NCM, the decoding accuracy for variants within each song was not different in the contrast and shuffled condition. (B) The decoding accuracy was very variable and did not show a consistent change across trials.

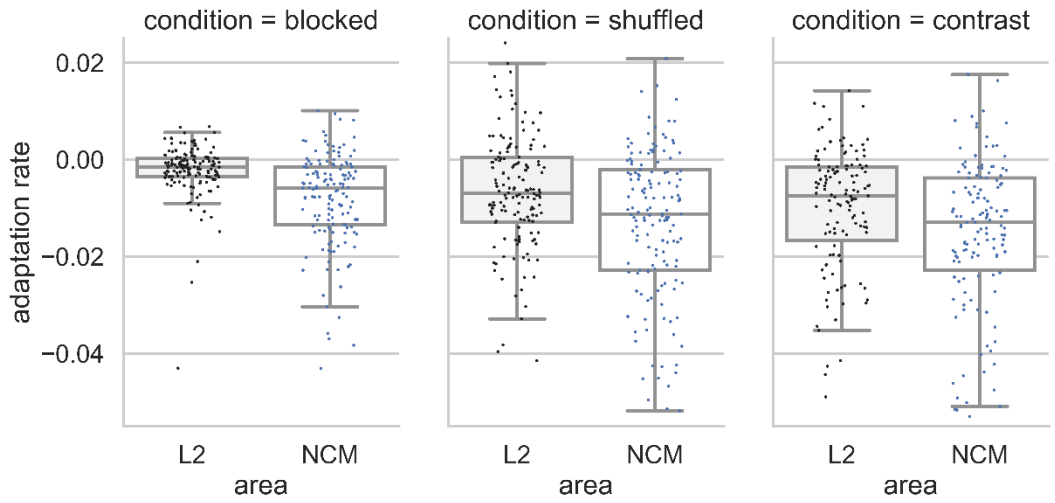


Figure 23 The average adaptation rate in L2 and NCM from trial 6 to trial 25 based on MUA. The adaptation rate was significantly more negative in NCM than in L2 in all three conditions.

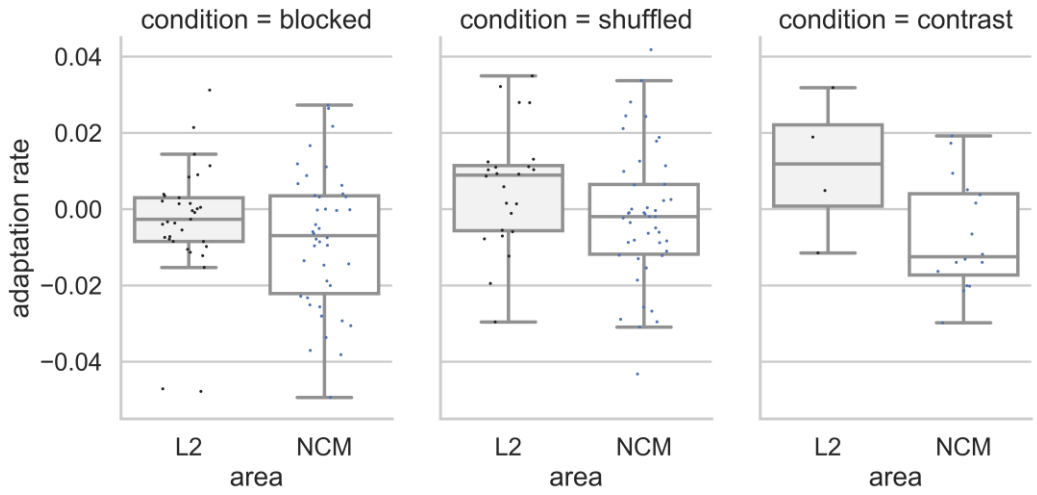


Figure 24 The average adaptation rate in L2 and NCM from trial 6 to trial 25 based on SUA. The adaptation rate was more negative in NCM than in L2 in all three conditions.

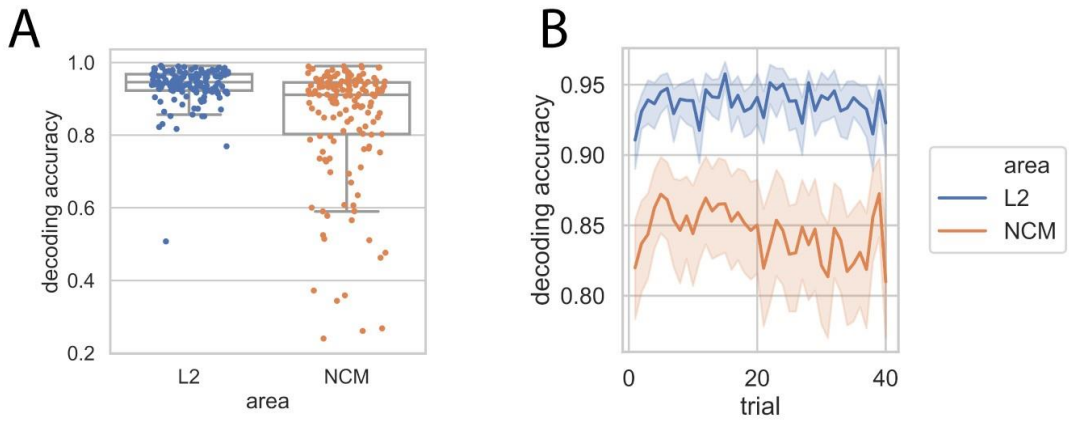


Figure 25 Decoding accuracy for different songs based on MUA in the contrast condition. (A) The average decoding accuracy in NCM was significantly lower than in L2. (B) In both L2 and NCM, the decoding accuracy increased during the first few trials. In L2, the decoding accuracy did not change much from trial 6 to trial 40. In contrast, the decoding accuracy in NCM decreased from trial 6 to trial 40.

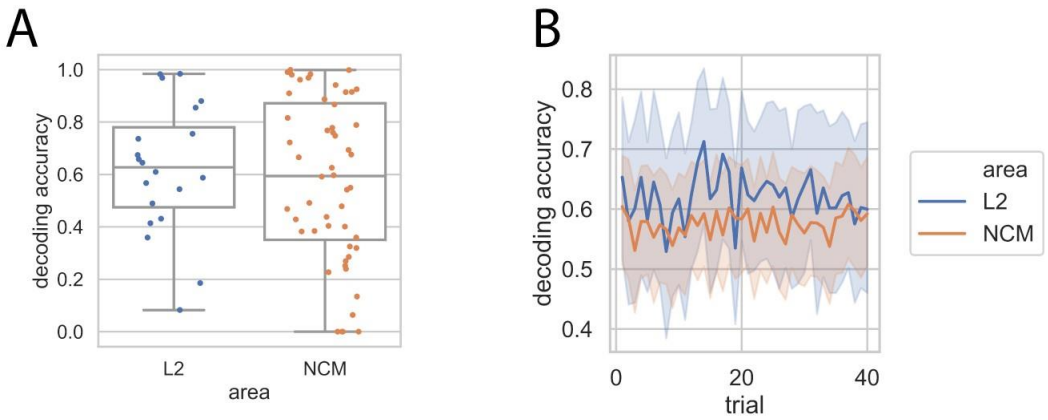


Figure 26 Decoding accuracy for different songs based on SUA in the contrast condition. (A) The average decoding accuracy for different songs (instead of variants within each song) in the contrast condition was similar in L2 and in NCM. (B) The decoding accuracy stayed stable across trials.

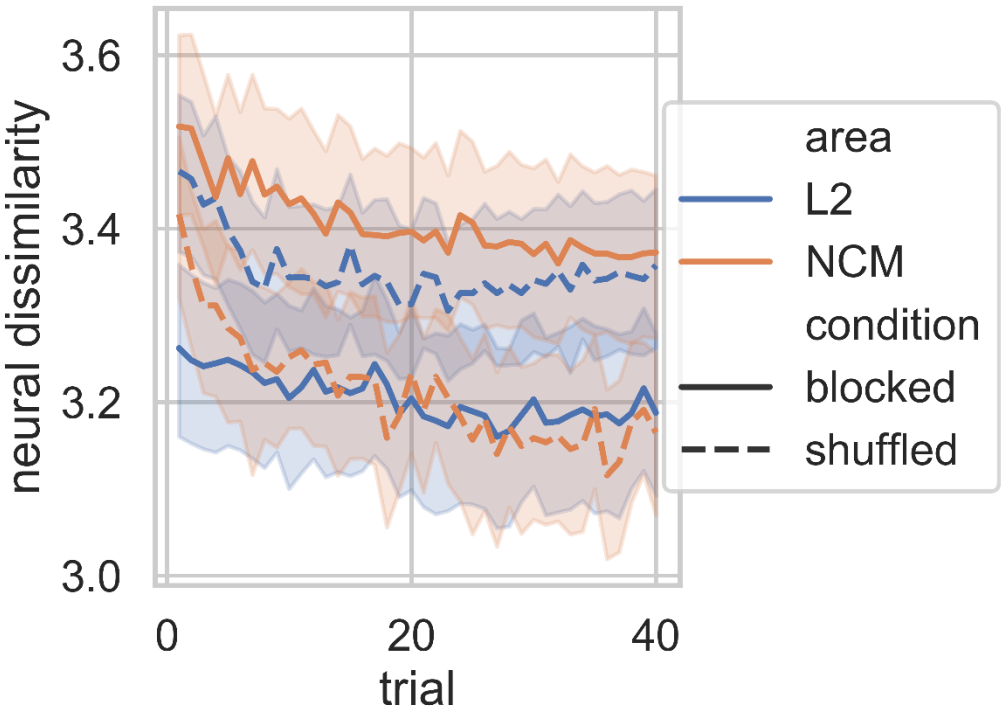


Figure 27 Change of neural dissimilarity across trials within each variant based on MUA. In both NCM and L2, neural dissimilarity decreased across trials.

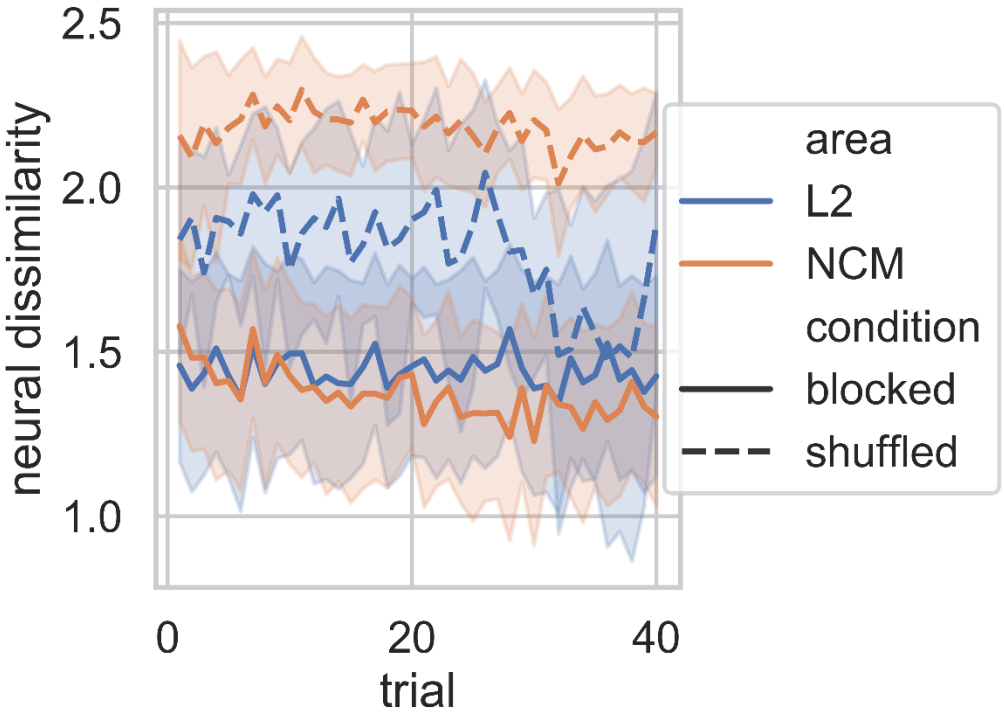


Figure 28 Change of neural dissimilarity across trials within each variant based on SUA. In both NCM and L2, the neural dissimilarity stayed stable across trials.