ABSTRACT

Title of thesis: STIMULUS TEMPORAL COHERENCE STRONGLY INFLUENCES RAPID PLASTICITY IN PRIMARY AUDITORY CORTEX UNDER GLOBAL ATTENTION

Yanbo Xu, Master of Science, 2013

Thesis directed by: Professor Shihab A. Shamma
Department of Electrical and Computer Engineering

Temporal coherence of stimulus features is a key property of sounds that emanate from single source. Consequently, it is important to understand how it may influence the direction and extent of the rapid plasticity postulated to occur during the streaming of concurrent sounds. We postulated that when animals listen attentively to coherent or incoherent stimuli, responses would adapt to effectively encode the correlational structure of the stimuli. In this study, ferrets were trained to attend globally to two-tone sequences which were played either simultaneously (SYNC) or alternatively (ALT) on a trial-by-trial basis, and to detect a transition to a random cloud of tones by licking a waterspout for reward. Neuronal activities were collected in the primary auditory cortex during performing the task and passively listening to the same stimuli sequences. Compared with the passive condition, neuronal responses changed distinctively between SYNC and ALT trials under the effect of attention. These results provide neuronal evidence for the role of stimulus temporal coherence in modulating responses during attentive listening to complex

sounds.

STIMULUS TEMPORAL COHERENCE STRONGLY INFLUENCES RAPID PLASTICITY IN PRIMARY AUDITORY CORTEX UNDER GLOBAL ATTENTION

by

Yanbo Xu

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2013

Advisory Committee:
Professor Shihab A. Shamma, Chair/Advisor
Professor Carol Espy-Wilson
Associate Professor Jonathan Z. Simon

Table of Contents

List of Figures

Chapter 1: Introduction

Our daily environment is never noise free, and the useful signal conveying information is always embedded in the background with all kinds of interference. Fortunately, human beings are adept at following the specific sound they are interested in. One representative example is the cock-tail party problem [1], in which the voice from a certain speaker one is paying attention to is interfered with voices from other guests, music, and other kinds of ambient noise. Though it seems effortless for one to conduct informative communications in such harsh environment, it is still quite challenging and beyond the ability of the best algorithms that have been proposed to tackle this problem. In the engineering community, computational auditory scene analysis (CASA) [2] has been a hot topic for several decades, aiming at better modeling the human auditory system in order to push the performance of computational models to match that of human beings. However, the performance gap between a computational model and a human being results from our limited knowledge about how our brain functions, and it might be necessary for us to have a full understanding of this complicated computational unit before we can reach our goal. The research in exploring the underlying neural mechanism of auditory scene analysis has drawn great effort around the world, ranging from single-unit electrophysiological recordings in animals [3] [4] [5], to human experiments using electroencephalography (EEG) [6] [7], magnetoencephalography (MEG) [8], functional magnetic resonance imaging (fMRI) [9] and other non-invasive techniques. Recent studies [10] [11] have demonstrated that responses in human auditory cortex encode critical features of attended speech in listening tasks with two simultaneous speakers.

Auditory streaming [12], a fundamental aspect of auditory scene analysis relates to the perceptual organization of sounds into "stream", has been a fundamental research direction [13] [14] [15]. A "stream" is a group of components that the listener perceives as a coherent entity in the auditory stimulus, and it could be just one sound or sounds from different sources. For example, in chorus, the voices from different singles including male and females are usually perceived as a stream since these voices are well-synchronized and temporal coherent. A classic psychoacoustic experiment on auditory streaming is composed of repeating two tones at different frequencies, A and B [1] [16]. By varying the frequency separation ΔF between A and B, two different percepts could be evoked. When ΔF is small, subjects tend to hear one stream; when ΔF is large, two separate streams are perceived, each containing the tones of the same frequency. A natural hypothesis from this simple experiment is that frequency separation is the main factor for auditory streaming, and it is consistent with the tonotopic organization of the auditory system of different kinds of species. Parallel sequences of repeating two tones with large frequency separation would excite two isolated population of neurons in auditory cortex, which could be correlated with streaming separation, and this "population separation" hypothesis [17] [18] was widely adopted in previous model for sound segregation.

Notwithstanding the simplicity in explanation, the "population separation" hypothesis failed to take into account the relative timing of sounds. If A tone and B tone are played synchronously no matter how large the frequency separation is, they form a coherent sound object, i.e. one stream even though they evoke responses the best in two different neuron population. Thus a model for streaming separation should not only consider the frequency separation, but also the temporal coherence. Recently, a model is proposed [19] [20], arguing the necessity of both temporal coherence and attention in the formation of auditory streaming. The natural stimuli consist of different kinds of features, pitch, timbre, location, and loudness and so on, much more complex than the two-tone sequence. When attention is paid for one specific feature, other features that are temporal coherent with the chosen one automatically bind together to form one coherent sound object. An example is the speech separation of co-channel speech mixture of one female and one male. A good candidate of feature to pay attention to is naturally the pitch. This coherence model successfully generalizes the hypotheses based on simple two-tone stimuli, and is suited for analyzing real world stimuli.

However, the neural mechanism of auditory streaming still remains unclear, and there is no direct neural evidence to support the coherence model. In this study, we train ferrets in a behavior task, in which they are required to pay attention to stimuli globally. The stimuli contain two parts, reference and target, and the ferrets learn to detect the change from reference to target. In the design of reference, we use the classic repeating two-tone sequences in which A and B are played either alternatively or synchronously, corresponding to the anti-coherent and coherent stimuli respectively. Our hypothesis is that under the effect of attention, the neural responses could rapidly adapt to encode the temporal correlation structure of the stimuli. Compared with neural responses in passive listening (no attention) with the same stimuli played, we observe distinct changes for anti-coherent and coherent stimuli in behavior, which might result from rapid reformation of neural connectivity under the effect of attention, driven by the intrinsic correlational structure in stimuli.

The organization of this thesis is as follows. In Chap. 2, we focus on the stimulus design and methods of data analysis. The results are presented in Chap. 3. In Chap. 4, several key issues are discussed. And we conclude this thesis with Chap. 5.

Chapter 2: Methods

2.1 Subject

A female ferret (*Mustela Putorius*) purchased from Marshall Farm was used in the experiment. The ferret only had free access to dry food in the weekdays, and access to water was restrained as reward in the behavior task. In the weekends, free access was provided to the ferret, and wet food was also given to help it regain weight. The weight of the ferret was carefully monitored, and maintained above 75% of its *ad libitum* weight. The care and use of animals in this study was consistent with NIH Guidelines. All procedures for behavioral testing of ferrets were approved by the institutional animal care and use committee (IACUC) of the University of Maryland, College Park.

2.2 Experimental Design

In order to explore the effect of global attention on the neural responses, the ferret was trained to engage in a behavior task. The task is cast into a reference-target paradigm [21], during each trial of which the ferret is required to pay attention to the audio stimuli, and response to the change from reference sound to target sound by licking a spout. A reward window is placed after the onset of target sound, and the ferret can the reward of water if it correctly detects the change and licks the spout during this window. However, if she licks before the onset of target sound, i.e. during the reference sound, the trial is aborted after the end of reference sound, and there is a delay for the beginning of the next trial as a punishment.

To relate the task with streaming perception, the reference sound is designed to consist of two repeating tone sequences, denoted as A tone sequence and B tone sequence. These two sequences are played either synchronously or alternatively, denoted as pattern SYNC and pattern ALT respectively. The pattern ALT is exactly the classic "ABAB" tone sequence frequently used to measure the threshold of streaming perception of human subjects in psychoacoustic experiments [1]. While the frequency of B tone is fixed, the frequency of A tone is varied in the tonotopic axis to create different frequency gaps between them. As confirmed by the psychoacoustic experiments, human subjects tend to perceive one stream when the gap is small and two separate steams when the gap is large. In [22], neural data from A1 was collected when the ferrets passively listened to the two tone sequences, and it is shown that it is not enough to differentiate between ALT and SYNC patterns by just considering the frequency separation. A computational model is proposed to take the temporal coherence into consideration to predict the stream perception. The reference part of the stimulus design in this study is almost the same with the second physiological experiment in [22]. ALT and SYNC trials are representatives of anti-correlated and correlated stimuli. To conduct the similar analysis of the effect of frequency gap of A tone and B tone on the neural responses, in this experiment, A tone is placed at distances of -2, -1, -0.5, -0.25, 0, 0.25, 0.5, 1, 2 octaves away to the B tone in the tonotopic axis. When the sequence of A is equal to B, in ALT, they form an iso-frequency sequence.

As an extension to [22], which only analyzed data under passive listening (without the effect of attention), ALT and SYNC patterns with various frequency separations are presented interleaved, followed by a target sound, requiring the ferret to pay attention to the stimulus in order to get water reward. The design of target sound is a random cloud of tones, making the transition from the reference to target easy to detect. The change from regularly repeating tones to random tones is so salient in the global statistics that the ferret is not required to pay attention selectively to one sequence to do the task. The frequencies of tones in the target sound are uniformly sampled to cover a range of 5 octaves, centered at the B tone.

The schematic diagram of the stimuli of the task is shown in Figure 2.1. In each tone sequence, the duration of a tone is 75ms, and the gap between two tones is 125ms, which is consistent with the setup in [22]. In ALT, the offset of the beginning of B tone sequence is 100ms. There are 64 tones in the target, and the duration of each tone is 25ms. Their occurrence in the time axis is random but balanced, and there are 16 tones per 200ms. 5ms cosine ramps are applied the start and end of each tone played in the reference and target. For each trial, there is 100ms pre-trial silence. The length of reference sound is varied trial by trial to prevent the ferret from using timing strategy. There are 3 kinds of reference lengths, 1.6, 3.2, and 4.8 seconds, corresponding to 8, 16, and 24 tones in each tone sequence. And the length of target is always 1.6 seconds.

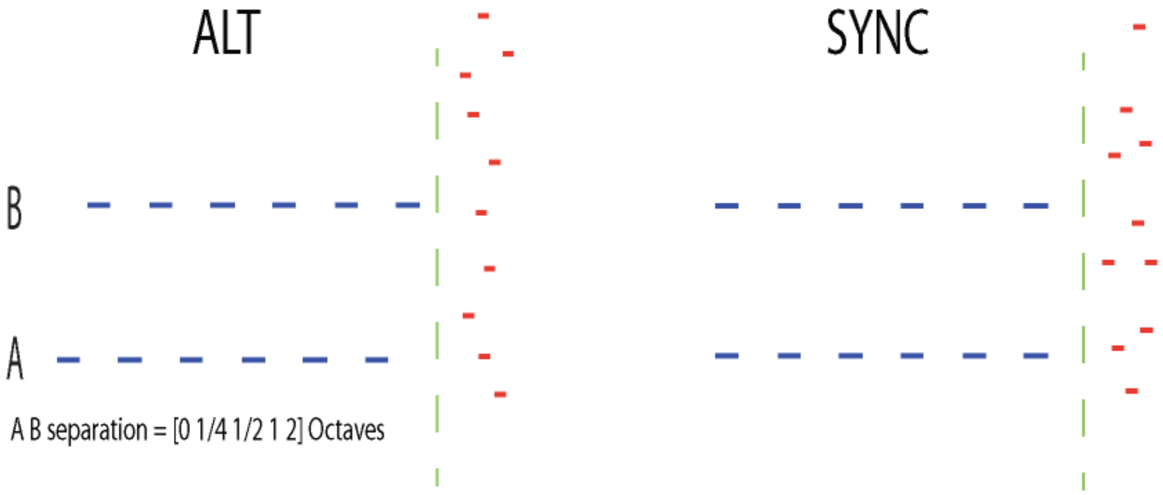

Figure 2.1: The schematic diagram of the stimulus design. The ALT and SYNC patterns are plotted on the left and right respectively before the green dash lines. There are 3 reference lengths, and here we only highlight the temporal organization of the two sequences in ALT and SYNC. An example of an exact spectrogram of target is shown in Figure 2.2.

2.3 Training Procedure

It took 3 months for the ferret to learn this task. In the beginning, the reference was played at 60dB below (hardly audible) the target, letting the ferret to form the association between the target and water reward. The intensity of reference was gradually increased to the same sound level as target to remove the intensity cue. In the first part of the training process, B tone was always fixed at 3kHz. Since in physiology, the frequency of B tone should be chosen according the best frequency of the neuron we find, in the second training stage, B tone was varied on a daily basis to help the ferret generalize to wider frequency range.

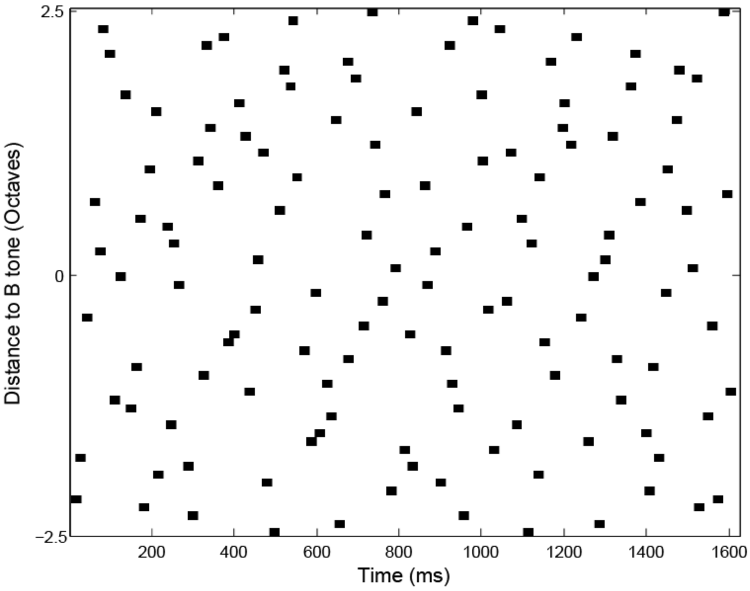

Figure 2.2: The spectrogram of a sample target.

2.4 Neurophysiological Recordings

After the ferret successfully learned the task, to ensure the stability of electrophysiological recordings, a titanium headpost was implanted above the ferret's skull through a surgery [21]. 2 to 4 independently moveable tungsten electrodes (FHC) were pushed to penetrate into the primary auditory cortex though small craniotomies made prior to the recording sessions which usually lasted 6 to 8 hours. Stimuli were played by a calibrated loud speaker which controlled by custom made Matlab program in a double-walled acoustic chamber (IAC). Electrophysiological data was recorded by an AlphaOmega system.

2.5 Data Analysis

The best frequency of a neuron was estimated by a sequence of tones at different frequencies in an online fashion. The B tone was always placed at the best frequency (BF) of the selected neuron. Since there was a slight mismatch between the true BF after spike sorting, in the final analysis, units whose BFs are within half octaves away from the B tone are selected to the following analysis. The data analysis is conducted on the two parts of the stimuli, responses to the reference and the target separately. There are in total $2 \times 9 = 18$ conditions (ALT and SYNC, and 9 frequencies of A tone), and each conditioned was repeated 9 to 12 times in which 3 kinds of trial lengths were uniformly sampled.

2.5.1 Analysis on Spikes from Reference

Auditory streaming is never static, and the dynamic evolvement along the time axis is essential for the formation of steaming perception. On the contrary, if only two pure tones are played either simultaneously or alternatively, there is no stream but just sound tokens. Thus it is necessary that in the stimulus design, two pure tones are repeated for enough times to allow the buildup of streams. In the first part of data analysis, we focus on how the neural responses change along the time axis. For ALT, there is no overlap between A sequence and B sequence, while for SYNC, they are totally overlapped. Hence in ALT trials, neural responses to A and B tone sequences are analyzed separately, which in SYNC trials, the neural responses to two sequences are analyzed as one object.

The typical response of neurons in A1 to pure tone stimuli is characterized by the onset and offset responses. Since each tone in reference has a duration of 75ms. The main contribution to responses at 10 - 40ms, and 80 - 100ms are neurons' intrinsic onset and offset responses. In order to focus on the interaction between neurons which is correlated with the temporal coherence between the two tone sequences, we specifically focus on the window around 40 - 70ms after the onset of each tone. The neural response in this window is more likely being affected by the interaction between neurons, which is consistent with the finding in [23].

Note that there are 9 different frequencies for A tone in each experiment session, and the range is from 0 to 2 octaves to B tone. In psychoacoustic experiments, the gap of 0.5 octaves between two tones is found to be a coarse threshold for streaming perception. Thus we generally divide these 9 pairs of tones into two groups, NEAR and FAR. The FAR group contains pairs with gaps larger than 0.5 octaves, i.e. 1 octave and 2 octaves, and the NEAR group contains the rest. Also note that there are 3 trials lengths for each pair, and the first 8 tones for each sequence are common for all these three lengths. So we first plot the average firing rate within the late window for each of the first 8 tones. Each unit is normalized by the maximum firing rate in the first 100ms of the iso-frequency pair of ALT pattern in the passive state. There is no particular physical meaning of this kind of normalization, and we just want to make the statistics for each unit comparable. To compare the firing rate between passive and active, we subtract the normalized firing rate curves for both NEAR and FAR groups.

After the rapid adaption from the onset of each trial, the average firing rate within the late window reaches a steady state response. Here in order to visualize the change in a finer scale, we average the responses in each 200ms period from the 5th note in each trial, and we call it the steady state Post-Stimulus Time Histogram (PSTH).

2.5.2 Analysis on Spikes from Target

To reconstruct the receptive fields (RFs) after the neuron adapts to the reference, we consider the correlation between the spike trains in target with the onset of each tone in target. Recall that there are 64 unique pure tones covering 5-octave range around the center at the B tone, one tone per semitone and representing 61 unique channels. In the reconstruction, each channel is treated independently, and in each channel, the cross correlation between onsets of tones and average firing rate is computed, normalized by the number of tones. This normalized cross correlation is very similar with spike triggered average (STA), but in an opposite way. Here, each event (the triggered spike in STA) is the onset of a pure tone in the channel we consider. The reconstructed RF describes the response pattern in two-dimensional space. Here, we show an example of the reconstructed RF of a single unit.

2.5.3 Same analysis on Local Field Potentials

We mainly use single unit data (SUA) to do the analysis in this study. Additionally, we also apply the same analysis on the local field potentials (LFPs) recorded simultaneously with spike data. While SUA represents the action potentials of indi-

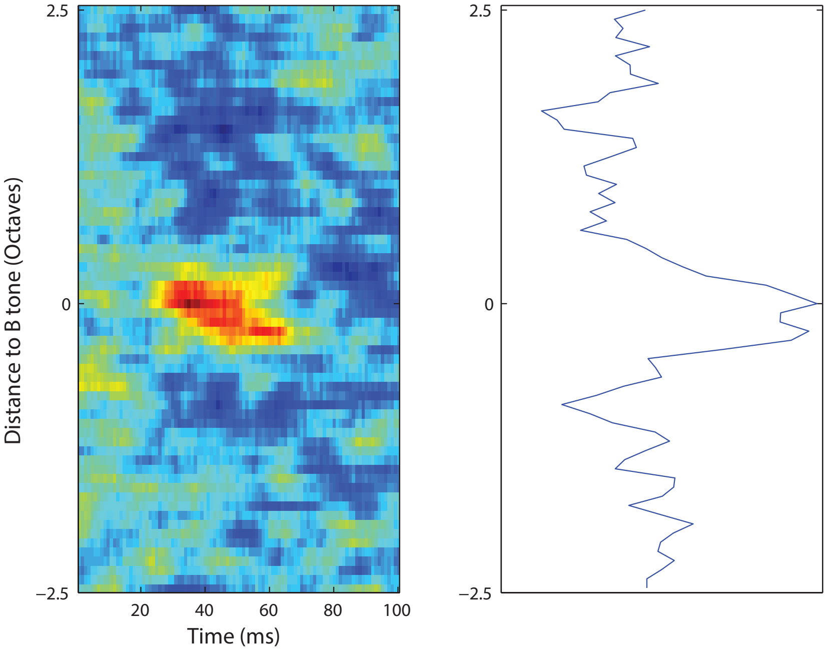

Figure 2.3: An example of the reconstructed RF of a single unit. Left: The reconstructed RF. Right: The averaged one-dimensional tuning curve by collapsing the RF along spectral axis.

vidual neurons, the components of LFPs are much more complicated. Simply put, they reflect the synchronized activities in related broaden volume around the tip of the electrode. The lower the frequency range, the wider the area the LFP covers but at the same time, the more complex the phenomenon is. In this study, we focus the frequency band from 30 to 200Hz, and divide it into 3 bands, 30-60Hz, 60-100Hz, and 100-200Hz, denoted as low gamma band, middle gamma band, and high gamma band. 4th order Butterworth IIR filters are designed, and MATLAB function *filtfilt* is adopted to obtain zero phase shift in each frequency band. Since LFPs are essentially bipolar (the spike count is always positive), instead of applying the analysis on the filtered signal, Hilbert Transform is applied to abstract the envelop of the oscillation which is correlated with the energy of overall neural activities.

Chapter 3: Results

In this chapter, the main results of analysis introduced in the Sec. 2.5. For SUA, there are $N_s = 86$ units. For LFPs, since each LFP corresponds to one electrode, the number of units is the same with multi-unit activity (MUA) $N_m = 47$.

3.1 Behavior Performance in the Recording Sessions

The behavior performance is measured by the discrimination rate, which is defined as the product of hit rate and false alarm rate. The distribution of discrimination rate across the days during which the physiology experiments were conducted is shown in Figure 3.1(a). Please note that for random guess, the discrimination rate is $0.5 \times 0.5 = 0.25$. As seen in Figure 2.2, the performance was way above chance level. Another way to visualize the performance is to summarize the time at which the ferret licked the spout the first time in each trial. The result is shown in Figure 3.1(b). We can see that there are 3 peaks, each corresponding to the target for each trail length respectively. The clear phase locking of the peaks to the onset of targets indicates that the ferret had a good understanding of the task, and that the performance was good.

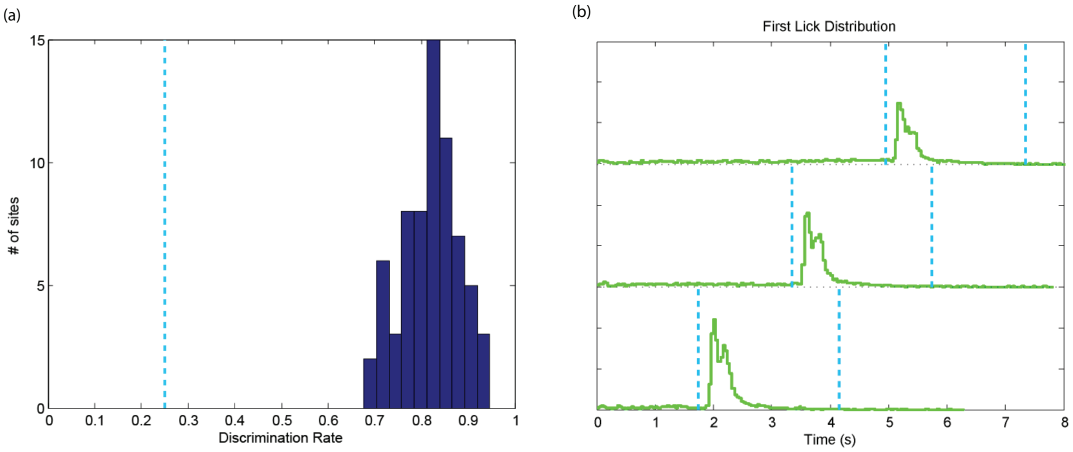

Figure 3.1: The behavior performance: (a) The distribution of Discrimination Rate of behavior performance. The dash blue line represents the performance at chance level. (b) The distribution of the time the first lick happened for three trial lengths. The two dash blue lines for each row mark the edges of reward window.

3.2 Data Analysis on electrophysiological data

3.2.1 Response Changes to Alternating and Synchronous Tone Sequences During Behavior

In the first row of Figure 3.2, the averaged firing rates (SUA) in the late window for each of the first 8 tones for both passive and behavior are shown, grouped by NEAR and FAR. The adaptation trends for both passive and active are very similar. The average firing rate goes down gradually as the repetition continues, and the adaptation quickly reaches a flat region after 3-4 notes, and we call this region the steady state response. The adaptation also agrees with the fact that it takes several repetitions for humans for form clear stream separation. For a better visualization, group NEAR and group FAR are not presented aligned, but shown paralleled.

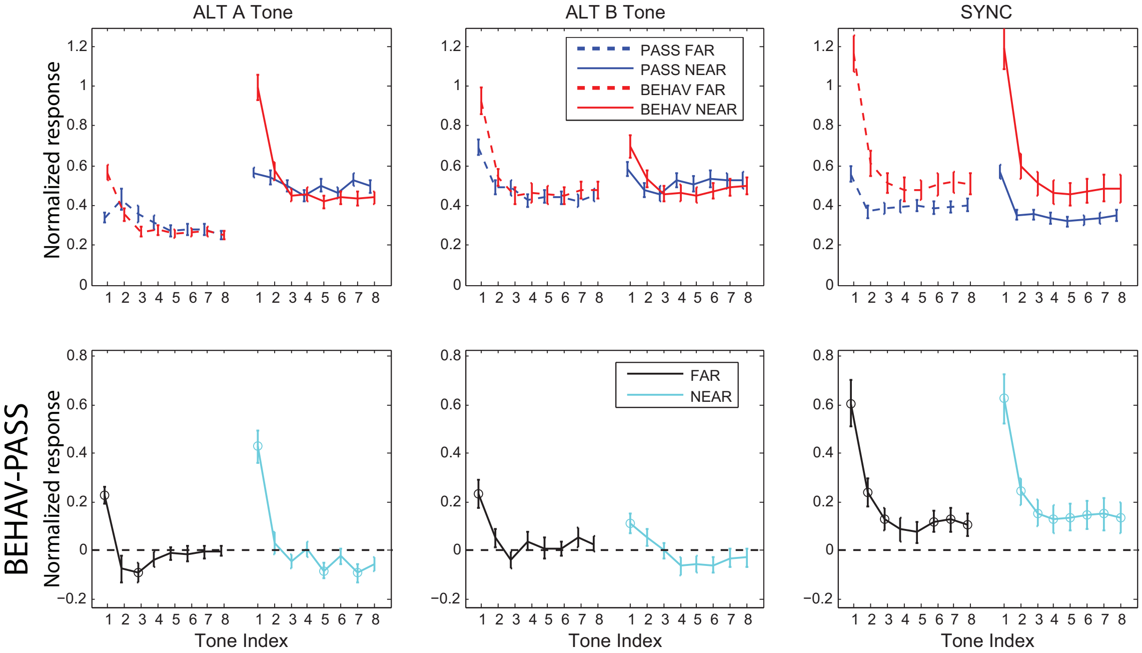

Figure 3.2: First row: The adaption of normalized average firing rates. Second row: The difference between passive and behavior. Colored circles indicate signicant (paired t-test, $p < 0.05$).

The second row of Figure 3.2, shows the difference between passive and behavior. There is a clear enhancement for both ALT and SYNC in the first several notes in behavior, and we postulate that this enhancement is mainly contributed by the effect of attention. What's interesting is that in the steady state response, There is a conspicuous difference in the changes for ALT and SYNC in NEAR. For SYNC, the responses are enhanced for both NEAR and FAR, and the changes in the last several tones, which represent the changes in the steady state response, are significant. While for ALT, the responses to both A tone and B tone are suppressed or not changed. This clear difference between ALT and SYNC supports the assumption that the late window encodes the effect of neuron interaction modulated by the correlational structure of components in the stimuli.

We also do the same comparison with MUA and LFPs in the three bands. The changes between passive and active in different scenarios are listed in Figure 3.3. We see that the changes in MUA are very similar with the ones of SUA. The high gamma band is also very similar with SUA, though the responses in behavior for both ALT and SYNC are enhanced, the enhancement for SYNC is much stronger that ALT. The middle band gamma begins to show some difference, and in low gamma band, the patterns of changes are quite different from SUA. So the high gamma band can provide similar information as MUA since it is nearer to the frequency band the spikes are identified. This provides a valuable alternative for neural response analysis when the quality of spikes is not good.

The analysis of PSTH of SUA is shown in Figure 3.4. The shapes of PSTHs are consistent to the grouping criteria. For ALT, in NEAR there are two salient components corresponding to the A tone and B tone in one period, while in FAR, the component for A tone almost disappears since A tone is far away from the BF of the unit, and sometimes it is out of the receptive field. For SYNC, in both NEAR and FAR, we can see it strong component which corresponds to the overlapped A tone and B tone in each period. The bars at each sample index are the standard errors.

The difference between passive and behavior is presented in the last column of Figure 3.4. The late window is shown as the grey green region. What's clear in the

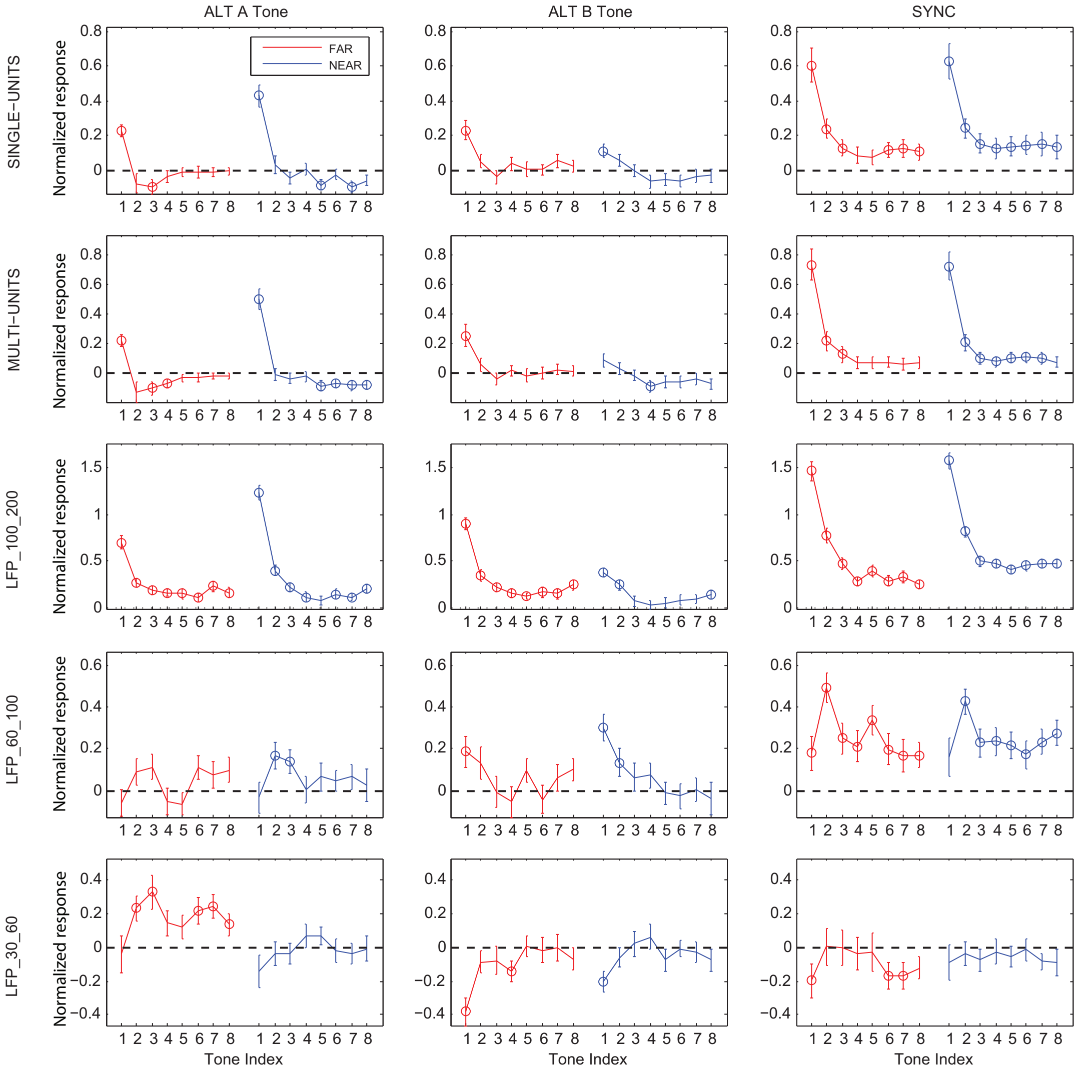

Figure 3.3: Response changes in different bands of data. First row: SUA. Second row: MUA. Third row: high gamma band. Fourth row: middle gamma band. Last row: low gamma band. Colored circles indicate signicant (paired t-test, $p < 0.05$).

behavior induced change is that in the shaded region, there is clear enhancement for SYNC and the enhancement is much stronger in NEAR, while for ALT, we can see clearly the suppression in NEAR, and the suppression region extends beyond the

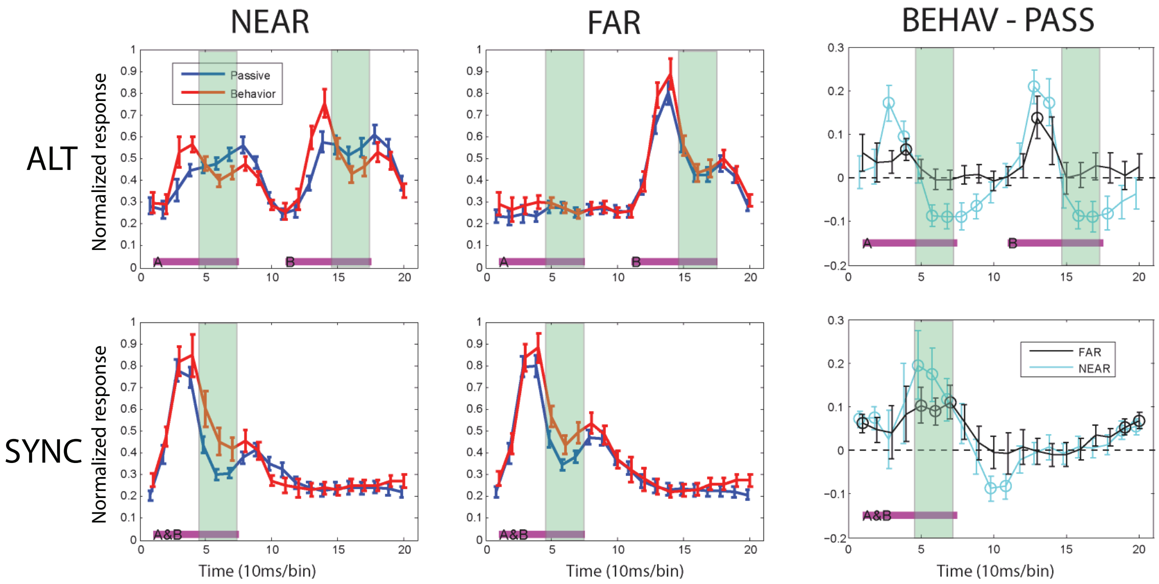

Figure 3.4: PSTH of SUA. First row: ALT. Second row: SYNC. First column: NEAR. Second column: FAR. Third column: behavior-passive. Grey green shade: the late window.

shaded region. Additionally, there is also clear enhancement at the onset of each tone for ALT, and suppression at the offset for SYNC. The neural mechanism and explanation are beyond the scope of this report, and we would further explore this in future research.

Besides SUA, we also analyze the LFP in high gamma band. The shade around the mean is the standard error. We can see similar trend as SUA. Even though the baseline change has been corrected, there is still overall enhancement after the onset of trials. The enhancement for SYNC is much stronger than ALT.

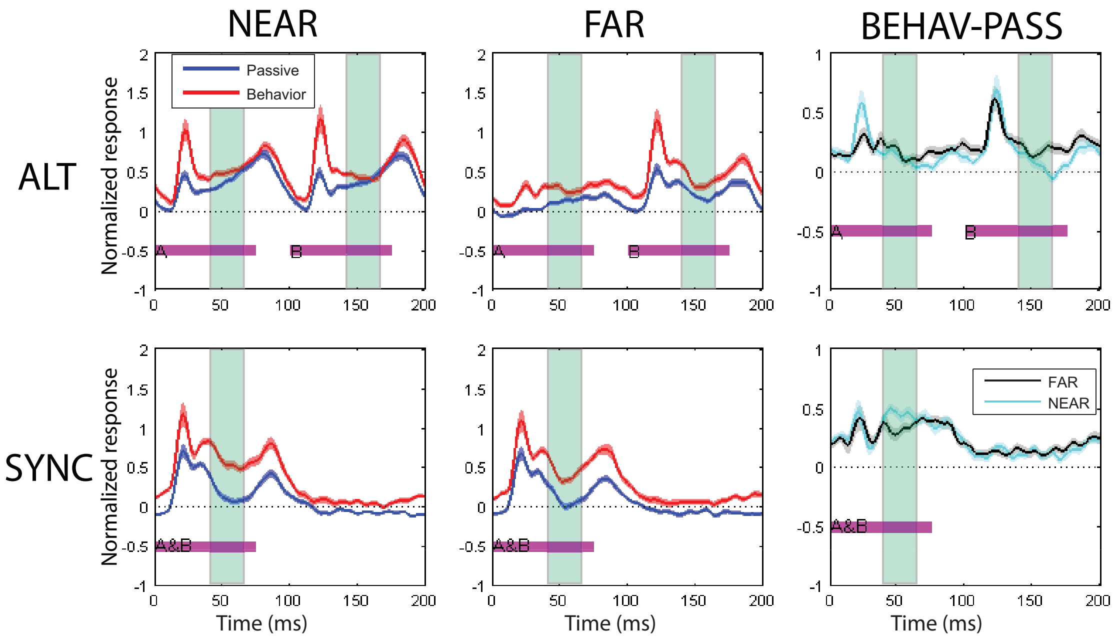

Figure 3.5: PSTH of high gamma band of LFP. First row: ALT. Second row: SYNC. First column: NEAR. Second column: FAR. Third column: behavior-passive. Grey green shade: the late window. Colored circles indicate signicant (paired t-test, $p < 0.05$).

3.2.2 Changes in the Reconstructed Receptive Fields

Firstly, we describe the results of SUA. The simplest way to group the trials is still by ALT and SYNC. Since the units pulled together have BFs near to the B tone, their BFs are almost aligned. In Figure 3.6, we can see that the responses are the strongest around the B tone, and responses are weak in the channels that are far away from B tone, which is consistent with one-dimensional tuning curve of neurons in A1. In time axis, the STRF has a peak around 25ms, which is caused by the physical delay in response to pure tone.

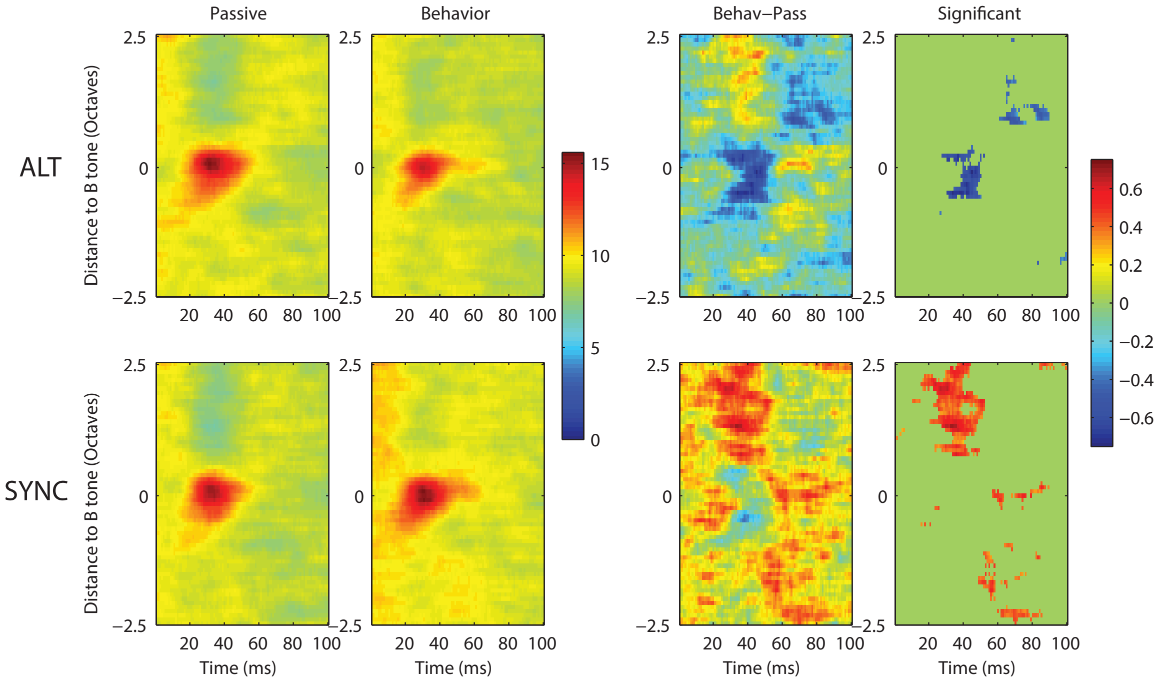

Figure 3.6: Changes in Reconstructed RFs of SUA. The left colorbar is shared by the left two columns, and the right colorbar the right two columns.

The last two columns show the change between passive and behavior and corresponding significant parts. The change of each unit is normalized to the standard deviation. The difference between ALT and SYNC is very salient. Around the B tone where the responses are the strongest, there is a big suppression for ALT, while for SYNC, there is also a slight suppression but it is not significant. Paired t-test is applied to each pixel of the RF at p-value = 0.05. The correction for multiple test is conducted for each pixel is that only when the four neighbor pixels are also significant by the t-test, the central pixel is considered as significant.

The effect of frequency separation is also investigated by further grouping the trials into NEAR and FAR. The changes between passive and behavior is shown in

Figure 3.7. It is easy to see that in FAR, there is almost no difference between ALT and SYNC, and neither of them has significant change around the B tone. However, there is a big difference in NEAR. The suppression for ALT is very strong, while for SYNC, there is no significant change. The comparison emphases that in A1, the effect is strong when A tone is near to the B tone, which is consistent to the result of analysis on data from reference.

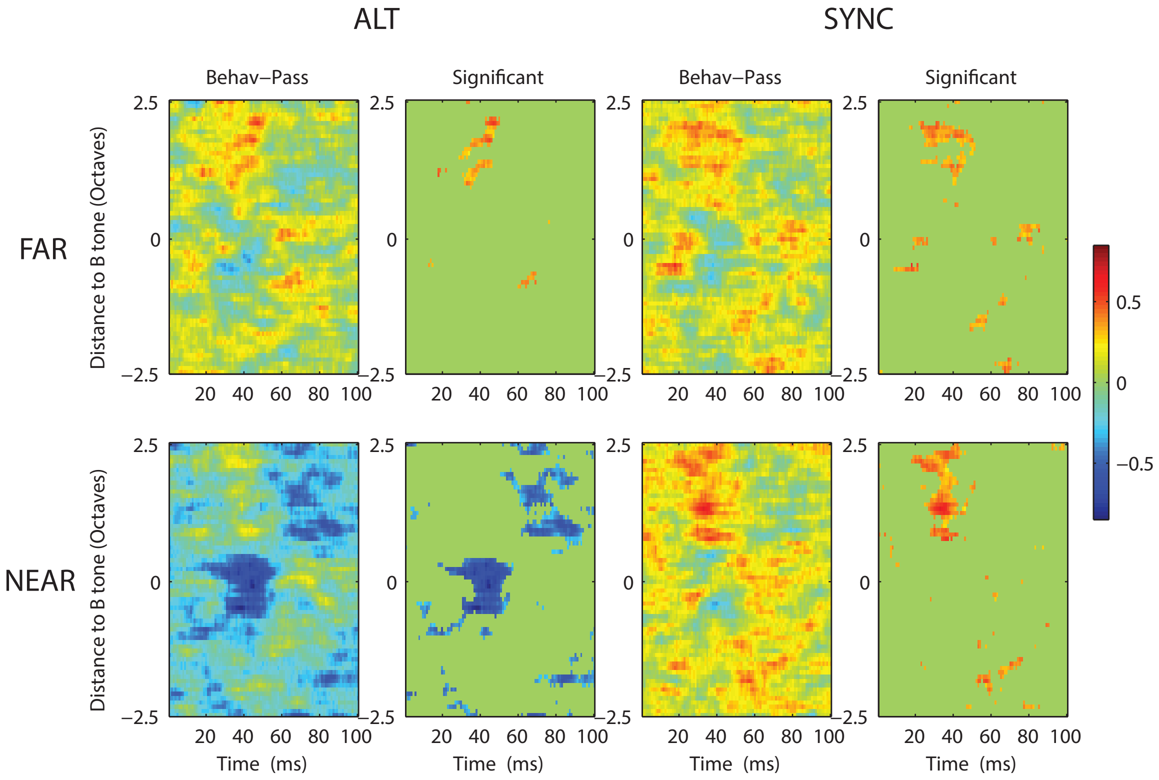

Figure 3.7: Changes in Reconstructed RFs of SUA further grouped by NEAR and FAR.

The RFs are also reconstructed with the high gamma band of LFP. Different from SUA, we observe consistent overall gains in the responses level. The histograms of average changes of RFs of SUA and high gamma band are shown in Figure 3.8.

It is clear to see that there is no significant gain between behavior and passive for SUA, while for high gamma band, the gains are significant in both ALT and SYNC. In order to make a meaningful comparison, the mean of each reconstructed RF is normalized to 0, and the result is shown in Figure 3.9. The changes after removing the mean are very similar with SUA.

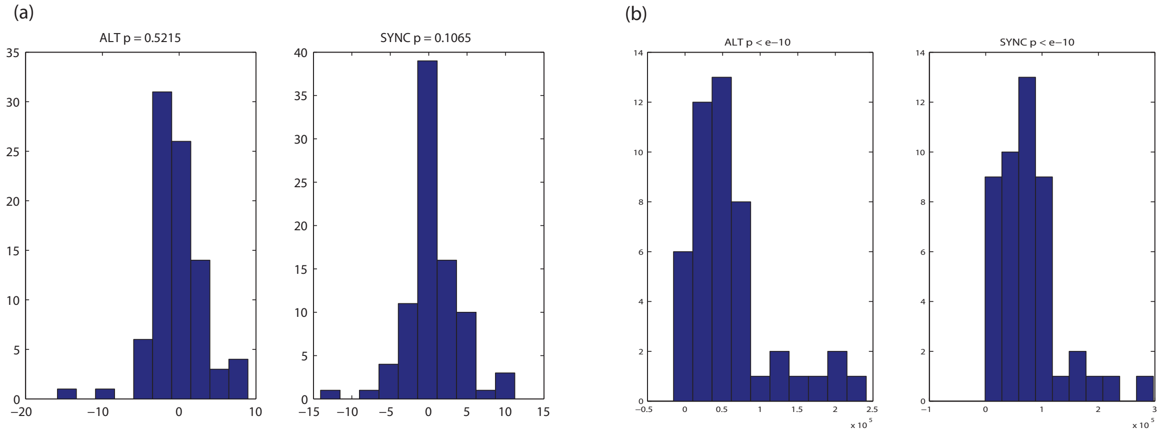

Figure 3.8: Histograms of average changes per unit in Reconstructed RFs. (a): Histograms of average changes per unit of SUA. The average changes are not significant. (b): Hisograms of average changes per unit of high gamma band of LFP. The average changes are significant.

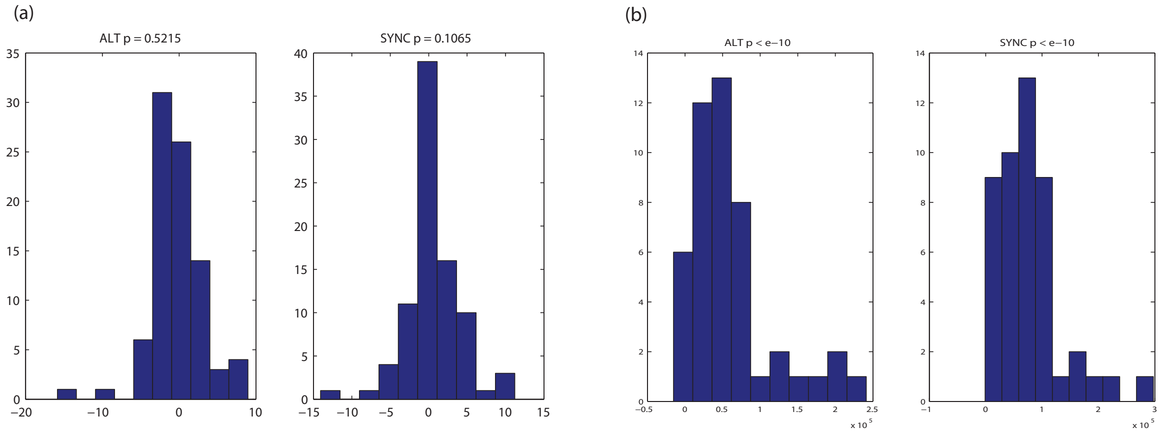

Figure 3.9: Changes in Reconstructed RFs of high gamma band of LFP. The left colorbar is shared by the left two columns, and the right colorbar the right two columns.

Chapter 4: Discussions

4.1 Rapid Plasticity in A1

With a clear tonotopic axis, tuning curves and BFs are the simplest statistics to characterize neurons in A1, and they can be easily estimated by pure tone sequences. In real life, the sound we hear is much most complex than pure tones. The rich spectral and temporal modulations in speech and music make them suitable at communicating information [24]. Various studies have explored how these modulations correlate with the quality and intelligibility of speech [25] [26]. At the very low level module in auditory cortex, neurons in A1 could be modeled as linear two-dimensional filters which are quasi-separable into two one-dimensional filters [27] determining neurons' responses to complex sound. The diverse characteristics of neurons in A1 could be represented by the frequency responses in the spectral and temporal filters. For real neural data, the spectrotemporal characteristics could be estimated by reverse correlation between spikes with specially designed stimuli [28]. TORC is frequently used to estimate the STRF of neurons in A1, and it is mathematically optimized to abstract the linear relationship. Nonlinear relationship could also be estimated by using more natural stimuli like speech [29], and it is shown that STRF estimated in this way has a better prediction power. Recently, a generalized nonlinear model [30] is proposed to model the excitatory and inhibitory components in STRF separately.

In the previous decade, various animal experiments have been conducted to explore the effect of attention on the neural responses in A1 [31] [32] [33] [34]and STRF is a natural candidate to describe the change induced by behavior when compared with passive condition. It has been demonstrated that responses of A1 neurons could be modulated by attention, and that the changes are related with the specific tasks. These changes might help subjects to better focus or ignore certain features in stimuli, leading to desirable behavioral reactions. In our study, the changes are analyzed from different perspectives. The salient difference between active and passive confirmed the effect of attention on the neural responses in A1, arguing for the necessity of the electrophysiological experiments with animals in active conditions when analyzing neural mechanism of auditory streaming. The distinct difference in changes between ALT and SYNC confirm our hypothesis that the temporal coherence is coded and reflected in the neural responses in A1. It is worth pointing out that the plasticity we observe in this study is very rapid. Recall that in the experiment design, ALT and SYNC trials were interleaved. Thus the plasticity is essentially evolving in the scale of seconds. From the analysis on the changes in the reconstructed receptive fields, we see no clear difference between different trial lengths. And it is also observed that the adaptation could reach steady state responses in several repetitions of tones in the reference. Thus this rapid plasticity can happen and fade away in less than 1 second.

4.2 Attention and Temporal Coherence for Auditory Streaming

It is argued that both attention and temporal coherence are both indispensable for the streaming perception. In [22], electrophysiological experiments on the passive listening ferrets led to results that questioned the previous argument about "population separation" hypothesis for streaming separation. Without taking temporal coherence into account, there is no clear difference between ALT and SYNC trials. However without the effect of attention, the study in [22] is not able to predict what will happen when ferrets are actively engaging in streaming tasks. Though many psychoacoustic experiments are trivial for human subjects to perform, the current non-invasive recording techniques like EEG, MEG are only capable of observing the group effect in a global sense. Either bad temporal resolution or bad spatial resolution renders these techniques weak in studying the rapid plasticity when brain is functioning. The exact neural network for streaming perception still remains unclear under the limitation of current technology. Nevertheless the invasive electrophysiological technique leads closer to the final goal, and allows researchers to investigate the neural mechanism in the single unit level. The limitation of traditional chronicle recording is the small number of channels that could sample the brain at the same time, but it is gradually improved by new technology like multi-channel array recoding and two photon imaging. In this study, we successfully train one ferret to pay attention to the stimuli, and results are very promising for future research in this direction.

The temporal coherence within each stream binds the components together and segregates them out of the background. Under the effect of attention, ALT and SYNC trials differ substantially in their induced neural representation in A1. The coherence of A and B sequences in SYNC leads to enhancement in neural responses while anti-coherence in ALT leads to suppression. Though various algorithms have been proposed to model the brain in streaming perception, the responses of real brains are essentially data driven. The intrinsic coherence property in stimuli can quickly modulate the neural connections, resulting in the rapid plasticity. Thus our study supports the theory and temporal coherence model introduced in [20].

4.3 Global Attention versus Selective Attention

From the classic psychoacoustic experiments using two tone sequences to the latest speech separation experiment on human subjects, streaming tasks have always been easy to conduct. The biggest advantages of using human subjects are the short training time and consistent and believable performance. However, it is extremely hard to train animals to performance similar tasks. To solve the cock-tail party problem, selective attention is required for a listener to focus on a particular speaker despite the noisy environment. Though this ability is trivial for most people with normal listener, it is impossible for animals to do as higher level knowledge is used besides the intrinsic temporal coherence in segregating and understanding one's speech. Even simple tasks like paying attention to one tone sequence in the mixture of alternating two tone sequences is very hard to train animals to perform. The problem is that there is no good way to make animals understand what the task is about, and it is hard to verify if animals are performing the task in the way we are interested instead of other irrelevant cues. In this study, we made a compromise to get around the hardness of directing animals to pay selective attention, and global attention which only requires the ferret to listen to the stimuli in a global sense is explored.

However, our observation is also restrained by the specific attention the ferret paid during experiment. In ALT, the two sequences competed with each other under global attention. It is not possible for a subject to pay attention to two separate streams at the same time. Thus when the gap between A and B is large, there might be a back and forth in the attention of the ferret, alternating between the two sequences. What's more, the reference of SYNC trials is essentially one stream, there is no well-defined streams in ALT trials. Without selective attention , A and B sequences were never separate. The general suppression in ALT trials could be the result of competition of the two sequences. In future research, selective attention should eventually be explored to fully verity the temporal coherence model. Our current hypothesis about selective attention is that in ALT, the attended stream will be enhanced while the ignored the sequence will be suppressed.

4.4 Auditory Streaming in A1 and Beyond

A1 is the lowest module in auditory cortex, and its main function is to decompose the input from the early auditory stage. The perception of streaming is a complex activity and sometimes requires high level knowledge to make full use available cues in the stimuli. Thus higher level modules in auditory cortex are also important in streaming separation. Since we restrain the analysis in A1, we could not have a full picture of the underlying mechanism. One clear limitation of our current study is that the strongest changes happened when A and B sequences are near which is a direct result of relatively narrow tuning of A1 neurons. Perceiving complex stimuli like speech and music requires broad spectral integration that beyond the functionality of A1. In future research, neural responses in secondary area would be sampled.

Chapter 5: Conclusions

In this study, we successfully trained one ferret with a global attention task. Based on the analysis with SUA, we observed large changes in firing rate between passive and behavior responses near the onset of the tone sequences for both ALT and SYNC which could be attributed to the effect of attention. The changes in firing rate to SYNC and ALT sequences significantly diverged during the later part ($>$ 4 tones) - or "steady state" portion of the sequences. During the steady sate, SYNC responses were enhanced compared with the passive state and ALT responses were suppressed (or no changes) compared to the passive state in the 40-70ms late window. And changes were most striking when frequency separation between the two tones was small (in NEAR group). The changes of RFs occurred very quickly (within the duration of a trial, or in less than 1-2 seconds). This extremely rapid plasticity is quite extraordinary, and occurred despite the fact that different trial conditions were interleaved in a random order. Additional analysis on the high gamma band of LFP shared similar trends with that of SUA. In summary, rapid task-related RF plasticity occurs within seconds over the course of single trials, and is profoundly influenced by stimulus temporal coherence.

In future research, we would like to explore the effect of attention beyond A1 and design selective attention tasks to further verify the temporal coherence model.

Bibliography

[1] Albert S Bregman. *Auditory scene analysis: The perceptual organization of sound*. The MIT Press, 1994.

[2] DeLiang Wang and Guy J Brown. *Computational auditory scene analysis: Principles, algorithms, and applications*. Wiley interscience, 2006.

[3] Yonatan I Fishman, David H Reser, Joseph C Arezzo, and Mitchell Steinschneider. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing research*, 151(1):167–187, 2001.

[4] Yonatan I Fishman, Joseph C Arezzo, and Mitchell Steinschneider. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *The Journal of the Acoustical Society of America*, 116:1656, 2004.

[5] Jagmeet S Kanwal, Andrei V Medvedev, and Christophe Micheyl. Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Network: Computation in Neural Systems*, 14(3):413–435, 2003.

[6] Joel S Snyder, Claude Alain, and Terence W Picton. Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of cognitive neuroscience*, 18(1):1–13, 2006.

[7] Joel S Snyder, W Trent Holder, David M Weintraub, Olivia L Carter, and Claude Alain. Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology*, 46(6):1208–1215, 2009.

[8] Alexander Gutschalk, Andrew J Oxenham, Christophe Micheyl, E Courtenay Wilson, and Jennifer R Melcher. Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *The Journal of Neuroscience*, 27(48):13074–13081, 2007.

[9] Hirohito M Kondo and Makio Kashino. Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *The Journal of Neuroscience*, 29(40):12695–12701, 2009.

[10] Nima Mesgarani and Edward F Chang. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397):233–236, 2012.

[11] Nai Ding and Jonathan Z Simon. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, 109(29):11854–11859, 2012.

[12] Ling Ma. *Auditory Streaming: Behavior, Physiology, and Modeling*. PhD thesis, University of Maryland, College Park, 2011.

[13] Israel Nelken et al. Processing of complex stimuli and natural scenes in the auditory cortex. *Current opinion in neurobiology*, 14(4):474–480, 2004.

[14] Claude Alain et al. Breaking the wave: effects of attention and learning on concurrent sound perception. *Hearing research*, 229(1-2):225, 2007.

[15] Christophe Micheyl and Andrew J Oxenham. Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. *Hearing research*, 266(1-2):36, 2010.

[16] Robert P Carlyon. How the brain separates sounds. *Trends in cognitive sciences*, 8(10):465–471, 2004.

[17] William Morris Hartmann and Douglas Johnson. Stream segregation and peripheral channeling. *Music perception*, pages 155–183, 1991.

[18] Daniel Pressnitzer, Mark Sayles, Christophe Micheyl, and Ian M Winter. Perceptual organization of sound begins in the auditory periphery. *Current Biology*, 18(15):1124–1128, 2008.

[19] Shihab A Shamma and Christophe Micheyl. Behind the scenes of auditory perception. *Current opinion in neurobiology*, 20(3):361, 2010.

[20] Shihab A Shamma, Mounya Elhilali, and Christophe Micheyl. Temporal coherence and attention in auditory scene analysis. *Trends in neurosciences*, 34(3):114–123, 2011.

[21] Stephen V David, Jonathan B Fritz, and Shihab A Shamma. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences*, 109(6):2144–2149, 2012.

[22] M Elhilali, L Ma, C Micheyl, A Oxenham, and S Shamma. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*, 61(2):317–329, 2009.

[23] Bryan A Seybold, Amelia Stanco, Kathleen KA Cho, Gregory B Potter, Carol Kim, Vikaas S Sohal, John LR Rubenstein, and Christoph E Schreiner. Chronic reduction in inhibition reduces receptive field size in mouse auditory cortex. *Proceedings of the National Academy of Sciences*, 109(34):13829–13834, 2012.

[24] Nima Mesgarani, Malcolm Slaney, and Shihab A Shamma. Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(3):920–930, 2006.

[25] Taishih Chi, Yujie Gao, Matthew C Guyton, Powen Ru, and Shihab Shamma. Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, 106:2719, 1999.

[26] M Elhilali, T Chi, and S Shamma. A spectro-temporal modulation index (stmi) for assessment of speech intelligibility. *Speech Communication*, 41:331–348, 2003.

[27] Taishih Chi, Powen Ru, and Shihab A Shamma. Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, 118:887, 2005.

[28] David J Klein, Didier A. Depireux, Jonathan Z. Simon, and Shihab A. Shamma. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *Journal of computational neuroscience*, 9(1):85–111, 2000.

[29] Stephen V David, Nima Mesgarani, and Shihab A Shamma. Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network: Computation in Neural Systems*, 18(3):191–212, 2007.

[30] Nadja Schinkel-Bielefeld, Stephen V David, Shihab A Shamma, and Daniel A Butts. Inferring the role of inhibition in auditory processing of complex natural stimuli. *Journal of Neurophysiology*, 107(12):3296–3307, 2012.

[31] Jonathan Fritz, Shihab Shamma, Mounya Elhilali, and David Klein. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature neuroscience*, 6(11):1216–1223, 2003.

[32] Jonathan Fritz, Mounya Elhilali, Shihab Shamma, et al. Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing research*, 206(1-2):159–176, 2005.

[33] Jonathan B Fritz, Mounya Elhilali, and Shihab A Shamma. Differential dynamic plasticity of a1 receptive fields during multiple spectral tasks. *The Journal of neuroscience*, 25(33):7623–7635, 2005.

[34] Serin Atiani, Mounya Elhilali, Stephen V David, Jonathan B Fritz, Shihab A Shamma, et al. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron*, 61(3):467, 2009.