

CONTEXTUAL AND TEMPORAL MODULATION OF AUDITORY
RESPONSES IN THE SONGBIRD FOREBRAIN

By

KAI LU

A dissertation submitted to the
Graduate School- New Brunswick
Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Dr. David S. Vicario

And approved by

New Brunswick, New Jersey

May, 2013

ABSTRACT OF THE DISSERTATION

Contextual and temporal modulation of auditory responses in the songbird forebrain

By KAI LU

Dissertation Director:

David S. Vicario

The traditional approach in studies of sensory coding in neurophysiology is to look for correlations between the firing properties of neurons and physical properties of stimuli. However, the current work shows that auditory responses in the auditory forebrain of awake animals in the passive hearing state was dynamically modulated by preceding acoustic experience at varying time scales: 1) Auditory responses to sequences of acoustic events were modulated by the order of events at a time scale of ten to hundreds of milliseconds. 2) Once a stimulus has been presented at a certain inter-stimulus interval (ISI) for several trials, a few seconds delay in the ISI changed the amplitude of the response. 3) Auditory responses were enhanced when a sound was presented along with context sounds from a different category, reflecting categorical prediction of upcoming sounds derived from recent acoustic history over a time scale from

seconds to minutes. 4) Familiarity of contextual sounds, lasting for minutes to hours, changed the strength of responses of other sounds and enhanced the responses to novel sounds in a familiar acoustic context. These results may shed light on three mechanisms that are essential for auditory object perception: 1) temporal integration of acoustic sequences, 2) categorical processing of salient stimuli with a continuum of features, e.g. vocalizations, and 3) segregation of auditory objects from natural acoustic scenes.

ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee, Dr. Randy Gallistel, Dr. John McGann and Dr. Marc Schmidt, for their time and for their sagely advice. I would especially like to thank my advisor, Dr. David S. Vicario, and deeply appreciate his support and advice to my career and for making it an enjoyable journey. I would also like to thank my colleagues and fellow graduate students for all of their help throughout these years. Finally, I would like to dedicate this dissertation to my parents. I feel sorry for not being with them in these years. Without their understanding and encouragement, this venture would not be possible.

Table of Contents

Abstract	ii
Acknowledgements.....	iv
Table of Contents	v
List of Tables	vi
List of Figures	vii
General Introduction	1
Experiment 1 (Contextual effects).....	13
Experiment 2 (Contextual effects2).....	42
Experiment 3 (Acoustic Features Contrast).....	45
Experiment 4 (Effects of familiarity).....	50
Experiment 5 (Timing effects).....	55
Experiment 6 (Sequential Effects).....	59
General Discussion	64
References	70
Figure and Table Legends.....	79
Figures and Tables	89
Appendix: Description of Non-thesis Research.....	119

List of Tables

Table 1	89
---------	-------	----

List of Figures

Figure 1.	90
Figure 2.	90
Figure 3.	91
Figure 4.	92
Figure 5.	92
Figure 6a.	93
Figure 6b.	93
Figure 7.	94
Figure 8.	95
Figure 9.	96
Figure 10.	97
Figure 11.	98
Figure 12.	99
Figure 13a.	100
Figure 13b.	100
Figure 14.	101
Figure 15.	101

Figure 16.	102
Figure 17.	102
Figure 18.	103
Figure 19.	104
Figure 20.	105
Figure 21.	105
Figure 22.	106
Figure 23.	107
Figure 24.	107
Figure 25.	108
Figure 26.	109
Figure 27.	109
Figure 28a.	110
Figure 28b.	110
Figure 29.	111
Figure 30.	102
Figure 31.	102
Figure 21b.	103
Figure 32.	113
Figure 33.	114
Figure 34.	114
Figure 35.	115
Figure 36.	115

Figure 37. 116

Figure 38. 117

Figure 39. 117

Figure 40. 118

Figure 41. 118

General Introduction

In nature, organisms must be able to recognize and discriminate salient acoustic information, ranging from communication signals produced by individuals of the same species to the sounds of a predator, in an acoustic environment filled with noise from many sources, e.g. wind or other animals. The energy from all of these different sound sources is summed to reach the ears of an organism as a single pressure signal that varies in time. For the organism to detect any individual acoustic event as a distinct object requires integrating information at multiple time scales and across a wide frequency range in order to extract specific patterns from variable backgrounds that contribute to the combined signal. Understanding how the auditory system achieves these tasks may have fundamental implications for how neural systems in general represent the external world.

However, our knowledge of the neural substrate of auditory object perception is very limited. The first challenge is to define an auditory object. In vision, what we see are interpretations of the physical world by the visual system, so a visual object could be considered to be a perceptual projection of a physical object. Similarly, in audition, an auditory object could be defined as a perceptual entity coming from a given sound source. Echolocation in bats gives the best-studied example in this sense (Suga and O'Neill, 1979). However, in a more general sense, sounds often transmit a signal, e.g. the sound of a threatening predator or a meaningful word in human speech. Thus, an auditory object could be defined as an invariant pattern shared by a class of sounds that carry the same piece of information (Griffiths and Warren, 2004), taking the perception of a given vowel as an example. From this point of view, perception of an auditory object not only involves detecting

individual features, but also binding multiple features together to form a perceptual entity and generalizing common features from different specific sounds within the same category.

While we know a lot about sensory transduction in the auditory periphery and the receptive fields of neurons at various levels of the ascending auditory projection, we do not have a good theory of the object processing that must emerge at the forebrain level (Nelken and Bar-Yosef, 2008). Until quite recently, with only a few exceptions (Wollberg and Newman, 1972; Chew et al., 1995; Wang et al., 1995, Rauschecker, 1998; Fritz et al., 2005), the cortex has been studied by using the same simple auditory stimuli that were effective for analyzing the auditory periphery (Mendelson and Cynader 1985; Schreiner, 1988; Shamma et al., 1993; He et al., 1997; Schreiner et al., 2000 for review). Although these studies reveal how auditory neurons represent the most basic sound features, such as pitch, frequency modulation, amplitude modulation and duration, we still know very little about how a complex natural sound, e.g. a vocalization, is represented as an auditory object in the brain and is successfully discriminated from the whole auditory scene.

To successfully represent an auditory object, such as a natural vocalization, the brain has to solve three problems: 1) it has to extract the invariant from the same class of sounds; 2) it has to separate the auditory object from the natural acoustic background; 3) it has to bind features across time into one perceptual entity. The first problem for auditory object perception is due to the variability of natural sounds. As discussed above, a given perceptual object may consist of a class of sounds. In human speech, people can speak the same word with a different pitch, accent or speed, etc (Goldstein, 2007). It has been found that human brains can solve this problem by sorting sounds into categories. Brains are sensitive to

variance of sounds across the boundary of different categories, but neglect the same amount of variance within the same category (Miyawaki et al., 1975). A study by Maye et al. (2002) shed light on how these kinds of perceptual categories develop in the infant's brain. Six- and 8-month-old infants were familiarized with sounds that varied on the entire continuum of voice onset time (VOT) of artificial syllables. However, one group of infants heard stimuli that had a bimodal distribution along an acoustic parameter, while the other group heard stimuli that had a unimodal distribution. The results show that only infants with the bimodal experience discriminated the sounds at the two ends of the continuum. Therefore, brains of infants can detect the statistical distribution of sounds and use this distribution to form perceptual categories. Interestingly, categorical perception is not a property unique to the human brain. Both mammals and birds have been found to be able to process sounds categorically (Kuhl, 1975, Prather, 2009, Vicario, 2004). In this context, it would be interesting to study whether animals' brains refine the perceptual categories by learning the statistical distribution of sounds as human infants do.

The second problem that brain has to solve in order to perceive an auditory object is to separate a sound from the natural acoustic background. In the natural acoustic world, no sound is presented by itself. Acoustic signals are embedded in the larger acoustic context and mixed with other sounds. Recognition of a sound depends on two complementary processes: 1) separating different acoustic objects from each other and from the background; and 2) integrating multiple acoustic components into the perception of a unitary acoustic object. This process is called auditory scene analysis (ASA); Bregman (1990) proposed that sounds can be decomposed into different streams or groups by two different

kind of process, dependent on whether experience and memory are involved. The first process use cues including sound location (computed from binaural phase differences), differences in fundamental frequency, non-matching harmonic components, asynchrony of sound onset and discontinuity of amplitude modulation and frequency contours (Bregman, 1990). Since the process of segregation that uses these cues depends on bottom-up and pre-attentive mechanisms, it is called “primitive” auditory scene analysis.

The other mechanism through which it is proposed that the sounds can be segregated is top-down and memory-based, called “schema-based” auditory scene analysis. In this process, segregating sounds may depend on learned control of attention. Although knowledge about the mechanism of “schema-based” auditory scene analysis is lacking, a body of evidence has shown that the statistical distribution of previously-experienced sounds may affect the strength of auditory responses. We proposed that this modulation of response strength by the statistical distribution of sounds may affect discriminability of sounds and could function as the underlying mechanism of “schema-based” auditory scene analysis (Lu and Vicario, 2011). Distribution-dependent modulation of auditory responses has been observed both in animals and in humans, from cellular level to whole brain level, as described in the following paragraphs.

In anesthetized cats, neurons of A1 responded more strongly to rarely presented sounds in an “oddball” paradigm than to the same sounds when they were common (Ulanovsky et al., 2003). A more recent study in the anesthetized songbird showed that the statistics of natural sounds may also affect auditory selectivity. Responses of auditory neurons in a songbird auditory forebrain area, the caudal lateral mesopallium (CLM), can be

more strongly modulated by natural song stimuli containing transitions that violate statistical expectation, as estimated from a large sample of conspecific song (Gill et al., 2008). This suggests that CLM responses to a given sound depend on the probability that this sound appears in the real acoustic world of the bird's experience, although a role for the bird's biological species cannot be excluded.

The most striking evidence that the response strength of auditory neurons depends on experience is from another auditory area, caudomedial nidopallium (NCM), in birds. When a specific novel song (the unique learned song of another individual) is presented repeatedly to an awake bird, both neural responses to the song and induction of ZENK (an immediate early gene) in NCM are reduced (Chew et al, 1995, 1996; Mello et al., 1992). This reduction is a form of stimulus-specific adaptation (SSA) with the following interesting properties: 1) The adaptation effect is stimulus-specific and the generalization effect is very weak. In other words, adaptation to a repeated song does not reduce the response to a different novel song presented subsequently, although both songs share similar spectro-temporal structures and are drawn from the class of zebra finch songs (Chew et al., 1995). Independent adaptation responses have been seen for as many as 16 songs (Chew et al, 1996). Therefore, the responsiveness of a neuron to each individual song can be modulated by SSA. A novel song is more likely to induce stronger responses in NCM neurons than a familiar song. 2) SSA in NCM occurs for most complex sounds, but it is also a long-term effect that lasts much longer for relevant stimuli. In contrast to the phenomenon - a short term effect over a few seconds - found by Ulanovsky et al.(2003), reduced responsiveness to familiar conspecific songs in NCM could be maintained for 24hrs or longer, while SSA to heterospecific vocalizations

lasted much less time - no longer than 6hrs (Chew et al., 1995). Therefore, SSA reflects the acoustic experience of the animal and is durable for salient sounds; it may be an important mechanism for modulating responsiveness to a given auditory object.

A related phenomenon from human research is the mismatch negativity (MMN), a component of event-related potentials that reveals statistically-dependent modulation of auditory responses at the whole brain level (Naatanen, 1992). MMN is induced by a deviant stimulus that followed a predictable series of unvarying stimuli and is considered to contribute to attentive novelty detection. Although, MMN may reflect activity from multiple brain areas, adaptation found at the neural level can explain some aspects of MMN and thus may play an important role in mechanisms of MMN (Nelken and Ulanovsky, 2007).

In all these examples, over different time scales, novelty increases response strength to an auditory stimulus, while familiarity decreases the auditory response. This mechanism may have essential ecological meaning in auditory processing. Repeatedly presented sounds that do not induce any behavioral consequence may have reduced salience over these repetitions, while novel sounds may signal changes and have potential behavior salience. Therefore, in a natural acoustic scene which consists of both novel sounds and non-salient familiar sounds, not all auditory objects would be treated equally by the auditory system. Stronger auditory responses to novel sounds would enhance the chance to detect novel stimuli and thus may benefit the survival of organisms. We consider this mechanism as a form of 'schema-based' auditory scene analysis.

In the examples discussed above, it should be noted that the familiarity of sounds is highly related to the predictability of the sounds. In Gill et al. (2008) the stimuli were directly manipulated to violate statistical expectation. In Ulanovsky et al. (2003) and Naatanen (1992) the response strength to the novel sounds was also dependent on the order in which the stimuli were presented. The more likely that a given sound can be predicted by the preceding sequence of sounds, the greater the reduction of responses that could be induced, and vice versa. Therefore the modulation of auditory responses depends on the acoustic history, in other words, the dynamic statistical distribution of sounds over a certain time range. The only apparent exception in the above examples is SSA in NCM. In NCM, it seems that all stimuli adapt independently; however, no direct test has ever been carried out to assess whether the order of stimulus presentation has an effect on auditory responses or on SSA in songbird NCM.

It is important to note that the predictability of sounds not only means predicting “what sound will be heard next”, but also includes predicting “when the next sound will be heard”. In all the experiments discussed above, the inter-stimulus intervals were fixed. Thus, what was directly manipulated was event uncertainty, but it has been found that temporal uncertainty can also enhance auditory responses. McCarthy and Donchin (1976) found that the ERP induced by machine-delivered stimuli at random intervals was larger than identical self-administered auditory stimuli. However, in this experiment, temporal uncertainty was not manipulated independently and may have interacted with event uncertainty. It would be interesting to test the effect of temporal uncertainty by itself on auditory responses. This question is not only important for auditory research, but also related to the fundamental

question of how the brain represents time. It has been recognized that understanding the neural mechanism of temporal representation in brains may be crucial for understanding basic principles of learning and memory (Balsam and Gallistel, 2008; Gallistel and King, 2009). Although behavioral studies have accumulated a body of evidence showing that animals can predict the timing of events (Balsam and Gallistel, 2008), relevant neural models are still limited. Moreover, because the representation of time may be distributed in multiple brain areas (Mauk 2004), convergent studies from different neural models and different sensory modalities is likely to be important for understanding this fundamental question. If the auditory system could be found to be sensitive to time variance at a sub-second scale, it would provide a potential neural model for studying the neural representation of time at a fine resolution, at least in the auditory domain.

The third problem for perception of an auditory object is integration of features distributed across the temporal dimension. Understanding the temporal integration at the scale of tens to hundreds of milliseconds is essential for understanding the mechanism of language decoding in the auditory system. Speech usually consists of sequences of sounds. Perception of a meaningful word in a spoken language not only requires identifying relevant acoustic features, but also the ability to order sounds in a sequence and detecting sound order in a sequence at a very fine time scale (McAdams and Bigand, 1993). Interestingly, processing sound sequences is not an ability unique to humans. The songbird is the most important animal model that can match humans on this aspect of processing. Birdsongs, like human language, are rapid sequences of sounds that obey complex syntactic rules (Zann, 1996; Okanoya, 2004; Gentner et al., 2006). Therefore, songbirds must have the ability to decode the

sequence of sounds in auditory domain (Rose et al, 2004; Plamondon et al. 2010).

Electrophysiological data have shown that auditory neurons in high vocal center (HVC) of songbirds are sensitive to the order of sounds (Margoliash and Fortune, 1992). Therefore, the auditory system of songbirds may provide a good model for understanding the neural mechanisms that underlie the temporal processing of sound sequences.

The current work uses the zebra finch auditory system to explore four independent, but related questions that are important for understanding neural mechanisms of auditory object perception. They are: 1) how brains process sounds categorically based on the statistical distribution of sounds; 2) how brains utilize auditory experience to achieve “schema-based” auditory scene analysis; 3) whether auditory neurons are sensitive to temporal uncertainty; 4) how brains process sequences of sounds at time scales of tens to hundreds of milliseconds. The first two questions are related to the central problem of auditory object perception: categorical processing may be the solution to the problem of variability of auditory objects; experience-based attention control may provide an important mechanism for segregating an auditory object from its background. These questions are fundamentally related. Both categorical processing and the role of auditory experience require some kind of representation of the statistical distribution of sounds. The third question, the effect of temporal uncertainty on auditory responses, is related to the mechanism of attention control although it depends on a representation of time. The fourth question is not only related to the question of sequential auditory segregation, but also important for finding the window over which auditory features are integrated in time.

The songbird auditory forebrain as a model system for auditory processing

The proposed studies will use the songbird auditory forebrain as a model to explore the four questions outlined above. Although the songbird forebrain does not have a layered cortex like that of mammals, anatomical homologies in the brainstem are well established and functional analogies can be documented. In the auditory forebrain (Figure 1), field L2 may be homologous to thalamo-recipient layer III-V of primary auditory cortex because it receives direct input from the auditory thalamus, called nucleus Ovoidalis (OV) in birds and homologous to the medial geniculate nucleus of mammals (Wild et al., 1993; Vates et al., 1996; Wang et al., 2010). Field L2 neurons have phasic auditory responses with defined frequency tuning curves and a tonotopic organization, very similar to the responses of A1 neurons (Muller and Leppelsack, 1985; Terleph et al., 2006). In addition, selectivity for some complex features has been described for field L (Leppelsack and Vogt, 1976; Sen et al., 2001; Woolley et al., 2009).

In the next stage of auditory processing, field L2 projects to closely associated fields L1 and L3, which in turn project to further auditory areas, including the caudomedial nidopallium (NCM) and caudal mesopallium (CM). NCM may be considered an analog of superficial layers of A1 (Wang et al, 2010) or of a secondary mammalian auditory region (Theunissen et al., 2000). There is no strong histological evidence supporting either of the opinions. NCM receives major inputs from the field L complex (Saini and Leppelsack, 1977; Mello et al 1992; Vates et al., 1996) and possibly some direct input from the auditory thalamus. NCM also reciprocally connects with CM. Although electrophysiological responses in NCM show a rough tonotopic organization that parallels that of field L (Muller and Leppelsack, 1985;

Terleph et al., 2006), tuning functions obtained with simple stimuli are broader and often multi-peaked. Furthermore, complex stimuli typically elicit stronger response than simple stimuli (Chew et al, 1996; Muller and Leppelsack, 1985).

Although NCM responds to many types of sounds, neurons there consistently give stronger responses to conspecific than heterospecific songs or other sounds (Chew et al, 1996), suggesting that it is specialized for these stimuli. Single unit data show that some neurons are responsive to certain elements in songs and some are selective for forward versus reversed playback of the same song (Stripling et al., 1997). These data are complemented by IEG studies showing greater IEG gene induction by conspecific than heterospecific songs (Mello et al., 1992). Auditory responses in NCM have another striking property: stimulus-specific adaptation (as described earlier), which is longer lasting for the unique songs of individual conspecifics than for other sounds (Chew et al, 1995). Taken together, these findings imply that NCM responds selectively not only to a particular class of sounds, but to individual exemplars of that class, suggesting that NCM neurons may represent auditory objects, rather than simple auditory parameters.

CM is considered to be an analog of the superficial layers of A1, based on anatomical evidence (Wang et al., 2010). The lateral part of CM (CLM) is reciprocally connected with Field L, while the medial part of CM (CMM) is reciprocally connected with NCM (Vates et al., 1996). CM shows a tonotopic organization (Muller and Leppelsack 1985) and higher selectivity for complex sounds than field L (Sen et al., 2001). Moreover, a recent study shows that CM neurons may be selective to target stimuli in associative learning (Gentner and Margoliash, 2003 Jeanne et al., 2011). CM neurons also show SSA similar to NCM

(unpublished data by Phan). These properties imply that CM may be also selective to individual auditory objects. Since Gill et al (2008) showed CLM neurons are sensitive to violations of statistical expectation, so we will focus on this part of CM in the current experiments. NCM and CLM are excellent neural models to explore the questions proposed earlier for four reasons. First, both areas show some degree of selectivity to individual auditory objects. Second, both areas show SSA and thus are sensitive to the acoustic history. Third, both areas show preference for conspecific songs over hetero-specific songs, so have the potential to process sounds categorically. Fourth, since at least some NCM neurons show selectivity for forward versus reversed playback of the same song, this area may have the potential to encode the order of sounds in a sequence.

Experiment 1

1. Introduction

The primary aim of the first experiment was to test the effect of acoustic context on auditory responses. As described above, it has been found that an odd ball sound in the context of another common sound induced higher responses in A1 of cats (Ulanovsky et al. 2003).

We first hypothesized that responses of neurons in the auditory forebrain of songbirds are also modulated by the probability of a sound in a context. Second, because stimulus-specific adaptation found in auditory forebrain of birds was persistent for hours or days, we assumed that SSA was a process independent to the probability of sounds in the short-term history. While the long-term properties of SSA may enable neurons to work as a counter of auditory events that functions to record the statistical distribution of these events on a large time scale, context-dependent auditory modulation may work as a short-term detector for unexpected sounds. Thus we hypothesized that context-dependent modulation on auditory responses and adaptation may be co-existed and parallel to each other. Third, since auditory forebrain of songbirds show preference in a particular perceptual category (viz. Conspecific songs) in terms of their response strength and endurance of long-term adaptation, we hypothesized that context-dependent auditory modulation in the auditory forebrain of songbirds may depend on some form of categorical processing. When a “deviant” sound is presented in a context of sounds from a different category, enhanced auditory responses may be induced. Furthermore, because this type of paradigm involves manipulating the relative frequency of

different stimuli in a series, it effectively changes the inter-stimulus interval (ISI) for any given recurring stimulus. Thus it introduces temporal as well as event uncertainty. Therefore, our experimental design for testing acoustic context effects needed to include the possible interaction of acoustic context with the ISI. Thus, we also hypothesized that changes in intervals between stimuli increase temporal uncertainty of stimuli and may enhance auditory responses. We have tested these interconnected hypotheses in two auditory forebrain areas, NCM and CLM of zebra finches.

2. Experimental design

Adult male zebra finches were surgically prepared for awake recording. Each bird was tested in 5 different acoustic context conditions, each consisting of 3 sessions: 1) a pre-adapting session, 2) a context-modulated session and 3) a post-context session. In the pre-adapting session, animals were presented with 20 repetitions a novel conspecific song (target) at a fixed ISI of 7 seconds. Measurement of responses to the target song enabled computation of an SSA function whose slope could be used to predict subsequent responses. Immediately following, in the context-modulated session, the target song was repeated for 19 trials in one of five acoustic contexts that varied in stimulus order and timing. Finally, in the post-context session, the same test song was presented for 20 trials (7s ISI). Comparison of the responses to the target song in the three sessions was used to quantify enhancement effects produced by context manipulations (detailed methods described below).

The stimulus order for five different context conditions is shown in Figure 2. In the first condition (ZCR, **Z**ebra finch song in a **C**anary context with **R**andom intervals), a pre-adapted

zebra finch song was presented as part of a set that included 7 novel canary songs as context; 19 repeats of the pre-adapted song and 20 repeats of each context stimulus were included, for a total of 159 trials presented in a randomly shuffled order. We used canary song as context sounds, because canary songs and zebra finch songs differ in their acoustic features (Figure 3) and neurons in NCM showed different responses to these two types of songs (Chew et al., 1996). Although the ISI was fixed at 7s, effectively, the intervals between repeats of the zebra finch song varied from 14 -161s. In the second condition (ZSR, a **Z**ebra finch song in a **S**ilent context with **R**andom intervals), all canary songs were replaced by silence, but all stimuli were presented in the same order as the first condition so that the intervals between the pre-adapted songs remained the same. In the third condition (ZCF, a **Z**ebra finch song in a **C**anary context with **F**ixed intervals), the pre-adapted song was presented with the 7 canary songs as in the first condition, but the intervals between the pre-adapted songs were fixed (56 seconds), while the order of intervening canary songs was randomized. In the fourth condition (ZSF, a **Z**ebra finch song in a **S**ilent context with **F**ixed intervals,), the pre-adapted songs were presented with 56 seconds fixed intervals as the third condition, but there are no canary songs between the zebra finch songs. In the fifth condition (ZZR, a **Z**ebra finch song in a **Z**ebra finch context with **R**andom intervals), the pre-adapted song was played in the context of 7 other zebra finch songs that were novel for the bird, and the order of stimulus presentation was randomized.

Based on this experimental design, we have four predictions. First, if responses to songs can be modulated by a context consisting of sounds from a different category (viz. canary songs), we will see larger enhancement effect on auditory responses to the pre-adapted

zebra finch target song in ZCR than when the context consists of zebra finch songs (ZZR).

Second, if increased temporal uncertainty can enhance auditory responses, we will also observe enhanced responses to the zebra finch songs when varying periods of silence come between each song presentation (ZSR). Third, if the effect of context and interval changes sum together, then the responses in ZCR should be higher than responses in ZCF and ZSR, because, in ZCR, both context and interval are uncertain and may induce higher responses, while in the other two conditions, only one factor would induce higher responses. Fourth, ZSF would show the lowest effect because, in ZSF, there is no context effect, and temporal uncertainty is only limited to the single transition from the first (7s ISI) to the second (56s ISI) session. Therefore, we expect enhancement of auditory responses in the five conditions to be ranked in the following order: ZCR>ZSR~ ZCF and ZZR>ZSF. Finally, if contextual and temporal modulations on auditory responses are independent of SSA, the reduction of responses in the last session relative to the first session due to stimulus repetition would not be affected by acoustic context and interval changes.

3. Materials and Methods

Subjects

The subjects were 11 adult zebra finches obtained from our aviary or from the Rockefeller Field Research Center. All animals were housed on a 12:12h light-dark cycle. Food and water were provided ad lib and the experiment conformed to an approved protocol from the Institutional Animal Care and Use Committee of Rutgers University.

Surgery

Two days before the electrophysiological recording, each animal was anesthetized with isoflurane (2% in oxygen). Then a head post was attached to the skull with dental cement. A chamber was formed with dental cement around a small craniotomy window that exposed the area of the bifurcation of the midsagittal sinus. The chamber was then sealed with silicone elastomer (Kwiksil, World Precision Instruments, USA).

Electrophysiology

Electrophysiological recordings were made in an acoustically isolated sound booth beginning 48h after the initial surgery, to allow for full recovery from anesthesia. The awake animal was immobilized in a comfortable tube and the implanted post was used to secure the head of the animal to the stereotaxic frame. Recordings were made at 16 sites (4 each in the left and right NCM, 4 each in the left and right CLM) using glass insulated platinum/tungsten microelectrodes (2-3 MOhms impedance, Ekhorn design, Thomas Recording, Germany) independently controlled by a multielectrode microdrive. Electrode signals were amplified and filtered (0.5-5kHz), then acquired at 25kHz using Spike 2 software (CED, England).

White noise stimuli with the amplitude envelope of canary song were presented to search for responsive sites that indicate the dorsal boundary of the auditory forebrain. Once all electrodes were placed at responsive sites, the context experiment was performed. The order of the five conditions was randomized for each subject. At the end the recording, eight small electrolytic lesions (20uA for 15s) were made to enable histological reconstruction of recording sites.

Histology

At the conclusion of the experiment, the animal was killed with an overdose of Nembutal, then perfused with saline and paraformaldehyde. Sagittal sections were cut from the fixed brains at 50um on a Vibratome, then stained with cresyl violet. Lesion sites in NCM and in CLM were confirmed histologically based on cytoarchtectonic landmarks.

Data analysis

For each channel, the root-mean-square (RMS) of the multi-unit neural activity was calculated both over a control window (the 500 ms period prior to stimulus onset) and over a response window (from stimulus onset to stimulus offset plus 100 ms) on each trial. The RMS was computed by squaring the value at each time point, taking the mean of these squares for each window, then taking the square root of each mean. This provides a method of rectifying the multi-unit activity and computing its average power. Responses to song stimuli were quantified as the difference between the control and response RMS measurements (Figure 4).

In order to remove artifacts due to movement of animals during recordings, we first took the regression for responses to the repeated song on trial 6-59 and took the residuals. Responses that were higher than or lower than 2.5 standard deviation of residuals were considered as outliers and were deleted.

The effect of different context manipulations on auditory responses was measured by quantifying how each response during the Context-modulated session deviated from the estimated distribution of responses derived from the responses in the Pre-adapting session and the Post-context session. This was computed as the “surprisal”, a measure from information theory (Levy, 2008), according to the following procedure: 1) The linear regression line for the responses in the Pre-adapting session was computed from the responses to the repeated “test” song during the linear portion of the adaptation function (**Figure 5**; trials 6-20). This line was extrapolated to estimate the response on the first trial of the Context-modulated session (Fig. 5, circle at trial 21). 2) The linear regression was computed from responses to the test stimulus in the Post-context session (trials 40-59), and then extrapolated backwards to estimate the expected response on the last trial (Fig. 2B, circle at trial 39) of the Context-modulated session. 3) The expected responses in the Context-modulated session were estimated by the line connecting the estimates for trials 21 and 39 (Fig. 2B, green line, called hereafter the interpolated regression). 4) The expected standard deviation of the responses about the interpolated regression line was estimated by the standard deviation of pooled residuals of the regressions of the Pre- and the Post-sessions; 5) An observed response that falls on the interpolated regression is the least surprising (most expected) response; the greater the deviation (d) of an observed response from this expectation, the more surprising it is. The degree to which it is surprising is an accelerating function of d , namely, $\log(1/P(d))$, where $P(d)$ is the probability density of d in the assumed-to-be normal distribution (see Figure 5). 6) Responses greater than expected are assigned positive surprisal, while responses less than expected are assigned negative

surprisal.. 7) The average magnitude of a surprisal is greater for diffuse distributions (distributions with large standard deviations) than for tight distributions, so we normalize the surprisals by subtracting from them the absolute value of the minimum surprisal, which is $\log(1/P(0))$. Normalization makes the surprisal of an observed response that exactly conforms to expectation 0, and it zeros the expectation of the signed surprisals when observed responses are drawn from the expected distribution. 8) Thus, our formula for the normalized signed surprisal of an observed response is : $DS = \text{sign}(d) * \log(1/P(d)) - \text{abs}(\log(1/P(0)))$.

If there is no enhancement effect due to context and temporal uncertainty, DS should approach zero. Modulations that increase responses will have $DS > 0$. In order to test whether the enhancement effect interacts with SSA, we also computed the Adaptation index (AI) for each condition at each recording site. First, we estimated the response amplitude of ending trial of context-modulated session (trial 39) and the start trial of the session (trial 21). Then the ratio of the response of trial 39 to trial 21 was taken (see Figure 5).

Single-unit analysis

To verify the results from multi-unit data, single-units were also isolated from the same electrode recordings. Single-units were detected by template-based digital clustering algorithms implemented in the Spike2 and verified as single-units by analysis of the inter-spike interval histograms. To be accepted, a unit had to have contamination rate lower than 1%. Contamination spikes were identified by ISpls less than 2ms (corresponding to spike rates $> 500\text{Hz}$), the criterion for contamination by another unit. The response amplitude

of each unit was quantified as the spike rate in the response window (from stimulus onset to stimulus offset plus 100 ms) for each trial. DSs for single units were calculated in the same way as for multi-units. Spectro-temporal receptive fields (STRFs) for single-units were calculated by a customized Matlab toolbox (STRFPak5.3. Theunissen and Gallant laboratories at UC Berkeley. URL: <http://strfpak.berkeley.edu/>). A STRF represents the spectro-temporal features of sounds that excite or inhibit a given neuron. It is calculated using the reverse correlation algorithm: a large, varied set of sound stimuli is presented and then sound components that precede spiking activity in a given neuron are averaged and adjusted by the autocorrelation function of the stimulus (Theunissen et al., 2000). A basic assumption of methodology of STRF is that neurons' firing rate has a linear relationship with acoustic features. . This assumption may be problematic because temporal pattern of spike trains may carry more information than spike rate. If an acoustic stimuli lead to changes of temporal patterns rather than spike rate in responses, then the strategy of averaging stimuli before a given spike would be failed. By now, the most success in calculating STRF with high predictability was achieved in midbrain or thalamo-recieipient layer of the forbrain (Theunissen et al., 2000; **Woolley et al., 2009**), in which auditory responses are more phasic than higher auditory areas, such as NCM. Moreover, these successful STRF calculations had been done in anesthesia status that made responses even more phasic. However, NCM showed tonic responses in awake animals even to simple stimuli (Terleph et al., 2006). Complex temporal patterns may exist in the tonic phase of auditory responses, based our preliminary observations. Thus, our analysis with STRF calculation is based on a restrained

model any may only provide a limited interpretation of response properties of auditory neurons in NCM and CLM.

Statistical approach

In data analysis, in order to provide an overview preferred by traditional electrophysiological approach, we first averaged DSs across trials on each sites and pooled DSs of all sites across birds. Many of our datasets do not, or may not, meet criteria for parametric tests. Therefore, appropriate non-parametric statistics were used. In a related decision, data are presented as cumulative frequency distributions which reveal much more detail than mean and standard error plots, as well as illustrating group differences. For Experiment 1, the DSs obtained for each recording site in each of the five conditions were treated as five repeated measures. The main effect was tested by the Friedman ANOVA, which does not require the data distribution to be normal. Post hoc differences between groups where data were matched (e.g. different conditions recorded at the same electrode site) were tested by the Wilcoxon matched pairs tests. In order to show how an effect is consistent from bird to bird, we also took the differences between conditions on each site and present the distribution of DS differences by sites within each bird by the box plot that shows how the effect vary across birds.

Second, in the further step, we conduct comparisons between different conditions on each single recording site in which trial-by-trial DSs in two conditions were treated as repeated measure and tested by Wilcoxon matched pairs tests. These tests showed whether trial-by-trial DSs of two conditions in one recordings site were significantly different

from each other. Percent of sites in each birds showing significant effect in each contrast are plotted in stack bar plot that shows how consistent of an effect across sites and across birds. Sites with similar patterns from different birds were classified together and then compared on other dimensions.

For single unit comparisons, where samples were not explicitly matched, we used the two sample Kolmogorov-Smirnov test. In order to find differences between NCM and CLM and differences between two hemispheres, we need to know the interaction between regional difference and conditions. As far as we know, there is no non-parametric approach available for testing the interactions of two factors, we decided to use two-way Repeated-Measures ANOVA, in which regional difference was treated as factors and DSs of five conditions were repeated measures.

4. Results

Main effects in DSs

In our recordings, 111 sites recorded in NCM and CLM that responded to auditory stimuli were analyzed. Figure 6a and 6b showed an example of histology confirmation of recording sites in NCM and CLM. Data were first analyzed as individual sites considered as independent observations. A two-way Repeated-Measures ANOVA showed no significant differences between NCM and CLM ($p=0.59$) and no interactions between regions and conditions ($p=0.17$). There was also no difference between two hemispheres ($p=0.633$) and no interactions between hemispheres and conditions ($p=0.3$). Thus data from NCM and CLM of both hemispheres were combined for the further analysis. The enhancement effect

quantified as DSs for sites are plotted as cumulative frequency distributions for each condition in Figure 7. The first four conditions in which zebra finch songs were embedded in the canary song or silence contexts showed strong enhancement effects (more than 90% of sites showed DS larger than zero), while the last condition in which the zebra finch song was presented with other zebra finch songs (ZZR) showed the lowest enhancement effect (black line, 60% sites showed DS larger than zero).

A Friedman ANOVA performed with these data showed significant differences across five conditions (Chi Sqr. ($N = 111$, $df = 4$) = 129.9, $p < 0.001$). Post hoc Wilcoxon tests showed that ZCR was significantly different from ZSR, ZCF and ZSF ($p < 0.001$ in all tests). ZSR, ZCF and ZSF were not significantly different from each other ($p > 0.5$ in all tests), but all of them were significantly different from ZZR ($p < 0.01$, as shown in Table 1). Therefore, DSs of the five conditions were ranked as $ZCR > ZSR$, ZCF , $ZSF > ZZR$.

The most interesting observation is that the effect of zebra finch song context was dramatically different from the effect of canary song context on auditory modulation. First, the largest difference observed in all comparisons was the contrast between the ZCR condition and the ZZR condition (the ratio of difference in the medians to the average interquartile interval is 0.58). Furthermore, zebra finch context showed a negative interaction with interval changes. DSs in ZZR were even lower than those in ZSF in which zebra finch songs were just presented in longer but fixed intervals in the context-modulated session. Rather than enhancing the effect induced by interval changes, the context of zebra finch songs actually significantly reduced the enhancement effect due to interval uncertainty. Therefore the

conspecific song context may be categorically different from hetero-specific song context in auditory processing.

Furthermore, our results showed an interaction of canary context and interval change on enhancement effect. Given that enhancement effect observed in ZSF was due to elongate intervals, either random intervals in ZSR or canary context added between the target song with fixed intervals in ZCF induced more enhancement than ZSF. Instead, only the condition with both canary context and random intervals in ZCR induced enhancement effect higher than ZSF.

Duplicability across birds

Matched contrasts based on averaged DS in each site are plotted as boxplot for each bird in Figure 8. The contrast between ZZR and other four conditions (Figure 8 D, G, I and J) were consistent across 9-11 birds, which confirmed the results by pooling all sites together. In contrasts between ZCR and ZSR, ZCF and ZSF (Figure 8 A, B and C), 7, 10 and 8 birds, respectively, showed preference for ZCR over other three conditions (measured in median), suggesting that there is individual differences across birds. But, the observation that ZCR showed higher DSs than ZSR, ZCF and ZSF is not likely due to chance.

Duplicability across sites

In further step, we also made comparisons trial-by-trial DSs between conditions on each site. Ten contrasts were made between each pair of the five conditions. In each of contrast, the percent of sites in each bird showed significant differences were shown in Figure 9. These trial-by-trial comparisons on each site showed a general picture that were consistent

with results based on analysis of averaged trial data. First, in the four contrasts with ZZR, no more than 8% of sites showed preference for ZZR. Second, in contrast between ZCR and ZSR, ZCF and ZSF, 38%-43% sites preferred ZCR. Therefore, our results are duplicable across sites.

Classification of sites into functional groups

Furthermore, in all contrasts, we found sites that showed significant effect in two opposite directions and these sites distributed in different birds. These observations suggested that neurons recorded from sites with opposite trends may be from different functional groups. Thus, we classified sites in each contrast that showed significant effect in the two directions. Then we compared the distribution of the two classes of sites in NCM and CLM. Chi square test of independence showed a significant interaction between regions and classes of sites in the contrast between ZCR and ZSF (Chi Sqr. (N=74,df=1)=6.84, $p<0.01$) and in the contrast between ZSR and ZSF (Chi Sqr. (N=68,df=1)=6.17, $p<0.05$). In the two contrasts, there are more sites in CLM that showed significant higher DSs in ZSF. Moreover, we compared the distribution of the two classes of sites in the two hemispheres. A significant interaction between hemispheres and classes of sites was found in the contrast between ZCR and ZCF (Chi Sqr. (N=64,df=1)=8.37, $p<0.01$). In this contrast, there are more sites in the left hemisphere that showed significant higher DSs in ZCF and opposite is true in the right hemisphere.

Enhancement effect is independent from adaptation

Our experiment used the predicted trajectory of stimulus-specific adaptation to calculate enhancements due to context manipulations. Thus, another important question is whether the enhancement effect is independent from or interacts with SSA. If they are not two independent processes, we would predict that the enhancement effect may slow down adaptation that would be produced by presenting the target song 19 or 20 times during the context-modulated period. In this case, the context that produces the largest enhancement effect would also show the smallest drop in response amplitude from the start of the context-modulated session to the end of the session. We can measure this drop by calculating the adaptation index (see Methods). We can then calculate the correlation coefficient between the adaptation index and the enhancement index site by site within each condition. Thus, there were 55 correlation coefficient values. They were nearly distributed symmetrically around zero (Figure 10). In two cases, there was a significant negative correlation (95% confident intervals lower than zero); in one case, a significant positive correlation (95% confident intervals higher than zero); and in the remaining 52 cases no correlation. The negative and positive correlations were not associated with any particular bird or conditions, thus may occur for random reasons. Therefore, we concluded that there was no systematic relationship between adaption and the enhancement effect and that the two processes were independent from each other.

Temporal characteristics of enhancement effect

Next, we analyzed temporal characteristics of the responses seen with context manipulations by computing the difference between the averaged temporal waveform of responses to the target stimulus in the Pre-adapting session and in the Context-modulated

session across all sites. We used the following procedure to accomplish this: 1) We first computed the moving average RMS (10ms window) of the multi-unit recording to produce a smoothed RMS waveform of the response to each stimulus at each site; 2) Averages of these waveforms were then computed both across the last 6 trials of the Pre-adapting session and across the first 6 trials of the Context-modulated session. We chose 6 trials for averaging, because testing showed this number maximized the difference in responses between two sessions. 3) Then the waveforms in the Pre-adapting session and in the Context-modulated session was averaged across all recoding sites respectively. Because the characteristics of stimuli and the response properties of sites were cancelled by averaging, comparison of the averaged waveform in the Pre-adapting session and the averaged waveform in the Context-modulated session provide us a temporal profile only associated with the enhancement effect ("enhancement profile").

We hypothesized that, first, if the enhancement effect is not due to adaptation or fatigue, we would expect that the latency of the onset and peak of the enhancement effect would be longer than the latency and peak of the auditory response. Second, because only temporal uncertainty is involved in inducing the enhancement effect in ZSR and ZSF and no analysis of acoustic information would be required to discriminate a difference, we expected the temporal profile of enhancement ("enhancement profile") for these two conditions would have a shorter latency and earlier peak than in ZCR and ZCF.

For each of the 5 conditions, the enhancement profile was computed as the difference between the averaged response waveforms in the Context-modulated session and the Pre-adapting session (**Figure 11A**). Because we found no differences between NCM and

CLM in the enhancement profile for any condition, data from the two brain areas were combined. The latency of the response waveform and the enhancement profile was defined as the time from stimulus onset until the signal crosses a threshold, computed as the maximum value of the 99% confidence interval for each signal during the control window. As shown in the example for ZCR (**Figure 11D**), the latency of the first increase in the enhancement profile was longer than the latency of the response waveform (27ms vs. 6ms). The peak also occurred later (112ms VS. 85ms). The first change in the enhancement profile occurred later than the latency of response waveform in all conditions (ZCR, ZSR, ZCF and ZSF) except ZZR, which showed no consistent change in the enhancement profile (**Figure 11B**), consistent with the dramatic difference between ZCR and ZZR observed in DSs. The differences in latency results suggest that the enhancement effect is not due to adaptation or fatigue. Instead, they are what one would expect if statistical inferences made across trials and the criteria for recognizing unexpected stimuli (e.g. context violations) are a top-down process. Such a top-down process may originate from the limbic system (Turk-Brown et al., 2009) or from the premotor system (Bar, 2003; 2004). The difference is also consistent with the observation in MMN that the peak of MMN is later than the peak of N1, a main component of ERP (Näätänen et al., 2005).

Another interesting observation is that the enhancement profiles in silent conditions (ZSR /ZSF) differ from the enhancement profiles with canary context (ZCR/ZCF) in three ways. First, the latency of ZSR was earlier than ZCR for 7 ms (**Figure 11F**). This observation supports the idea that early enhancement in ZSR/ZSF is due to uncertainty about when the stimulus will occur, resulting in a rapid detection of stimulus onset, while, for ZCR/ZCF,

detecting an acoustic violation of the context requires more time. Second, the enhancement profiles of ZSR/ZSF had a shorter duration than in the canary context, with a decay to zero 181/161 ms earlier than that of ZCR/ ZCF (**Figure 11C**). This is consistent with the idea that processing of acoustic features in the canary context continues during the evolving stimulus, while temporal uncertainty is largely detected at stimulus onset. Third, the enhancement profiles of ZSR/ ZSF have two peaks (**Figure 11E**). The first peak was 27/30ms after stimulus onset and the second peak was 105/98ms after stimulus onset, suggesting that there might be two sources that induced enhancement under conditions of temporal uncertainty. The first peak may be due to a local processing, or to a “startle effect” induced by a noradrenergic input that originates in the brainstem and reaches the auditory forebrain in parallel with the main auditory input. In contrast, the later peak at 105/98ms is similar in timing to the peaks for ZCR/ZCF and may due to the detection of temporal uncertainty through a top-down process.

Single Unit

Twenty nine single-units in the ZCR condition and 24 single-units in the ZZR condition were isolated from these recordings. First, in order to understand the tuning properties of recorded neurons, STRFs were calculated from responses of neurons to the eight zebra finch songs in the ZZR condition. STRFs of these neurons showed varied complex structures. Some STRFs displayed a single broad-band structure (Figure 12a and 12b), implying that the neurons tuned to a broad-band acoustic event that is a typical feature in zebra finch songs. Others displayed two structures separated in spectral and/or in temporal dimensions (Figure 12c and 12d), implying that these neurons could integrate information of features separated in frequency or in time.

Second, DSs were calculated for single-units for comparison with results obtained from multi-units. Increased firing rate in the context-modulated session could be observed both in the roster plot (Figure 13a) and in the plot of spike rates by trials (Figure 13b) in the ZCR condition. As shown in Figure 14, the DSs of these enhanced responses were significantly larger in the ZCR condition than in ZZR (Two sample Kolmogorov-Smirnov test: $P=0.022$).

5. Discussion

In this experiment, zebra finch songs were embedded in five different contexts and presented to awake animals. Significant increases in auditory responses were observed in four of five conditions. Only the zebra finch song heard in the context of other zebra finch songs induced very little or no enhancement effect.

Effect of long intervals

The first interesting question is why presenting the target song at fixed but very long intervals (ZSF: ISI = 56s) induce an enhancement effect. In the experimental design, we intended to use this condition as the control condition and supposed that fixed intervals would not induce enhancement. However, our results showed that more than 90% of sites in the ZSF condition showed DSs higher than zero. The temporal profile of the enhancement effect in ZSF also showed similar latency and peak timing as the other three conditions that show enhancement, implying similar underlying mechanisms. The simplest explanation for enhancement in ZSF was that the longer interval (56s in the context-modulated session vs. 7s in pre-adapting session) caused recovery from adaptation thus higher responses. But this explanation is not true for two reasons. First, the measurement of the enhancement effect

was designed not to be affected by adaptation. DSs were calculated from the differences between observed responses and the estimates predicted from the first and last session. If the delayed interval led to recovery from adaptation, then the responses of the last session would also increase and so the estimates of the responses to songs in longer intervals would also be higher. Thus recovery from adaptation would only change the slope of adaption but would not change the difference between the observations and the estimates. Second, as we reported in our results, there was no systematic relationship between DSs and AIs, thus, enhancement of auditory responses in the context-modulated session did not interact with adaptation. In other words, there was no recovery of adaptation in ZSF. Therefore, we hypothesized that enhancement of responses in ZSF was due to increased temporal uncertainty. It may be more difficult to predict the timing of the upcoming stimulus in 49s ISI than in 7s ISI, although both of them are fixed intervals. It has been documented that the ability to predict the timing of upcoming events obeys Weber's law (Gibbon, 1977). Thus, the longer the ISI, the larger the range of uncertainty in the animal's ability to estimate timing.

Effect of the canary context and random intervals

If we consider the enhancement effect observed in ZSF as the baseline enhancement caused by temporal uncertainty of long intervals, then the next question is whether the random intervals and canary context in ZCR, ZSR and ZCF cause further enhancement. First, the fact that ZSR was not different from ZSF suggests that random intervals could not independently induce more enhancement effect than fixed long intervals. It is possible that the enhancement effect induced at 56s long interval reached the ceiling of the effect that can be induced by temporal uncertainty. Second, the observation that ZCF was not different from

ZSF suggests that canary context could not independently induce more enhancement effect than fixed long intervals. It may be because that adding 7 canary songs in the fixed 56s interval between two target songs did not change the predictability of the target songs. In contrast, because DSs in ZCR was larger than DSs in ZSR, ZCF and ZSF, we conclude that the interaction of the canary context and random intervals induce more enhancement effect than random intervals, fixed long intervals, canary context in fixed intervals. Furthermore, the enhancement profiles for conditions with canary songs showed different temporal characteristics, which suggested that different neural processes could be involved in induction of the enhancement effect. Therefore, we concluded that the canary context may play a different role in inducing the enhancement effect.

Effect of the zebra finch context

Another question is whether the ZZR condition induced a significant enhancement effect. In our results, about 45% of sites showed DSs lower than zero. However, we did not know whether this was due to sampling fluctuation. In order to answer this question, we need to compare ZZR with a condition in which a zebra finch song was only presented in constant intervals (different from ZSF that had a 7s ISI and a 56s ISI). Since such a condition did not exist in the current experiment, we had to use data from other experiments to estimate DSs in this condition. Data obtained from 22 birds in other experiments were used in this analysis. Each bird was presented with 3 zebra finch songs in blocks of 70 repeats (ISI 6s). Electrophysiological recordings were made using the same methods as in the current experiment. A Monte Carlo approach was used to analyze data from these birds. One hundred random samples were made from the data set. To create each sample, we first

selected 11 birds randomly from the data set. Then, for each bird, one of the three stimuli was randomly selected. Neural responses to the first 59 trials of this stimulus from all sites were analyzed. DSs were computed in the same way as the current experiment. At the end, DSs from each bird were averaged. Thus, from same number of birds, we computed a random sample of DSs during a mock “context-modulated” session when no change in stimulus context or timing actually occurred. Then these DSs were compared with DSs collected from the ZZR condition in the current experiment. Mann-whitney u test showed that only 11 out of 100 random samples were significantly lower than ZZR ($p < 0.01$). We therefore believe that the ZZR does not induce significant enhancement. This conclusion is consistent with the observation in enhancement profiles in ZZR, in which no enhancement in responses was observed.

Differences between the canary context and zebra finch context

It appears that the canary and zebra finch contexts can have opposite effects on auditory responses. If we consider enhancement in the ZSF condition as a baseline induced just by the temporal uncertainty of long intervals, then the canary context in random intervals (ZCR) significantly increased the enhancement effect above this baseline. This observation is reasonable, because the canary context with random intervals did make the prediction of upcoming events even more difficult. On the contrary, the zebra finch context in random intervals actually eliminated baseline enhancement. As we discussed before, DSs of ZZR only showed very weak differences from DSs obtained from responses to a zebra finch song presented by itself at constant intervals. This phenomenon was interesting because it seemed that a zebra finch context eliminated uncertainty related to both timing and acoustic

content. This observation may have very important implication. It suggested that the auditory system may predict the upcoming events categorically! We speculated that when a zebra finch song was presented in the canary context, the auditory system would predict that, just about every 7 seconds, there would be one canary song, so arrival of a zebra finch song would be a violation of the prediction and would induce enhancement in responses. On the other hand, as a zebra finch song was played in the zebra finch context, the auditory system would predict that a zebra finch song would occur every 7 seconds. Thus, in thus a case, occurring of the pre-adapted zebra finch song in the context would be not a violation and no enhancement would be induced. Therefore, although the temporal uncertainty of the pre-adapted zebra finch song was same in ZCR and ZZR conditions, the canary context and zebra finch context could lead to opposite predictions, which in turn lead to opposite modulation of auditory responses.

If such an explanation is right, this might be the first observation at the neural level that the sensory system of animals can make categorical predictions. There are three important points that are worth consideration. First, the prediction is very likely made in the auditory system, because the temporal profile of enhancement showed a 20ms latency. In other words, the enhancement effect begins at the same time that the auditory forebrain starts to respond. Such a short latency tends to support the possibility of local processing. Second, if the auditory system can make predictions, it must have a memory at a scale of a few seconds to minutes. At the time that a sound is coming, the prediction of the upcoming sound should be at least dependent on what occurred on the last trial 7 seconds earlier. Considering that the prediction is very likely based on accumulation of knowledge of past events, rather than

just the very last trial, the memory may have to be taken forward in time for even longer periods. As has been reported in MMN research, two or three trials of the standard sound presented before the deviant sound were necessary to inducing MMN to the deviant sound (Näätänen, 1992). Similarly, the enhancement effect may also require a few trials to set up an expectation, which may require memory up to minutes. Based on this reasoning, we speculate that the enhancement effect may be generated in NCM and CM, because neural responses in these two areas show long-term stimulus-specific adaptation, a form of long-term memory. By contrast, there is no evidence showing that the regions in the ascending auditory pathway, Field L, MLD etc., have similar long-lasting adaptation. Thus, these downstream areas are also unlikely to be the generator of the enhancement effect. The third point that makes our observation interesting is that the prediction made in the auditory system may be categorical. Both in ZCR and ZZR, the 7 context songs were not only different from the pre-adapted song, but also different from each other. All these songs in the context-modulated session occurred with equal probability, so there was no standard song. Thus the auditory modulation is not based on the prediction of one standard sound. Instead, it is based on the prediction of a category of sounds. In ZZR, when a song is presented that is in the same category as the context songs, responses are the same as when it is presented by itself at a constant interval. If the auditory system did not treat each stimulus as a member of a category, the predictability of the pre-adapted song in ZZR would be as low as in ZSR condition and the enhancement effect would be induced. But this was not the case.

Related phenomena

Keeping these ideas in mind, we would compare our observations with similar phenomena found in human and animals. The first important phenomenon is MMN. It has been well documented for decades that if a deviant sound with low probability (10%) is presented along with a standard sound (90% probability and different from the deviant sound in frequency or amplitude) in random order, the deviant sound can induce higher ERPs than the standard sound in human subjects. MMN could be induced no matter whether subjects attended to stimuli or not. MMN has a latency of 100-200ms from stimulus onset and has been tracked to two main sources: the frontal cortex and the auditory cortex (for review, see Näätänen et al., 1995). An important fact is that MMN could be induced with ISIs as long as 10s (Bottcher et al., 1992). Cowan (1984) described two types of auditory memory: a short form of memory with life time of about hundreds of milliseconds participating in temporal integration and a long-term auditory store lasting 2-10s or “roughly an order of magnitude greater”. MMN observed with long ISI requires this long-term storage (Näätänen et al., 1995). Ulanovsky et al., (2003) reported a phenomenon at the neural level that was very similar to MMN. They presented 100 deviant sounds and 900 standard sounds (different in frequency) in random order and recorded from A1 of anesthetized cats. Then they exchanged the probability of the two sounds in the next session, so that the sound with 10% in the previous session had 90% probability in the succeeding session and served as a standard sound. It was found that neural responses was much higher for the same sound when it was in the low probability condition than in the high probability condition, but this effect disappeared when the ISI was longer than 2 second. The authors interpreted the phenomenon as a form of stimulus-specific adaptation (SSA) and suggested that it may function as the upstream

mechanism of MMN. Most recently, a similar phenomenon had been described in the auditory forebrain of songbirds (Beckers and Gahr, 2010). In this study, different male zebra finch calls were used as the deviant sound (10% probability) and the standard sound (90% probability). Recordings were made in the auditory forebrain of anesthetized female zebra finches. It was found that the neural responses to the deviant sound were higher than those to the standard sound. Again, such an effect was only observed when the ISI was shorter than 2.5 second. These observations were very similar to those reported in Ulanovsky et al., (2003). The authors explained the effect as probability-dependent enhancement on auditory responses that may be related to MMN. Although evidence supporting the idea that responses of auditory neurons are modulated by the statistics of sounds has accumulated in recent years, some fundamental questions are unresolved. First, there is conceptual confusion between SSA and probability-dependent enhancement in responses. Although the two phenomena are usually considered to be the same, they are probably dependent on two independent processes, as we found in the current experiment. While SSA is the decrease of responses that may be only related to the count of encountering a specific event over some (long-term) history, the enhancement effect may related to the predictions based on the probability of a sound in short-term experience. In Ulanovsky et al., (2004), researchers found that both the short-term and the long-term history had effects on auditory responses. On the one hand, the response amplitude reduced with the increasing count of occurrence of the standard sound over the whole test session (long-term history), which was typically explained as the result of adaptation. On the other hand, they also found that the probability of the deviant sound within the course of a few trials modulated the responses of the deviant sound

(short-term history). However, the phenomenon was exclusively interpreted as SSA.

Furthermore, on the methodological level, it is difficult to separate the measurement of SSA from the measurement of the enhancement effect in this oddball paradigm, in which the manipulation of the probability was not independent to the manipulation of the count of a sound. Thus, the differences in responses to the same sound when it was rare and when it was common included both the short-term and the long-term components (see analysis in Ulanovsky et al., 2003, 2004), and thus was combination of SSA and an enhancement effect. Although, the authors tried to understand the short-term effect by the analysis of local history tree of stimuli presentation, the interpretation was still unable to separate effects of probability from the effect of the count of occurrence. In Beckers et al., (2010), the authors measured the difference in responses to a rare call and a common call. Because the comparison was not made between the responses to the same sound in common condition and in rare condition, the interpretation was even more complicated. First, the tuning properties of neurons may cause the differences in response between two sounds. Second, even if the tuning properties to the two sounds were same, the differences could be exclusively explained as: the more a sound occurred during the testing, the lower the responses to it. . Therefore we proposed in the current experiment a new paradigm to test the effect due to the probability of a sound, which was independent of effects of the count of the sound in the long-term history. We used DSs to measure the enhancement effect over the whole context-modulated session, which indicated the increase in responses compared to the trajectory of response adaptation, thus excluded the effect of adaptation.

The second problem in interpreting previous work is to determine the relationship between enhancement in auditory responses observed at the neural level and MMN in humans. While MMN could be induced at ISI as long as 10s and had been proposed as a form of “a long auditory store” (described by Cowan 1984), no enhancement in auditory responses was observed when the ISI exceeded 2s in Ulanovsky et al., (2003) or 2.5s in Beckers et al., (2010). Thus, these previous observation at the neural level could not explain the long-term effects found in MMN. However, in the current experiment, we found enhancement effect in 7s ISI. Although we did not test the effect of different ISIs on the enhancement effect, preliminary data shows the same enhancement effect in 8 ISI. Thus, as discussed, the enhancement effect in the current experiment reflected a “long-term” auditory memory on the same scale as what was observed in MMN. Moreover, it is interesting to ask why previous works only found enhancement in responses at short ISIs. The different observation in Ulanovsky et al., (2003) could be due to the differences between animal models. But, Beckers et al., (2010) had recorded approximately from the same areas in auditory forebrain of songbirds as we did. We hypothesized that the real factor led to the different observations may be the state of the animals. Animals were in halothane anesthesia in Ulanovsky et al., (2003) and isoflurane anesthesia in Beckers et al., (2010). It was very like that anesthesia affected the memory process and thus the enhancement effect was only expressed at short ISI. It would be interesting to repeat the experiments of Ulanovsky et al., (2003) and Beckers et al., (2010) in awake animals and repeat ours in anesthetized animals. This question is important, because many electrophysiology experiments were conducted on anesthetized animals. However, if memory is a variable involved in an experiment and could

be impaired by anesthesia, one should consider whether anesthesia is a suitable choice. We have done some preliminary experiments using isoflurane or urethane to anesthetize animals. In these pilot birds, we find that auditory responses in NCM are significantly reduced by anesthesia state. In addition, we have yet to observe either clear SSA or the enhancement effect at 7s ISI. Therefore, compare to previous work in A1 of mammals and in the auditory forebrain of birds, our results provide the clearest evidence that the probability of sounds modulated auditory responses at a time scale comparable to the effect found in MMN in human research, and thus reflect “a long auditory store”. Although direct comparison between MMN and the enhancement effect observed in our experiment was difficult, because MMN were obtained at the whole brain level and at least had two sources (one was from the auditory cortex; the other one from the frontal cortex), it was reasonable to speculate that auditory neurons in human brain may share similar sensitivity to the probability of sounds and may utilize this probabilistic knowledge to modulate auditory responses at the time scale from seconds to minutes and such a mechanism may be underlying the phenomenon found in MMN research.

In addition, the current work suggests that the auditory system can extract statistical features from multiple objects, a process which may play an important role in auditory perception. Information from the sensory world changes rapidly and is often ambiguous. Under these conditions, rapid perception of “global image features” may be important for survival of an organism, because it allows rapid identification of the category of an object, before detailed information is available for reliable object recognition (Oliva and Torralba 2006). This form of perception has been described as “statistical summary perception” or

“perception averaging” in human psychophysical studies, in which subjects automatically extract features, e.g. average size, position, inclination, facial expression, from a group of objects across time and space (Ariely, 2001; Alvarez & Oliva, 2008; Miller & Sheldon, 1969; Haberman & Whitney, 2007). In the natural acoustic environment of songbirds, salient signals such as vocalizations usually occur in a noisy acoustic background. Details of a sound may be masked by other signals or noises. However, having an ongoing average of the statistics of background sounds may constantly update the inference or prediction of probability of a up-coming sound (Think about Bayesian inference). As the result, inference or prediction would assist detection and categorization of a given sound in an ambiguous acoustic environment. This process may function through a top-down influence on auditory system. Such a top-down process may originate from the limbic system (Turk-Brown et al., 2009) or from the premotor system (Bar, 2003; 2004). Further studies with simultaneous recording from both the auditory system and the limbic system/premotor system would may reveal how statistics of acoustic environment and prediction affect auditory perception.

Experiment 2

1. Introduction

In the first experiment, we found a dramatic difference in the enhancement effect between ZCR and ZZR and we interpreted it as evidence of categorical processing. We concluded that, in the ZCR condition, neurons process canary songs and zebra finch songs as members of different categories and that the brain predicted that there would be a canary song every 7

seconds, so occurring of a zebra finch song was a violation of the prediction and caused the enhancement effect. However, an alternative hypothesis would be that the enhancement effect in ZCR was due to an innate bias in selectivity to conspecific songs. It had been reported that NCM neurons show stronger responses to conspecific over heterospecific songs (Chew et al, 1995). Thus the enhancement effect may be interpreted as a pre-existing “attention” to zebra finch songs, perhaps innately determined. We tested this alternative by reversing the context of presentation, so that, in this new version, a pre-adapted canary song became the target that was tested in a context of zebra finch songs. If no enhancement effect could be induced, then the enhancement effect in ZCR condition could be explained as a pre-existing bias to conspecific songs.

2. Experimental design and methods

The paradigm used in Experiment 1 was repeated in this experiment. Six adult male Zebra finches were tested in this experiment. Each of them was tested in three different conditions. In the first condition, a novel canary song was presented as the target song for 20 repeats in the pre-adapting session. Then, in the context-modulated session, the canary song was presented for 19 trials together with 20 repeats of each of 7 zebra finch songs in shuffled order. Finally, in the post-context session, the same target song was presented for 20 trials. Thus we named this condition as CZR condition (a **C**anary song in a **Z**ebra finch context with **R**andom intervals). Two other conditions were ZCR and ZZR, which were tested exactly as described in Experiment 1 and served as controls in the current experiment. We predicted that the canary song in zebra finch context would induce same DSs as ZCR and would be significantly higher than ZZR. Surgery, electrophysiological recordings and data analysis

followed the same methods as Experiment 1. DSs were quantified and compared across three conditions. Because there were only six birds used in this experiment, sufficient single-unit data was not available for analysis.

3. Results

In this experiment, 47 sites recorded in NCM and CLM that responded to auditory stimuli were analyzed. Data were first analyzed by sites as Experiment 1. A Friedman ANOVA performed with these data showed significant differences across three conditions (Chi Sqr. ($N = 47$, $df = 2$) = 30.9, $p < 0.001$). As expected, both CZR and ZCR conditions showed DSs that are significantly higher than those of the ZZR condition (Wilcoxon tests: $p < 0.001$ for both CZR and ZCR). But CZR and ZCR were not significantly different from each other (Wilcoxon tests: $p = 0.526$). DSs for sites were plotted as cumulative frequency distributions for each condition in Figure 15. A two-way Repeated-Measures ANOVA showed no significant differences between NCM and CLM ($p = 0.060$) or any interaction with three conditions ($p = 0.180$). We also did not find a significant difference between two hemispheres ($p = 0.479$) or interaction with three conditions ($p = 0.392$).

Matched contrasts based on averaged DS in each site are plotted as boxplot for each bird in Figure 16. In the comparison between CZR and ZCR, four birds showed preference for CZR over ZCR, while two other birds showed preference for ZCR over CZR suggesting that there may be no difference between the two conditions. On the contrary, all birds showed preference for ZCR and CZR over ZZR. Thus the difference between CZR and ZZR is highly duplicable across birds.

When comparisons of trial-by-trial DSs were made on each site, the percent of sites in in each bird that showed significant differences in the comparison between two conditions were shown in Figure 17. In comparison between CZR and ZCR, totally 29% sites show significant higher enhancement effect for CZR than ZCR, while 34% sites showed opposite effect. On the contrary, in comparisons with ZZR, no more than 9% sites preferred ZZR. This duplicability across sites is relatively low compared to the duplicability observed in Experiment 1 (76% sites prefer ZCR over ZZR).

4. Discussion

Based on observations in this experiment, a canary song imbedded in a zebra finch context or a zebra finch song in a canary context induced the same level of enhancement in responses. This result supported the hypothesis that the enhancement effect was due to the violation of ongoing prediction, rather than to a pre-existing bias to conspecific songs.

Experiment 3

1. Introduction

In Experiment 1, we hypothesized that the enhancement effect reflected a kind of categorical processing, reflected by the difference between ZCR and ZZR. Experiment 2 supported that this process was not due to a pre-existing bias to zebra finch songs. Thus, what make the zebra finch context and the canary context categorically different in inducing the enhancement effect are probably their acoustic features. As shown in Figure 3, canary

songs usually include two types of notes: 1) whistles in narrow frequency range or with broad-band harmonically structures and 2) trills with high frequency of rhythmic structure (10-40 Hz). In contrast, zebra finch songs consist of broad-band harmonic notes, with varying durations. High frequency rhythms are absent in zebra finch repertoires. Although repertoires of the two species differ both in spectral and temporal dimensions, it possible that differences in just one dimension would enable the auditory system to differentiate the two types of songs and set up a prediction based on some form of categorization in that dimension. Thus, this experiment was designed to determine whether it is simply the differences in rhythm between a zebra finch or canary “oddball” song and context songs of the other species that causes the enhancement effect. To test this variable in isolation, we used synthetic sounds whose features we could control.

2. Experimental design and Methods

Stimuli used in the experiment were sequences of artificially synthesized zebra finch notes. Each artificial note was a harmonic stack with the same spectral timbre and amplitude envelope derived from an average of zebra finch calls so that they could induce responses in most neurons of the auditory forebrain. We only manipulated two acoustic features: 1) the duration of each note 2) the fundamental frequencies (pitches) of notes. By manipulating the duration of each note, we actually produced sequences that differed in rhythmic structure. The oddball sound and context sounds were made discriminable by their rhythms. Three conditions were tested. In the first condition, the seven context sounds were sequences with notes with durations of 30-170ms in 20ms increments, so the rhythm of context sounds ranged from 5.9-33.3 Hz. The duration of the notes in the oddball sound was 390ms, a few

times longer than the duration of notes in the context sounds, with a slower rhythm at 2.56 Hz. Examples of the oddball sound and the context sounds are shown in Figure 18. We assumed that although each sound would be presented with the same probability, the auditory system would be able to set up a prediction for fast rhythm sounds based on their distribution on the temporal dimension. Thus occurrence of a slow rhythm sound (the oddball) would violate this prediction and would induce the enhancement effect. In the second condition, a fast rhythm oddball sound (10 Hz) would be presented with slow rhythm context sounds (2.2-3.1 Hz). We hypothesized that these two conditions would induce the same enhancement effect. In the control condition, either a fast rhythm sound would be presented with fast rhythm context sounds or a slow rhythm sound would be presented with slow rhythm context sounds. In this case, the “oddball” sound consisted of notes whose duration was approximately equal to the median of durations that used for making context sounds, so that “oddball” was at the center of the distribution of context sounds in the temporal dimension. Thus, we expected no enhancement effect in the control condition.

In addition to manipulating the duration of notes to make the oddball sound different from context sounds in rhythm, we also varied the pitches for each sound so that they were distinguishable from each other. We manipulated this feature in the spectral dimension to avoid generalization in pitches across all stimuli. Fundamental frequencies of all sounds varied from 450-915, in 15 Hz increments. The pitches of oddball sounds in all conditions were between 660 and 705Hz, so they were near the center of the distribution of pitches. Thus, we would not expect the pitches of the oddball sounds to induce any enhancement effect.

Five adult male zebra finches were tested in this experiment. We followed the paradigm used in the first two experiments. First, the oddball sound was presented for 20 trials in the pre-adapting session. Then this sound was presented with seven context sounds (19 trials for the oddball sound, 20 trials for each context sounds) in random order. Finally, the oddball sound was presented by itself for 20 trials. Surgery, electrophysiological recordings and data analysis followed the methods of Experiment 1. DSs were quantified and compared across three conditions.

3. Result

In this experiment, 44 sites recorded in NCM and CLM that responded to auditory stimuli were analyzed. DSs for sites are plotted as cumulative frequency distributions for each condition in Figure 19. A Friedman ANOVA showed significant differences across three conditions (Chi Sqr. ($N = 44$, $df = 2$) = 16.0, $p < 0.001$). Post hoc Wilcoxon tests showed that DSs of the control condition were significantly lower than the condition having slow rhythm in the fast rhythm context ($p = 0.001$) and the condition having the fast rhythm in the slow rhythm context conditions ($p < 0.001$). But the slow rhythm in the fast rhythm context and the fast rhythm in the slow rhythm context were not significantly different from each other ($p = 0.086$). We did not find a significant difference between NCM and CLM (two-way Repeated-Measures ANOVA: $p = 0.752$) or any interactions with conditions ($p = 0.805$). We also did not find a significant difference between two hemispheres (two-way Repeated-Measures ANOVA: $p = 0.953$) or any interactions with conditions ($p = 0.827$).

However, as shown in boxplot for each bird in Figure 20. In comparison between the condition having slow rhythm in the fast rhythm context and the control condition, four birds showed higher DSs in the condition having slow rhythm in the fast rhythm context than the control condition, while one bird showed opposite effect. Same result was also observed in the comparison between the condition having the fast rhythm in the slow rhythm context conditions and the control condition. Thus, our manipulation of rhythms of stimuli may produce the enhancement effect, although individual differences may exist.

When comparisons of trial-by-trial DSs were made on each site, the percent of sites in each bird showing significant differences in each contrast were shown in Figure 21. In comparison between the condition having slow rhythm in the fast rhythm context and the condition having the fast rhythm in the slow rhythm context conditions and the control condition, totally 15% sites showed significant higher DSs in the first condition and 33% sites showed significant higher DSs in the second condition. In comparison between the condition having slow rhythm in the fast rhythm context and the control condition, totally 26% sites showed significant higher DSs in the first condition, while totally 9% showed opposite effect. In comparison, the fast rhythm in the slow rhythm context conditions and the control condition, totally 50% sites showed significant higher DSs in the first condition and only 6% showed opposite effect. Thus the duplicability across sites for the first condition is low, while the duplicability in the second condition is comparable with the ZCR-ZZR difference observed in Experiment 1. Therefore, we concluded that there is a weak enhancement effect that observed in the two conditions in which the oddball stimulus and context stimuli had different rhythms, although inconsistency existed in our observations.

4. Discussion

We speculated that the inconsistency of the enhancement effect in the artificial oddball paradigm may be due to an incorrect assumption that the distance in temporal dimension of stimuli had a linear relationship with the duration of each note. For example, in condition one, the frequency of slow rhythm sound was only 3.33 Hz lower than the slowest sound in the context. However, some differences between context sounds were even larger. For instance, the fastest context sound had a rhythm 13.3 Hz higher than next slower sound. Thus, the context sounds may sound more different from each other than from the “oddball”. Although the rhythm of the oddball was at one end of the range in the temporal dimension, the context sounds may not have been close enough in timing to be treated as members of the same category and thus set up a prediction. Therefore, in future experiments, we will manipulate the rhythm directly, which may allow a more effective manipulation of the distance between stimuli in the temporal dimension.

Experiment 4

1. Introduction

In Experiment 1, we proposed that adaptation and enhancement of auditory responses were two different processes. Enhancement of auditory responses reflects the violation of prediction based on short-term acoustic experience. In contrast, adaptation was simply related to the number of occurrence in long-term acoustic experience. In Lu and Vicario (2011), we proposed that adaptation may play an important role in auditory scene analysis. It

may increase the contrast between familiar sounds and novel sounds, thus increasing the chance of a novel sound being perceived in a familiar acoustic environment. The aim of the current experiment was to test this hypothesis.

2. Experimental design and methods

In Experiment 1, we found that zebra finch song in a zebra finch context did not induce the enhancement effect. We speculated that a zebra finch context set up a prediction of one zebra finch song every 7 seconds. Thus, arrival of the pre-adapted zebra finch song was consistent with the prediction and did not induce any enhancement. However, if the zebra finch context consisted of songs that were familiar to the animal, the auditory system might predict a familiar song every 7 seconds. We hypothesized that, in such a case, arrival of a novel zebra finch song would be a violation of the prediction and would induce an enhancement effect. Therefore, in this experiment, we tested whether familiarity of context songs would increase the enhancement effect in the ZZR condition.

Two groups of animals were tested in this experiment. In the experimental group, 7 adult male zebra finches were first tested in a ZZR condition, in which the context was 7 novel zebra finch songs, as in Experiment 1. Then the same context songs were presented to the animal for 50 repeats in random order. At the end, the same animals were tested in ZZR condition again, in which a novel pre-adapted zebra finch song was presented with the now familiar context songs in the context-modulated session. DSs obtained before the training and after the training were compared. In the control group, 6 adult male zebra finches were first tested in the ZZR condition, in which both the pre-adapted song and the context songs

were novel. Then 7 canary songs were presented to the animal for 50 repeats in a random order. Finally, the same animals were tested in the ZZR condition again. But, in this group, the novel pre-adapted zebra finch song was presented with 7 other novel context songs in the context-modulated session. We predicted that exposure to familiar context songs would increase DSs in the experimental group, but exposure to irrelevant canary songs would have no effect on DSs in the control group. All experimental procedures and data analyses were identical to previous experiments.

3. Result

In the experimental group, 67 sites recorded in NCM and CLM that responded to auditory stimuli were analyzed. Data were first analyzed with individual sites considered as independent observations. Wilcoxon tests showed that DSs obtained in ZZR after the training were significantly higher than DSs obtained before the training ($p < 0.001$). DSs for sites are plotted as cumulative frequency distributions in Figure 22. We did not find a significant difference between NCM and CLM (two-way Repeated-Measures ANOVA: $p = 0.839$) or any interactions with training ($p = 0.488$). We also did not find a significant difference between hemispheres (two-way Repeated-Measures ANOVA: $p = 0.161$) or any interactions with training ($p = 0.482$).

Duplicability of the effect of training across birds was confirmed in the boxplot of matched contrasts based on averaged DS in each site (Figure 23). Five out of six birds showed that three quarter of matched contrast by sites were higher than zero. Thus, familiarity with context songs induced enhancement effect, although inconsistency exists.

When comparisons of trial-by-trial DSs were made on each site, the percent of sites showing significant differences in each bird were shown in Figure 24. Totally 41% sites showed significant higher enhancement effect after training, while 15% sites showed opposite effect.

In contrast, no significant differences were observed in DSs obtained before and after exposure to canary songs in the control group (Figure 25), Wilcoxon tests by site: $p=0.963$). As showed in the boxplot (Figure 26), three out of 6 birds showed higher DSs (measured in median) after training, two of them showed lower DSs after training and one showed no change in DSs. The comparison based on trial-by-trial DSs (Figure 27), showed 20% sites showed significant higher DSs after training and 21% showed lower DSs after training. Therefore, there was no effect of exposure to canary songs on enhancement effect.

4. Discussion

Observations in this experiment were consistent with our predictions that familiarity of context songs significantly increased the enhancement effect in ZZR. The observation in the control group also ruled out the possibility that the increase in DSs was due to any non-specific effect. One question may be raised about these results. As discussed before, we assumed that, in ZZR, all zebra finch songs may be treated as members of the same category so that the prediction may be made categorically. But, a potential contradiction was that familiarity of the context songs caused the target zebra finch song to be treated as a violation of the context sounds, although their acoustic features should place them in the same perceptual category. Our best explanation is that adaptation reduced response to

context sounds so that the differences in response strength between the novel pre-adapted song and the familiar context song made them perceptually different.

To test whether this is true, we calculated the D-prime (D') for all songs in the context modulated session before and after exposure to context songs. D' was used to measure the selectivity of an individual neuron for one stimulus (A) over another stimulus (B). It was calculated by the following formula (Solis and Doupe, 1997):

$$D' = (\text{mean of responses to A} - \text{mean of responses to B}) / \sqrt{(\text{variance of responses to A} + \text{variance of responses to B})}$$

A positive D' means that the neuron prefers stimulus A in its responses. To calculate D', we first took the mean and variance of responses to each song at each site (obtained from ten trials). Then D' for each song over other seven songs were calculated following the formula, so that there were 7 D's for each of 8 songs in the context-modulated session. At the end, 7 D's for each song were averaged. Thus, we had 8 averaged D's for each recording site in one session. These means reflected whether this song elicited responses stronger or weaker than other songs on average. We plotted the cumulative frequency distributions for all averaged D's obtained from pre-training session in Figure 28a and those obtained from post-training session in Figure 28b. We found that, before the training, the d-prime of the pre-adapted song only differed from d-primes of three out seven context songs (tested by Wilcoxon test). In contrast, after exposure to context songs, more than 85% of d-primes for pre-adapted songs were higher than zero, while d-primes for context songs remained same. Therefore, exposure to context songs increased the contrast in responses between the pre-adapted song and

context songs. This contrast in responses may have made perception of the pre-adapted song dramatically different from context songs, which may have led to the enhancement effect in the ZZR condition after training. A speculative way of describing this is that, like spectral features and perhaps rhythm, the degree of familiarity is in some sense another perceptual dimension of a stimulus. In any case, we hypothesized that some version of this mechanism may contribute to schema-base auditory scene analysis and increase the chance of perception of novel sounds in a familiar acoustic environment.

Experiment 5

1. Introduction

In Experiment 1, we explored whether temporal uncertainty could induce the enhancement effect independently. Our results were not clear. When data were analyzed by individual site, DSs of the ZSR condition were significantly larger than those of the ZSF condition. In contrast, when data were averaged for each bird, the difference between ZSR and ZSF was not significant. In addition, in Experiment 1, the random intervals in ZSR were only longer intervals than the 7 seconds used in the pre-adapting session. Thus we did not know whether random intervals had effect on enhancement of responses. An alternative explanation would be that the amplitude of responses was positively correlated with the length of intervals, rather than temporal uncertainty. To rule out this alternative, we need to test the enhancement effect in a balanced design, in which there were equal numbers of increased intervals and decreased intervals, so that the averaged interval remained same.

Furthermore, there was another problem that complicated the interpretation of results of ZSR. The random intervals in ZSR varied from seconds to minutes. A long period of silence may change the state of the animal. For example, if the animal went to sleep after a long period of silence, arrival of a song may increase responses due to a startle effect. Thus, we need to manipulate interval changes over relatively small scales (a few seconds) and test whether brain is sensitive to smaller interval changes.

Two experimental paradigms were carried out with the same group of animals. In the first paradigm, we presented a song for 104 trials. The standard ISI for song presentation was 8 seconds, but 1 out of 8 songs occurred after a 12s ISI for a total of 13 long intervals. The two different intervals were presented in pseudo-random order, so that there were at least two standard intervals between any two long intervals. This paradigm was repeated for two other songs in independent sessions. We predicted that the standard 8s ISI would set up a prediction and so the occurrence of a 12s ISI would be a violation and thus cause enhancement in responses, which could not be explained as a startle effect. To quantify this increase, we took the ratio of the response amplitude of the trial following the 12s ISI to the response amplitude of the trial before this longer interval, defined as interval-varying ratios. If brains were sensitive to timing change, this ratio would be larger than one. As the control, the ratio of response amplitude of the 8s ISI trial before the 12s ISI to responses of the previous trial were taken and defined as a control ratio. Because there was no interval change between the two trials, control ratios would be expected to be about 1.0. Then the ratios of all 13 interval changes and control ratios were averaged for each song. Then they were averaged across three sessions for each site. Because the oddball intervals only caused 4s

delay, if our prediction was fulfilled, we could rule out the startle effect as an alternative explanation and conclude that auditory responses was sensitive to timing changes.

In the second paradigm, a song was first presented in a pre-adapted session for 20 trials with 8s ISI. Then it was immediately presented in an interval-varying session for 30 trials, in which ISI switched between 6s and 10s (+2 and -2s from the initial ISI) in random order. Then, the song was presented in a post-modulated session for 20 trials. Following the procedure used in Experiment 1, DSs were calculated as the mean difference from the expected responses in the interval-varying session. As a control, a song was presented for 70 trials at 8s ISI and DSs were again calculated as the mean difference from predicted on trials 21-50. Three songs were tested for the effect of interval changes and three songs for the control respectively.

Five adult male zebra finches were tested in this experiment. The three songs for the first paradigm and the six songs for the second paradigm were presented in a random order. Surgery and Electrophysiological recordings followed the same procedures used in Experiment 1. A Wilcoxon test was used to compare the interval-varying ratios and control ratios for Paradigm 1. The same test was used to compare DSs obtained from interval-varying sessions and those from the controls in Paradigm 2.

3. Results.

In Paradigm 1, 57 sites recorded in NCM and CLM that responded to auditory stimuli were analyzed. A Wilcoxon test showed that interval-varying ratios were significantly higher than control ratios (by site: $p < 0.001$). Interval-varying ratios and the control ratios for each

site are plotted as cumulative frequency distributions in Figure 29. Therefore, we concluded that 50% increase (4s) in 8s ISI could induced enhancement in auditory responses. We did not find a significant difference between NCM and CLM ($p=0.220$) or interaction with intervals ($p=0.217$). We also did not find a significant difference between two hemispheres ($p=0.513$) or interaction with intervals ($p=0.447$).

However, as shown in boxplot for each bird in Figure 30. In 4 out of 5 birds, interval-varying ratios were significantly higher than control ratios (measured in median). Thus the effect was duplicable across birds, but inconsistency existed.

When comparing interval-varying ratios and the control ratios trial-by-trial on each site (Figure 31). Surprisingly, only 5 sites showed significant higher responses for song following long intervals. To understand why duplicability was low, we calculated cohen's d for each site. We found that, although 80% sites showed averaged higher interval-varying ratios than averaged control ratios, cohen's ds calculated on trial-by-trial bases were lower than 0.5 in 80% sites. The small effect size may be due to the limited sample size (12 comparison pairs) in trial-by-trial comparisons.

In Paradigm 2, 57 sites recorded in NCM and CLM that responded to auditory stimuli were analyzed. We did not find significant differences between DSs from interval-varying sessions and DSs from the control condition (Wilcoxon: by site: $p=0.303$). DSs obtained from each site for the two conditions are plotted as cumulative frequency distributions in Figure 32. The boxplot for each bird (Figure 33) showed positive DSs in three birds and native DSs in two other birds (measured in median). We also made trial-by-trial comparisons of DSs

between the interval-varying condition and the control condition (Figure 34). Totally 19% sites showed significant higher DSs in the condition with varying intervals, while 21% sites showed significant higher DSs in the control condition. Therefore, there was no enhancement effect observed in Paradigm 2.

4. Discussion

In this experiment, we found that 4 seconds increase to an 8s ISI on random trials caused an increase in auditory responses. However, the same kind of interval change in a balanced design did not induce the enhancement effect measured as DSs. We had two hypotheses. First, the amplitude of auditory responses was only positively correlated with the length of ISI, rather than temporal uncertainty. But this explanation was contradicted by the observation that ZSR showed larger DSs than ZSF in analysis by sites. The second possibility was that DS was not a measurement sensitive to the small enhancement effect induced in interval changes at a scale of a few seconds. In Paradigm 1, we found that the difference in medians between interval-varying ratios and control ratios was only 0.03, implying that increase of responses induced by 4s delay were only about 3% on a trial by trial basis. Therefore, to explore this question, we may have to use larger interval changes in a balanced design.

Experiment 6

1. Introduction

As we discussed before, perception of songs may require processing acoustic sequences. STRFs obtained in Experiment 1 also implied that, in a higher auditory area like NCM or CLM, a neuron might be able to integrate temporal information over a hundred millisecond scale (Figure 12d). Therefore, this experiment tested whether auditory neurons in NCM and CLM can detect a change in the order of a sequence of sounds over a time scale of tens to hundreds of milliseconds.

2. Experimental design

Stimuli used in this experiment were sequences of artificial zebra finch notes. Each sequence consisted of four notes with the same duration, but different pitches (harmonic stacks with different fundamental frequencies), assembled into a continuous sequence with no silent gaps. Each set of notes was assembled in 4 different orders, such that no two sequences shared the same note transition. In each testing session, a given sequence was presented 40 times to produce adaptation to that stimulus. Then, a sequence with the same notes but in a different order (examples of stimuli in Figure 35) was presented for 20 trials, followed by a switch to another order, until all four were presented in each session. We expected that, if neurons in the bird's auditory forebrain could only detect the acoustic features of notes, sequences of the same notes presented in different orders would be processed as a collection of stimuli, resulting in responses at the same level as responses to the initial, pre-adapted order. But if auditory neurons could detect the change of order, strong

responses would be induced to the first few trials of a new order. Three conditions were tested to explore the time scale at which sounds in a sequence were integrated (the test paradigm is shown in Figure 36). The three conditions differed in the duration of the individual notes: 60ms, 100ms and 150ms. We expected that, as the duration of each single note became shorter and the notes became closer to each other, the sequence of sounds would tend to be processed as a single object, and the contrast between these different objects would show the largest effect of order changes. In contrast, we thought that, in sequences with longer notes, individual notes would tend to be processed separately, so this condition would show the smallest order effect. But the difference between object changes (due to order of internal elements) and note order may be just one of time scales so it would imply a difference in processing at different scales.

3. Materials and methods

Five adult male zebra finches were tested. In each bird, the three conditions (note durations) were tested in random order. The procedures for the surgery and electrophysiology have been described in Experiment 1. The discriminability of orders was quantified by a novelty index (NI), which was defined as the response magnitude to each new orders divided by the response to the preceding order at the transition point of two different orders (illustrated in Figure 37). First, responses to the last three trials of each order were averaged. Then the responses to the first three trials of the next order were averaged. Finally, the ratio of the averaged responses to the earlier order to that of the later order was taken. If the change of order was not detected, the trials with the two sequences would be processed as ongoing repetitions of the same stimulus. Thus responses to the later order should be the

same or lower (due to SSA) than the responses to the earlier order and $NI \leq 1$. On the contrary, if the change of order was detected, NI should be greater than one. Since four orders were presented for the same four notes, three NIs were obtained from three transitions. The three NIs were averaged for each condition at each site. A control index (CI) was used for comparison; CI was defined as the ratio of the mean response on the last three trials of a given order divided by the mean response on the three trials that preceded the last three. Because there was no order difference between the last three trials and the preceding ones, CI should approach one, and the variance of CI reflects sampling errors in the current experiment. NIs and CIs were first averaged across three different conditions for each site for comparison. Then we took differences between NIs and CIs as normalized NI, which were used for comparison across three conditions and to examine differences between NCM and CLM.

4. Results

The main assessment of order change discrimination was based on the comparison between CI (note order did not change) and NI (change in note order) across all sites from the five subjects. The overall results are shown in Figure 38 as cumulative frequency distributions. The difference between the distributions of CI and NI was significant (Wilcoxon test: $p < 0.01$). In other words, order changes induced a significant increase in auditory responses. Matched contrasts of averaged CI and NI on each site are plotted in Figure 39. All birds showed preference for NI (measured in median), so the effect was consistent across birds.

A further comparison showed a significant difference across the three duration conditions (Friedman ANOVA: Chi Sqr. ($N = 69$, $df = 2$) = 7.07, $p < 0.029$). However, post hoc testing showed that none of the three conditions significantly differed from each other in their normalized NIs, except that the comparison between 60ms and 100ms conditions showed a marginal effect (Wilcoxon: $p = 0.052$). Normalized NIs in the three conditions obtained from each site are plotted as cumulative frequency distributions in Figure 40. Matched contrasts of averaged normalized NI on each site are plotted in Figure 41. In the contrast between the 60ms condition and the 100ms condition, 4 of 5 birds preferred the 60ms condition (measured in median), suggesting that the effect is duplicable across bird. We also did not find significant differences in normalized NIs between NCM and CLM (Two sample Kolmogorov-Smirnov test: $p = 0.20$) and interactions between regions and durations (two-way repeated measure ANOVA: $P = 0.49$).

5. Discussion

In sum, we found that auditory neurons in NCM and CLM were sensitive to order changes in an acoustic sequence, but it was still not clear whether the duration of individual notes had effect on sensitivity to order changes. There was a conceptual problem in this experiment. The increased responses to new order of sounds may be explained as sensitivity to the changes at the onset and offset of the sequence, rather than the order. Further experiments will use recycled sound sequences as stimuli, in which the onset and offset of the sequence would remain the same and only the sequence in the middle would change. If this new design shows the same results as the current experiment, we can exclude the alternative

explanation. Further experiments are also needed to clarify whether and how durations of notes affect sensitivity to order changes.

General Discussion

First, the current work suggests two methodological improvements that should be made in order to study the kinds of enhancement effects observed here. First, we need to measure SSA and the enhancement effect independently. Previous work with the oddball design measured the combination of both effects when comparing responses to the standard sounds and the deviant sounds, which complicated the interpretation. Another drawback of previous studies was due to the fact that animals were typically tested in the anesthetized state. Even under anesthesia, very short-term forms of adaptation or enhancement occur (as measured by oddballs). However, other types of adaptation and enhancement may depend on memory, e.g. longer term and category effects. Since anesthesia may impair memory, it could thus limit the results, or hide interesting phenomena in these types of experiments. Therefore, future studies should include experimental designs that address these methodological problems.

Second, the current work has several implications for understanding the neural mechanism of auditory object perception. As we discussed in the introduction, auditory object perception requires 1) separating a sound from its background, 2) extracting invariant variables from the same class of sounds and 3) integrating acoustic features across time.

Observations in the current work may shed light on the three aspects.

The first implication from the current work is that the enhancement effect can modulate auditory processing and thus may facilitate segregation (detection) of novel sounds from a familiar acoustic environment (Experiment 1). The enhancement effect segregating of an auditory object from acoustic environment may depend on three factors: 1) differences in acoustic features between the auditory object and its context; 2) probability of auditory object in the context in a short-term history; 3) familiarity of the auditory system to the acoustic context. If a sound is different from the acoustic context in its acoustic features and has low probability (as in ZCR in Experiment 1 or CZR in Experiment 2), enhancement in auditory responses may facilitate detection of this sound. However, if the sound is in the same perceptual category with context sounds, the familiarity of the context sounds accumulated in a long-term history still drive enhancement of responses to the novel sound (Experiment 4). Both the feature-based and the familiarity-based enhancement effects may be part of the mechanism of the schema-based auditory scene analysis (Lu and Vicario, 2010). The traditional “cocktail party effect”, in which one auditory stream is distinguished in a noisy context, has been analyzed in terms of timing and spatial cues that separate the streams (Bragman, 1990). To these cues, we may add the probability (in a short-term history) and familiarity (in a long-term history) of the context, and thus a role for memory; for instance, the voice of a stranger may be distinguished from familiar voices due to enhanced auditory responses.

The second implication is that a form of categorical perception may be involved in auditory processing, because the auditory system may be able to predict the perceptual category of the upcoming sound (demonstrated by the differences between ZCR and ZZR in

Experiment 1). Processing and predicting sounds categorically require generalization of similarity or extraction of invariant variables from acoustic features based on statistics of sounds. The ability to make perception and predictions at the categorical level would allow inferences about acoustic events in categories that have never been encountered and could thus improve organism's adaptation in the natural environment. In further experiments, we could manipulate acoustic features on one dimension and test whether a categorical boundary can be established or shifted on the manipulated acoustic dimension for other stimuli.

The third implication is that auditory neurons may be able to integrate features across time into a holistic perceptual object. First, Experiment 6 showed that the auditory forebrain is sensitive to the order of sounds in an acoustic sequence. Second, we found STRFs of neurons in the auditory forebrain may be tuned to a pair of acoustic features separated in time (Figure 12d). It will be worthwhile to test whether there is a functional group of neurons that respond to multiple acoustic structures separated in time. If responses of this type can be documented in a population of neurons, they may contribute to temporal integration over hundreds of milliseconds.

In addition to these implications concerning sound segregation, categorical perception and temporal integration, the current work also raises an important question that brings together two important topics in neuroscience: sensory coding and memory. On the one hand, the traditional approach in studies of sensory coding in neurophysiology is to look for correlations between the firing properties of neurons and physical properties of stimuli. In auditory research, a lot of studies have been done to understand how neural activity codes

pitch, frequency contours and amplitude modulations, etc. Although it was understood that most of these features were dynamic in time, researchers tended to consider auditory neurons as filters with fixed responses to these features. On the other hand, most neurophysiological studies on memory in sensory systems are conducted in a behaviorist approach. Animals were extensively trained to make discriminations between stimuli or establish associations between a sound as CS with a shock as US and then the sensory maps, the tuning properties or the response strength in sensory systems were tested for changes (Recanzone et al., 1993; Polley et al., 2006; Fritz et al., 2005; Gentner and Margoliash, 2003; for a review, see Weinberger, 2004). However, cognitive studies of priming effects showed that acquiring information about a sensory object may not need extensive training. Instead, a few trials of exposure to a stimulus (without any task requirement) may cause significant changes in object recognition or discrimination in subjects (for a review, see Bowers and Marslek, 2003). These effects, some of which may be considered “priming effects” have been recognized as an important form of implicit memory, but attracted little interest from neurophysiologists.

The current work has shown that memory and sensory coding may be two tightly related issues. Auditory processing in the passive hearing state may dynamically update memory, which in turn may modulate the sensory coding at different time scales. First, auditory responses to sequences of acoustic events were modulated by the order of events at a time scale of ten to hundreds of milliseconds (Experiment 6), which probably involves the short memory auditory store described by Cowan (1984). Second, once a stimulus has been presented at a certain ISI for several trials, a few seconds delay in the ISI changed the

amplitude of the response (Experiment 5). Third, auditory responses reflected predictions about the category of expected sounds derived from recent acoustic history over a time scale from seconds to minutes (Experiment 1). Finally, familiarity of contextual sounds changed the strength of responses of other sounds, an effect that could last for minutes to hours (Experiment 4). In general, the results support the view that neuronal activity in response to sounds reflects not just the stimulus, but also sequential effects that depend on short-term memory and/or the expected stimulus inferred from “a long auditory store”. This expectation is constantly updated by what has been heard and when it was heard. Therefore, we believe that a form of implicit auditory memory interacts dynamically with auditory coding in NCM and CLM. Information about sound statistics can be accumulated during passive exposure to repeated stimuli and such information may enable the auditory system to make predictions, which may not only enhance detection of novel sounds, but also facilitate recognition and discrimination of experienced auditory objects. Furthermore, to understand how information is coded in the responses of sensory neurons, we also need a model that includes not only the interpretation of sound features but also the influence of other factors, such as familiarity of sounds in long-term history, the probability of sounds in short-term history and sequential effects at the hundred millisecond scale. One pioneering study obtained data consistent with this kind of thinking: the responses of neurons in CLM were better explained by a model that included not only stimulus features, but also whether transitions between those features violated the normative statistics of those transitions in conspecific song (Gill et al., 2008).

In sum, the current work contributes to the understanding of auditory processing under the influences of the short-term and the long-term acoustic history. We clarified the confusion

between SSA and probability-dependent enhancement effects in previous studies and provide evidence supporting the potential roles of categorical processing and familiarity in auditory scene analysis. The current work also stresses the importance of bringing together the research of sensory memory and neural coding, and of including memory as a factor in future studies of auditory processing.

References

Alvarez, G.A., & Oliva, A, 2008. The representation of simple ensemble visual features outside the focus of attention. *Psychological Science* 19:392–398.

Ariely D, 2001. Seeing sets: Representation by statistical properties. *Psychological Science* 12, 157–162.

Balsam, P. & Gallistel, C. R. 2009. Temporal maps and informativeness in associative learning. *Trends in Neurosciences*, 32(2), 73-78.

Bar M, 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. *J Cogn Neurosci* 4:600-9.

Bar M, 2004. Visual objects in context. *Nat Rev Neurosci* 5:617-29.

Beckers GJ, Gahr M. 2010. Neural processing of short-term recurrence in songbird vocal communication. *PLoS One*. 17;5(6):e11129.

Böttcher-Gandor C, Ullsperger P. 1992. Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval. *Psychophysiology*. 29(5):546-50.

Bowers and Marslek, 2003. *Rethinking Implicit Memory*. Oxford University Press.

Bregman AS, 1990. *Auditory Scene Analysis*. MIT Press Cambridge, Massachusetts.

Chew SJ, Mello C, Nottebohm F, Jarvis E, Vicario DS, 1995. Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proc. Natl. Acad. Sci. U S A.*:3406-10.

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

Cowan N. 1984. On short and long auditory stores. *Psychol Bull.* 96(2):341-70.

Fritz JB, Elhilali M, Shamma SA, 2005. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J. Neurosci.* 25(33):7623-35.

Gallistel, C. R. & King, A. P. 2009. *Memory and the computational brain: Why cognitive science will transform neuroscience.* New York: Wiley/Blackwell.

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. 2006. Recursive syntactic pattern learning by songbirds. *Nature.* 27;440(7088):1204-7.

Gentner TQ, Margoliash D, 2003. Neuronal populations and single cells representing learned auditory objects. *Nature* 424(6949):669-74.

Gibbon J. 1977. Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84(3), 279-325.

Gill P, Woolley SM, Fremouw T, Theunissen FE, 2008. What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. *J. Neurophysiol.* 99(6):2809-20.

Goldstein, 2007. *Sensation and Perception.* Wadsworth Publishing.

Griffiths TD, Warren JD, 2004. What is an auditory object? *Nat. Rev. Neurosci.* 5(11):887-92.

Haberman, J, Whitney, D, 2007. Rapid extraction of mean emotion and gender from sets of faces. *Current Biology* 17: R751–R753.

He J, Hashikawa T, Ojima H, Kinouchi Y, 1997. Temporal integration and duration tuning in the dorsal zone of cat auditory cortex. *J. Neurosci.* 17(7):2615-25.

Jeanne JM, Thompson JV, Sharpee TO, Gentner TQ. 2011. Emergence of learned categorical representations within an auditory forebrain circuit. *J Neurosci.* 31(7):2595-606.

Kuhl, P. K. & Miller, J. D., 1975 Speech perception by the chinchilla: voice–voiceless distinction in alveolar plosive consonants. *Science* 90, 69–72.

Leppelsack HJ and Vogt M, 1976. Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology.* Volume 107, Number 3, 263-274.

Lu K, Vicario DS. 2011. Toward a neurobiology of auditory object perception: what can we learn from the songbird forebrain. *Current Zoology* 57 (6): 671–683.

Margoliash D, Fortune ES., 1992. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *J Neurosci.* 12(11):4309-26.

Mauk MD, Buonomano DV., 2004. The neural basis of temporal processing. *Annu Rev Neurosci.* 27:307-40.

Maye, J., Werker, J. F. & Gerken, L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.

McAdams, S., Bigand, E., 1993. *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford University Press.

McCarthy, G., Donchin, E., 1976. The Effects of Temporal and Event Uncertainty in Determining the Waveforms of the Auditory Event Related Potential (ERP). *Psychophysiology*. 13(6):581-90.

Mello CV, Vicario DS, Clayton DF, 1992. Song presentation induces gene expression in the songbird forebrain. *Proc. Natl. Acad. Sci. U S A*. 89(15):6818-22.

Mendelson JR, Cynader MS, 1985. Sensitivity of cat primary auditory cortex (AI) neurons to the direction and rate of frequency modulation. *Brain Res*. 327: 331–335

Miller AL, Sheldon R, 1969. Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology* 81:16–21.

Miyawaki, K., Liberman, k., Jenkins, J., Fujimura, O. An effect of linguistic experience: the discrimination of /r/ and /l/ by native speakers of Japanese and English. *Percept. Psychophys.* 18, 331–340 (1975).

Muller CM, Leppelsack HJ, 1985. Feature extraction and tonotopic organization in the avian forebrain. *Exp. Brain. Res*. 59:587–599.

Näätänen R. 1995. The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear*. 16(1):6-18.

Naatanen, 1992. Attention and brain function. Psychology Press.

Näätänen R, Jacobsen T, Winkler I, 2005. Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology*. 42:25-32.

Nelken I, Bar-Yosef O, 2008. Neurons and objects: the case of auditory cortex. *Front. Neurosci*. 2(1):107-13.

Nelken, I., Ulanovsky, N., 2007. Mismatch Negativity and Stimulus-Specific Adaptation in Animal Models. *Journal of Psychophysiology*. Volume 21, Issues 3-4, Pages 214-223

Okanoya K. 2004. The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. *Ann N Y Acad Sci*. 1016:724-35.

Oliva A, Torralba A, 2006. Building the gist of a scene: the role of global image features in recognition. *Prog Brain Res* 155:23-36.

Plamondon SL, Rose GJ, Goller F. 2010. Roles of syntax information in directing song development in white-crowned sparrows (*Zonotrichia leucophrys*). *J Comp Psychol*. 124(2):117-32.

Polley DB, Steinberg EE, Merzenich MM, 2006. Perceptual learning directs auditory cortical map reorganization through top-down influences. *J. Neurosci*. 26(18):4970-82.

Prather JF, Peters S, Nowicki S, Mooney R. , 2009. Neural correlates of categorical perception in learned vocal communication. *Nature Neuroscience*. 12(2): 221-228.

Rauschecker JP, 1998. Cortical processing of complex sounds. *Curr. Opin. Neurobiol*. 8(4):516-21.

Recanzone GH, Schreiner CE, Merzenich MM, 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13(1):87-103.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., Bialek, W, 1999. *Spikes: Exploring the Neural Code (Computational Neuroscience)*. The MIT Press.

Rose GJ, Goller F, Gritton HJ, Plamondon SL. 2004. Baugh AT, Cooper BG. Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature*. 9;432(7018):753-8.

Saini KD, Leppelsack HJ, 1977. Neuronal arrangement in the auditory field L of the neostriatum of the starling. *Cell Tissue Res.* 176(3):309-16.

Schreiner CE, Read HL, Sutter ML, 2000. Modular organization of frequency integration in primary auditory cortex. *Annu. Rev. Neurosci.* 23:501-29.

Schreiner CE, Urbas JV, 1988. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res.* 32(1):49-63.

Sen K, Theunissen FE, Doupe AJ, 2001. Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* 86(3):1445-58.

Sen K, Theunissen FE, Doupe AJ, 2001. Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* 86(3):1445-58.

Shamma SA, Fleshman JW, Wiser PR, Versnel H, 1993. Organization of response areas in ferret primary auditory cortex. *J. Neurophysiol.* 69: 367–383.

Solis MM, Doupe AJ. 1997. Anterior forebrain neurons develop selectivity by an intermediate stage of birdsong learning. *J Neurosci.* 17(16):6447-62.

Stilp C, Rogers T, Kluender K. 2010. Rapid efficient coding of correlated complex acoustic properties. *Proc Natl Acad Sci U S A.* 14;107(50):21914-9.

Stripling R, Volman SF, Clayton DF, 1997. Response modulation in the zebra finch neostriatum: relationship to nuclear gene regulation. *J. Neurosci.* 17(10):3883-93.

Suga N, O'Neill WE, 1979. Neural axis representing target range in the auditory cortex of the mustache bat. *Science.* 206(4416):351-3.

Terleph TA, Mello CV, Vicario DS, 2006. Auditory topography and temporal response dynamics of canary caudal telencephalon. *J. Neurobiol.* 66(3):281-92.

Theunissen FE, Sen K, Doupe AJ, 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci.* 20(6):2315-31.

Theunissen FE, Shaevitz SS, 2006. Auditory processing of vocal sounds in birds. *Curr Opin Neurobiol.* 16(4):400-7.

Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK, 2009. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci* 10:1934-45.

Ulanovsky N, Las L, Farkas D, 2004. Nelken I. Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci.* 17;24(46):10440-53.

Ulanovsky N, Las L, Nelken I, 2003. Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6(4):391-8.

Vates GE, Broome BM, Mello CV, Nottebohm F, 1996. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J. Comp. Neurol.* 366(4):613-42.

Vicario DS., 2004. Using learned calls to study sensory-motor integration in songbirds. *Ann N Y Acad Sci.* 1016:246-62.

Wang X, Merzenich MM, Beitel R, Schreiner CE, 1995. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiol.* 74(6):2685-706.

Wang Y, Brzozowska-Prechtl A, Karten HJ, 2010. Laminar and columnar auditory cortex in avian brain. *Proc. Natl. Acad. Sci. U S A.* 107(28):12676-81.

Weinberger NM. 2004. Specific long-term memory traces in primary auditory cortex. *Nat Rev Neurosci.* 5(4):279-90.

Wild JM, Karten HJ, Frost BJ, 1993. Connections of the auditory forebrain in the pigeon (*Columba livia*). *J. Comp. Neurol.* 1;337(1):32-62.

Wollberg, Z., and Newman, J. D, 1972. Auditory cortex of squirrel monkey: response patterns of single cells to species-specific vocalizations. *Science* 175, 212–214.

Woolley SM, Gill PR, Fremouw T, Theunissen FE, 2009. Functional groups in the avian auditory system. *J Neurosci.* 29(9):2780-93.

Woolley SM, Gill PR, Fremouw T, Theunissen FE, 2009. Functional groups in the avian auditory system. *J. Neurosci.* 29(9):2780-93.

Zann R, 1996. *The Zebra Finch: A Synthesis of Field and Laboratory Studies* (Oxford Ornithology Series). Oxford University Press.

Figure and Table Legends

Table 1. P-values for all Post hoc comparisons in DSs by sites in Experiment 1.

Figure 1. Ascending auditory pathways in songbirds are indicated by arrows. Auditory nuclei of avian hindbrain innervate MLd (homologue of the inferior colliculus). MLd innervates OV (homologue of the medial geniculate). OV projects to forebrain field L2 (analogue of lay IV of A1). Field L2 innervates L1, L3, NCM and CM (thought to be analogues of superficial layers of A1). (modified from Fig1, Theunissen and Shaevitz, 2006)

Figure 2. Schematic of the order of stimulus presentation in Experiment 1. Each row represents one test condition (labeled at left). The stimulus order in the pre-adapting session and the post-context session is the same for all conditions (red boxes). In ZCR and ZSR, the oddball song was presented with the same timing, but the context songs between oddballs in ZCR were replaced by same length of silences in ZSR. The same manipulation was made for ZCF and ZSF. In ZZR, the pre-adapted song was presented with the same timing as ZCR, but context songs were replaced by seven zebra finch songs. The oddball song denoted by “zf” is a different song in each condition.

Figure 3. Spectrograms of a canary song (top panel) and a zebra finch song (bottom panel). X axis represents the time in milliseconds. Y axis represents the frequency. The darkness represents the intensity of sounds. Songs of the two species differ in both their spectral and temporal structures.

Figure 4. Example of electrophysiological recordings and RMS computation. The bottom panel shows the spectrogram of the song presented to the animal. Raw neural activity

recorded from three electrodes (green traces) during a playback trial is shown, together with the moving RMS (window= 50ms) calculated for each recording (black traces). RMS of the baseline was calculated over the 500ms before each stimulus (black box). RMS of the response was calculated over the stimulus period plus 100ms (red box). Responses to each song were defined as the difference between the response RMS and the baseline RMS.

Figure 5. Illustration of Delta Surprisal (DS) computation. Each dot in the figure represents a response to the test song on one trial. Black dashed lines indicate the borders between the three sessions. Solid black lines are regression lines obtained from responses (blue dots) in the first and last sessions. The solid green line connecting the end points of these regressions estimates the mean of the distribution of responses expected during the Context-Modulated session if the context had no effect on the course of adaptation in the response to the test song. The observed responses in the Context-modulated session are shown as red dots. The standard deviation of the expected distribution is estimated as the standard deviation of the residuals in the Pre-adapting and Post-context sessions. The extent to which the observed responses are unexpected is measured by their surprisal, which is $1/\log(P(d))$, normalized by the surprisal of the expected response, which is $\log(1/P(0))$. Note that the expectation of the signed surprisals is 0

Figure 6a. Example of electrolytic lesions in **NCM** for confirmation of recording sites. Two lesions were made by this electrode. First lesion was made at the recording site (the upper dark area), then the electrode was advanced for 500um and the second lesion was made (the lower dark area). The two lesions allow confirmation of both the recording site and the angle of the electrode penetration.

Figure 6b. Example of electrolytic lesions in **CLM** for confirmation of recording sites. The recording was made at the upper lesion site.

Figure 7. Cumulative frequency distributions of DSs for the five conditions tested in Experiment 1. X axis represents the DS values. Y axis represents the cumulative frequency for each x value as a percent of the total. The order of the five lines shows the rank of DSs observed from the five conditions.

Figure 8. Boxplots of matched contrasts between conditions in each bird. Ten panels represent ten comparisons that were made between each two out of the 5 conditions. The title of each panel indicates the two conditions that were compared. In each panel, the distributions of differences of DSs between two conditions in 11 bird were represented by 11 boxes. The end of whiskers represents the minimum and maximum values. The upper border and lower border indicate the upper quartile and lower quartile, respectively. The red line in the middle of the box indicates median. The red cross represents outliers.

Figure 9. Percent of sites in each bird that show significant differences in trial-by-trial DSs between conditions. Ten panels represent ten comparisons that were made between each two out of the 5 conditions. The title of each panel indicates the two conditions that were compared. Each stacked bar shows the percent of sites showing positive differences (blue) and negative differences (red) in each bird.

Figure 10. The cumulative frequency distribution of correlation coefficients between AIs and DSs has a median of -0.04 and shows that there is no systematic relationship.

Figure 11.

A: The enhancement profile (black trace) for ZCR is displayed as the difference between the averaged response waveforms in the Pre-adapting session (blue trace) and the Context-modulated session (red trace). The vertical dashed line indicates the onset of the stimulus.

B: Comparison between the enhancement profiles for ZCR (red trace) and for ZZR (black trace). ZZR shows no enhancement effect (note: gain is higher than in A).

C: Comparison between the enhancement profiles for ZCR (red trace) and ZSR (blue trace). The solid vertical lines indicate the time after the stimulus onset when the profile decays to zero. This occurs earlier for ZSR (486ms) than for ZCR (667ms).

D: The early phase of the averaged response waveforms and the enhancement profile for ZCR, from Figure 5A (at a finer temporal scale). The solid vertical lines indicate the latency of responses (red) and the latency of the enhancement profile (black; 20ms later than the red line). Dashed red line with triangle indicates the peak of the responses (85ms after the stimulus onset). The arrow and dashed black line with triangle indicate the peak of the enhancement profile (112ms after the stimulus onset). Both the latency and the peak of the enhancement profile were later than the latency and the peak of the responses.

E: Plot of the early phase of the enhancement profiles in ZSR (blue trace) and ZSF (pink trace) shows two peaks (indicated by arrows and the dashed lines) in the enhancement profiles in the two conditions. For ZSR the peaks are at 27 and 105ms after stimulus onset, while, for ZSF, the peaks are at 30 and 89ms after stimulus onset.

F: Comparison of the initial phase of the enhancement profiles in ZCR (red trace) and in ZSR (blue trace), magnified from Figure 5C. The solid vertical lines indicate the latency for ZCR (26ms) and for ZSR (20ms). The dashed vertical lines indicate the peaks for ZCR (112ms) and for ZSR (105ms). Both the latency and the second peak occur earlier for ZSR than for ZCR.

Figure 12. Spectro-temporal receptive fields (STRFs) obtained from single-units in ZZR. The STRF in Figure **12a** indicates that this neuron tends to be excited by sounds within 3-7kHz frequency range and tends to be inhibited by sounds below 2kHz. The STRF in Figure **12d** indicates that this neuron tends to be excited by a high frequency sound (5-6kHz, beginning 150ms ago) followed by a low frequency sound (3-5kHz) with 50 ms interval.

Figure 13a. The lower panel shows the raster plot of responses of a neuron to a pre-adapted song in the ZCR condition. Each blue dot represents a spike. The red vertical line indicates the stimulus onset. The black vertical dashed line represents the offset of the stimulus. The three horizontal dashed lines indicate the end of the three sessions. The upper trace shows the waveform of the song that induced the responses. Temporal structure of spike trains in response to the song was observed across three sessions. Increased spikes were observed during the context-modulated session.

Figure 13b. Plot of spike rate of the single-unit activity displayed in figure 13a. Blue dots represent spike rate in the context-modulated session. Red dots represent spike rate in the pre-adapting session and the post-context session. Reduced responses in the first session

demonstrated SSA. Increased activity in the context-modulated session demonstrated the enhancement effect.

Figure 14. Cumulative frequency distributions of DSs obtained from single-units in ZCR (n=29) and ZZR (n=24).

Figure 15. Cumulative frequency distributions for the three conditions tested in Experiment 2. ZCR and CZR were not different from each other, but both were different from ZZR.

Figure 16. Boxplots of matched contrasts between conditions in each bird. Three comparisons were made between each two of the three conditions (represented in each panel). The title of each panel indicates the two conditions that were compared. Each box represented the distribution of differences of DSs between two conditions in each bird. Both CZR and ZCR were consistently different from ZZR.

Figure 17. Percent of sites in each bird that show significant differences in trial-by-trial DSs between conditions. Three comparisons were made between each two of the three conditions (represented in each panel). The title of each panel indicates the two conditions that were compared. Each stacked bar represents the percent of sites showing significant differences in each bird. In contrast of CZR/ZZR and ZCR/ZZR, differences were consistent across most sites.

Figure 18. Examples of stimuli used in Experiment 3. The top example is the waveform of an oddball sound with slow rhythm. The two examples below it are context sounds with fast rhythms.

Figure 19. Cumulative frequency distributions of DSs by sites for three conditions tested in Experiment 3. The control condition was significantly different from the other two conditions.

Figure 20. Boxplots of matched contrasts between conditions in each bird. The title of each panel indicates the two conditions that were compared. Each box represented the distribution of differences of DSs between two conditions in each bird.

Figure 21. Percent of sites in each bird that show significant differences in trial-by-trial DSs between conditions.

Figure 22. Cumulative frequency distributions of DSs for ZZR tested after and before training with zebra finch context songs in Experiment 4. Each boxplot represents the distribution of differences in DSs in each bird obtained after training and before training. Increased DSs were seen after training.

Figure 23. Boxplots of matched contrasts between post-training and pre-training session in each bird. Each box represented the distribution of differences of DSs between the session after training with zebra finch songs and the session before training with zebra finch songs in each bird. The training effect was consistent across 6 out of 7 birds.

Figure 24. Percent of sites in each bird that show significant differences in trial-by-trial DSs after and before training.

Figure 25. Cumulative frequency distributions of DSs for ZZR tested after and before exposure to canary songs in Experiment 4. DSs were not change after training.

Figure 26. Boxplots of matched contrasts between post-training and pre-training session in each bird. Each box represented the distribution of differences of DSs between the session after training with canary songs and the session before training with canary songs in each bird. Three birds showed small increase in DSs, while three birds showed decrease in DSs.

Figure 27. Percent of sites in each bird that show significant differences in trial-by-trial DSs after and before exposure to canary songs.

Figure 28.

A: Cumulative frequency distributions of D' for 7 context songs (blue lines) and the pre-adapted song (red line) in the **pre-training** ZZR. The red line was not distinguishable from the blue lines.

B: Cumulative frequency distributions of D' for 7 context songs (blue lines) and the pre-adapted song (red line) in the **post-training** ZZR. After training, the red line showed a categorical difference from the blue lines.

Figure 29. Cumulative frequency distributions of responses ratios by sites tested in the Paradigm 1 of Experiment 5. More than 85% of sites showed a significant increase in responses on the trials following increased intervals (blue line). Responses on the trial without interval change showed no increase and thus the ratios were symmetrically distributed around zero (red line).

Figure 30. Boxplots of matched contrasts between conditions in each bird. Each box represented the distribution of differences between interval-varying ratios and control ratios in each bird.

Figure 31. Percent of sites in each bird that show significant differences between interval-varying ratios and control ratios.

Figure 32. Cumulative frequency distributions of DSs by sites for the interval-varying condition and the control condition tested in Paradigm 2 of Experiment 5. Interval changes in balanced design did not significantly change DSs compared with the control condition.

Figure 33. Boxplots of matched contrasts between conditions in each bird. Each box represents the distribution of differences of DSs between the interval-varying condition and the control condition in each bird.

Figure 34. Percent of sites in each bird that show significant differences in trial-by-trial DSs between conditions.

Figure 35. Examples of stimuli used in experiment 6. The lower two panels show two sequence stimulus sequences assembled from the same four notes in different order, with stimulus spectrographs shown above. These examples have note durations of 150 ms, but sequences with note durations of 60 and 200ms were also tested (see Fig. 40).

Figure 36. Example of a testing order in Experiment 6. The three tables represent the tests for the three duration conditions. Each letter in a box represents a different artificial note and the order of the letters represents the sequence. Each box in a table represents the sequence

tested in a particular session. In this example, the 60 ms condition was tested first and the 200 ms condition was tested last. The order of conditions was varied between birds.

Figure 37. Example of order-selective responses showing how the Novelty index (NI) and the Control index (CI) were calculated. Each red dot represents the response RMS to a stimulus. The x axis indicates the trial number. Each blue dashed line indicates a transition from one order to another. The NI is the ratio of the mean response on the three trials in the black box to the mean response in the green box, while the CI is the ratio of the mean response on the three trials in the green box to the mean response in the blue box. As shown in this figure, responses in the black box are higher than those in the green box, suggesting a response to the order change.

Figure 38. Cumulative frequency distributions of NIs (Novelty indices) and CIs (Control indices) by sites tested in Experiment 6. More than 90% of sites show NIs higher than 1.0, compared with CIs that were symmetrically distributed around zero.

Figure 39. Boxplots of matched contrasts between NIs and CIs in each bird. Each box represented the distribution of differences of NIs and CIs in each bird.

Figure 40. Cumulative frequency distributions of normalized NIs by sites tested for three note durations in Experiment 6. The sequence with 100 ms note duration showed lowest novelty effect.

Figure 41. Boxplots of matched contrasts between conditions in each bird. The title of each panel indicates the two conditions that were compared. Each box represented the distribution of differences of NIs and CIs in each bird.

Table 1

	ZCR	ZSR	ZCF	ZSF
ZCR				
ZSR	0.006*			
ZCF	0.001*	0.739		
ZSF	0.001*	0.512	0.624	
ZZR	0.001*	0.001*	0.001*	0.001*

Figure 1

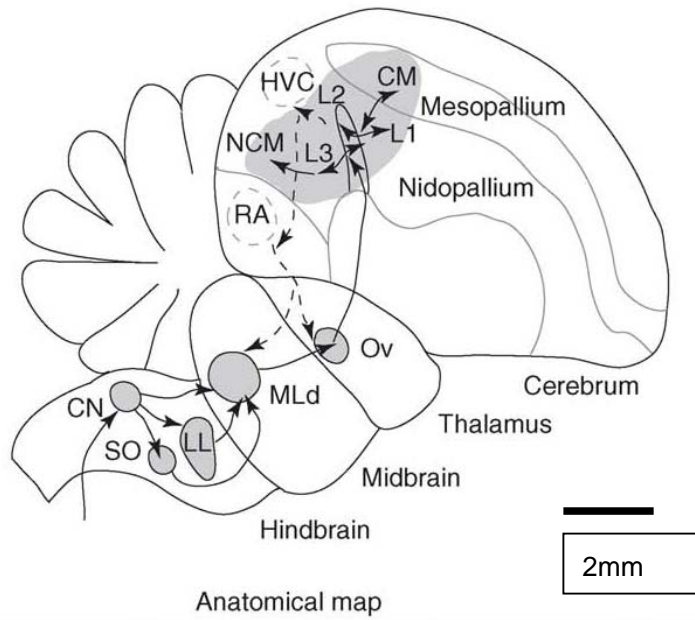


Figure 2

ZCR:	zf...zf,zf,	c1,c7,c6,c4,c5,zf,c3,c1,zf,c2,c6,c3,c3,c1,c6,c1,zf,c5,zf,c2,c2,c5,c2,c2...	zf,zf...zf
ZSR:	zf...zf,zf,	SI,SI,SI,SI,SI, zf, SI,SI, zf, SI,SI, SI,SI,SI,SI,SI, zf, SI,zf, SI,SI,SI,SI,SI ...	zf,zf...zf
ZCF:	zf...zf,zf,	c3,c6,c2,c5,c7,c1,c4,zf,c2,c6,c7,c1,c3,c2,c3, zf,c4,c4,c3,c5,c6,c1,c7,zf...	zf,zf...zf
ZSF:	zf...zf,zf,	SI,SI,SI,SI,SI,SI,SI, zf, SI,SI,SI,SI,SI,SI,SI, zf, SI,SI,SI,SI,SI,SI,SI, zf...	zf,zf...zf
ZZR:	zf...zf,zf,	z1,z7,z6,z4,z5,zf,z3,z1,zf,z2,z6,z3,z3,z1,z6,z1,zf,z5,zf,z2,z2,z5,z2,z2....	zf,zf...zf
	pre-adapting session	context-modulated session	post-context session

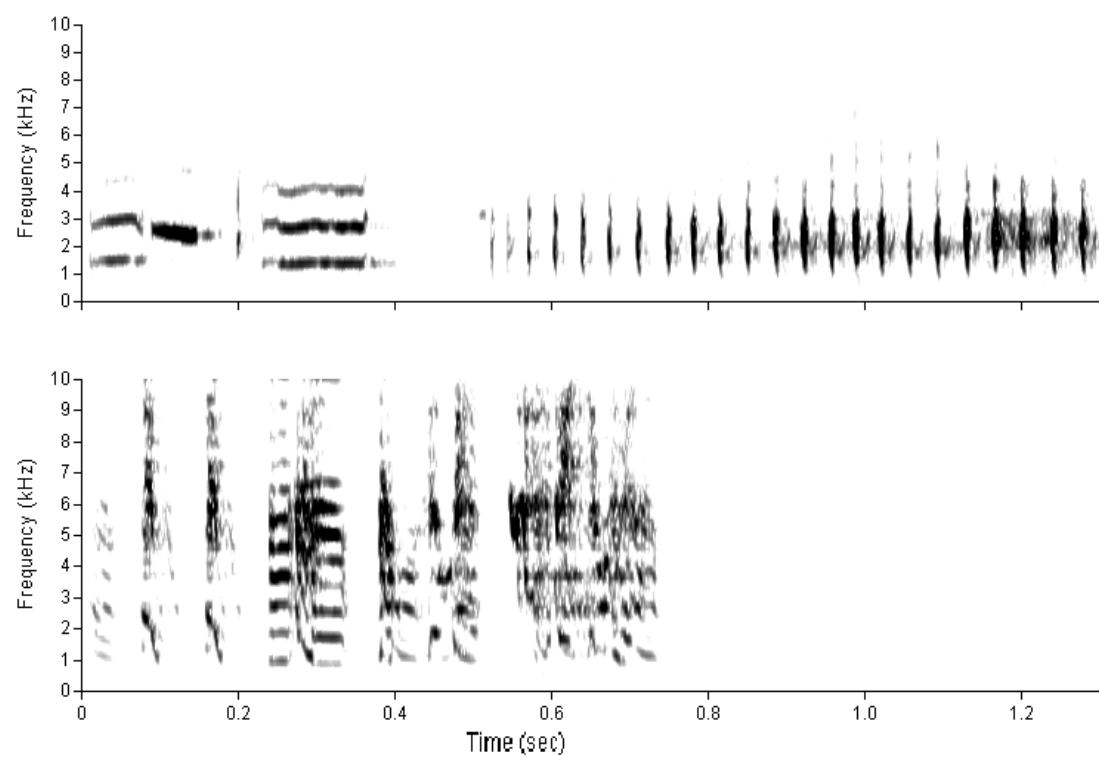
Figure 3

Figure 4

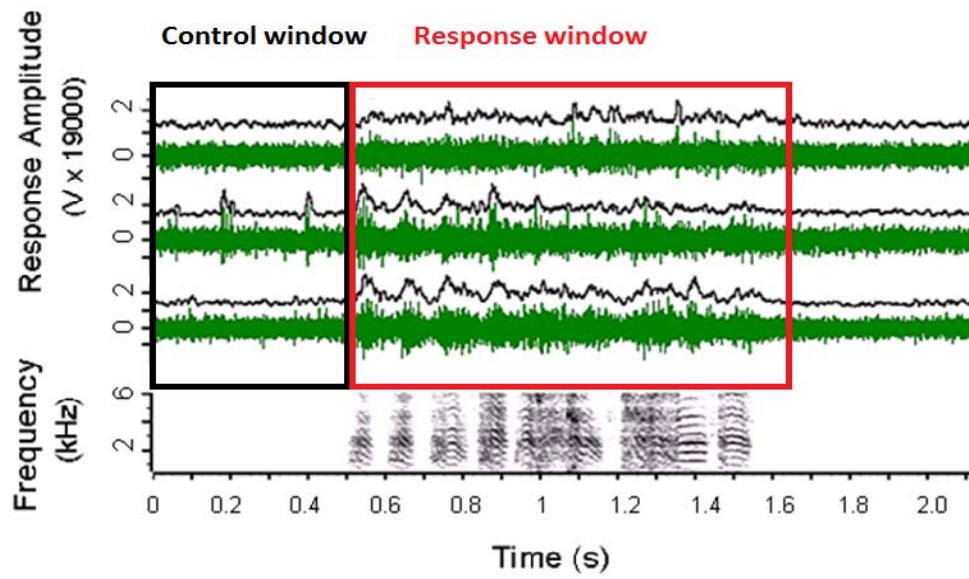


Figure 5

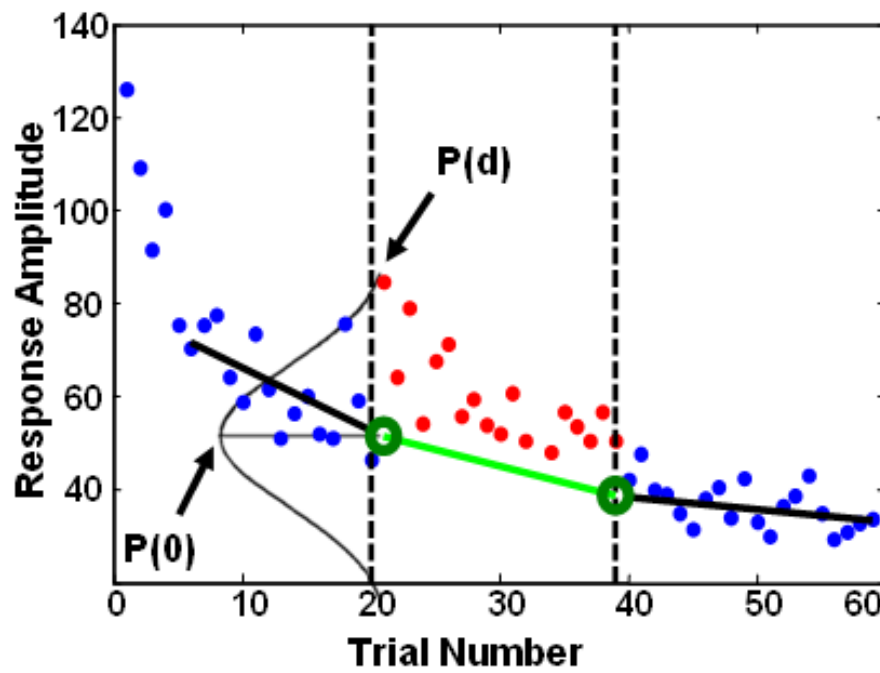


Figure 6a

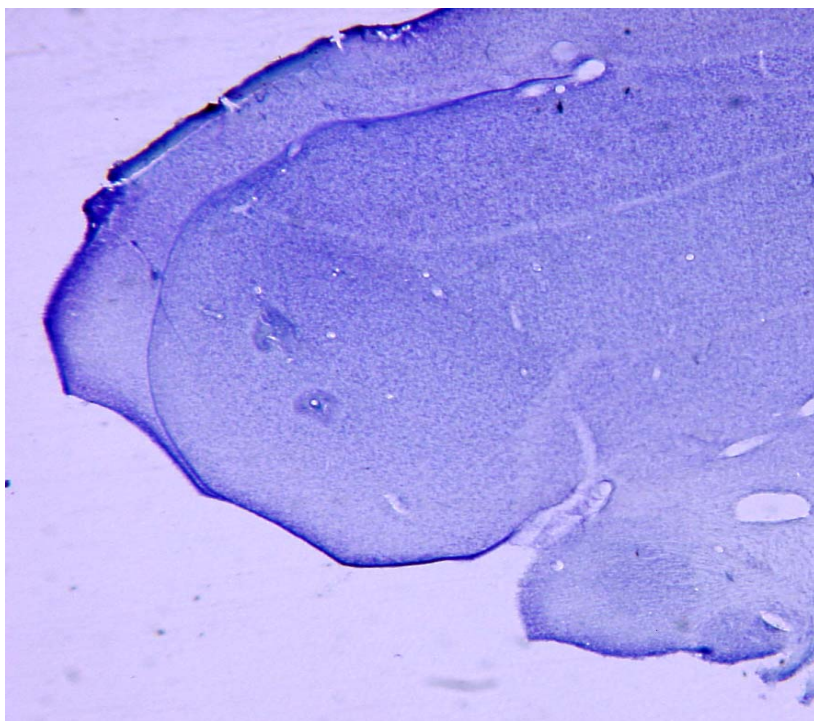


Figure 6b

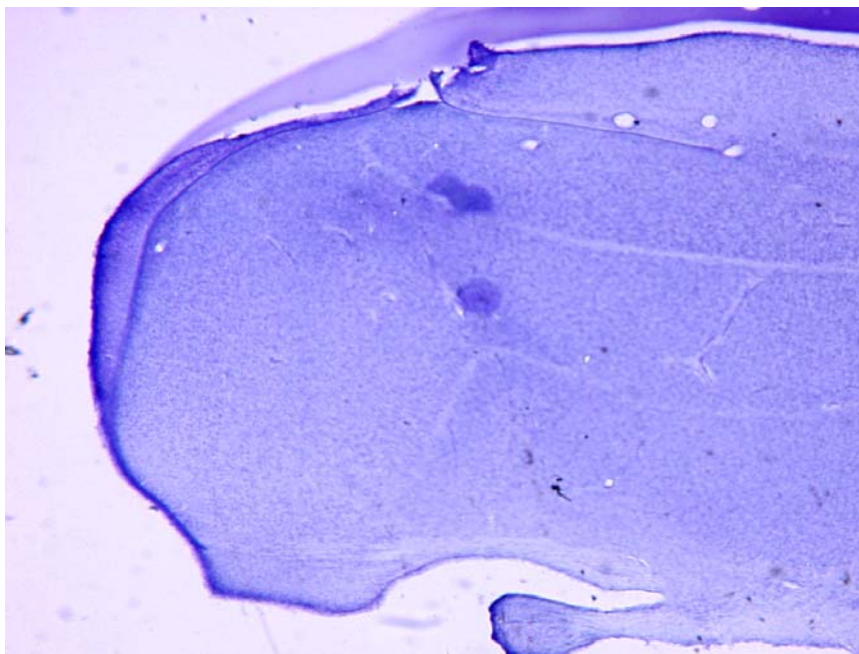


Figure 7

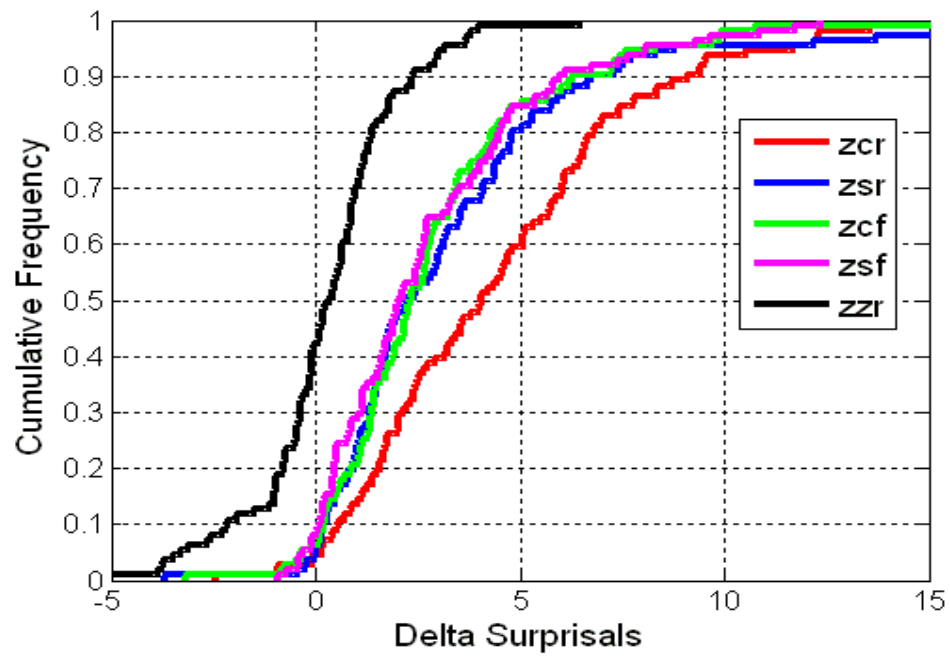


Figure 8.

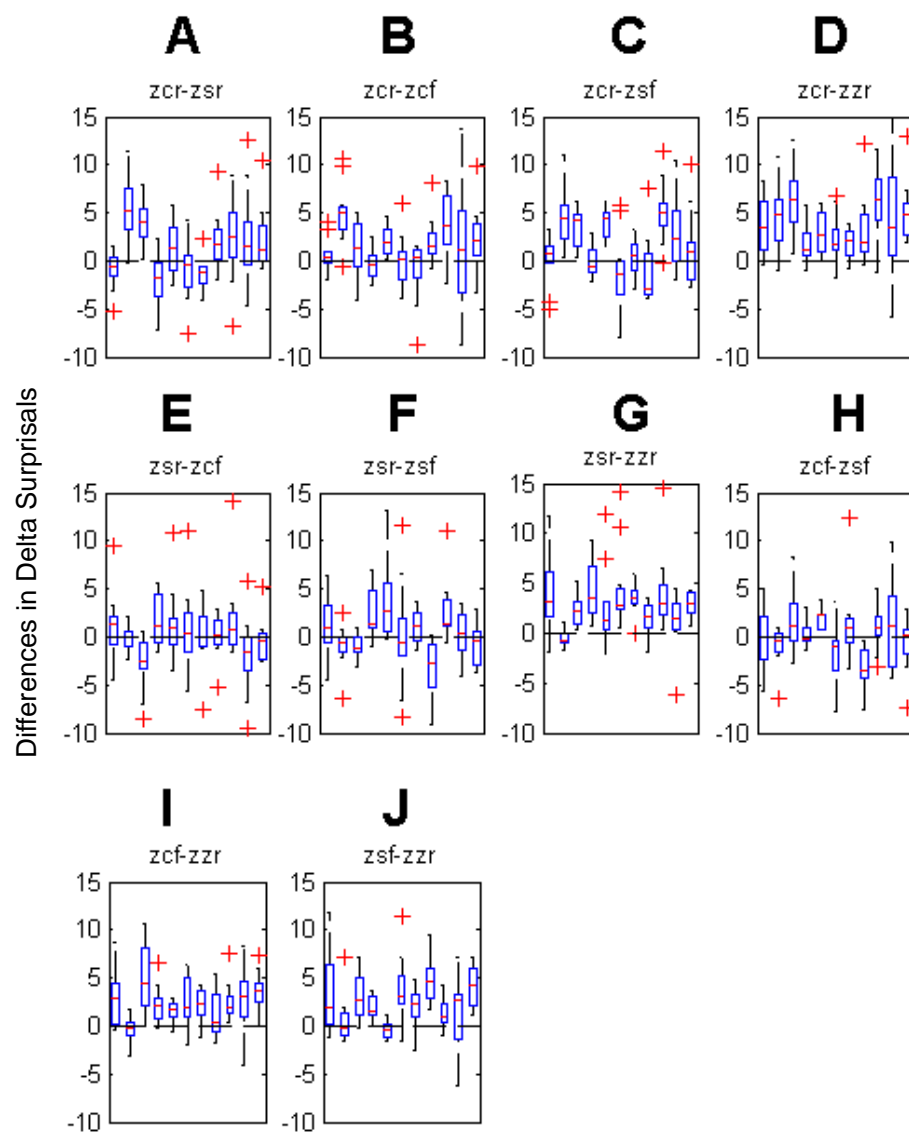


Figure 9.

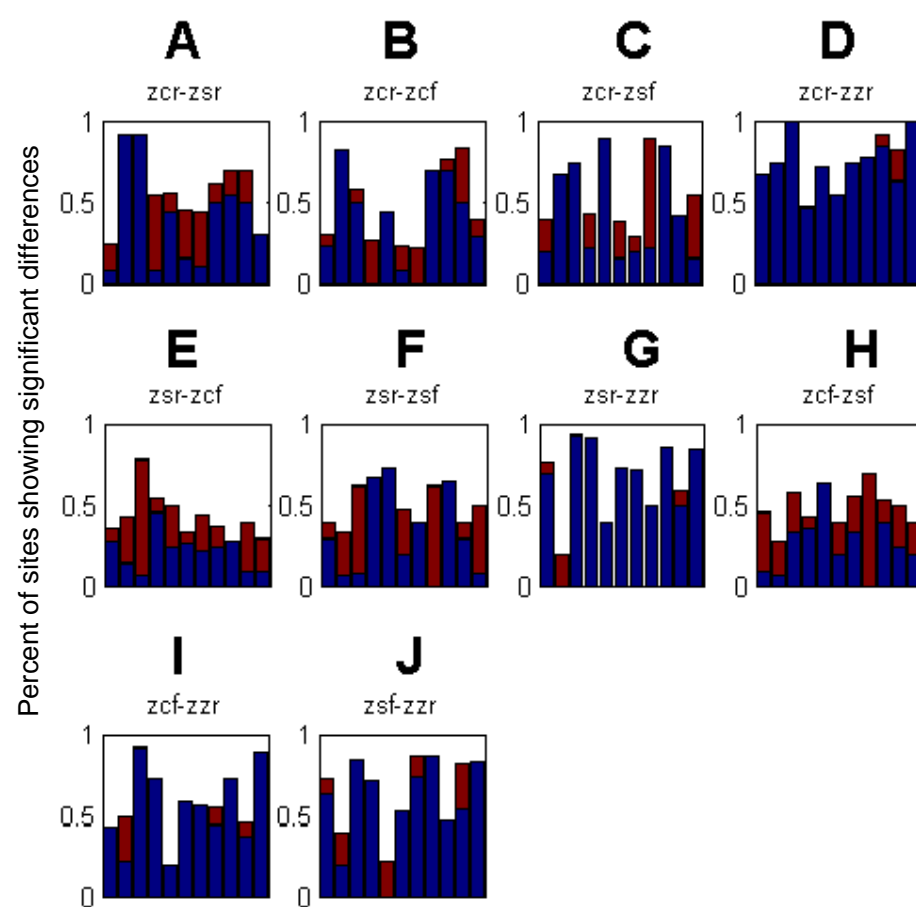


Figure 10

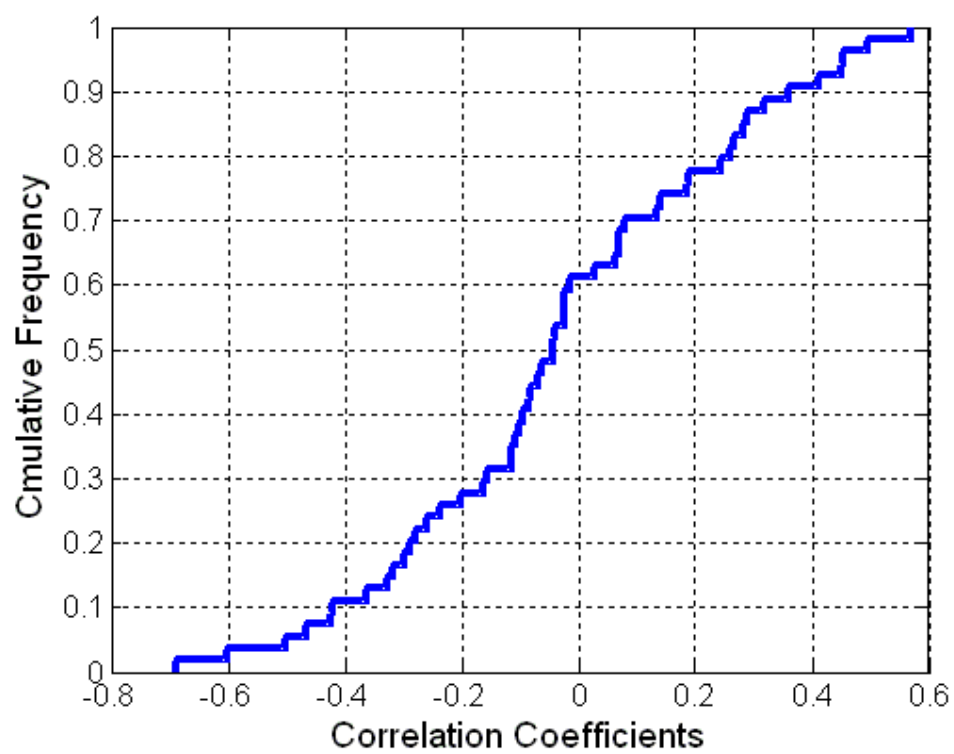


Figure 11

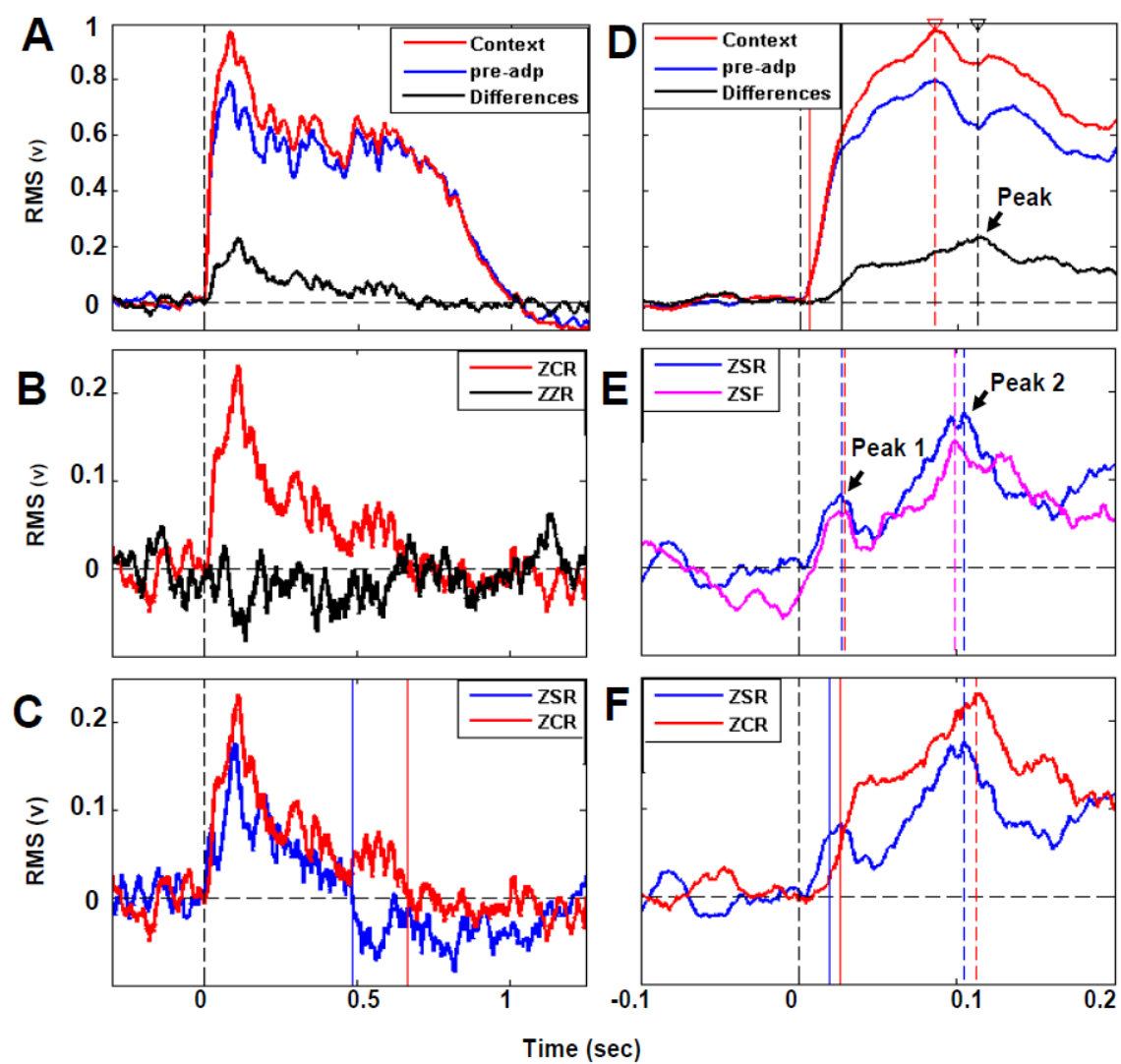


Figure 12

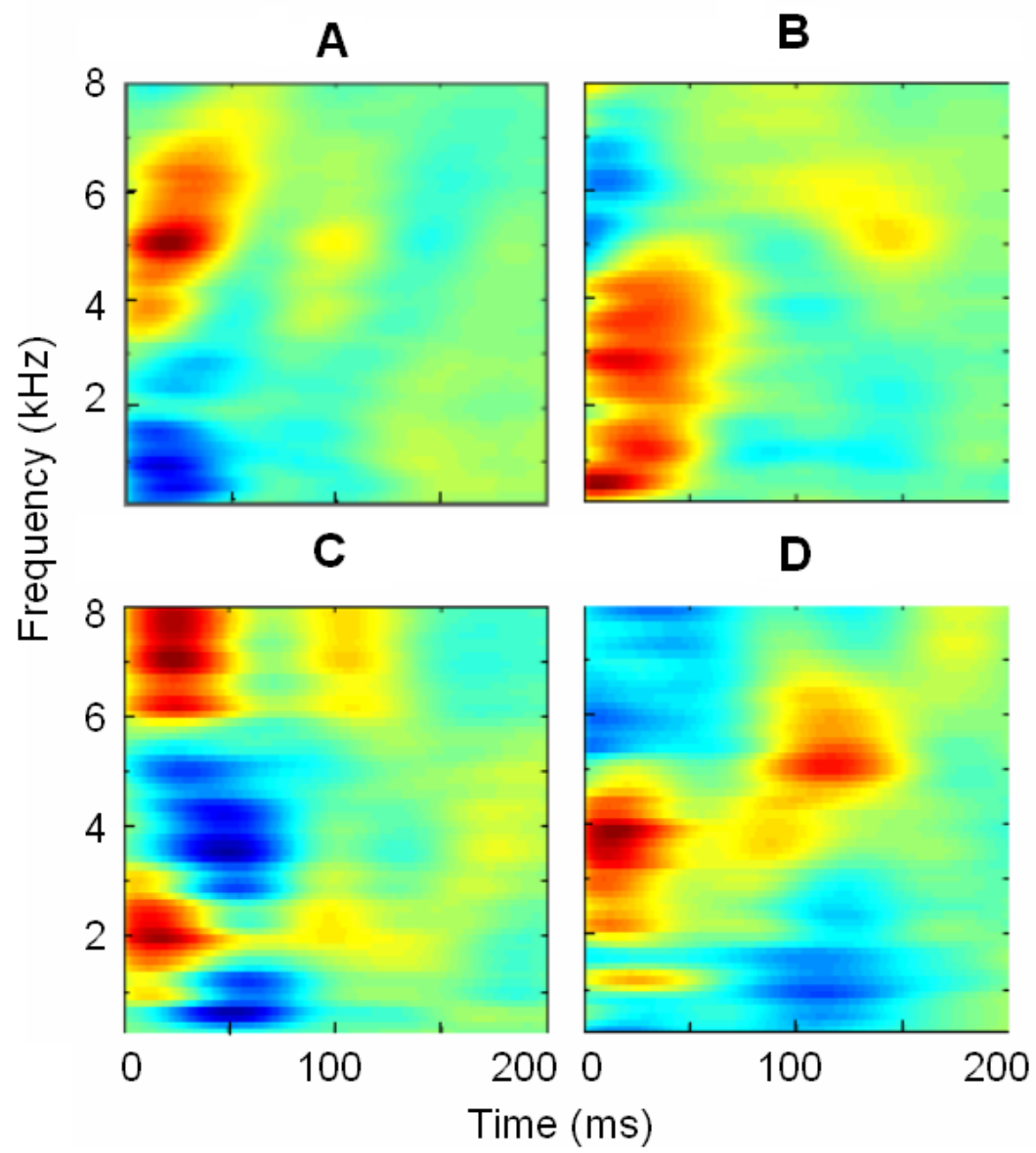


Figure13a

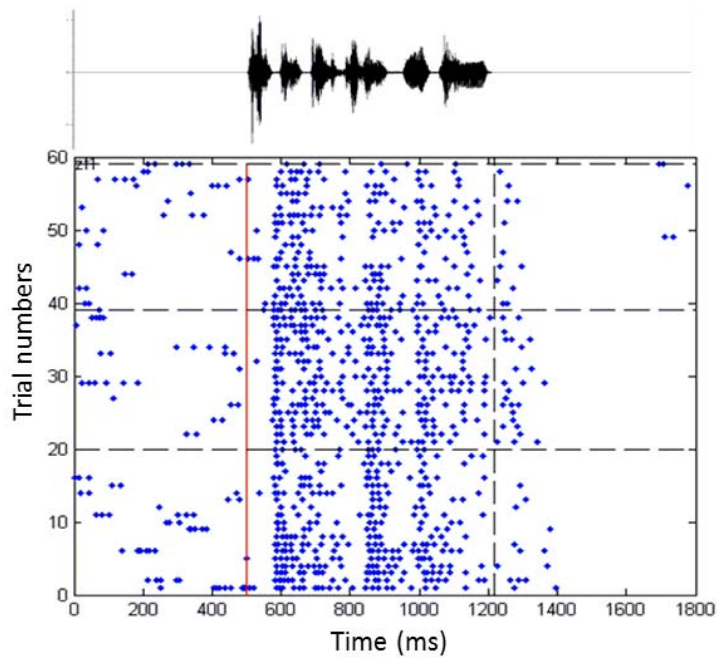


Figure13b

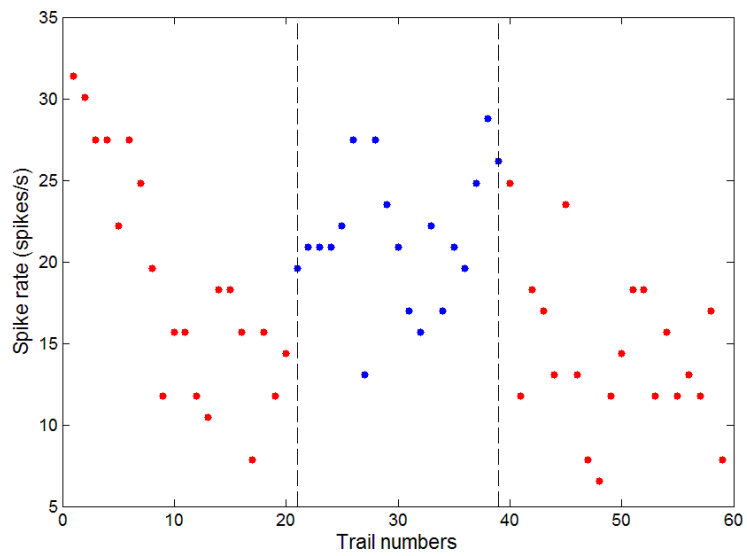


Figure 14

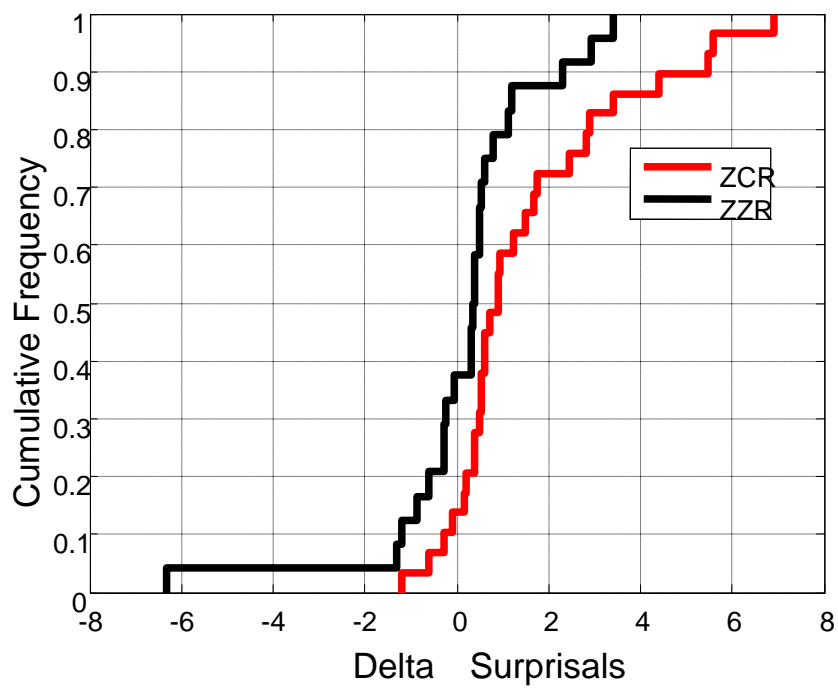


Figure 15

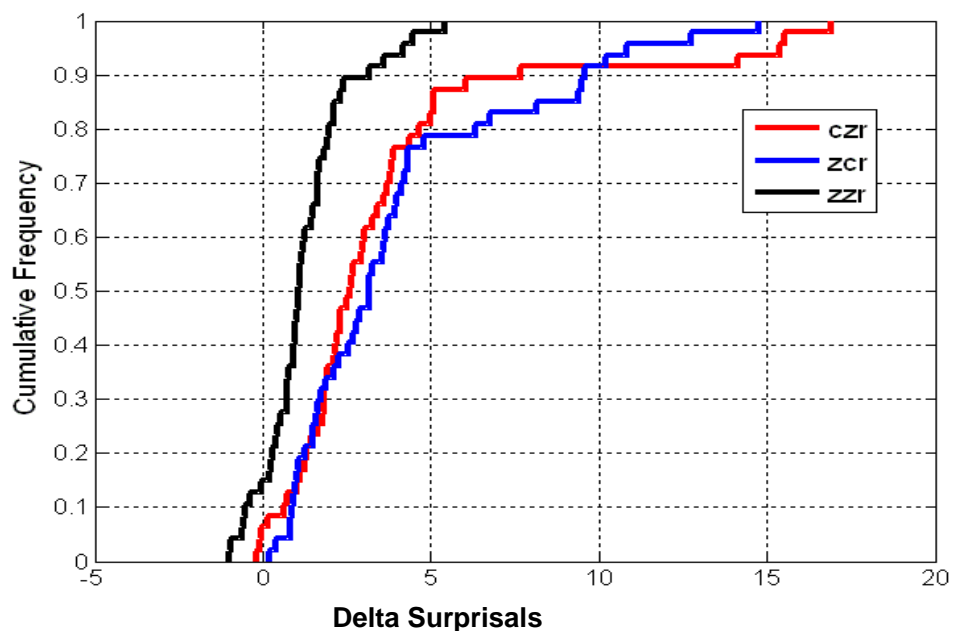


Figure 16

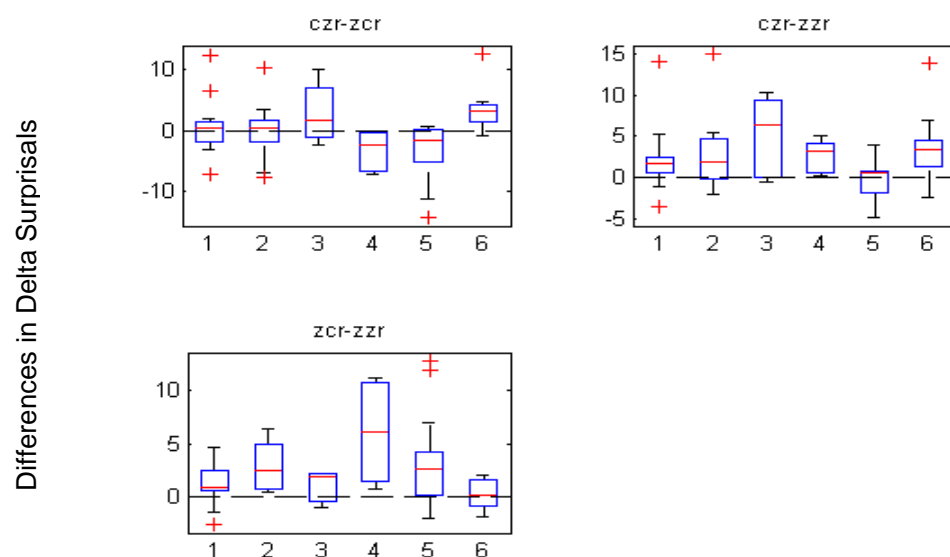


Figure 17

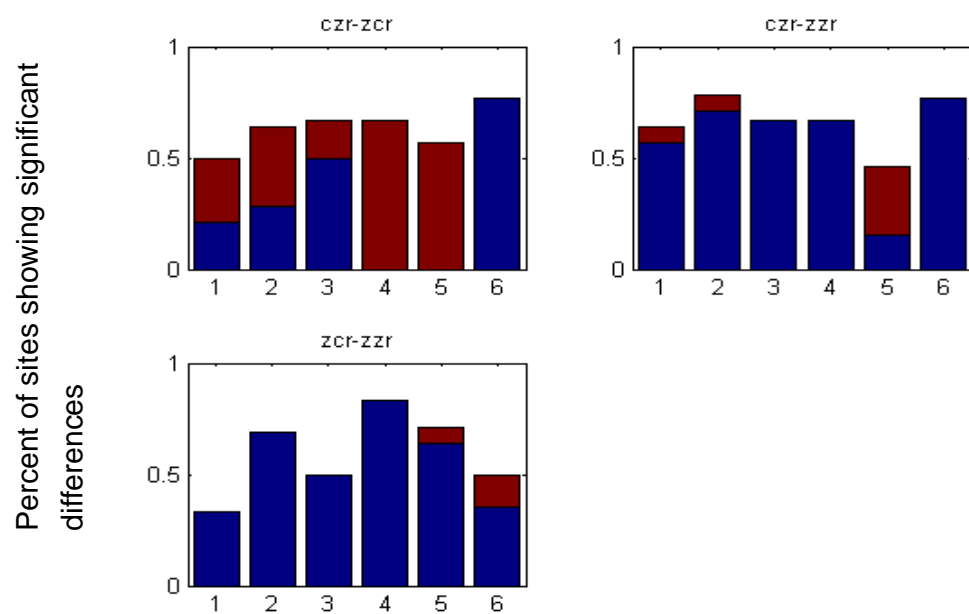


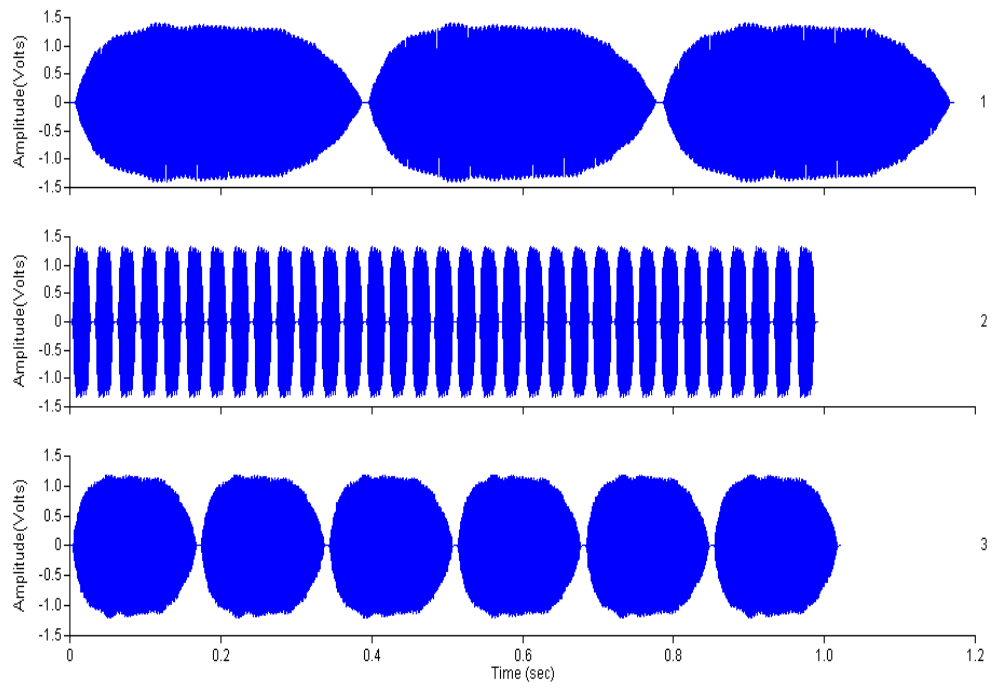
Figure 18

Figure 19

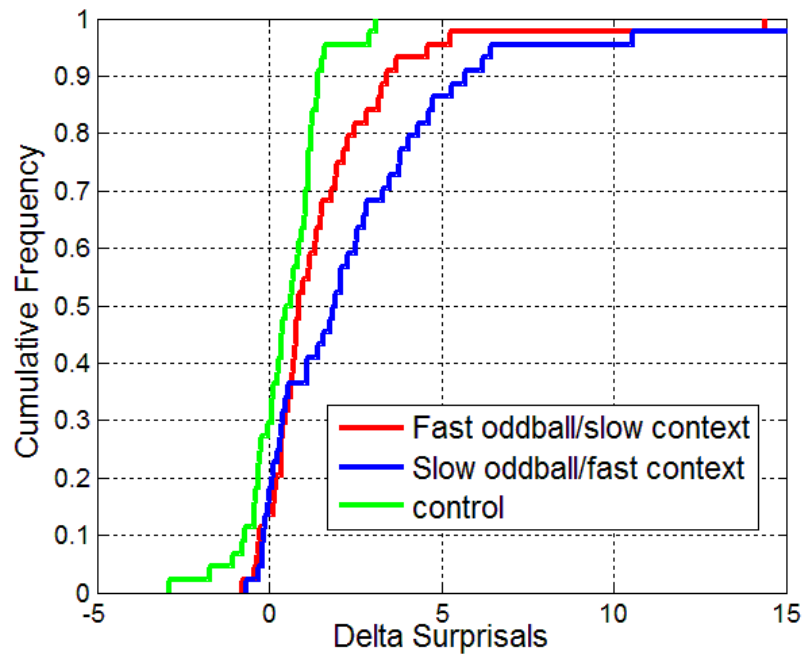


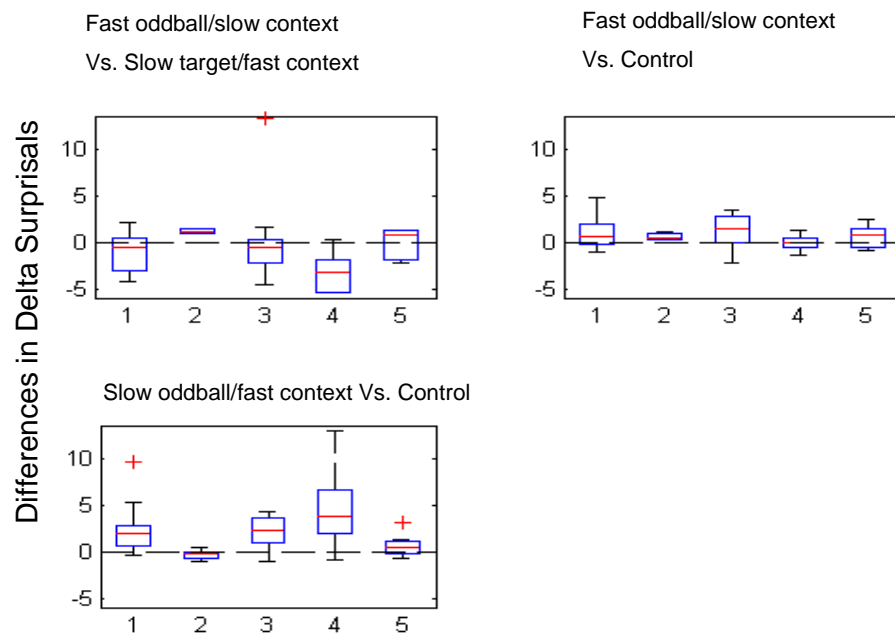
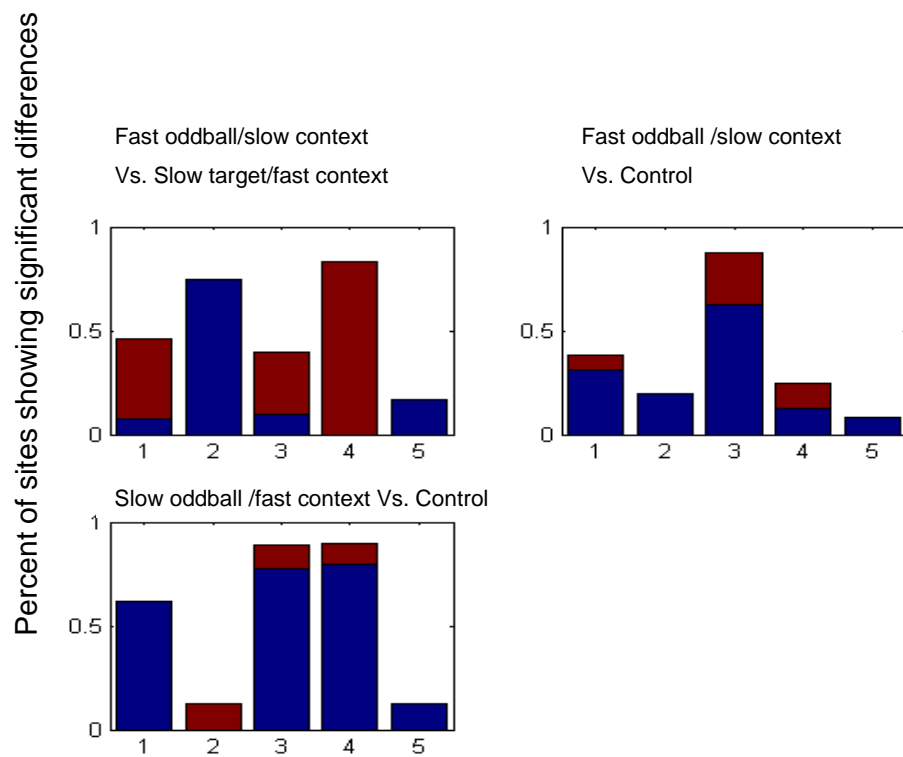
Figure 20**Figure 21**

Figure 22

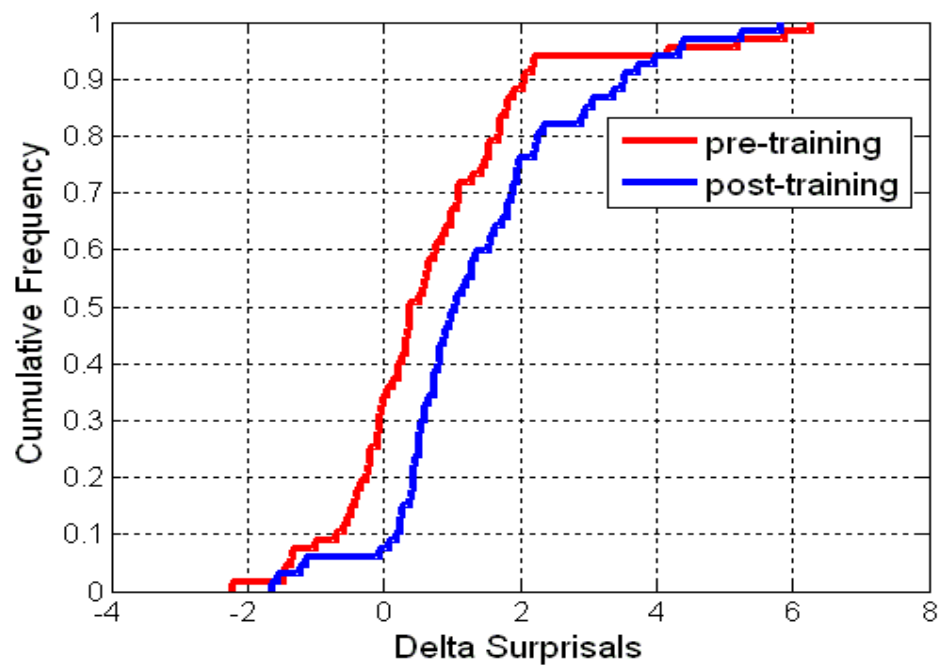


Figure 23

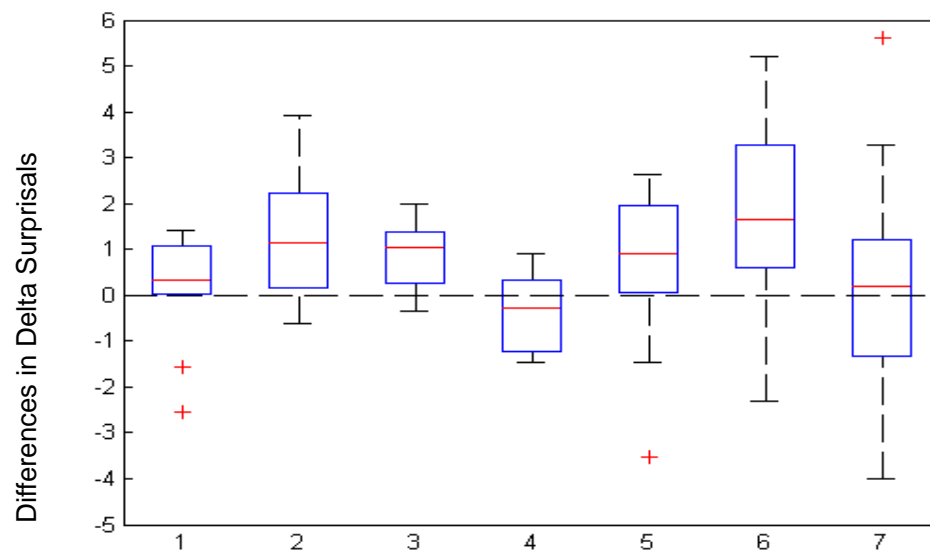


Figure 24

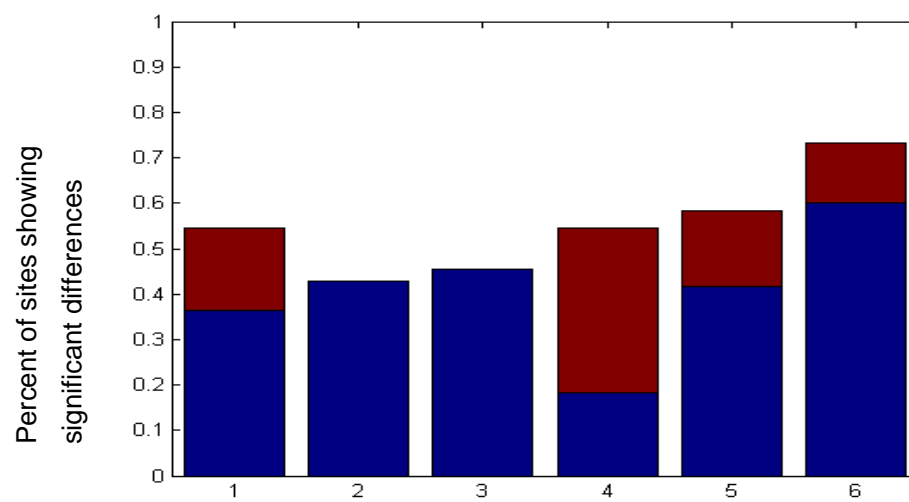


Figure 25

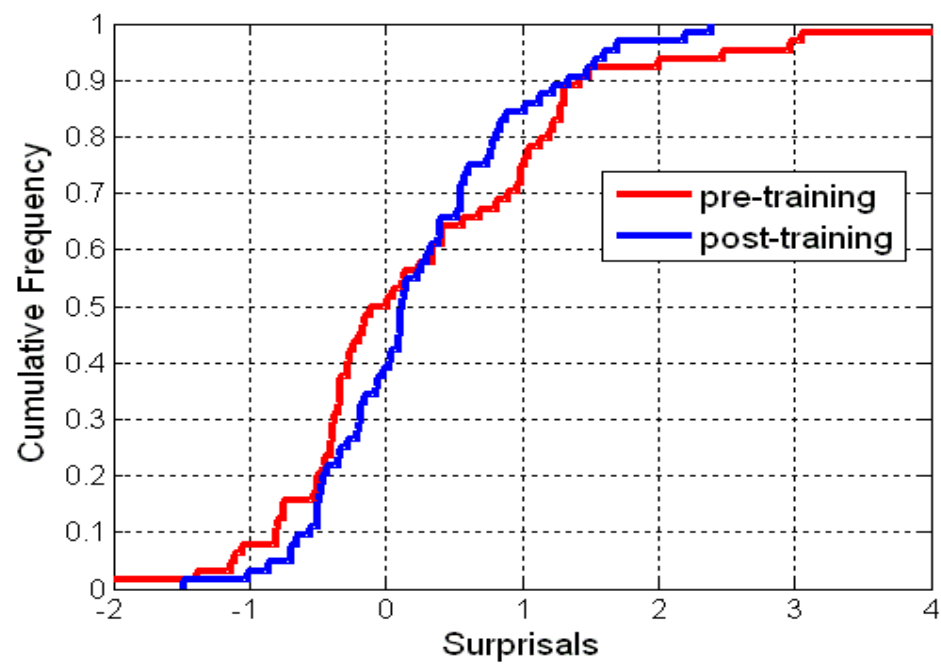


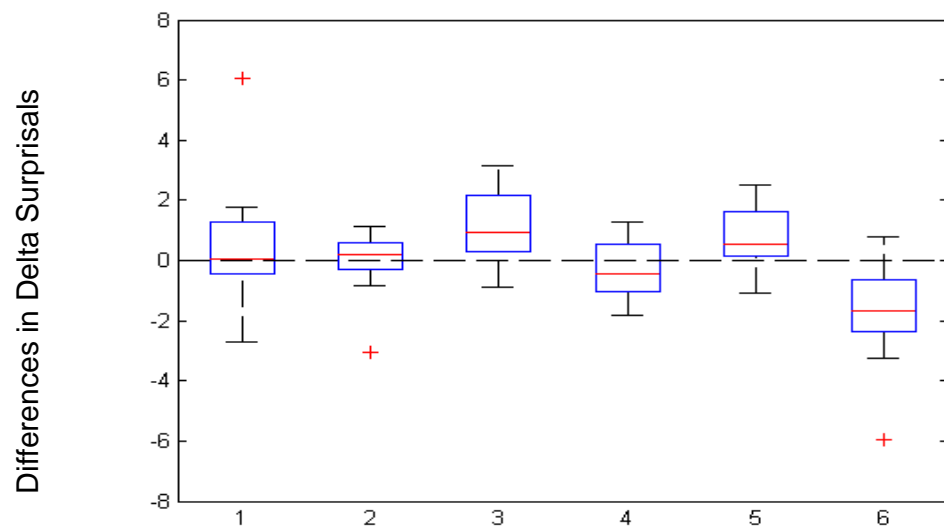
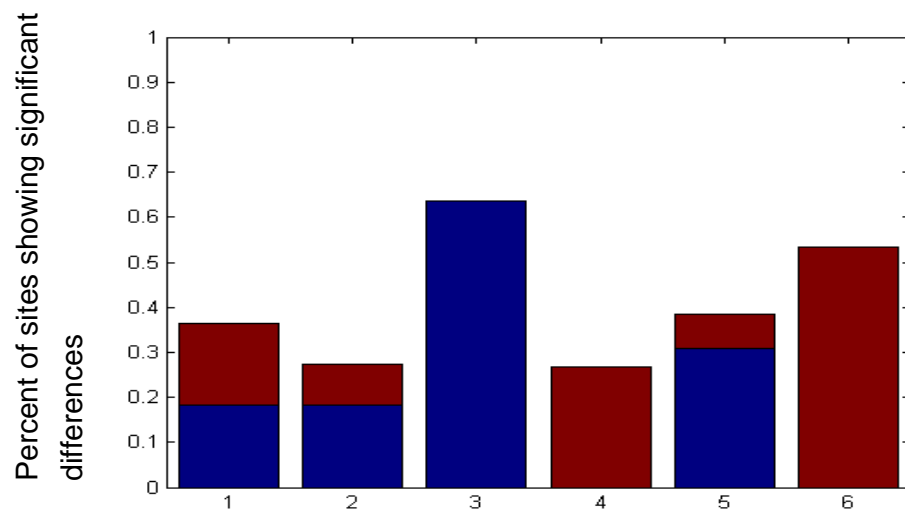
Figure 26**Figure 27**

Figure 28a

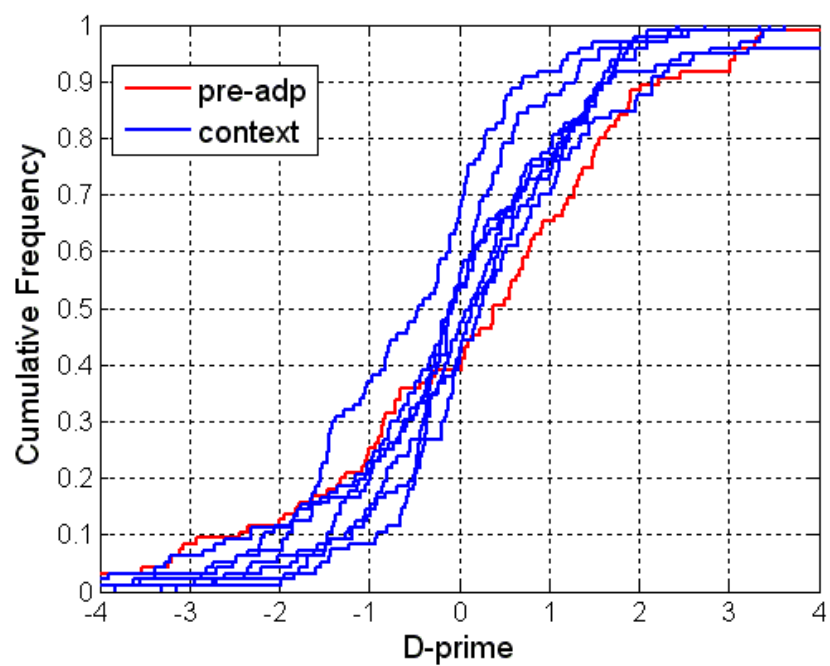


Figure 28b

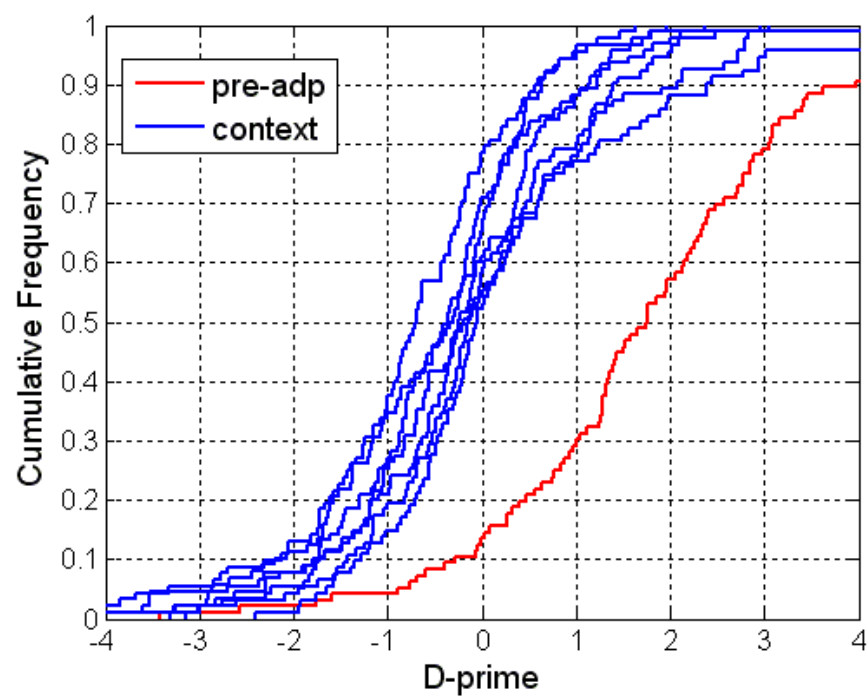


Figure 29

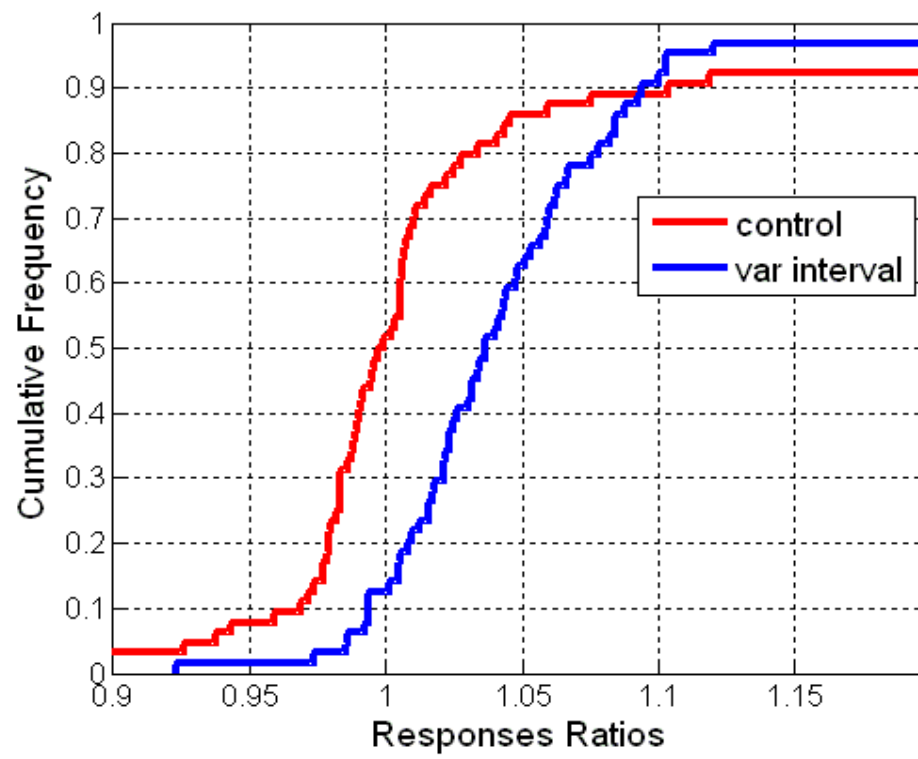


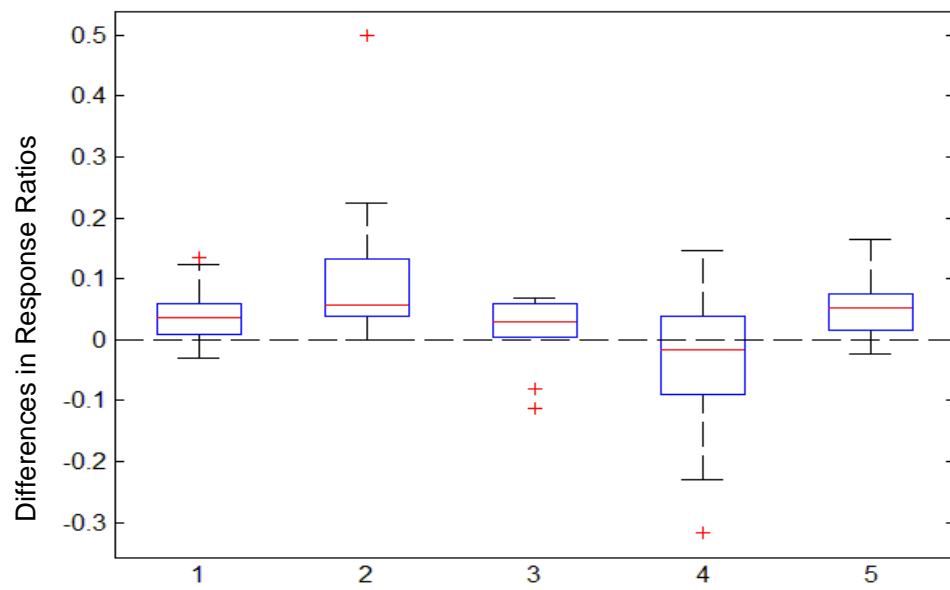
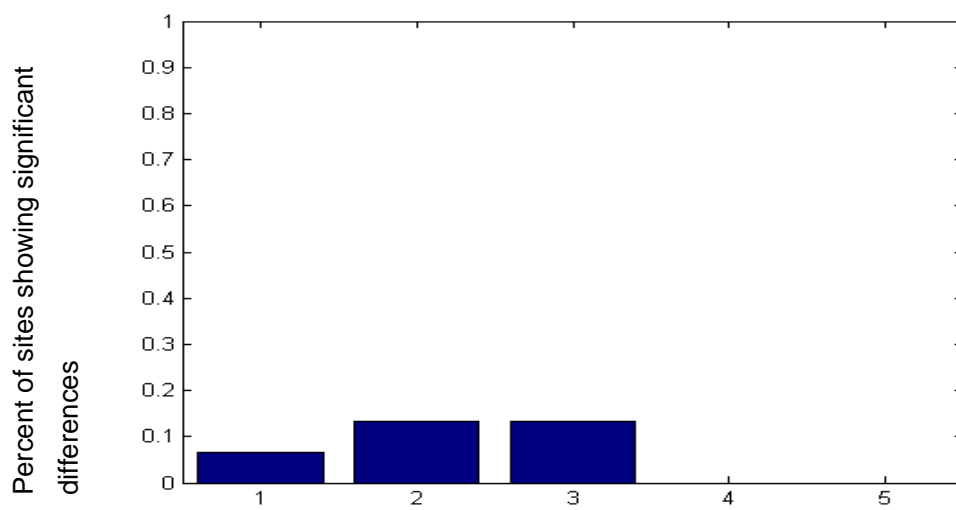
Figure 30**Figure 31**

Figure 32

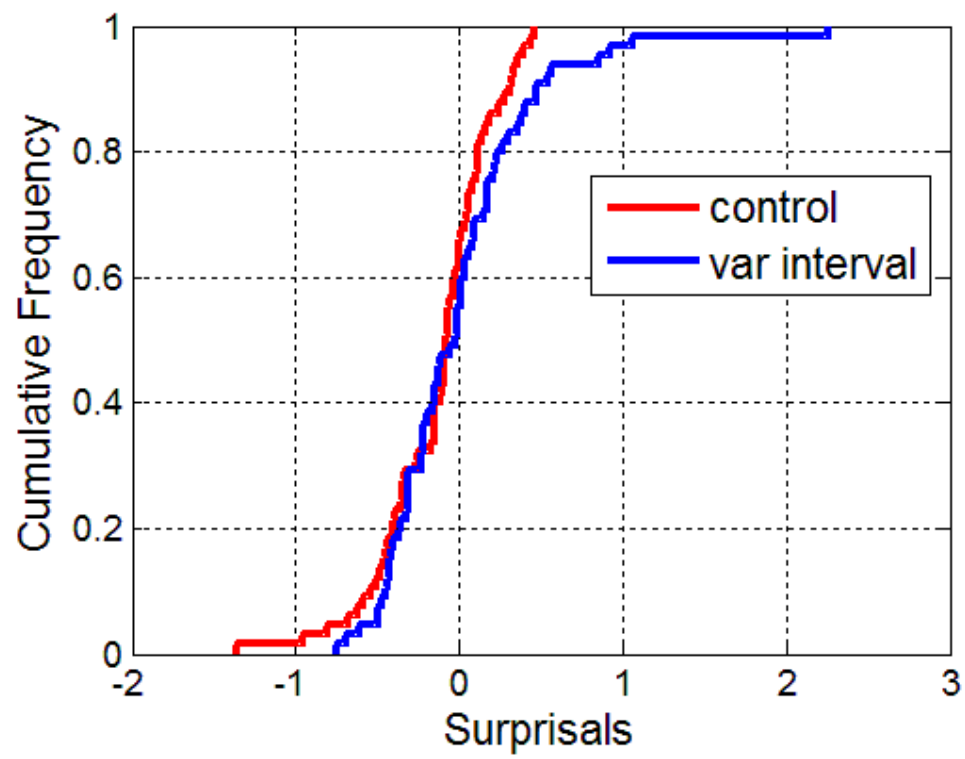


Figure 33

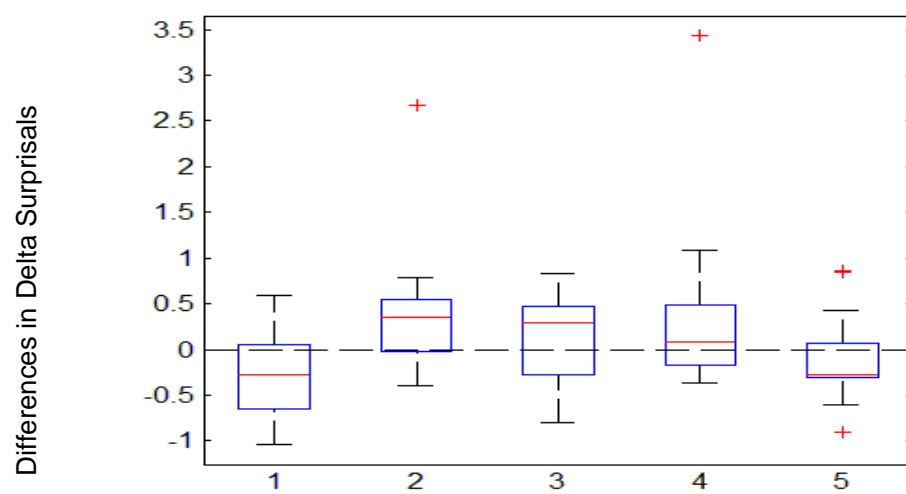


Figure 34

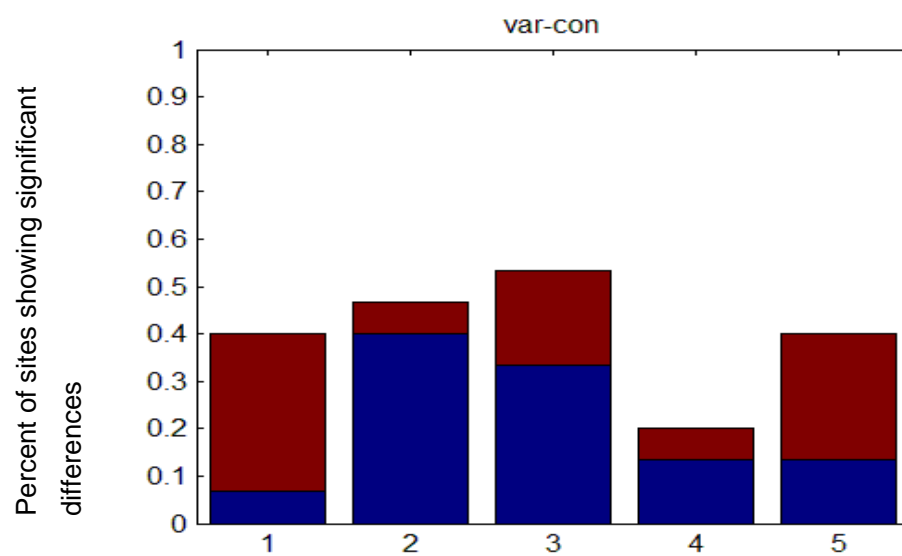


Figure 35

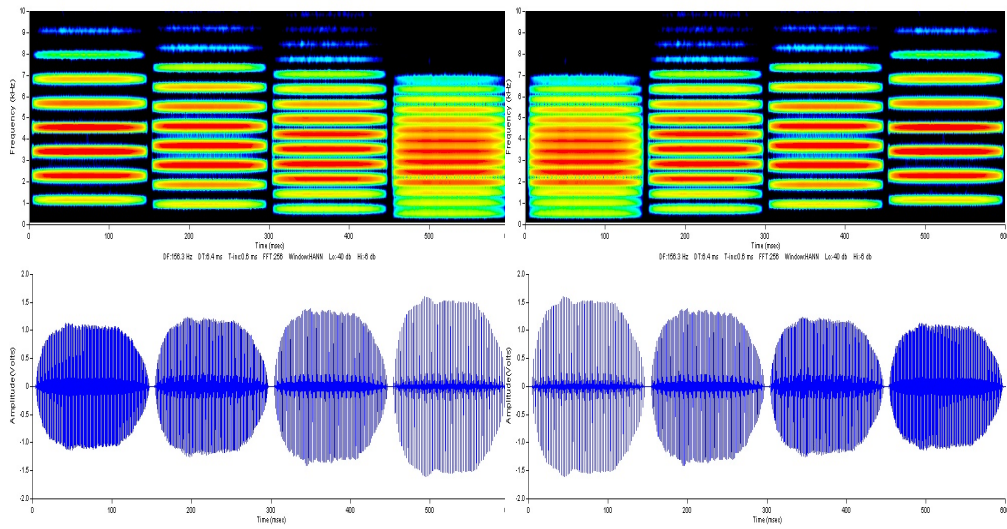


Figure 36

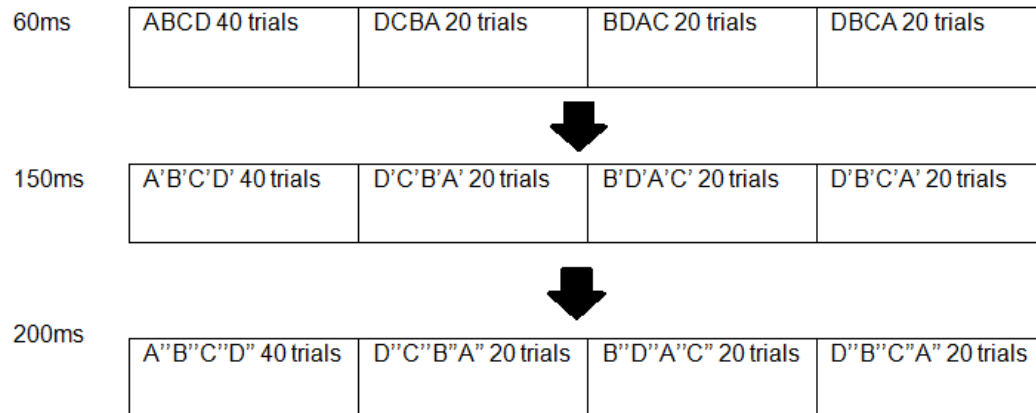


Figure 37

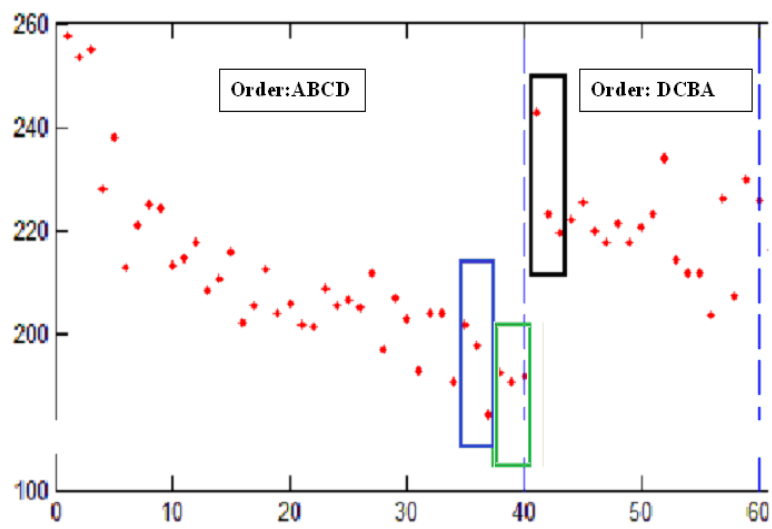


Figure 38

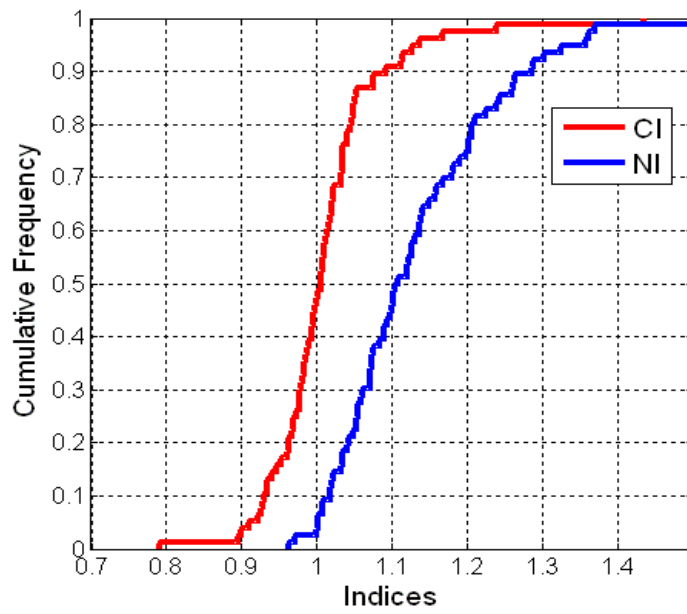


Figure 39

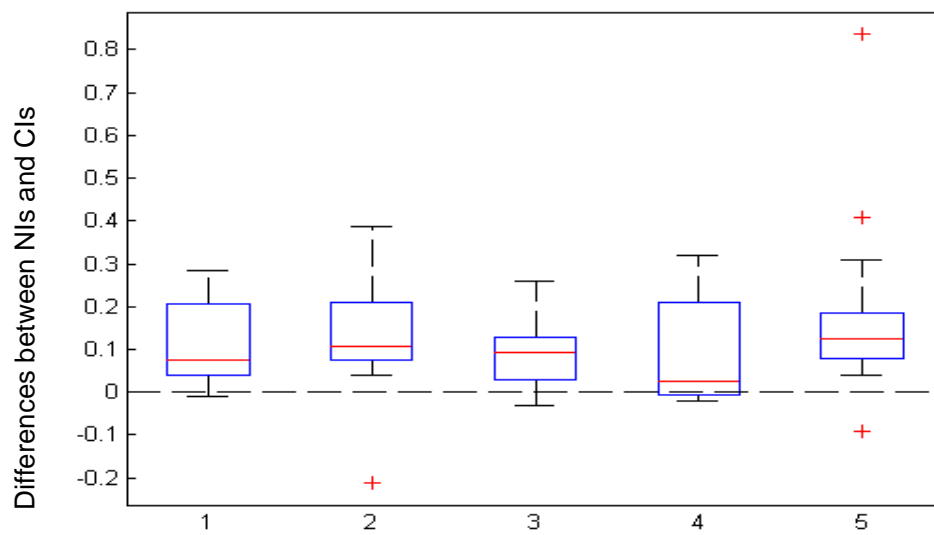


Figure 40

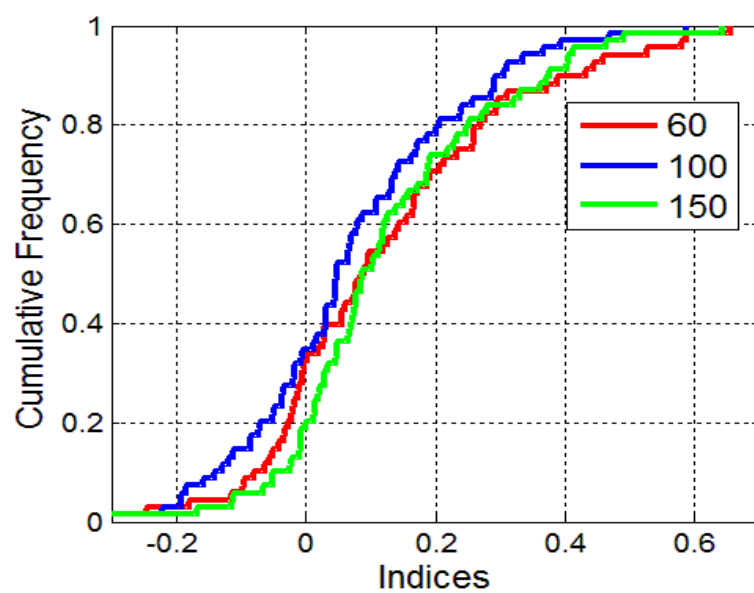
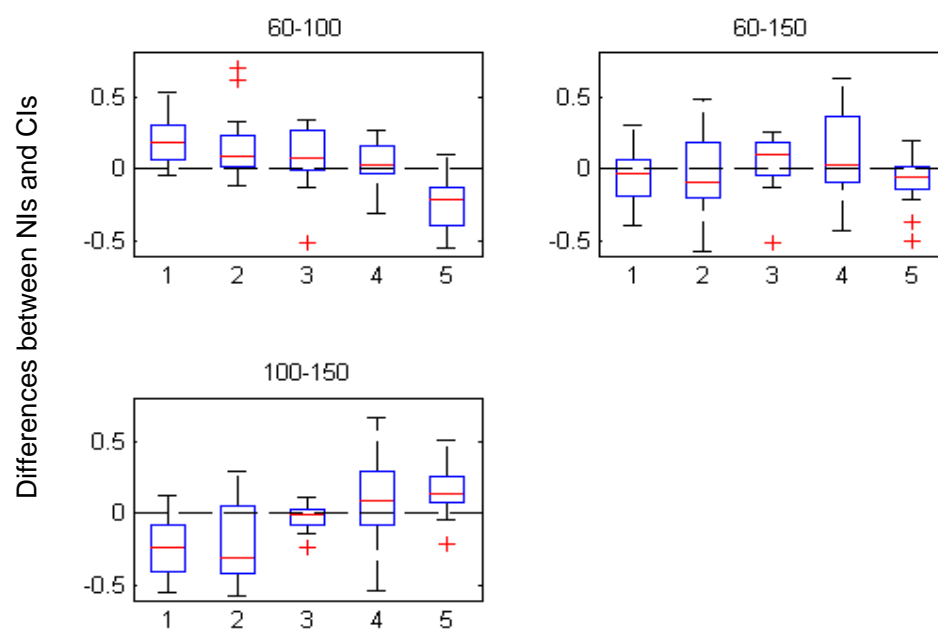


Figure 41



Appendix: In addition to thesis research, I have worked on these other projects:

Project1:

Recent Experience and Season Modulate Auditory Tuning in Canary Caudomedial Nidopallium

The tuning properties of NCM neurons differ between canaries and zebra finches. To test whether these differences reflect recent experience, three groups canaries of each sex were housed for 9 days in different conditions: conspecific aviary (normal condition), zebra finch aviary (cross condition) or in isolation box (isolation condition). Extracellular multi-unit electrophysiological responses to simple pure tone stimuli (250-5000Hz) in NCM of awake birds were recorded. Both phasic and tonic responses were measured, and auditory tuning width was quantified as the number of contiguous frequencies at least 1 SD above baseline. Compared to normal birds, tuning was narrower in cross-housed birds and wider in isolated birds. This effect was more pronounced in female canaries than in males. Tuning width was narrower early in the year and broader later in the year and this effect was most pronounced in cross-housed males. These findings demonstrate that tuning properties of NCM neurons are not fixed, but change in response to recent experience. The acoustic and social environment and season-related hormonal change may have important effects on auditory tuning and thus affect the perception of complex songs.

This work was published as:

Terleph TA, Lu K, Vicario DS. (2008) Response properties of the auditory telencephalon in songbirds change with recent experience and season. PLoS One. 6:3:e2854.

Project 2:

Glycine transmission in the avian auditory telencephalon

Glycine is predominantly expressed in the spinal cord and brain stem and thought to be scarce in the cortex of adult mammals. But earlier work by our collaborators, Mello and Lovell, found that wide GlyR β expression in the songbird telencephalon, including NCM and motor nuclei. In contrast, GlyR $\alpha 2$ expression was found to be much more limited overall, with very low expression in vocal nuclei, but significant expression in auditory forebrain, including NCM. These data suggest that a functional glycine receptor – assembled as a heteromer of α and β subunits – may be present in NCM. Therefore, we tested the effect of glycinergic agents on multi-unit activity in NCM of awake, restrained zebra finches (N=3). Unilateral microinjection of strychnine, a glycine receptor antagonist, increased the frequency of spontaneous bursting activity and the size of phasic auditory responses to tone and song stimuli on the injected side. In contrast, microinjection of glycine depressed spontaneous activity and reduced auditory responses. These effects are consistent with the presence of inhibitory neurotransmission that depends on functional glycine receptors, and suggests that inhibitory processes involving glycine, in addition to GABA, may contribute to auditory processing in NCM.

This work was presented as a poster:

Lu, K., Velho, T.A.F., Mello, C.V. and Vicario, D.S. (2008) Adrenergic modulation may contribute to long-term memory in songbird auditory telencephalon. Soc. Neurosci. Abstr.

Project 3:

Noradrenergic dependence of long-term adaptation induced by learned vocalizations

The ability to form long-term memories of sensory events is a fundamental property of the brain. Noradrenaline (NE) is thought to play important roles in the acquisition and retrieval of such memories, but the precise mechanism remains unclear. Early works by our collaborators, Velho and Mello, showed that noradrenergic transmission is required for the expression of the activity-dependent gene, ZENK. Since ZENK has been considered to be important for long-term memory formation, we hypothesized that NE exerts its long-term effects by coupling the electrophysiological responses and gene transcription responses that follow sensory stimulation. Therefore, we test the effect of local injection of α -adrenergic blocker on the long-term maintenance of adaptation of responses to song in NCM. We found that local α -adrenergic transmission is required for long-term, but not short-term, adaptation. This result supports our hypothesis that the noradrenergic system modulates long-term changes in song-responsive neurons by coupling the electrophysiological and gene expression responses to song. We suggest that this mechanism may be a major determinant of long-term sensory memories in the vertebrate brain.

This work was accepted as a publication:

Velho, T.A.F., Lu, K., Ribeiro, S., Pinaud, R., Vicario, D.S., and Mello, C.V. Noradrenergic modulation of song-induced gene expression and long-term adaptation (Accepted).