

©2016

MINGWEN DONG

ALL RIGHTS RESERVED

CONTEXT-DEPENDENT ENCODING IN SONGBIRD AUDITORY FOREBRAIN

By

MINGWEN DONG

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

David S. Vicario, PhD.

And approved by

New Brunswick, New Jersey

May 2016

ABSTRACT OF THE THESIS

Context-Dependent Encoding in Songbird Auditory Forebrain

By MINGWEN DONG

Thesis Director:

David S. Vicario, PhD.

Deviants are stimuli that violate the ongoing sequence or distribution of sensory events; they are potentially salient and the sensory system constantly monitors them. However, detecting a deviant is not a trivial process and requires to compare the current stimulus with prior memories or predictions. Traditionally, deviance detection has been studied in both humans and animals by presenting pure tones in a paradigm in which a rare tone stimulus (the oddball) occurs at random in a repeated sequence of a different tone (the standard). The current study seeks to more fully investigate the process and neural substrate of auditory deviance detection by using several paradigms, including both an extension of the traditional oddball approach that uses complex sounds as stimuli and also a new context paradigm. Multi-unit auditory responses to these stimuli were recorded from the auditory forebrain of awake male zebra finches. Results show that an oddball effect (larger responses to a sound when it is deviant than when it is common) can be elicited with complex stimuli like zebra finch calls (as well as with tones), and that the effect magnitude increases as common and deviant stimuli become more different acoustically. These results are consistent with a simple form of stimulus-specific

adaptation that generalizes to similar sounds. However, the order in which blocks of stimuli were presented changed the size of the oddball effect, suggesting a role for a memory of stimulus patterns that persists over longer durations and across many intervening stimuli. In the new context paradigm, where deviance could be defined mathematically, the neural response to a given stimulus depended on the larger context in which it was presented, again suggesting a perceptual learning effect. These experiments advances the study of deviance detection by using neural data to identify and separate the longer term effects of stimulus familiarity and pattern from the immediate effects of presentation order and relative frequency that are studied in simple oddball paradigm.

Acknowledgement

I would like to thank my advisor Dr. David S. Vicario for his guidance throughout the experiment and completion of this thesis. I would thank my committee members Dr. John McGann and Dr. Charles Randy Gallistel for their comments and suggestions. I would also thank Dr. Mimi L. Phan, Brittany Bell, Efe Soyman, and Lillian Yang for their great help and friendship in the Vicario lab. In the end, I would thank my parents and friends for being so understanding and supportive.

Table of Contents

Title page
Abstract.....	ii
Acknowledgement.....	iv
Table of contents	v
List of illustrations	vi
Introduction.....	1
Methods.....	5
Results	13
Discussion	16
Figures.....	25
References.....	35

List of illustrations:

Figure 1. Example of thresholded multi-unit activity and histology section	25
Figure 2. Oddball and context experimental paradigm.....	26
Figure 3. Spectrograms of auditory stimuli used in the experiment	27
Figure 4. Calculation of delta surprisal	28
Figure 5. Stimulus type affects the magnitude of oddball effect	29
Figure 6. Effects of ISI and stimulus complexity on stimulus-specific adaptation	30
Figure 7. Delta surprisal induced by different context condition	31
Figure 8. Bayesian and prediction error-based surprise	32
Figure 9. Neural network stimulation of stimulus-specific adaptation	33
Supplementary Figure 1. Distribution of delta surprisals	34

INTRODUCTION

Deviants are stimuli that violate the ongoing sequence or distribution of sensory events. They are potentially salient because they suggest something new is happening, which might be dangerous and/or require attention. Although sensory systems readily detect deviant stimuli, the neural mechanism of deviance detection is still unknown.

Traditionally, deviance detection has been studied using the oddball paradigm, in which a given stimulus is presented at different probability in two blocks (low probability as oddball/deviant and high as standard) and the difference in the magnitude of a response is measured. Depending on the specific question, different subjects, stimuli, and measurements were used.

In human studies, the stimulus, when deviant, induced a larger negative deflection in the electroencephalogram (EEG) than when common. The difference began ~150ms after stimulus onset, and is referred to as mismatch negativity (MMN) (Näätänen R. Attention and brain function. Hillsdale, NJ: Lawrence Erlbaum; 1992). Recently, it was reported that some earlier (~30ms after stimulus onset) and later (300-400ms after stimulus onset) EEG responses may also signal deviance (Slabu et al., 2010; Ruhnau, Philipp, et al., 2013). Based on these results, Lieder and other (2013) proposed two possible hypotheses to account for these effects: “prediction error” and “Bayesian surprise”. The former quantifies deviance by how unexpected the event is (Tribus, 1961), while the latter quantifies by how much the event changes one’s belief (Barto, Mirolli, & Baldassarre, 2013; Mars et al., 2008). Both of these models assume the subjects are actively predicting the next incoming stimulus and updating their belief about the world. These results suggested ways of thinking about the computational problems the brain

may be solving when a stimulus is identified as deviant, but didn't suggest the underlying neural mechanisms.

In electrophysiological studies, Ulanovsky and colleagues (2003) studied neural responses in the cat primary auditory cortex (A1) using an oddball paradigm with pure tones as stimuli. Their results showed that the neural response to a stimulus was stronger when it was an oddball than when it was a standard (Ulanovsky et al., 2003). The proposed mechanism is the stimulus-specific adaptation (SSA) hypothesis, that is, neural responses adapt to a repeated stimulus but the adaptation doesn't generalize to acoustically different stimuli. Later, oddball effects were also found in non-lemniscal parts of the inferior colliculus (IC) and medial geniculate nucleus (MGN) (Ayala, Pérez-González, & Malmierca, 2015; Malmierca, Anderson, & Antunes, 2015). Surprisingly, the oddball effect in IC and MGN didn't depend on activity in A1, even though A1 projects densely back to IC and MGN (Malmierca et al., 2015). One limitation of these experiments is that only pure tones were used as stimuli and that they only define "deviance" operationally in their task. In addition, this paradigm typically used inter-stimulus-intervals (ISI) of <1s, and oddball effects were weak or absent for ISIs >2-3s (Ulanovsky et al., 2003). Nelken (2013) reported that filtered white noise and word-like stimuli can also induce oddball effects, but no detailed statistics were included. Beckers and colleagues (2012) used complex zebra finch calls in a similar oddball paradigm and showed the oddball effect existed in caudal medial mesopallium (CMM, similar to a secondary auditory cortex in mammals) but not in Field L (analog to primary auditory cortex in mammals). However, these birds were anesthetized and they didn't compare the oddball effects elicited by calls and (more traditional) tones.

To further investigate deviance detection, a former lab member Lu (2013) did an experiments using complex stimuli (zebra finch and canary songs) in a changing context paradigm. Results showed that neural response to a stimulus was stronger when it's embedded in a sequence of acoustically-contrasting stimuli than when in a sequence of similar stimuli. This context-dependent effect was seen in both caudomedial neostriatum (NCM, similar to a secondary auditory cortex in mammal) and CMM (unpublished, Lu, 2013). The inter-stimulus interval (ISI) was 7 seconds, which is longer than those typically used in oddball experiment (< 1 second) (Ulanovsky et al., 2003; Xu, Yu, He, & Nelken, 2014; Yaron, Hershenhoren, & Nelken, 2012). This showed that the songbird auditory system could detect deviance over a much longer timescale when stimuli are complex. However, all contexts induced some significant response enhancement, suggesting that the paradigm may not have fully controlled for subtle effects of stimulus distribution and familiarity.

In summary, the computational definition of deviance and the study of how the nervous system detects it should be better linked. In order to address this need and to investigate how the auditory system detects complex deviant stimuli need, the current study carried out two experiments in awake zebra finches. First, in an oddball paradigm, both tones and complex vocalizations were used to vary the degree of acoustic similarity between standard and deviant stimuli. This tested whether the oddball effect decreases as the similarity increases, as predicted by the SSA hypothesis. Second, we designed a new type of the context experiment, in which a rare target stimulus was presented in a sequence of varying context stimuli. This experiment varied the acoustic similarity and familiarity of context and target stimuli, again with the prediction that response

enhancement to the target will become bigger when similarity decreases. This design enabled the component of deviance that is due to the novelty of the target stimulus to be separated from the component that is due to the immediate acoustic contrast at the transition between common context stimuli and the rare target stimulus. Furthermore, the degree of deviance in the paradigm was mathematically defined, calculated, and evaluated based on the neural data to show how deviance detection may be investigated at both neural and computational level.

METHODS

Subjects:

This study included three experiments and each used 6 adult (> 130 days) male zebra finches. All birds were housed in a general aviary with other zebra finches at Rutgers university under a 12h:12h light/dark cycle and provided with enough water and food. 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. The feathers on the scalp were removed and 0.04cc 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 0.04cc Metacam (5mg/mL) for post-operative analgesia and was closely monitored for recovery.

Electrophysiological recording:

After two days' recovery, the bird was restrained in a custom tube, and fixed to the stereotaxic frame by clamping the previously implanted pin. Then, a small craniotomy exposed the dura over the recording area. Sixteen electrodes (Type ESI2ec, impedance: 2– 4 M, Thomas Recording) were lowered into the field L/CMM (1 mm lateral from

midline, 1.5mm rostral to Y sinus) or NCM (1mm lateral from midline, 1mm rostral to Y sinus) of the two hemispheres (8 electrodes per side). 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. A power 1401 (CED, Cambridge, England) was used for both stimulus presentation and neural recording. Neural activities were amplified ($\times 19,000$), filtered (0.5 - 5 kHz bandpass), digitized (25 kHz), and stored for further analysis.

The waveforms were thresholded (2.5 standard deviation above the mean) to obtain multi-unit spike trains (Figure 1a). Responses to each stimulus were computed by subtracting the firing during a baseline period ($\frac{1}{4}$ of ISI before stimulus onset, smoothed across trials) from the firing during the stimulus period (stimulus duration plus 0.1sec).

Auditory stimulus presentation:

A speaker placed 30 cm in front of the bird was used for stimulus presentation. All stimuli were equated for RMS amplitude, with peak amplitude of 65dB SPL (“A” scale). Depending on the experiments, stimuli were presented either in oddball paradigm or context paradigm as described below. Because multiple stimulus sets were presented, the presentation sequence was counterbalanced across different birds.

1. Oddball experiment

- a. Oddball paradigm: two stimuli (S1 & S2) were presented in two blocks. In the first block, 20 repetitions of S1 (the oddball) and 180 repetitions of S2 (the standard) were presented in a shuffled order. In the second block, S2 became the deviant and S1 became the standard (Figure 2a). Between the

two blocks, no stimuli were played for around 60 seconds so that we could load the next block of stimuli and neurons could fully recover from short-term adaptation from the previous block. Note that, S1 was the deviant in the first block (1st deviant), while S2 was the deviant in the 2nd block (2nd deviant).

- b. Stimuli: before playing the experimental stimuli, a tuning set (19 tones equally spaced from 500 to 5000 Hz) was played to calculate the best frequency of each recording site. Then, six sets of experimental stimuli (3 types, each with 2 different sets) were played in the counterbalanced order across birds.

Pure tones: two pure tones were used as the oddball and standard, whose frequency $f1$ and $f2$ were determined by the best frequency (BF) of the target sites and a normalized frequency difference factor Δ (Ulanovsky et al., 2003). $f1$ and $f2$ last 260 ms (including a 10 ms ramp at both ends) and satisfy the following equations:

$$\Delta = 0.1 = \frac{f2 - f1}{\sqrt{f2 + f1}},$$

$$f1 + f2 = 2 * BF$$

Natural calls/reversal: male zebra finch call and its reversal in time domain were used as stimuli. These calls last around 260 ms and sound clearly different from corresponding reversals (Figure 3a). Compared with pure tones, call and reversal contain the same frequency but different spectro-temporal structures.

Doublets: two different female zebra finch calls connected with a 50 ms

silence in between were used as stimuli. Standard and oddball were made of the same calls with opposite order (Figure 3b). For example, if doublet (C1-C2) was deviant, then doublet consisting of the same calls but in opposite order (C2-C1) would be standard. Doublets last around 550 ms and the only difference within each pair was the order of the two calls. In this way, oddball and standard share all the frequency components and most of the spectro-temporal structures.

- c. Manipulations: two groups of zebra finches were used in the oddball experiment. In the first group (6 birds, 2 excluded because of technical difficulty), pure tone and call/reversal were presented with either a 0.736 or 2.208 sec ISI. In the second group (6 birds), all stimuli were presented with a 1.2sec ISI.
- d. Measurement: the magnitude of oddball effect is measured by the stimulus-specific adaptation index (SSAI) (Ulanovsky et al., 2003),

$$SSAI(S) = \frac{d(S) - s(S)}{d(S) + s(S)}$$

where $d(S)$ and $s(S)$ are the average neural response to the deviant and standard, respectively. The larger the SSAI is, the stronger the oddball effect.

Because birds in the second group all have 6 pair of SSAI values for each electrode (3 types of stimuli, each has 2 different sets), SSAIs from the same condition (the same column in Table 1) were averaged to measure the oddball effect. Note that both Field L and NCM have a tonotopic organization; thus each recording site often only responded to a limited

range of frequencies and the two tone sets thus could target the preferred frequencies of different sites. As a result, there is not always an average across stimuli sets for tones.

	1st as Deviant			2nd as Deviant		
SSAI	tone	call/reversal	doublet	tone	call/reversal	doublet
set1	f1	call	forward	f2	reversal	backward
set2	f2	reversal	backward	f1	call	forward

Table 1: stimulus sets in oddball experiment.

2. Context experiment

a. Context paradigm:

Stimuli were played in three blocks: pre-context, context, and post-context. In the pre-context block, one target stimulus and 7 background stimuli were played: 40 repetitions each in a shuffled order. In the context block, the target stimulus and 7 context stimuli were played 20 times each in a randomized order. Finally, in the post-context block, stimuli in pre-context block were played again but only for 20 repetitions each (Figure 2b). Three blocks of stimuli were played consecutively without extra gaps in between.

b. Stimuli:

In this experiment, the target stimulus and background stimulus were always syllables from male zebra finch songs, while context stimuli were one of the following types:

silence: silence intervals.

canary: syllables from canary song.

diffZF: zebra finch syllables that were different from those background zebra finch syllables in pre- and post-context block.

control (sameZF): context stimuli were the same zebra finch syllables that were used in the pre- and post-context block as background stimuli. Because no change was made across different blocks, control condition shouldn't induce any surprisal.

The duration of syllables ranged between 200 ms and 300 ms for both zebra finch (Figure 3c) and canary syllables (Figure 3d).

c. Measurement:

Delta surprise (DS) measured how much actual neural responses deviate from expected responses. To calculate DS, the responses to the target stimulus were first ranked based on their presentation order. Then, the expected response to the 1st and last target stimulus in the context block were extrapolated by correlating the neural responses in pre- and post-context block with their ranks. The line connecting the two dots were then used to estimate the expected response to the target stimulus as if context stimuli were the same as background stimuli. By using the standard deviation of the residuals from regression and the difference between the actual neural response and the expected neural response, DS for each target trial during context block could be calculated using the following formula:

$$DS = \text{sign}(d) * \{ \log_2 \frac{1}{P(d)} - \log_2 \frac{1}{P(0)} \}$$

where d is the difference between actual response and expected response, $P(d)$ is the probability of observing d in a normal distribution $N(0; \text{standard deviation})$. The standard deviation was calculated by using the residuals from the regression (Figure 4).

The general surprisal effect induced by a stimulus set was calculated by using the median of 21 delta surprisal from the in-context target trials (the mean is highly correlated with the median).

Because each context condition included two different stimulus sets (with different target, background, and context stimuli), each electrode of a bird has 8 median DS corresponding to 8 different stimulus sets (different target, background, and context stimuli). A context condition that is presented for the first time is labeled as 1st. Similarly, a previously heard context condition is labeled as 2nd.

Histology:

After the experiment, several electric lesions (20 uA, 10 seconds) were made for later histological verification of the recording sites and the bird was returned to its home cage. 2 days later, the bird is sacrificed with an overdose of pentobarbital (390 mg/ml, 2ml), and perfused with 0.9% saline and 3.3% paraformaldehyde. After several day's fixation, the brain was cut into 50nm sagittal slices using a Vibrotome and stained with Cresyl Violet. In the end, the stained slices were visualized with a microscope and recording sites were inferred from the previously made lesions (Figure 1b).

Data analysis:

First, electrodes with no response to the stimulus ($\text{abs}(\text{response}) < 10$ spikes/sec) were excluded. Then, electrodes with missing values were also discarded to enable within-subject comparisons. Because the measurements in both oddball and context experiment are not normally distributed and the sphericity assumption is not met, we used the non-parametric Wilcoxon matched pairs test for within-subjects comparisons and the Kolmogorov-Smirnov test for between-subject comparisons. The criterion of statistical significance was set at $p < 0.01$.

RESULTS

Oddball experiments:

The oddball experiment tested the SSA hypothesis by varying the acoustic similarity between the oddball and standard stimuli. Because the SSA hypothesis assumes the oddball effect is mainly driven by non-overlapping pathways (e.g. different tonotopic channels for pure tone stimuli with different frequencies), it predicts that acoustically more similar standard and deviant stimuli will show a smaller oddball effect.

1. Effect of varying stimulus similarity on oddball responses.

Figure 5 shows that tones, call/reversal, and doublets in turn induced smaller oddball effects ($n = 64$, $p < 0.01$ for all three comparisons) when they were 1st presented as deviant. When presented 2nd as deviant, tones still induced significant oddball effect but calls/reversal and doublets didn't ($n = 64$, $p < 0.01$ for tones; n.s. for calls/reversal and doublets). The results are consistent with the SSA hypothesis when stimuli were presented 1st as deviant but not for complex stimuli when presented 2nd as deviant.

2. Effect of ISI, stimulus type, and sequence on oddball responses.

Surprisingly, when longer ISIs ($\geq 1.2s$) were used, all stimuli induced stronger oddball effect when presented 1st as deviant than 2nd as deviant ($n = 64$, $p < 0.01$ for all three comparisons) (Figure 6). For an ISI of 0.736s (the typical value in previous oddball experiments in animals), calls/reversal and doublets showed a significant sequence effect ($n = 18$, $p < 0.01$) but tones didn't ($n = 18$, n.s.; Figure 6). The Sequence effect was measured as the difference between SSAI when a stimulus was presented 1st as deviant and when 2nd as deviant. Though the sample

size is small when ISI is 0.736 second, this preliminary result suggests ISI may interact with stimulus type and affect the sequence effect. It also suggests why sequence effects may not have been observed in traditional oddball experiments using short ISIs and simple tone stimuli.

Context experiments:

The context experiment has two goals: First, it again tests the validity of the SSA hypothesis by varying the similarity between context stimuli and the target stimulus; second, it allows quantifying deviance both in neural data and computationally. It improves on the earlier context experiment (Lu, 2013) by controlling for target stimulus probability and context stimulus familiarity.

1. Effect of target and context stimulus similarity on response enhancement.

The median delta surprisals from the two different stimulus sets of the same context type were averaged to measure context-specific response enhancement. In the control condition, the target stimulus was presented in the same group of other stimuli throughout the recording, there is no change in the context phase and thus the neural response should be the same as expected response, with no response enhancement. Indeed, results showed just that: no significant response enhancement in the sameZF context ($n = 84$, $p > 0.01$, Figure 7; Bayes Factor = 11.6, supporting null hypothesis). This validates the method for extrapolating expected neural responses in the other context conditions.

In contrast, canary context induced a significant response enhancement, and silence context induced an even larger enhancement ($n = 84$, $p < 0.01$ for both canary and silence context, Figure 7). Both of these effects are consistent with

earlier work (Lu, 2013), although ISI is much shorter and presentation probability of target stimulus doesn't change across pre-, context, and post-context blocks. A smaller but still significant enhancement was seen for diffZF context stimuli ($n = 84$, $p < 0.01$). Overall, the results showed that diffZF, canary, and silence context in turn elicited larger response enhancement ($n = 84$, $p < 0.01$ for all comparisons).

2. Effect of context familiarity on response enhancement

Furthermore, the canary context showed a significant familiarity effect (Figure 7 & supplementary Figure 1); that is, it elicited a stronger response enhancement when presented for the 1st time than for the 2nd time ($n = 83$, $p < 0.01$); also, the enhancement for the 2nd canary was greater than that seen for diffZF ($n = 83$, $p < 0.01$). There was a trend for a familiarity effect in the silence context ($n = 83$, $p = 0.08$), but no significant familiarity effect for diffZF or sameZF contexts ($n = 83$, $p = 0.27$ and 0.94 , respectively).

DISCUSSION

This study has three major results. Firstly, 1) as predicted by the SSA hypothesis, both complex stimuli and pure tones appeared more deviant when they shared less acoustic structures with the standard/context stimuli; 2) we observed a sequence effect in the oddball experiment; 3) in the context experiment, less acoustically similar target and context stimuli induced stronger response enhancement (consistent with SSA hypothesis). Besides, in the canary context experiment, we quantified the deviance of the target stimulus based on the stimulus presentation order and tried to correlate it with delta surprisal, a measurement of neural surprisal. Here, we discuss what these results imply and how computational modeling helps interpret them.

Similar deviant and standard/context elicited weaker response enhancement:

In the SSA hypothesis, the ratio of non-overlapping to overlapping synaptic channels determine the magnitude of the oddball effect (Mill, Coath, Wennekers, & Denham, 2011b) (Ulanovsky et al., 2003). The larger the ratio, the bigger the oddball effect. Because pure tones, calls/reversal, and doublets have successively smaller and smaller non-overlapping to overlapping ratio, they should also in turn elicit smaller oddball effects. Our results were consistent with this prediction when stimuli were presented as deviant in the 1st block. In the case of doublets, because standard and deviant only differed in the order of the calls (Figure 3b), it's surprising that significant oddball effect still existed and suggests that the auditory system may bind and perceive the two calls together. Alternatively, because a stimulus can induce an inhibitory after-effect (observation in pilot study, unpublished), the neural response to the 2nd call in the doublet may be inhibited by that to the 1st call and consequently depletes the synaptic channels

much less (of course, one could speculate that this inhibitory effect is the neural manifestation of a binding process). Because the 1st call elicited oddball effect while the 2nd call weakened it, the net effect was the small yet significant oddball effects induced by doublets. Lastly, the longer absolute duration of the doublet stimuli used here may reduce the oddball effect further because neurons have less time to recover from adaptation (given a the same ISI).

In the context experiment, as context stimuli change from sameZF (control), diffZF, canary, to silence, the ratio of non-overlapping to overlapping synaptic channels decreased in turn and the SSA hypothesis provides an explanation for why they induced smaller and smaller response enhancement.

However, the pattern of results was not as clear when stimuli were presented as deviant in the 2nd block because of the significant sequence effect observed under some conditions (discussed below).

Sequence effect: deviance, novelty, and surprise in the oddball experiment:

In the oddball experiment, tones, calls/reversal, and doublets all induced significantly stronger oddball effect when they were presented in the 1st block than in the 2nd block for ISIs ≥ 1.2 s. This sequence effect has several possible explanations.

First, a form of SSA has been shown to operate in zebra finches over a much longer timescale than that observed in the oddball experiment in rodents (Sek Jin Chew, Mello, Nottebohm, Jarvis, & Vicario, 1995; S. J. Chew, Vicario, & Nottebohm, 1996; Nelken, 2004). This long-term adaptation makes the neurons respond less to both specific stimuli and stimuli of the same class (sharing acoustic features) in the 2nd block than to those in the 1st block. Consequently, the response difference between oddball in the 1st

block and standard in the 2nd block will be bigger than the difference between the oddball in the 2nd block and the standard in the 1st block. This results in the larger oddball effect in the former case than the latter one.

Second, because standard and oddball both activate their overlapping synaptic channels (frequency channels for tones and calls/reversal, spectro-temporal channels for doublets), the adaptation in overlapping synaptic channels may last longer and reduces the overall responses to stimuli in the 2nd block. Then, similar to the first explanation, the oddball effect will become smaller for stimuli presented as deviant in the 2nd block. However, this cross-adaptation hypothesis is unlikely because neurons should fully recover from short-term adaptation during the 60-second silence after the 1st block of stimuli (see Method).

Last, the sequence effect may have occurred because deviance in the oddball experiment contains two different components: novelty and surprise. Conceptually, a novel stimulus is being heard for the first time (thus not in memory) while a surprising stimulus is something that is unexpected, even if it's not novel (Barto et al., 2013). Mathematically, surprise has been quantified by using Bayesian framework (Baldi & Itti, 2010; Itti & Baldi, 2005), information theory (Monsalve, Frank, & Vigliocco, 2012; Roark, 2011; Tribus, 1961), or “description” (Palm, 2012). Novelty has been quantified based on clustering (Markou & Singh, 2003a), content-addressable neural network (Kohonen et al., 1977; Markou & Singh, 2003b), or statistical outlier approach (Duda, Hart, & Others, 1973). Notice that the last definition of novelty, statistical outlier, is more like surprise than novelty because it is based on comparison with prior prediction rather than memory (Barto et al., 2013). These definitions suggest that, though novelty and

surprise are related, they require different computations and thus may involve different neural mechanisms. With this distinction between novelty and surprise in mind, an oddball is both novel and surprising in the 1st block but only surprising in the 2nd block. The standard is novel in the 1st block but not in the 2nd block. If both novelty and surprise enhance neural responses, the asymmetry of novelty may have caused the sequence effect observed in the oddball experiment.

Familiarity effect:

Among the 4 conditions of the context experiment, only canary condition showed a significant familiarity effect; i.e., the canary context elicited a larger delta surprisal when presented for the 1st time than that for the 2nd time. Compared with the sequence effect, the familiarity effect operates at even longer timescale because three other stimulus conditions were played in between and last around 30 minutes in total. Because mechanistic hypotheses like SSA cannot explain this phenomenon, we propose a statistical learning hypothesis based on two kinds of mechanisms. The bird learns:

1) The acoustic features of the canary context. In the 1st canary trial block, the zebra finch hears canary syllables (which contain a different combination of acoustic features from zebra finch syllables) for the very first time. Even though the 2nd canary context block contains different specific canary syllables, they are now of the familiar canary type and there is thus less contrast with the target zebra finch syllable.

2) The transition probabilities between different kinds of stimuli (Lu & Vicario, 2014). Because the bird heard the “canary to zebra finch” transition pattern from the 1st canary context block, the same transition pattern in the 2nd canary context became less unexpected and thus less surprising.

Exploratory models to bridge computational and neural surprise:

In the typical oddball experiment, the oddball stimulus is called a deviant, but it's unclear whether novelty, surprise, or both have contributed to the oddball effect. In contrast, in the present context experiment, the response enhancement can be attributed to surprise but not novelty because the bird has already heard the target stimulus in the pre-context block. To quantify the surprise using information theory (Tribus, 1961) and Bayesian inference (Baldi & Itti, 2010), the following simplifications and assumptions were made:

- 1) Only modeled the surprise in the canary context because zebra finch and canary syllables are acoustically very different (Figure 3c & d). No distinction was made within stimulus classes (zebra finch or canary syllables) and thus the canary condition was simplified to be a Bernoulli processes.
- 2) Bird initially has a uniform prior and updates it by using Bayes rule whenever a stimulus was observed.
- 3) Prior contains a set of models M_i that predict how likely the next stimulus is a zebra finch or canary syllable:

$P_i(\text{zf})$: probability of observing a zebra finch syllable;

$P_i(\text{canary}) = 1 - P_i(\text{zf})$: probability of observing a canary syllable.

- 4) Calculation of surprise:
 - a) Information-based surprise was calculated by using the most likely model

(P_{ml}) in the current prior (Tribus, 1961):

$$\text{surprise} = -\log_2(P_{ml}(S)),$$

where, S is either a zebra finch syllable or canary syllable.

- b) Bayesian surprise was defined as the Kullback–Leibler divergence between the prior and posterior (Baldi & Itti, 2010):

$$surprise = KL(prior, posterior),$$

where posterior is the new belief about the models after observing the stimulus.

In this preliminary analysis, we further assumed that the bird has a permanent memory within each experiment, and each model in the prior has a different but fixed $P(zf)$, and the bird uses stimulus distribution but not transition probability between syllables to infer the $P(zf)$. Then, the ideal surprise for each target stimulus in the context block was calculated using both information-based and Bayesian surprise.

Since each target trial also had a delta surprisal calculated from neural data, we could test our models (Bayesian surprise and prediction error-based surprise) by checking how well they predict the bird's actual surprise (delta surprisal). For most birds, the correlation between the ideal surprise and delta surprisal was not significant (data not shown). However, this was non-surprising (no pun intended) because the model was probably too simple in this preliminary version. Had the model included transition probability and acoustic similarity between syllables, the neural surprise may have been better predicted. Nonetheless, this approach and analysis show how one can study surprise (deviance) detection from both computational and mechanistic perspectives by using the context paradigm. Even if one only studies surprise (deviance) detection at mechanistic or computational level, the context paradigm has less confounding factors (e.g., novelty/familiarity) and allows trial-by-trial analysis compared with the oddball paradigm.

Preliminary attempt at quantifying the SSA hypothesis:

Most discussions about the SSA hypothesis were qualitative rather than quantitative (Khouri & Nelken, 2015; Ulanovsky et al., 2003; Yaron et al., 2012). Consequently, it is not clear whether SSA could actually generate the kinds of responses to deviants actually observed in the experiments. Mill and colleagues proposed the first SSA-based neural network model that simulates the oddball experiment (Mill, Coath, Wennekers, & Denham, 2011a; Mill et al., 2011b; Mill, Coath, Wennekers, & Denham, 2012). Though the model showed oddball effects, it only took pure tones as input stimuli.

In an attempt to elaborate this kind of model to accept complex stimuli, we added a cochlea model (Zilany, Bruce, & Carney, 2014) in front of the two-layer neural network proposed by Mill and colleagues (Figure 9a). The cochlea model (Zilany et al., 2014) decomposes the complex stimuli into spike trains and feeds these spikes into the input neuron group. The input neurons are connected to the output neurons using “depressing synapses” (Mill et al., 2011b) in an all-to-all manner. The depressing synapse is excitatory but decreases its efficacy every time after the presynaptic neuron fires a spike and recovers slowly afterwards (Destexhe, Mainen, & Sejnowski, 1998; Tsodyks & Markram, 1997). The output neuron group consists of adaptive exponential integrate-and-fire (AdEx) neurons, which increase their threshold whenever they fire a spike and slowly recover afterwards (Brette & Gerstner, 2005). Most parameters of this modified model are the same as those in (Brette & Gerstner, 2005; Destexhe et al., 1998; Mill et al., 2011b; Tsodyks & Markram, 1997) and perturbed with random noise. However, to obtain the desired results, it is necessary to manually tune the recovery time constant for the depressing synapse and AdEx neurons to be ~ 1.2 s. When the model was fed a sequence

of zebra finch (Figure 3c) and canary syllables (Figure 3d), the output neurons responded less when the zebra finch syllable was presented consecutively than when interleaved with canary syllable (average over 20 simulations, Figure 9b & c). This shows that SSA can induce oddball effects even for complex stimuli if the right parameters are used. But it's unclear whether a 1.2s recovery time constant for depressing synapse and AdEx neurons is biologically plausible or not.

Finally, it should be noted that an SSA-based neural network model is potentially a mechanistic model while a Bayes rule-based surprise model operates at the computational level. Consequently, they're not mutually exclusive (Khouri & Nelken, 2015) and may just be descriptions of the same phenomenon at different levels.

In summary, the current study shows neural responses are context-dependent. Specifically, neurons respond more strongly when a rare stimulus (simple or complex) was presented with acoustically different stimuli than when presented with similar stimuli. This response difference may provide a deviance signal that enables the bird to shift attention, acquire new information about the world, and potentially make adaptive reactions. These effects can occur for stimulus patterns over time scales that suggest a contribution of perceptual learning, not merely immediate adaptation. Furthermore, this new context paradigm separates the effect of surprise from that of novelty, and thus the results suggest that novelty and surprise may be two different kinds of deviants, possibly produced by different brain mechanisms.

Future directions:

To clarify what computations the neuron may be performing, I would first like to further develop the prediction and Bayesian surprisal model by including transition probability and acoustic similarity between syllables.

In addition, since we still don't know whether SSA and Bayes rule-based surprise model are complementary, I will test it with the context paradigm. In the new experiment, target zebra finch syllables will be first played together with background canary song syllables in the pre-context block, then presented consecutively in the context block, and finally played with background canary syllables again in the post-context block. In this way, Bayesian and prediction surprise model will predict the zebra finch syllables in the context block to be surprising because it rarely occurred consecutively before; but SSA hypothesis will predict a decrease in response. If we assume neurons respond to surprise by increasing their responses, the relationship between the surprise model and SSA hypothesis will be clarified because the experiment results will support only one of the models.

Lastly, it's also worth exploring the anatomical origin of oddball, sequence, similarity and familiarity effects in different brain regions. I don't have enough histological data to analyze regional differences yet.

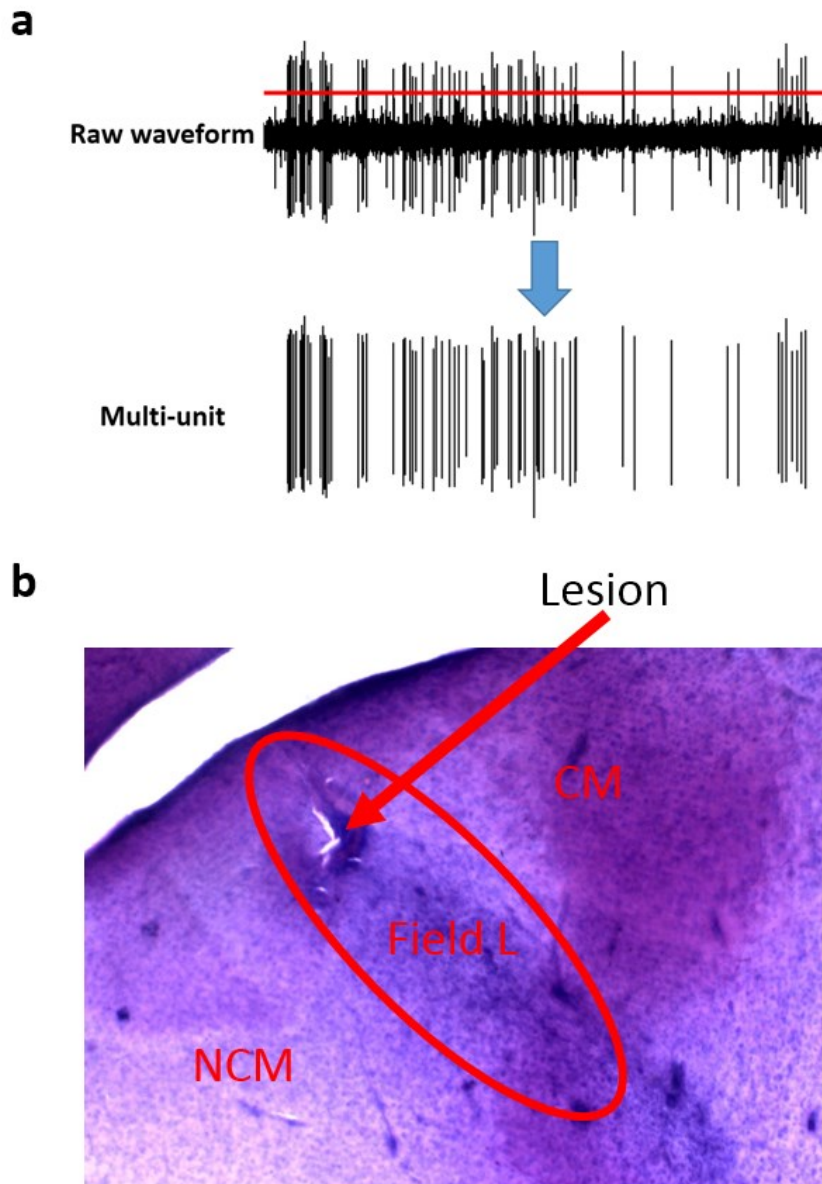
Figures:

Figure 1: a) Example of thresholded multi-unit activity. b) Cresyl violet stained section showing a recording site in Field L.

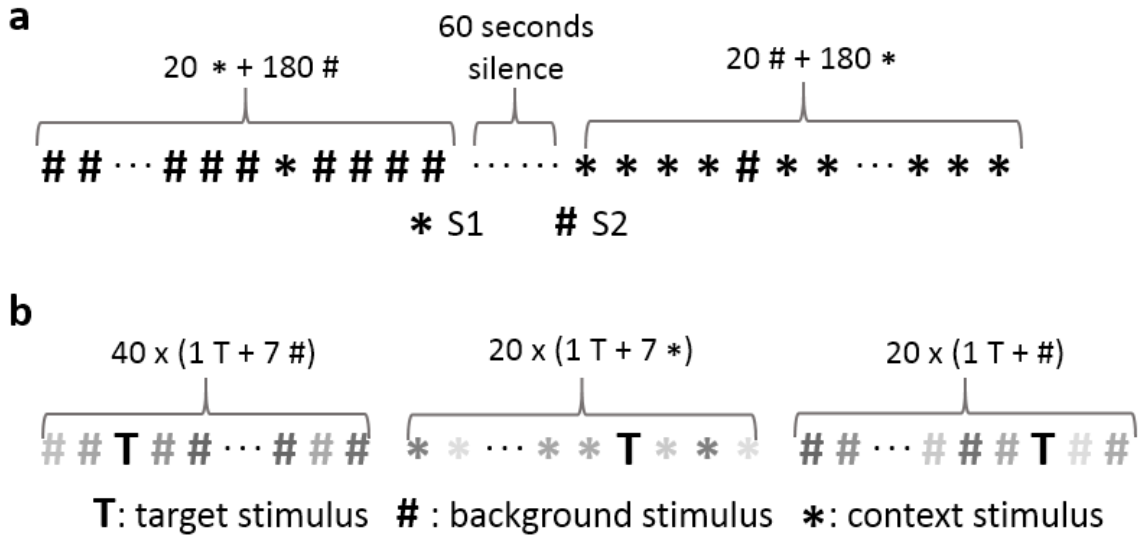


Figure 2: a) In the oddball paradigm, two stimuli were presented with different probability in two blocks. When a stimulus was presented with low probability (e.g., S1 in the 1st block), it was called oddball/deviant. When a stimulus was presented with high probability, it was called standard (e.g., S2 in the 1st block). Depending on whether a stimulus was oddball in the first or second block, it's denoted as 1st as Deviant and 2nd as Deviant, respectively. b) In the context paradigm, the target zebra finch syllable (T) was presented with the same probability (1/8) across different blocks. In the pre-context phase, T occurred in a shuffled sequence with background zebra finch syllables (#); in the context phase, T occurred in a sequence of context stimuli (*, of different types, see methods); in the post-context phase, T again occurred in a sequence of the same background zebra finch syllables.

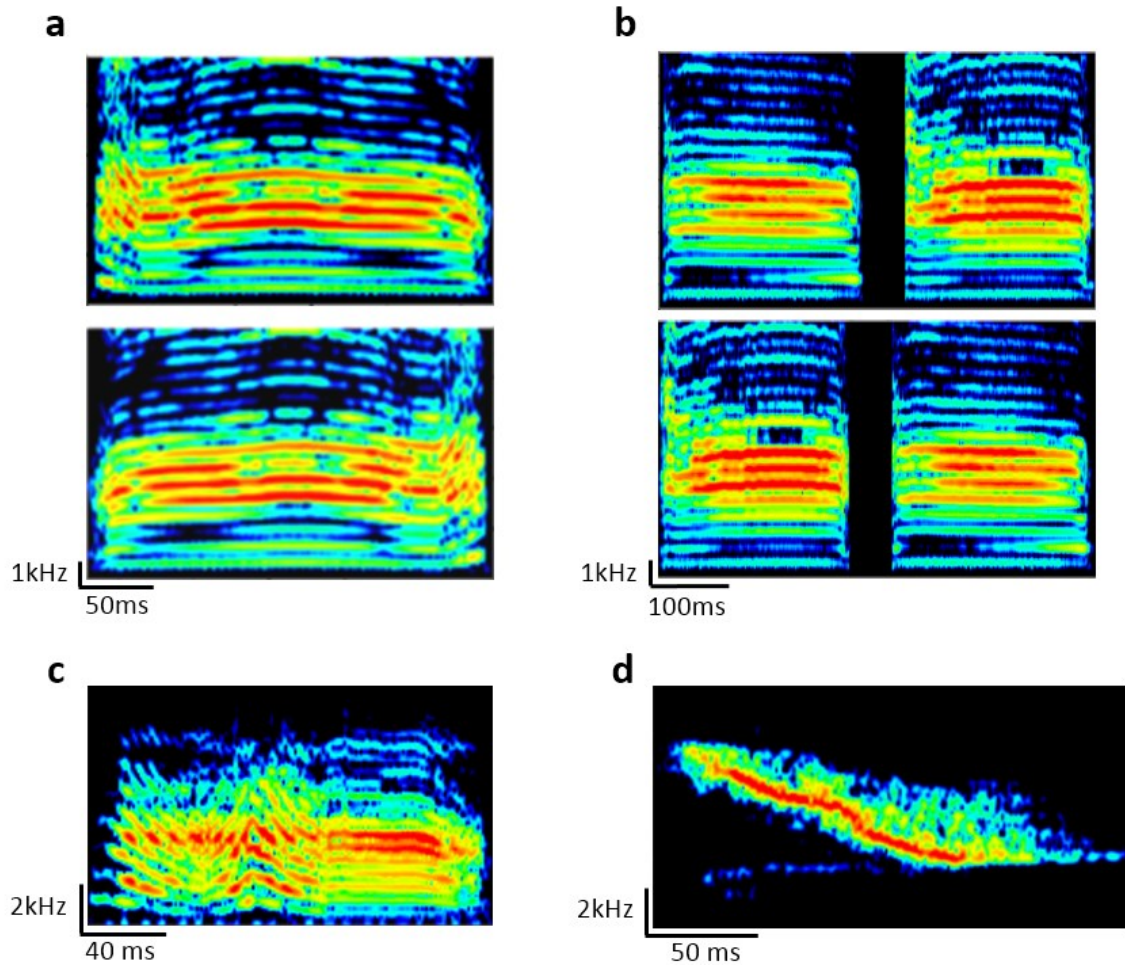


Figure 3: Spectrogram of example stimuli used in this study. a) Male zebra finch call and its reversal in time domain. b) Doublets that consist of two different female zebra finch calls with a 50 ms silence in between. c) Syllable from zebra finch song. d) Syllable from canary song.

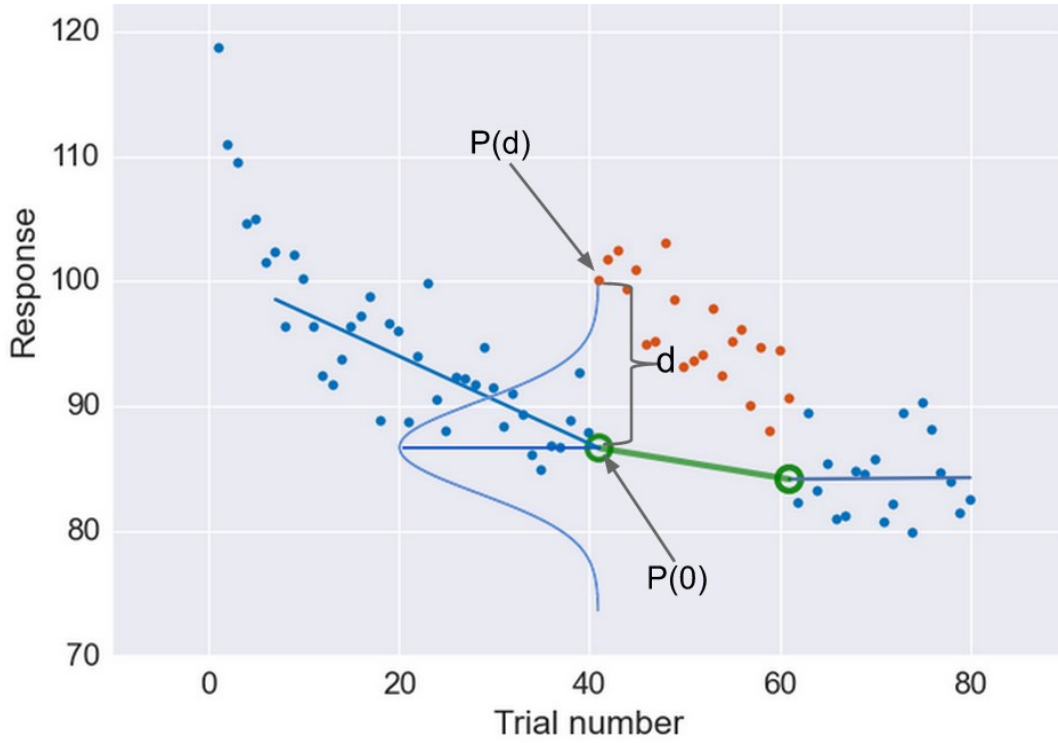


Figure 4: Calculation of delta surprisal (DS). The actual neural responses to the target stimulus (TS) were first ranked by presentation time (blue: TS in the pre- and post-context block; red: TS in the context block). Then, the expected responses to the 1st and last TS trial (green circles) in the context block were estimated by regressing neural response with trial number in the pre- and post-context block. By connecting these two extrapolated points (green line), the expected responses to the target stimulus in the context block were estimated. In the end, by assuming the deviance (actual response - expected response) follows a normal distribution (mean = 0, standard deviation is estimated as that of the residuals from regression), the DS was calculated as:

$$DS = \text{sign}(d) * \left\{ \log_2 \frac{1}{P(d)} - \log_2 \frac{1}{P(0)} \right\}$$

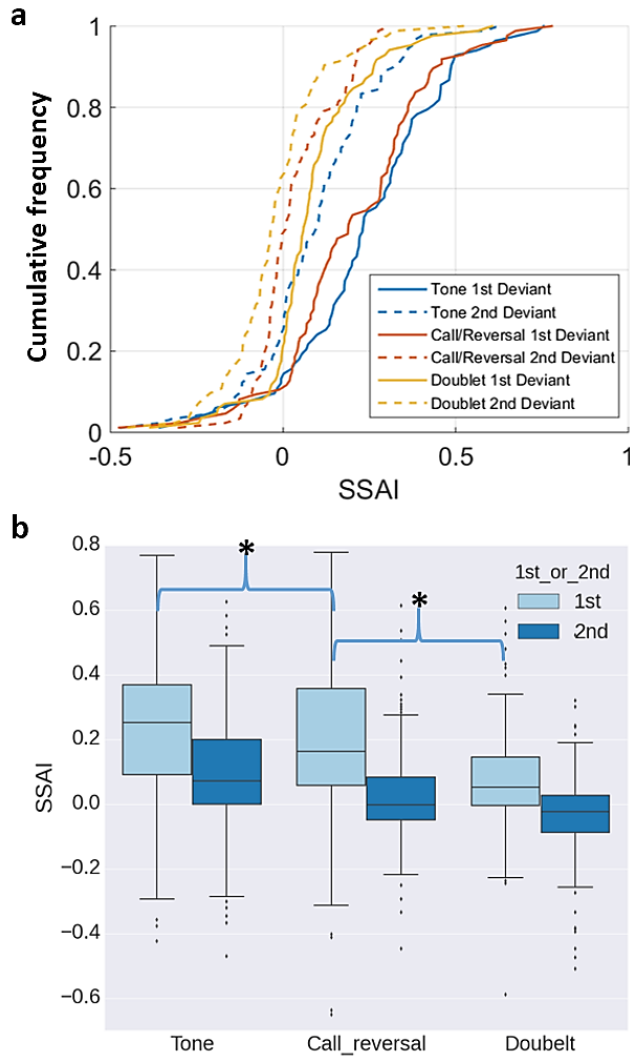


Figure 5: Stimulus type affects oddball effect. a) Cumulative frequency plot of SSAI induced by different stimulus type in different blocks. Solid line shows SSAI when a stimulus is the oddball in the 1st block, while the dashed line is when a stimulus occurs as deviant in the 2nd block. All three types of stimuli elicited significant oddball effects when presented in the 1st block. However, when presented as oddball in the 2nd block, only tones still showed a significant oddball effect, while calls/reversal and doublets did not. b) A box plot of the same data. As shown by the asterisk, tones, calls/reversal, and doublets in turn induced a smaller oddball effect when presented as the oddball in the 1st block.

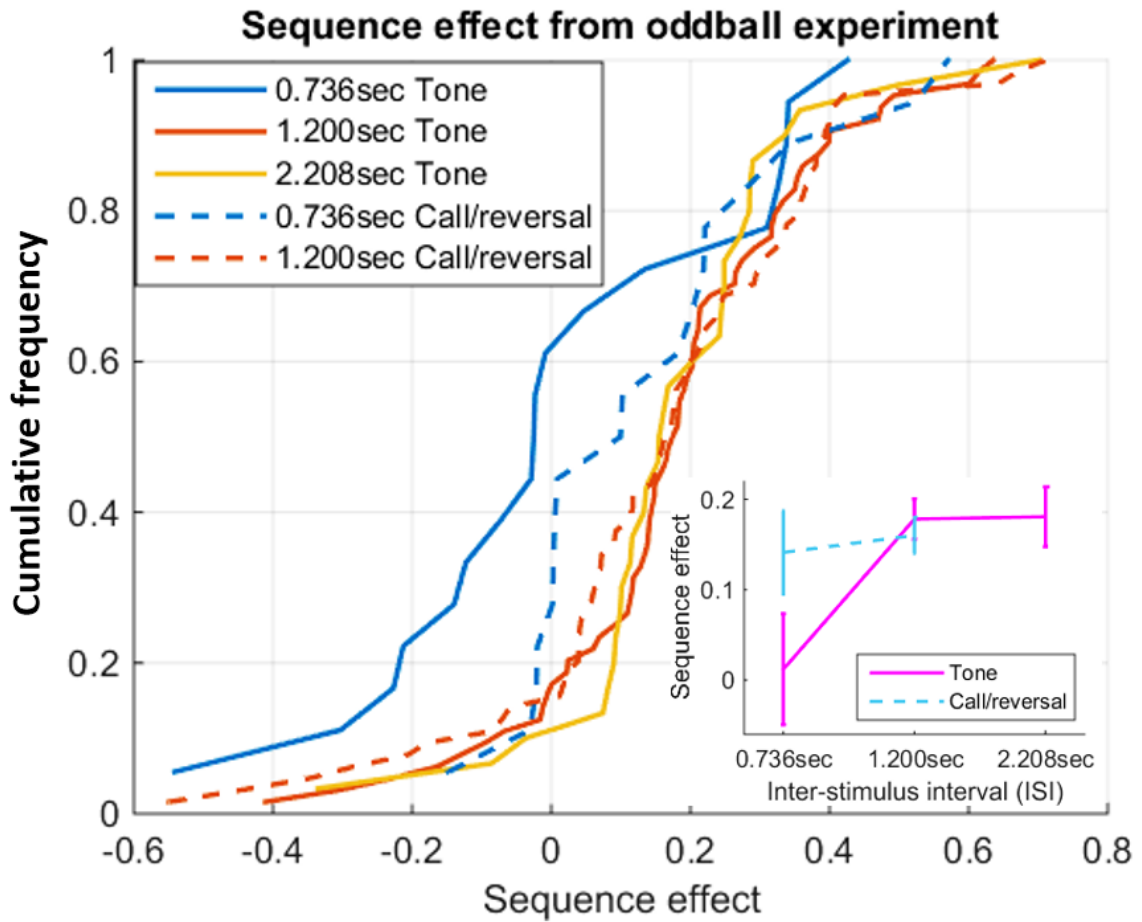


Figure 6: Effect of ISI and stimulus complexity on SSA. The sequence effect was quantified by subtracting the SSAI induced by the oddball in the 2st block from that in the 1nd block. The cumulative frequency plot showed sequence effect from two sets of birds using different inter-stimulus interval (ISI). Sequence effect didn't exist when tones were played with a 0.736 second ISI but occurred in all other conditions. This suggests ISI and stimulus type may affect sequence effect in an interacting way. The small plot showed that sequence effect existed at both ISIs for calls, but not for tones.

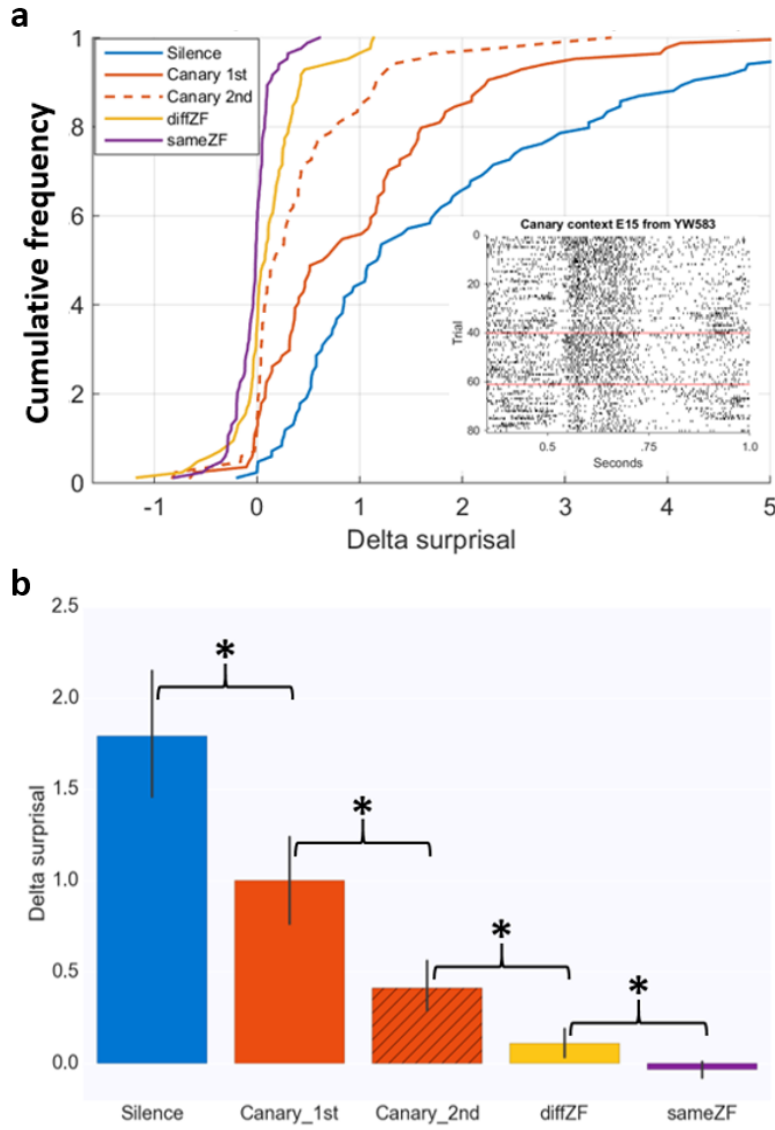


Figure 7: Delta surprisal (DS) induced by different context condition. a) The cumulative frequency plot of DS from different context conditions. Because presentation order didn't cause significant differences in DS for silence, diffZF, and control (sameZF) condition, the DS's obtained from two stimulus sets were averaged within each condition. The inset raster plot is an example showing how responses in context block differ from those in pre- and post-context block. b) The bar plot shows that silence, 1st canary, 2nd canary, and diffZF in turn induced smaller DS, while control (sameZF) condition didn't induce surprisal at all.

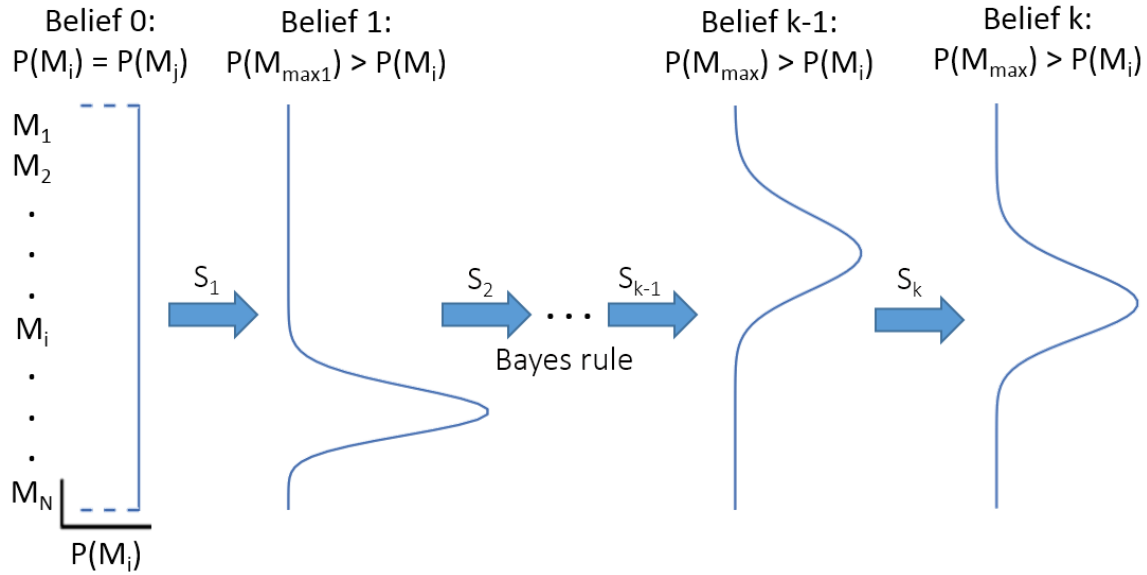


Figure 8: Bayesian and prediction error-based surprise. This assume that bird constantly predicts the identity of incoming stimuli based on its internal models (M_i), which give predictions about how likely the next stimulus is zebra finch or canary syllable. Every time a stimulus (S_k) is observed, the probability distribution of M_i (belief) is updated via Bayes' rule. Bayesian surprise induced by a stimulus (S_k) is quantified as the Kullback–Leibler divergence between the belief before and after the stimulus presentation.

Conceptually, the bigger the change in the bird's belief, the bigger the surprise.

Prediction error-based surprise (using information theory) is quantified as the negative logarithm of the prediction ($P(S_k | M_{\max})$) from the most likely model (model with biggest $P(M_i)$). If we assume the bird initially has a uniform prior (belief 0), we could calculate the Bayesian and prediction surprise for every target stimulus trial. By comparing which surprise model gives better predictions for delta surprise, we could compare their validity. Unfortunately, in the preliminary analysis, both models failed to give significant predictions for delta surprisal.

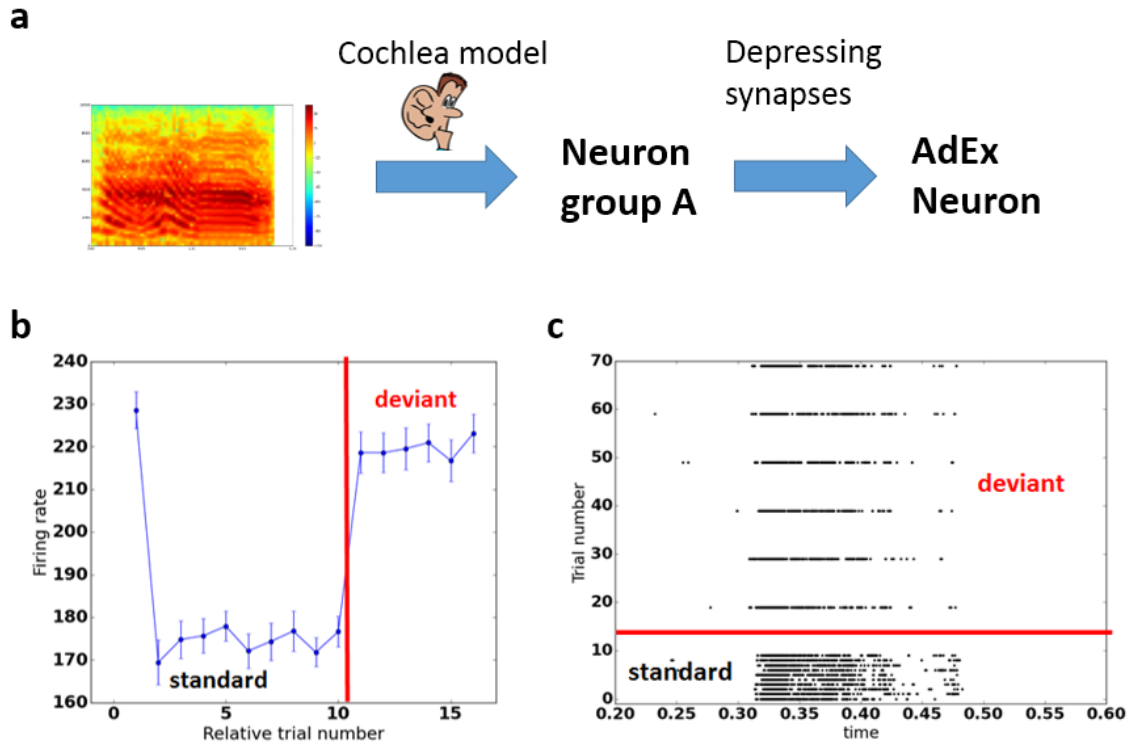
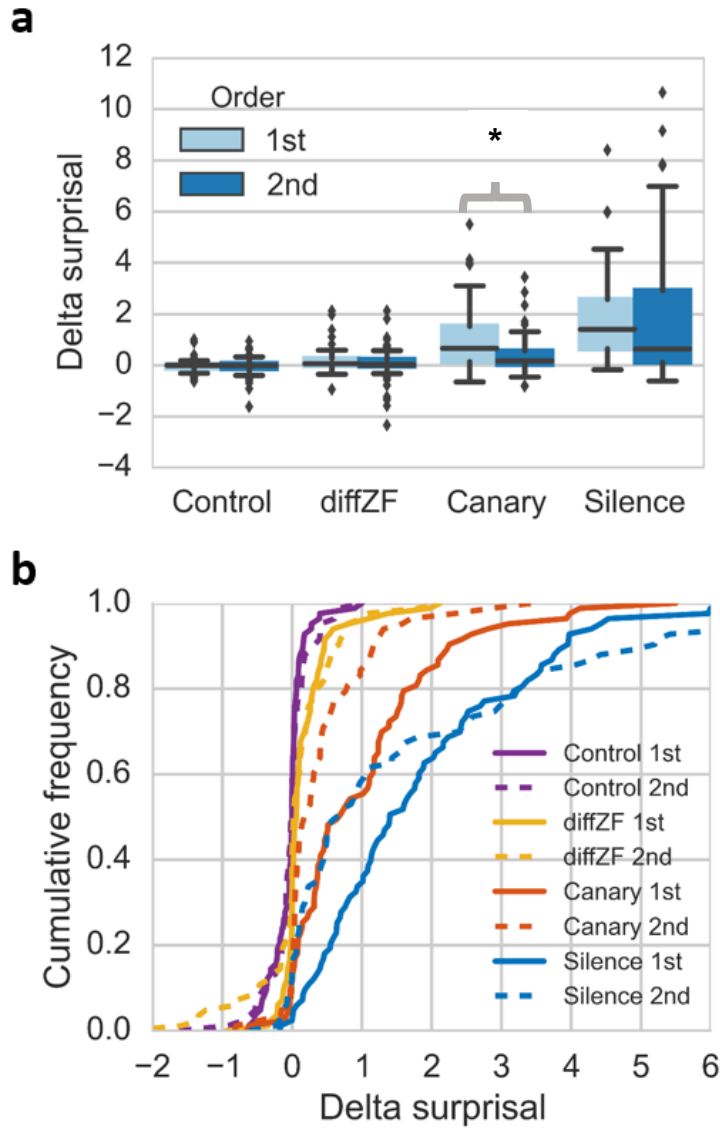


Figure 9: Neural network simulation results. a) A schematic illustration of the neural network model. b) Results from 20 simulations showed the AdEx neurons in the output neuron group responded stronger when a zebra finch syllable was played with canary syllables than when presented alone. c) An example raster plot.



Supplementary Figure 1: distribution of delta surprisals for all 8 different stimulus sets in the context experiment (without averaging). Figures at the top and bottom show the same data in different format: boxplot and cdf plot.

References:

- Ayala, Y. A., Pérez-González, D., & Malmierca, M. S. (2015). Stimulus-specific adaptation in the inferior colliculus: The role of excitatory, inhibitory and modulatory inputs. *Biological Psychology*.
<http://doi.org/10.1016/j.biopsycho.2015.06.016>
- Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks: The Official Journal of the International Neural Network Society*, 23(5), 649–666.
- Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or Surprise? *Frontiers in Psychology*, 4, 907.
- Beckers, G. J. L., & Gahr, M. (2012). Large-Scale Synchronized Activity during Vocal Deviance Detection in the Zebra Finch Auditory Forebrain. *Journal of Neuroscience*, 32(31), 10594–10608.
- Brette, R., & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94(5), 3637–3642.
- 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.
- Chew, S. J., Vicario, D. S., & Nottebohm, F. (1996). A large-capacity memory system that recognizes the calls and songs of individual birds. *Proceedings of the National Academy of Sciences of the United States of America*, 93(5), 1950–1955.
- Destexhe, A., Mainen, Z. F., & Sejnowski, T. J. (1998). Kinetic models of synaptic transmission. *Methods in Neuronal Modeling*, 2, 1–25.
- Duda, R. O., Hart, P. E., & Others. (1973). *Pattern classification and scene analysis* (Vol. 3). Wiley New York.
- Itti, L., & Baldi, P. (2005). A principled approach to detecting surprising events in video. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 631–637 vol. 1).
- Khouri, L., & Nelken, I. (2015). Detecting the unexpected. *Current Opinion in Neurobiology*, 35, 142–147.
- Kohonen, T., Lehtiö, P., Rovamo, J., Hyvärinen, J., Bry, K., & Vainio, L. (1977). A principle of neural associative memory. *Neuroscience*, 2(6), 1065–1076.
- Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J., & Stephan, K. E. (2013). Modelling Trial-by-Trial Changes in the Mismatch Negativity. *PLoS Computational Biology*, 9(2), e1002911.
- 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 of the United States of America*, 111(40), 14553–14558.
- Malmierca, M. S., Anderson, L. A., & Antunes, F. M. (2015). The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: a potential neuronal correlate for predictive coding. *Frontiers in Systems Neuroscience*, 9.
<http://doi.org/10.3389/fnsys.2015.00019>
- Markou, M., & Singh, S. (2003a). Novelty detection: a review—part 1: statistical

- approaches. *Signal Processing*, 83(12), 2481–2497.
- Markou, M., & Singh, S. (2003b). Novelty detection: a review—part 2:: neural network based approaches. *Signal Processing*, 83(12), 2499–2521.
- Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-Trial Fluctuations in the Event-Related Electroencephalogram Reflect Dynamic Changes in the Degree of Surprise. *Journal of Neuroscience*, 28(47), 12539–12545.
- Mill, R., Coath, M., Wennekers, T., & Denham, S. L. (2011a). Abstract stimulus-specific adaptation models. *Neural Computation*, 23(2), 435–476.
- Mill, R., Coath, M., Wennekers, T., & Denham, S. L. (2011b). A Neurocomputational Model of Stimulus-Specific Adaptation to Oddball and Markov Sequences. *PLoS Computational Biology*, 7(8), e1002117.
- Mill, R., Coath, M., Wennekers, T., & Denham, S. L. (2012). Characterising stimulus-specific adaptation using a multi-layer field model. *Brain Research*, 1434, 178–188.
- Monsalve, I. F., Frank, S. L., & Vigliocco, G. (2012). Lexical Surprisal As a General Predictor of Reading Time. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 398–408). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology*, 14(4), 474–480.
- Nelken, I., Yaron, A., Polterovich, A., & Hershenhoren, I. (2013). Stimulus-Specific Adaptation Beyond Pure Tones. In B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, & H. E. Gockel (Eds.), *Basic Aspects of Hearing* (pp. 411–418). Springer New York.
- Palm, G. (2012). *Novelty, Information and Surprise*. Springer Science & Business Media.
- Roark, B. (2011). Expected surprisal and entropy. *Oregon Health & Science University, Tech. Rep, 1*. Retrieved from <http://www.cslu.ogi.edu/people/roark/techrpt-CSLU-11-004.pdf>
- Tribus, M. (1961). *Thermostatics and thermodynamics: an introduction to energy, information and states of matter, with engineering applications*. van Nostrand.
- Tsodyks, M. V., & Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences of the United States of America*, 94(2), 719–723.
- Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, 6(4), 391–398.
- Xu, X., Yu, X., He, J., & Nelken, I. (2014). Across-ear stimulus-specific adaptation in the auditory cortex. *Frontiers in Neural Circuits*, 8, 89.
- Yaron, A., Hershenhoren, I., & Nelken, I. (2012). Sensitivity to Complex Statistical Regularities in Rat Auditory Cortex. *Neuron*, 76(3), 603–615.
- Zilany, M. S. A., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America*, 135(1), 283–286.