

## ABSTRACT

Title of Dissertation:                   AUDITORY CORTICAL RESPONSE TO  
SPECTROTEMPORALLY DYNAMIC  
STIMULI DURING PASSIVE LISTENING  
AND BEHAVIOR

Ji Liu, Doctor of Philosophy, 2022

Dissertation directed by:           Professor Patrick O. Kanold, Department of  
Biomedical Engineering, Johns Hopkins  
University  
Professor Daniel A. Butts, Department of  
Biology, University of Maryland

Our sensory system is bombarded with information that can change whimsically and yet we make sense of the flow of the information effortlessly. How does the brain encode such richly dynamic stimuli? Specifically, how does the auditory system encode rich spectral and temporal aspects of the stimulus and how does it depend on the behavioral state of the animal? My study aims to answer these questions within the scope of mouse auditory cortex (ACX) using imaging techniques on various scales. Firstly, I studied how the ACX encodes one temporal aspect of the

sound, specifically the onset and the offset. I found that offset responses dominated ACX at high sound levels and their strength depended on auditory cortical fields. Moreover, ACX neurons likely inherit their offset responses from thalamocortical input, which is further processed by local cortical microcircuit. Second, I studied the spectral tuning properties of layer 2/3 neurons in mouse ACX using two-tone stimuli. This study revealed the complex inhibitory sideband structures not only in excitatory and inhibitory neurons, but also in feedforward input from auditory thalamus. These complex structures showed a higher degree of feature selectivity of auditory neurons beyond what is predicted by conventional tuning, and thus auditory cortical responses are highly dependent on the spectral context. These two studies focused on passive listening, but cortical responses could depend on the behavioral state of the animal. The predictive coding theory proposes that sensory cortical responses are a form of error response signaling when sensory input failed to conform with predictions from higher order brain areas. Thus, to study the encoding of spectrotemporally dynamic stimulus under active engagement and to test the predictive coding theory, I designed a novel behavior paradigm that allowed the animal to interact with the sound stimulus and studied the cortical responses to not only the combination of sensory information and the animal's action but also the introduced perturbation. Together, this dissertation combined advanced imaging techniques and innovations in experimental designs to provide new insight into how ACX encodes sound stimulus under various scenarios.

AUDITORY CORTICAL RESPONSE TO SPECTROTEMPORALLY DYNAMIC  
STIMULI DURING PASSIVE LISTENING AND BEHAVIOR

by

Ji Liu

Dissertation 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  
Doctor of Philosophy  
2022

Advisory Committee:

Professor Daniel Butts, Chair

Professor Patrick Kanold

Professor Jonathan Simon

Professor Behdash Babadi

Professor Matthew Roesch, Dean's Representative

© Copyright by

Ji Liu

2022

# Dedication

*To my parents and dearest friends*

# Acknowledgements

This work would be impossible without the many people who supported me in various ways over the years.

Above all, I would like to thank my mother, Meilian Mao, and my father Youzhi Liu, who have given me unconditioned support throughout my PhD studies.

I want to thank my mentor, Prof. Patrick O. Kanold, who provided me with guidance and cultivated my scientific thinking. I am grateful for the freedom I was provided to explore my unique ideas and approach to auditory neuroscience. My achievements would be impossible without the rigorous training I received from Prof. Kanold.

I want to thank the current and past members of Kanold lab. They include Xiangying Meng, Dan Winkowski, Rongkang Deng, Bingham Xue, Aminah Sheikh, Krystyna Solarana, Zac Bowen, Kevin Armengol, Didhiti Mukherjee, Minzi Chang, Travis Babola, Clare Choi, Peter Jendrichovsky, HiJee Kang, Katherine Maximov, Lillian Choi, from whom I received immense help for my projects as well as for my life.

I want to thank Prof. Daniel Butts for providing me with great input on my first published work and other projects, and also for being my on-campus advisor after the lab transitioned to a new institution. I want to thank Matt Whiteway, whose work inspired me and provided me with immense help on my first paper.

I want to thank Prof. Behtash Babadi for the powerful new analysis tools that provided more rigor to our scientific findings and also for being part of my

dissertation committee. I want to thank Alireza Sheikhattar, who helped incorporating these analyses into my studies.

I want to thank Prof. Jonathan Simon and Prof. Matthew Roesch for your support as members of my thesis committee, and for providing advice on my projects.

# Table of Contents

Dedication .....	ii
Acknowledgements .....	iii
Table of Contents .....	v
List of Tables .....	vii
List of Figures .....	viii
List of Abbreviations .....	xi
Chapter 1. Introduction.....	1
Functional organization of auditory cortex.....	1
Receptive fields in auditory neurons.....	3
Behavioral dependence of auditory responses .....	5
Predictive coding and auditory error response.....	7
Outline of chapters.....	8
Chapter 2. Parallel processing of sound dynamics across mouse auditory cortex via spatially patterned thalamic inputs and distinct areal intracortical circuits .....	11
Abstract .....	11
Introduction.....	12
Methods.....	14
Results.....	35
Discussion.....	75
Chapter 3. Diversity of receptive fields and sideband inhibition with complex thalamocortical and intracortical origin in L2/3 of mouse primary auditory cortex...	83

Abstract .....	83
Introduction.....	84
Methods.....	87
Results.....	99
Discussion.....	136
Chapter 4. Interactive auditory task reveals complex sensory-action integration in mouse primary auditory cortex .....	144
Abstract .....	144
Introduction.....	145
Methods.....	148
Results.....	163
Discussion.....	201
Chapter 5. Summary and Discussion.....	206
Summary.....	206
Discussion and Future Directions .....	207
Bibliography .....	214

# List of Tables

Table 3-1 Basic response properties across cell types .....	106
Table 3-2 Proportion of neurons within each FRA types with sideband inferred ....	114
Table 3-3 P values, Wilcoxon signed-rank test, width of tuning curve vs. width of sideband inhibition and effect size measured by Hedges' g .....	114
Table 3-4 Wilcoxon rank sum test p values and Hedges' g, facilitation vs suppression across cell type and FRA types .....	126

# List of Figures

Figure 2.1 Both On-R and Off-R show global tonotopy .....	38
Figure 2.2 Tone onset and offset differentially activate large areas of ACX .....	39
Figure 2.3 On/Off-R is not related to the animal's arousal state .....	40
Figure 2.4 ACX organization is stereotypical across mice.....	42
Figure 2.5 Widefield image segmentation using an Autoencoder reveals ACX areas with distinct On/Off selectivity.....	46
Figure 2.6 Behavior of autoencoder fitting.....	48
Figure 2.7 On/Off-R show areal differences in amplitude and spatial distribution....	50
Figure 2.8 L2/3 neurons show distinct On-R and Off-R and are differentially distributed across ACX areas.....	54
Figure 2.9 Neuronal Off-R is time locked to tone offset. ....	55
Figure 2.10 Cellular and neuropil signals from different ACX fields show similar On/Off-R selectivity to that seen in widefield imaging .....	56
Figure 2.11 Local field potential (LFP) in A1 shows bias towards Off-R in strength and spatial spread.....	58
Figure 2.12 GC analysis reveals distinct On/Off sub-networks .....	64
Figure 2.13 MGB injections of AAV-GCAMP6s label terminals in A1.....	68
Figure 2.14 MGB terminals in A1 largely show either On-R or Off-R and Off-R terminals show higher local signal correlations.....	69
Figure 2.15 PV and SOM neurons show distinct temporal dynamics in response to prolonged tones.....	73

Figure 2.16 Circuit model of separate On/Off processing stream from auditory thalamus to A1 .....	74
Figure 3.1 Inhibitory sideband can be inferred using TT stimuli combined with 2-photon imaging of GCaMP6s across neural populations.....	102
Figure 3.2 Inhibitory sidebands can be inferred with TT stimuli in PV and SST interneurons.....	105
Figure 3.3 The FRAs of individual neurons show a large variability.....	109
Figure 3.4 Classification of FRA shapes reveals distinct receptive field types.....	110
Figure 3.5 All cell types and FRA types show broader inhibitory sidebands than tuning curves.....	116
Figure 3.6 BF of tuning curves were not different across cell types .....	118
Figure 3.7 Inhibitory sidebands show more local heterogeneity than tuning curves	121
Figure 3.8 The presence of the second tone decorrelates neuronal responses.....	123
Figure 3.9 Nonlinear frequency interactions show prominent suppression among all cell types while the relative facilitation/suppression strength depends specific cell/FRA type combination.....	126
Figure 3.10 Nonlinear frequency interactions as a function of response timing .....	129
Figure 3.11 MGB terminals show inhibitory sideband structures .....	133
Figure 3.12 MGB terminals show narrow tuning and prominent sideband inhibition .....	135
Figure 4.1 Mice were trained to perform a novel interactive behavior paradigm ....	165
Figure 4.2 A1 responses can be explained by various factors .....	169

Figure 4.3 Projection of feature sensitive neurons in a low dimension space and reward neuron traces .....	172
Figure 4.4 Action sensitive (R) neurons showed sound dependent activities.....	177
Figure 4.5 A subset of action-sound jointly sensitive neurons encode the spectrum boundary of the task paradigm.....	182
Figure 4.6 “Delay-Sound” (DS) perturbation induced error responses in a subset of A1 neurons .....	188
Figure 4.7 “Stop-frequency” (SF) perturbations induced DCP responses at frequencies outside target zones .....	196
Figure 4.8 Feature sensitive neurons form local clusters.....	199

# List of Abbreviations

A1	Primary auditory cortex
ACX	Auditory cortex
BF	Best frequency
CF	Characteristic frequency
dB	Decibels
FOV	Field of view
FRA	Frequency response area
GCaMP	GFP-calmodulin fusion protein
GECI	Genetically encoded calcium indicator
L2/3	Layer 2/3
L4	Layer 4
ROI	Region of interest
SPL	Sound pressure level
2P	2-Photon

# Chapter 1. Introduction

One of the fundamental goals of neuroscience is to understand how the brain utilizes the sensory information and guides the behavior of the organism. To achieve this goal, it is paramount that we understand how the brain represents the outside world and how that representation might depend on the organism's interaction with the world. In this dissertation, I investigated how the mouse auditory cortex (ACX) encodes temporal and spectral information of the sound during passive listening using imaging techniques at different spatial scales. Furthermore, I investigated how ACX represented both sound and action while the animal performed a novel interactive behavioral paradigm. These studies aim to shed light on the functional organization of ACX under various conditions and beyond the already known tonotopic structures. An outline of each chapter is included at the end of the introduction.

## **Functional organization of auditory cortex**

Mouse ACX is consisted of multiple auditory fields, each of which is organized tonotopically, i.e., the preference for the frequency of the sound is mapped topographically onto the surface of cerebral cortex (Decharms and Merzenich, 1996; Guo et al., 2012; Issa et al., 2014; Kanold et al., 2014; Tsukano et al., 2015; Romero et al., 2020). This property is inherited from the ascending auditory pathway (Hackett et al., 2011), which traces back to the tonotopic organization of the cochlear (Russell and Sellick, 1977). In primary auditory cortex (A1), the caudolateral site prefers low frequency while the dorsomedial site prefers high frequency (Guo et al., 2012; Issa et al., 2014). This one-dimensional mapping allows the frequency of the sound stimulus

to be decoded by location of the neuron activated on the cortical surface. However, more recent imaging studies at cellular resolutions revealed that although tonotopy exists on the large scale of millimeters, individual neurons along A1 tonotopy exhibited considerable heterogeneity in their frequency selectivity (Bandyopadhyay et al., 2010; Rothschild et al., 2010). These studies suggest that frequency tuning of individual neurons in mouse A1 not only depends on the tonotopic gradient but is also heavily influenced by the structure of the local microcircuit.

Nevertheless, sound stimulus contains not only spectral information but also rich temporal modulation. Thus, auditory neurons are also sensitive to the amplitude modulation (AM) rate (Schreiner and Urbas, 1986, 1988). It has also been shown in several species that the organization of the rate of AM modulation is orthogonal to the tonotopic axis (Schreiner and Langner, 1988; Langner et al., 2002; Baumann et al., 2011). These results suggest that the temporal preference as well as tonotopic gradient form a two-dimensional grid to encode the joint feature selectivity. In addition to AM, frequency modulation (FM) is also common in naturalistic stimuli, e.g., animal vocalization, and thus is also represented by the auditory system (Mendelson and Cynader, 1985). In mouse ACX, one region at the boundary between A1 and anterior auditory field (AAF) has been found that seems to be specialized in encoding FM sweeps as it showed less sensitivity to tones and other stimuli (Issa et al., 2017), although other regions have also been shown to be sensitive to FM sweep rate, including AAF and ultrasonic field (UF) (Stiebler et al., 1997; Trujillo et al., 2011; Honma et al., 2013). These studies show that ACX is intrinsically sensitive to the dynamic spectrotemporal information in the sound stimulus and its organization

with regard to AM or FM are overlaid with its tonotopic axis, creating neurons diversely selective to the combinations of different features.

Another fundamental aspect of the temporal information of sound is its onset and offset. While most of the studies focused on the onset responses, a few studies on offset responses have revealed that ACX also encode the termination of sound (Recanzone et al., 2000; Qin et al., 2007; Fishman and Steinschneider, 2009; Scholl et al., 2010; Baba et al., 2016). Large scale imaging of ACX using flavoprotein showed that one region near A1 might be specialized in encoding tone offset (Baba et al., 2016). Within A1, individual neurons showed distinct onset/offset response pattern, possibly due to highly nonlinear interactions within its input (Deneux et al., 2016). Offset responses were also shown to be functionally separate from onset responses as they were relayed by different set of synapses in A1 (Scholl et al., 2010), potentially having segregated origins in auditory thalamus (He, 2002). Finally, using widefield and 2-photon imaging, it was shown that offset response is also tonotopically organized and it activates larger areas over ACX (Liu et al., 2019). In summary, multiple feature selective maps coexist within ACX, and they combine to create neurons with distinct stimulus selectivity.

## **Receptive fields in auditory neurons**

Due to the tonotopic organization of ACX, auditory cortical neurons typically have best frequency (BF) or characteristic frequency (CF), which they are most sensitive to or respond with the largest strength. These receptive fields typically have

a “V” shape, where the tip of the “V” is centered at the CF, while the bandwidth of the receptive field broadens as the sound level increases. Thus, auditory neurons can be modeled as bandpass filters.

However, auditory receptive fields showed considerable changes along the auditory ascending pathway. As one moves towards ACX, the neuronal responses become increasingly less temporally faithful, due to the low pass filtering by the cellular membrane across several synapses, while the spectrotemporal feature selectivity becomes more complex (Sharpee et al., 2011). Atencio et al. (2012) showed that in inferior colliculus, the auditory responses can be characterized by a single filter while in ACX the responses increase in their receptive field dimensions. The multiple receptive field dimensions also showed cooperative nonlinearities, i.e., the nonlinearities between different receptive field dimensions are not independent of each other (Atencio et al., 2008). Thus, the responses of ACX neurons can be highly nonlinear.

These complex nonlinearities might be the result of complex interactions between excitatory and inhibitory input a neuron receives. A key property of the receptive field is that they often exhibit lateral inhibitions, i.e., the frequency away from the one that the neuron prefers often produces inhibition in the same neuron. Throughout the ascending auditory pathway, such lateral inhibitions could be found in cochlear nucleus (Greenwood et al., 1976; Nelken and Young, 1994), the auditory midbrain (Brimijoin and O’Neill, 2005; Mayko et al., 2012), the auditory thalamus (Schreiner, 1981) and ACX (Sutter and Schreiner, 1991; Nelken et al., 1994; Sutter et al., 1999; Li et al., 2014b; Kato et al., 2017). The origin of lateral inhibition in the

auditory system lies in the nonlinear mechanical property of the cochlea (Ruggero et al., 1992). However, in ACX, input from local inhibitory neurons might also contribute to the lateral inhibition (Li et al., 2014b; Kato et al., 2017; Lakunina et al., 2020). Further, the mismatch between the receptive field of excitatory and inhibitory synaptic input could contribute to the diversification of suprathreshold neuronal responses (Tao et al., 2017). Nevertheless, as different studies utilized different methods to measure the lateral inhibition, there is still a lack of consensus with regard to the relative contribution from various sources to this phenomenon in ACX. Thus, further studies are necessary to characterize lateral inhibition in ACX, not only in excitatory and inhibitory neurons but also in thalamic feedforward input.

## **Behavioral dependence of auditory responses**

How does auditory response depend on the behavioral state of the organism? Human studies suggest that active engagement or attention can enhance the representation of external stimulus in the brain (Hillyard et al., 1973; Petkov et al., 2004). Ferret ACX neurons engaged in a detection task showed rapid increase in the sensitivity towards the target tone in their receptive fields (Fritz et al., 2003). Long range cholinergic input to ACX could serve as the source of such plasticity as pairing stimulation of cholinergic nucleus basalis with tone presentations induced receptive field plasticity where the excitatory synaptic input evoked by the paired tone increased (Froemke et al., 2007). Thus, a recurring theme involving attention is its boost of the representation of relevant sensory information in the brain.

However, a more recent study found a suppressive effect of task engagement on the tone responses in rat ACX (Otazu et al., 2009), while another imaging study found that the relative proportion of suppression and facilitation depended on the neuronal types, with the majority of excitatory neurons showed suppression instead of facilitation during active engagement while two classes of interneurons, parvalbumin positive (PV) and somatostatin positive (SST) neurons showed mostly facilitation during tasks (Kuchibhotla et al., 2017). Thus, although a subset of auditory neurons indeed showed attention related facilitation of sensory processing, the effects of attention on other neurons can be diverse and are dependent on their molecular identity.

In addition to task engagement, other behavior such as locomotion also strongly influence auditory cortical responses. It has been shown that when sound stimuli were presented during locomotion, the responses were suppressed compared when the animal was quiescent. (Schneider et al., 2014; Zhou et al., 2014; Williamson et al., 2015). The suppression of responses in ACX likely involves the long range projection from the pre-motor cortex onto local PV neurons (Schneider et al., 2014). Furthermore, when a tone is coupled to the action of the animal, e.g., running, lever-pushing etc., ACX also selectively suppresses the response to the paired tone, suggesting a reduction in stimulus saliency due to the predictability (Rummell et al., 2016; Schneider et al., 2018). These studies suggest that mouse ACX integrates both sensory and behaviorally related information, and thus is an ideal locus for studying their interactions.

## **Predictive coding and auditory error response**

The predictive coding theory has become popular in recent years as it is acclaimed to offer a unifying framework to explain a variety of phenomena including associative learning, synaptic plasticity rules, repetition suppression (Friston, 2005). Specifically, it makes the hypothesis that the brain maintains a generative model of the external world and uses incoming sensory information to update its internal representations (Heilbron and Chait, 2017). When incoming sensory events violate existing model, an error signal is generated and propagated to higher order areas. And thus, these error signals are key aspects of the theory

The auditory system has become a good candidate for testing the predictive coding theory as phenomena such as stimulus specific adaptations (SSA) and mismatch negativity (MMN) can be readily interpreted as error signals. Animal studies have primarily relied on SSA to evoke error signals, and it has been argued that SSA and MMN manifest the same predictive coding mechanism albeit on different temporal and spatial scales (Carbajal and Malmierca, 2018). However, although SSA can evoke robust suppression of the repeated tones, a similar phenomenon to the suppression of the tone when it is behaviorally predictable (Rummell et al., 2016; Schneider et al., 2018), SSA is likely to still result mostly from a feedforward mechanism, i.e., synaptic depression. Thalamocortical synapses exhibit high level of depression with stimulation rate higher than 1Hz (Bayazitov et al., 2013), which is consistent with the ‘driver’ like depressing synapses from ventral MGB projecting to primary ACX (Lee and Sherman, 2010). Moreover, similar level

of SSA has been found in both excitatory and inhibitory neural populations in mouse ACX, suggesting that SSA might be due to adaption in the thalamic input (Chen et al., 2015). Finally, as SSA is present in both anesthetized and awake animals (Ulanovsky et al., 2003; Malmierca et al., 2009), it further argues against using SSA as a paradigm to study predictive coding as it likely lacks a top-down component. We argue that prediction should be an intrinsic aspect of action, and thus it should be placed in a framework that involves the active engagement of the animal. In the visual system, the predictive coding has been framed under the paradigm of visual flow where the locomotion of the animal is coupled to the visual stimulus (Keller et al., 2012). This paradigm showed that both local and global mismatch between the animal's action, i.e., running speed, and the visual flow feedback evoked error responses in the visual cortex of the animal (Keller et al., 2012). Further, such responses were altered if the animal was reared in an environment that had abnormal visuomotor coupling (Attinger et al., 2017). These studies present strong evidence that the error responses depend on the experience of the animal, and thus evoking bona fide error responses requires manipulating the animal's interaction with the outside world through training or rearing under special conditions.

## **Outline of chapters**

To investigate how mouse ACX encodes both spectral and temporal aspects of sound stimulus and how the representation depends on the behavioral state of the animal, I used imaging techniques of various spatial resolutions to both investigate

large scale as well as microscopic scale functional organizations of different features in mouse ACX.

In Chapter 2 of this dissertation, I investigated the encoding of sound offset across mouse ACX. This research has been published in Cell Report (Liu et al., 2019). We used both widefield imaging as well as 2P imaging to delineate the offset responses profile across ACX fields. Using a novel widefield imaging segmentation technique, we found that offset responses were organized also according to tonotopic axis, and that they activated larger areas over the ACX. The relative strength of onset and offset responses depended on the particular ACX field, while A1 showed the strongest bias toward offset response. 2P imaging largely recapitulated the results from widefield imaging and showed that the majority of ACX neurons encoded either onset or offset responses. In addition, inhibitory neurons showed distinct temporal response profiles that might help amplify offset responses in excitatory cells. Lastly, thalamocortical feedforward input also showed offset responses, albeit with a weaker offset response bias, suggesting further processing by the cortical microcircuits. These results suggest that onset and offset responses were processed by parallel networks that might have their origins in the auditory thalamus.

In Chapter 3, I characterized the sideband inhibition in L2/3 of mouse A1. Specifically, I designed a variant of the traditional two-tone paradigm, which was adaptable to measure sideband inhibition in large populations of neurons, especially combined with 2P imaging. Using this approach, we showed that L2/3 excitatory neurons not only had distinct frequency response area (FRA) shapes, but also showed distinct profiles of sideband inhibition. These shapes further corresponded to

differential levels of nonlinear interactions. We also showed that the signal correlations between neurons were decreased when presented with two-tone instead of single pure tones, suggesting a sparser representation of spectrally rich stimulus. Finally, we showed that sideband inhibitions also exist in PV and SST positive interneurons, as well as in thalamic input, suggesting that sideband inhibitions are largely inherited from thalamocortical feedforward input or local excitatory populations.

In Chapter 4, I investigated A1 cortical responses while the animal actively engaged in an interactive behavioral paradigm. This paradigm is innovative because it allows the animal to modulate one attribute of the sound, i.e., its carrier frequency. After extensive training, we imaged the neuronal activity in both L2/3 and L4 of the behaving animal, which showed complex interaction between sound and action. Especially in L2/3, neuronal activities began to encode action with selectivity. We further introduced perturbations of the coupling between sound and action, and we observed evoked error responses in both L2/3 and L4. Together, the results suggest that A1 neurons learnt to represent the general rule of the paradigm by signaling its violation through error responses.

In Chapter 5, I summarize the results and discussed potential future work that could extend beyond the scope of this thesis.

# Chapter 2. Parallel processing of sound dynamics across mouse auditory cortex via spatially patterned thalamic inputs and distinct areal intracortical circuits

*This chapter has been published in Cell Report (Liu et al., 2019). JL, POK designed research. JL performed experiments and analyzed the data. MW, DB, and JL designed and implemented the autoencoder analysis algorithm. AS and BB implemented Granger analysis. POK supervised research. JL and POK wrote the paper.*

## Abstract

Natural sounds have rich spectral and temporal dynamics. Multiple auditory cortical (ACX) areas spatially represent spectral information via large-scale maps. However, the representation of temporal aspects of sounds, e.g. sound offset, is unclear. We performed multiscale imaging of large-scale and cellular neuronal activity as well as thalamic activity evoked by sound onset and offset in awake mouse ACX. We find that known and novel higher ACX areas differed in onset (On-R) and offset-responses (Off-R). Most excitatory layer 2/3 neurons showed either On-Rs or Off-R and ACX areas were characterized by differing fractions of On-R/Off-R neurons. Somatostatin (SOM) and parvalbumin (PV) interneurons showed distinct temporal dynamics, potentially amplifying Off-Rs. Functional network analysis showed that ACX areas contained distinct parallel On- and Off-networks. Thalamic (MGB) terminals showed either On-Rs or Off-Rs indicating a thalamic origin of On/Off-R pathways. Thus,

ACX areas spatially represents temporal features and contain circuit specializations for temporal processing. Our findings suggest that spatial convergence and co-activation of MGB inputs initially determines the cellular On/Off-preference which is refined upon by distinct intracortical connectivity.

## **Introduction**

Natural sounds have rich spectral and temporal dynamics, and neuronal populations along the auditory processing stream encode both spectral and temporal information. Sound onset and offset are fundamental dynamic features of sound, to which single neurons at multiple levels of the auditory system respond (Hillyard and Picton, 1978; Henry, 1985; He et al., 1997; Kopp-Scheinflug et al., 2011), including the auditory cortex (ACX) (Recanzone, 2000; He, 2001; Qin et al., 2007; Fishman and Steinschneider, 2009; Scholl et al., 2010; Baba et al., 2016). While off-responses (Off-R) have been suggested to be responsible for duration coding (He, 2001), they, together with on-response (On-R), encode the basic cues (onset/offset) for auditory scene analysis (Bregman, 1994). Thus, besides elucidating the encoding of both sound onset and offset, revealing the underlying cellular networks is essential for understanding auditory processing.

The ACX contains multiple functional areas and the spatial organization of ACX with respect to On-Rs has been extensively studied. On a large scale (hundreds of microns), there are clear tonotopic maps, which are due to topographic thalamocortical projections (Merzenich et al., 1975; Stiebler et al., 1997; Guo et al., 2012; Issa et al., 2014; Kanold et al., 2014; Tsukano et al., 2015), while on a finer

scale 2-photon imaging studies in mouse primary ACX (A1) revealed a diverse tonotopic organization of On-Rs in superficial layers (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Winkowski and Kanold, 2013; Kanold et al., 2014). In contrast, the spatial organization of Off-Rs in ACX is less well understood. Widefield flavoprotein imaging revealed the existence of an area adjacent to A1 that is specialized in processing tone offset regardless of tone frequency in anesthetized mice (Baba et al., 2016). On a finer scale, neurons in mouse ACX show distinct On/Off-R patterns (Deneux et al., 2016), and inputs carrying On-Rs and Off-Rs are proposed to originate in non-overlapping synaptic circuits (Scholl et al., 2010). These findings at different scales raise the possibility that On- and Off-Rs reflect distinct parallel pathways not only within A1 but also across ACX, and that On- and Off-Rs might be differentially represented in the cortical space. Here, we tested these hypotheses by investigating the spatial representation and functional microcircuits contributing to On-Rs and Off-Rs on multiple spatial scales in ACX.

Since multiple cortical areas contribute to auditory processing, we first performed widefield imaging of GCaMP6s in awake mouse. For unbiased identification of ACX areas we developed an automated image segmentation algorithm based on the temporal responses. With this unbiased method we detected known ACX areas but also identified novel higher ACX areas. The identified ACX areas differed in their response properties to tone onset and offset indicating that temporal selectivity might underlie the functional streams of analysis in ACX. Both On-Rs and Off-Rs showed tonotopic organization. To identify the cellular and circuit substrate of the differing areal selectivity we performed 2-photon imaging of ACX neurons. The majority of

individual excitatory layer 2/3 neurons showed either On- or Off-Rs and ACX areas were characterized by differing fractions of On- or Off-responsive neurons. Selective imaging of parvalbumin (PV) and somatostatin (SOM) interneurons showed differential On-R and Off-R dynamics suggesting suppression of PV neurons by SOM neurons during prolonged tone presentation, consistent with known cortical connection (Cottam et al., 2013; Pfeffer et al., 2013) and potentially exerting disinhibiting effect on local excitatory neurons. This disinhibition could serve to selectively amplify cortical Off-R. To further characterize areal differences, we analyzed both intrinsic and ascending functional connectivity. Functional connectivity analysis using pairwise activity correlations and Granger causality showed that ACX areas varied in their intrinsic network structure. Imaging of medial geniculate body (MGB) terminals demonstrated that the most individual MGB terminals showed either On- or Off-Rs suggesting a thalamic origin of the parallel cortical On/Off-R circuits. Spatial analysis suggested that spatial convergence and co-activation of MGB inputs determines cellular On/Off-preference. Inhibitory circuits then further selectively emphasize these preferences. Together our results demonstrate that ACX fields differentially process sound onset/offsets via parallel and spatially patterned projections from the MGB as well as differing intrinsic areal connectivity.

## **Methods**

All procedures were approved by the University of Maryland's Animal Care and

Use Committee.

### Animal

We crossed CBA/CaJ mice with Thy1-GCaMP6s (JAX stock #024275, GP4.3, (Dana et al., 2014)) to obtain F1's since C57BL/6 are homozygous for *Cdh23* allele *ahl*, which causes them to suffer from aging related hearing loss, while CBA/CaJ mice are homozygous for *Ahl+*, which spare them from the phenotype (Kane et al., 2012). F1's thus have no hearing loss and yet have uniform expression of GCaMP6s under Thy1 promotor in excitatory neurons. We used adult mice of both sexes whose ages range from 2 to 4 months old. For imaging PV or SOM neurons, we crossed Thy1-GCaMP6s mice with PV-cre (JAX #008069) or SOM-cre (JAX #013044) mice and injected ~30nl of AAV1.Syn.Flex.mRuby2.GSG.P2A.GCaMP6s.WPRE.SV40 (Addgene68720) into the left ACX of the F1 animals. Such generated animals express innate GCaMP6s in Thy1 pyramidal cells while expressing GCaMP6s and mRuby in either PV and SOM interneurons.

### Chronic window implant

2-3 hours before surgery, 0.1cc dexamethasone (2mg/ml, VetOne) was injected subcutaneously to reduce brain swelling during craniotomy. Anesthesia was induced with 4% isoflurane (Fluriso, VetOne) with a calibrated vaporizer (Matrx VIP 3000). During surgery, isoflurane level was reduced to and maintained at a level of 1.5%-2%. Body temperature of the animal was maintained at 36.0 degrees Celsius during surgery. Hair on top of head of the animal was removed using Hair Remover Face

Cream (Nair), after which Betadine (Purdue Products) and 70% ethanol was applied sequentially 3 times to the surface of the skin before the central part is removed. Soft tissues and muscles were scraped to expose the skull. Then a custom designed 3D printed stainless headplate was mounted over left auditory cortex and secured with C&B-bond (Parkell). A craniotomy with a diameter of about 3.5mm was then performed over left auditory cortex. A three layered cover slip was used as cranial window, which is made by stacking 2 pieces of 3mm coverslips (64-0720 (CS-3R), Warner Instruments) at the center of a 5mm coverslip (64-0700 (CS-5R), Warner Instruments), using optic glue (NOA71, Norland Products). Cranial window was quickly dabbed in kwik-sil (World Precision Instruments) before mounted onto the brain with 3mm coverslips facing down. After kwik-sil cured (2-5min), C&B-bond was applied to secure the cranial window. Synthetic black iron oxide (Alpha Chemicals) was then applied to the hardened surface. 0.05cc Cefazolin (1 gram/vial, West Ward Pharmaceuticals) was injected subcutaneously when entire procedure was finished. After the surgery, the animal was kept warm under heat light for 30 minutes for recovery before returning to home cage. Medicated water (Sulfamethoxazole and Trimethoprim Oral Suspension, USP 200mg/40mg per 5ml, Aurobindo Pharms USA; 6ml solution diluted in 100ml water) substituted normal drinking water for 7 days before any imaging was performed.

### Widefield imaging

Mice were affixed to a custom designed head-post and restrained within a plastic

tube. The head of the animal was held upright. Imaging was performed using Ultima-IV two photon microscope (Bruker Technologies) with an orbital nosepiece such that the illuminance light is roughly perpendicular to cranial window (rotation angle was ~60 degrees). As a result, the anterior-posterior axis was not parallel to the edge of the images. 470nm LED light (M470L3, Thorlabs Inc.) was used to excite green fluorescence. Images were acquired with StreamPix 6.5 software (Norpix) at 10Hz and 100ms exposure time. In StreamPix software, we specified the image size to be 400 by 400 with a spatial binning of 3.

#### Acoustic stimulus

Pure tones were generated with custom MATLAB script. Each tone lasted 2 seconds with linear ramps of 5ms at the beginning and at the end of the tone. The amplitudes of the tones were calibrated to 75Db SPL with a Brüel & Kjær 4944-A microphone. During sound presentation, sound waveform was loaded into RX6 multi-function processor (Tucker-Davis Technologies (TDT)) and attenuated to desired sound levels by PA5 attenuator (TDT). Then the signal was fed into ED1 speaker driver (TDT), which drove an ES1 electrostatic speaker (TDT). The speaker was placed on the right-hand side of the animal, 10cm away from the head, at an angle of 45 degrees relative to the mid-line. The presentation of tones with various combination of frequencies and sound levels are randomized and controlled by a custom MATLAB program. The silent period in between the 2-second tones was randomly chosen from a uniform distribution between 3 and 3.5 seconds. Frequencies

of the tones vary from 4kHz to 83.0kHz with logarithmic spacing and with a density of 2.28 tones per octave. Sound levels vary from 5dB SPL to 65dB SPL with a step of 15dB. Each stimulus was repeated 10 times. In total, the widefield imaging session for for each animal lasted ~45min. For 2-photon imaging, 9 tones with equal logarithmic spacing between 4 and 64kHz were used at a single level of 60dB SPL. The tone duration was 2 second and repeated 10 times.

### Image preprocessing

We performed three preprocessing steps before using autoencoder for image segmentation. First, we downsampled the original image (400 by 400) using MATLAB (2015b) using the MATLAB built-in function ‘imresize’, by a factor of 4. The resultant image size was 100 by 100. Next we performed whitening of the image sequence. We first re-shaped each image into column vectors, then we stacked them horizontally. Let  $I_t$  denote the column vector corresponding to image at time  $t$ ,  $M$  be the stacked matrix, and  $N$  be the total number of images:

$$M = [I_1, I_2, \dots, I_N]$$

We then subtracted the time average image ( $\langle I \rangle_t$ ) from all images:

$$\hat{M} = M - \langle I \rangle_t \times \underbrace{[1, 1, \dots, 1]}_N$$

We then performed singular value decomposition on sample covariance matrix of  $\hat{M}$ :

$$[U, S, V] = SVD(\hat{M} \times \hat{M}' / N)$$

Then we obtained the whitened images using the following equation:

$$\tilde{M} = U \times (S^{-1} + \lambda) \times U' \times \hat{M}$$

where  $\lambda$  is the regularization term. We picked  $\lambda$  by first plotting the sorted eigenvalues in  $S$  in logarithmic space and usually a fast initial drop off and a following relatively flat region can be observed. We picked  $\lambda$  close to the turning point to preserve relevant variance and to avoid amplifying noise. We then fed  $\tilde{M}$  into autoencoder algorithm.

### Image Segmentation with constrained autoencoder

We used a dimensionality reduction technique to perform automatic image segmentation such that pixels with strong temporal correlations across the set of images were grouped together into single components (ROIs), following the formulation of Whiteway and Butts (2017). To perform this dimensionality reduction, we used an autoencoder neural network. The goal of this constrained autoencoder is to adjust the weights between the input layer and the hidden layer and those between the hidden layer and the output layer such that the output matches the input as closely as possible. For each time point  $t$ , the autoencoder takes the vector of pixel values  $\mathbf{y}_t \in \mathbb{R}^N$  and projects it down onto a lower dimensional space  $\mathbb{R}^M$  using an *encoding matrix*  $W_1 \in \mathbb{R}^{M \times N}$ . A bias term  $\mathbf{b}_1 \in \mathbb{R}^M$  is added to this projected vector, so that the resulting vector  $\mathbf{z}_t \in \mathbb{R}^M$  is given by

$$\mathbf{z}_t = W_1 \mathbf{y}_t + \mathbf{b}_1$$

The autoencoder then reconstructs the original activity  $\mathbf{y}_t$  by applying a *decoding matrix*  $W_2 \in \mathbb{R}^{N \times M}$  to  $\mathbf{z}_t$  and adding a bias term  $\mathbf{b}_2 \in \mathbb{R}^N$ , so that the reconstructed activity  $\hat{\mathbf{y}}_t \in \mathbb{R}^N$  is given by

$$\hat{\mathbf{y}}_t = W_2 \mathbf{z}_t + \mathbf{b}_2$$

Since the dimensionality of  $\mathbf{z}_t$  is typically much smaller than that of  $\mathbf{y}_t$ ,  $\mathbf{z}_t$  should capture variations in  $\mathbf{y}_t$  that are shared across many pixels. The entries of  $W_2$  then describe how each pixel is related to each dimension of  $\mathbf{z}_t$  (see Figure 2.5C).

The weight matrices and bias terms, grouped as  $\Theta = [W_1, W_2, \mathbf{b}_1, \mathbf{b}_2]$ , are simultaneously fit by minimizing the mean square error between the observed activity  $\mathbf{y}_t$  and the predicted activity  $\hat{\mathbf{y}}_t$ :

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \frac{1}{2} \sum_t \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2^2$$

To further enable interpretability of the results, we constrained the weights  $W_2$  to be non-negative, as one could flip the signs of both spatial and temporal components arbitrarily. This also ensured that all pixels in a given ROI always increase or decrease in intensity together, depending on the sign of  $\mathbf{z}_t$ . We also tied the weights such that  $W_2 = W_1^T$ . Thus, there was essentially only one spatial weight matrix.

This version of the autoencoder is closely related to principal components analysis (PCA) (Bengio et al., 2013). However, PCA is an inadequate technique for automatic image segmentation since it did not in general result in spatially localized ROIs, due to the orthogonality constraints imposed by the PCA model. A similar approach to our non-negatively constrained autoencoder is to use non-negative matrix factorization (NNMF) on the

preprocessed image sequence. NNMF constrains both the spatial maps and the temporal activations to be non-negative, whereas the RLVM just constrains the spatial maps to be non-negative. The NNMF ROIs also failed to be spatially localized. Finally, in order to solve the constrained minimization problem above we used the spectral projected gradient method, a constrained variant of gradient descent (Schmidt et al., 2009).

To perform image segmentation with this method we must first specify the number of ROIs (the dimensionality of  $\mathbf{z}_t$ ). We determined the appropriate number of ROIs using cross-validation by first fitting the parameters of the autoencoder on 75% of the frames from the image sequence (training data), and then reconstructing the remaining 25% of the images (testing data) using the autoencoder. We then calculated the correlation between the true and reconstructed images on the testing data, as a measurement for goodness of fit. In Figure 2.4A, we show that with an increasing number of ROIs, the correlation from the testing data increases monotonically, and roughly plateaus after ~50 ROIs. We also performed fitting on the entire image sequence and plot the correlation (Figure 2.4A, blue curve). A similar monotonic increase is observed, and with 50 or more ROIs, the correlation value is above 0.8, which is agreeable considering that the full image sequence consisted of more than 28,000 images. Another criterion we utilized to choose the number of ROIs was the total spatial area covered by the ROIs. An increasing portion of the total area is covered with an increasing number of ROIs, (Figure 2.4B), and total area covered by 50 ROIs are close to maximum coverage. Given these results, we typically used 50 ROIs in the autoencoder.

### WF On-R and Off-R amplitude

To determine response amplitude, first the temporal trace from each trial was normalized to percentage change with respect to baseline fluorescence:

$$\text{normalized trace at time } t = \frac{F_t - F_0}{F_0}$$

where  $F_0$  is the baseline determined by finding the most frequent value in the histogram of the trace assuming stability. For On-R amplitude, we averaged the normalized trace from 200-500ms after tone onset with the baseline from normalized trace subtracted. For Off-R, we averaged the normalized trace from 200-500ms after tone offset and subtracted the average from the same trace 0-200ms right before tone offset. The 200-500ms window was sufficient to capture the rising phase as well as the peak of the increase in fluorescence in typical On/Off-R.

### Field Parcellation

We assigned ROIs to different auditory fields based upon known tonotopic structure revealed with optical approach (Issa et al., 2014; Tsukano et al., 2015). ACX of mice contains several auditory fields, including A1, AAF and Ultrasonic Field (UF), which are characterized by the presence of tonotopic gradients in the On-R (Stiebler et al., 1997). Tonotopy also exists in secondary area A2, albeit on a compressed scale (Issa et al., 2014). First, we identified A1 and UF ROIs based on their two tonotopic axes, one from the caudal side to dorsomedial side (low to high) and the other one, sharing the same low frequency area, from caudal to ventrolateral

side (Issa et al., 2014). The example A1 and UF ROIs (Figure 2.5I-O) show progression of frequency selectivity along the two tonotopic axes. We use ‘UF’ and ‘high A1’ to distinguish between the two spatially distinct areas that are high frequency selective, while they are both considered primary auditory cortices. We also found a subset of ROIs located dorsoposterior to A1 which we assigned as DP. They showed relatively weak On-Rs but prominent Off-Rs (Figure 2.5M). We performed parcellation of ROIs in all animals studied, and the similar spatial layout of A1, UF, AAF, A2 and DP can be robustly observed.

#### Signal correlation among ROIs

We used corrected signal correlation (SC) for all our calculation due to the limited number of repeats and the strong tendency of close-by pixels to covary in time (Rothschild et al., 2010; Winkowski and Kanold, 2013). The basic idea is that the uncorrected SC equation contains products of responses from the two ROIs in question on the same trial, and these terms also appear in noise correlation equation. Thus, these products represent to some extent the covariation of ROIs regardless of stimulus presentation, and thus should be excluded from SC calculation. The denominator in the equation was adjusted accordingly to take into account the reduction of number of summation in the nominator.

In Figure 2.7G, H, we calculated SC among selected ROIs that were dorsally located with respect to A1 and UF respectively. These ROIs have centers within ~450um to the A1 and UF ROIs in the rostrocaudal direction but dorsally located.

Then we calculated pairwise SCs among all these ROI pairs and plotted them as a function of distance (Figure 2.7I).

### On- and Off-tonotopy

To establish On- and Off-tonotopy, threshold of WF On-R and Off-R were first manually determined (Figure 2.2, white solid lines). Then WF images with baseline subtracted following tone onset or offset were obtained at identified threshold. Next a homomorphic filter was applied to the images to correct for unevenness of illumination. Then 95 percentile contour lines of the responses were extracted and overlaid to demonstrate systematic movement of activation area as a function of different tone frequencies (Figure 2.1C, D, Figure 2.4).

### 2-Photon imaging of mouse ACX

A week after the cranial window implant, the animals were head-fixed in custom designed holder while 2-second long tones were presented in a similar fashion as in WF experiment. Field of views were placed in A1, AAF, A2 and DP region with a depth of around 150 $\mu$ m and with a size of 369 $\mu$ m x 369 $\mu$ m. The imaging was performed with a B-SCOPE (Thorlabs Inc.) with the microscope body tilted around 45 to 50 degrees while the mouse head was held upright. The excitation wavelength was 920nm and images were collected with ThorImage software (Thorlabs Inc.) at a

frame rate of 30Hz. A 16x Nikon objective was used (NA 0.80). For terminal imaging, the average imaging depth was around 140um, comparable to cellular data.

### 2-Photon imaging data analysis

First motion correction was performed with TurboReg plugin (Thevenaz et al., 1998). In a subset of experiments, the motion correction was performed using the Suite2P package (Pachitariu et al., 2016). ROIs were drawn manually using a custom written GUI. A ring was placed on each cell soma to extract raw fluorescence trace while a circular region of 20μm radius was used to extract nearby neuropil signal (excluding soma). We used the following equation to correct neuropil contamination of cell:

$$F_{corrected}(t) = F_{cell}(t) - 0.8 \times F_{neuropil}(t)$$

The coefficient of correction (0.8) was measured with the collected 2P dataset by taking the ratio of the intensity non-radial blood vessel and the intensity of adjacent neuropil containing no neurons. To calculate  $\Delta F/F$ , the baseline of each cell was determined by constructing a histogram of all fluorescence intensity over time and by finding the peak of the histogram and the corresponding fluorescence intensity value, which we used as the estimate of fluorescence baseline. This procedure is based on several assumptions. First, we assume the baseline is constant over time, which we generally found to be true given our relatively short imaging sessions (~9 min). Second, we assume that the response in ACX is sparse (Hromádka et al., 2008) and thus baseline value should be observed the most often, which will be reflected as the

peak in histogram. This procedure is generally robust and generates  $\Delta F/F$  change over a reasonable range. If this procedure found negative baseline values, suggesting the soma fluorescence was lower in intensity than surrounding neuropil, then these cells were excluded from further analysis. Then, we calculate  $\Delta F/F$  using the following equation:

$$\frac{\Delta F}{F}(t) = \frac{F_{corrected}(t) - baseline}{baseline}$$

To determine whether a cell is significantly responding to sound onset or offset, we first determine the response amplitude in the  $\Delta F/F$  trace by finding the maximum change within 1 second after sound onset or offset and average over a small window ( $\pm 2$  frames) around the maximum time point to account for the noisy fluctuation in the trace. Then the 95 percent confidence interval (CI) of the median of the response amplitude was constructed through a bootstrapping procedure (resampling 1000 times) and if the lower CI bound exceeded 1.5 times the standard deviation of the baseline fluctuation (5 frames or  $\sim 150$ ms before sound onset/offset) then the cell was considered significantly on/off-responsive. The response significance was determined separately for each frequency and sound level combination and separately for On-R- and Off-R. Neuropil and MGB terminal signals were processed with the same procedure. Unlike cellular ROIs, MGB terminal ROIs were obtained with Suite2P in an automated fashion.

For classifying different types of responses (Figure 2.15E), we performed k-means clustering on averaged responses (across repetitions) to each frequency, pooling these traces from Thy1, PV and SOM neurons. The clustering is only

confined to statistically significant responses. We used correlation as the distance measure and thus the clustering disregarded absolute amplitude of the traces. We chose 5 clusters to sufficiently encompass the different response types encountered.

### Off-R Bias Index (OBI)

OBI is calculated by first averaging On-R and Off-R for responding neurons over frequency and repeats, and then calculated with the following equation:

$$OBI = \frac{\langle R_{off} \rangle - \langle R_{on} \rangle}{\langle R_{off} \rangle + \langle R_{on} \rangle}$$

where the angle brackets denote average over tone frequency and repeats.

### Injection of GCaMP6s virus in MGB

AAV1.hSyn1.mRuby2.GSG.P2A.GCaMP6s.WPRE.SV40 (Addgene 50942) virus was injected into MGB for axon terminal imaging in ACX. Micropipettes pulled with a long tapering tip (>3mm) were used for injection with Nanoject II (Drummond Inc.). The location of the left MGB was determined using mouse brain atlas (AP: 3.2mm; ML 2.1mm; DV: 3.0mm). Anesthesia was induced with 4% isoflurane and maintained at 1.5%. The skin over the skull was cut open and a small craniotomy was made to allow penetration from the dorsal side and the micropipette was lowered vertically into MGB. 150-200nl of undiluted virus was injected over 5min. After the injection, the skin was sutured back. 3-4 weeks after the injection, the cranial window was implanted over the left ACX as previously described.

## Granger Causality analysis

The notion of causality proposed by Granger (Granger, 1969) aims at capturing the two fundamental principles of temporal predictability and the precedence of cause over effect. In order to capture the functional dependencies within a neuronal ensemble and the sparsity of interactions, we employ sparse multivariate autoregressive models. We introduce a novel measure of GC which accounts for sparse interactions, estimate the model parameters using fast optimization methods, and perform statistical tests to assess the significance of possible GC interactions, while controlling the false discovery rate (FDR) to avoid spurious detection of GC links.

We used the same framework as in for our Granger Causality (GC) measurement (Sheikhattar and Babadi, 2016). In order to infer GC patterns for the two On/Off conditions, we divide the corresponding responses to the onset and offset inputs, and pool across all the tone frequencies, thereby treating them as implicit repetitions to the same stimuli condition. In what follows, we present our modeling, parameter estimation and GC inference procedure.

Modeling: Consider a sequence of calcium indicator fluorescence measurements from a set of  $C$  neurons indexed by  $c = 1, 2, \dots, C$  within a slice, denoted by  $\{y_{r,n}^{(c)}\}_{r=1:R, n=1:N}^{c=1:C}$  over time bins  $n = 1, \dots, N$ , and across  $R$  trial repetitions indexed by  $r = 1, \dots, R$ . We adopt a sparse vector autoregressive (VAR) framework (Valdés-Sosa

et al., 2005) for modeling the slow-decaying and transient dynamics of the calcium fluorescence signals as well as the cross-dependencies among the neurons.

Suppose that the fluorescence observation vector of neuron  $(c)$  at the  $r$ -th repetition is represented by  $\mathbf{y}_r^{(c)} := [y_{r,1}^{(c)}, \dots, y_{r,N}^{(c)}]'$ , and let  $\bar{\mathbf{y}}^{(c)} := [\mathbf{y}_1^{(c)'}, \mathbf{y}_2^{(c)'}, \dots, \mathbf{y}_R^{(c)'}]'$  denote the zero-mean total observation vector, containing the set of all observation vectors  $\mathbf{y}_r^{(c)}$  from all trials  $r = 1, \dots, R$ .

The effective neural covariates taken into account in our models are each neuron's self-history of activity and the history of activities of other neurons in the ensemble. We consider a lag of  $L$  samples within which the possible neuronal interactions may occur. Then, we segment  $L$  into  $M$  windows of lengths  $W_1, W_2, \dots, W_M$  such that  $\sum_{i=1}^M W_i = L$ . Let  $b_m := \sum_{\ell=1}^m W_\ell$  for  $m = 1, \dots, M$ , and  $b_0 = 0$ . Let

$$\mathbf{h}_{r,n,m}^{(c)} := \frac{1}{W_m} \sum_{k=n-1-b_m}^{n-1-b_{m-1}} \mathbf{y}_{r,k}^{(c)}, \quad (1)$$

represent the average activity of neuron  $(c)$  within the  $m$ -th window lag of length  $W_m$  with respect to time  $n$  and at trial  $r$ . We can then define the vector of history covariates from neuron  $(c)$ , effective at time  $n$  and trial  $r$  as  $\mathbf{h}_{r,n}^{(c)} := [h_{r,n,1}^{(c)}, h_{r,n,2}^{(c)}, \dots, h_{r,n,M}^{(c)}]'$ . Next, let  $\mathbf{x}_{r,n} := [\mathbf{h}_{r,n}^{(1)'}, \mathbf{h}_{r,n}^{(2)'}, \dots, \mathbf{h}_{r,n}^{(C)'}]'$  denote the vector of covariates from all neurons at time  $n$  and trial  $r$ .

In order to represent the covariates in a more compact form, we consider the  $N \times MC$  matrix  $\mathbf{X}_r := [\mathbf{x}_{r,1}, \mathbf{x}_{r,2}, \dots, \mathbf{x}_{r,N}]'$  which contains in its rows the covariate vectors at all times  $n = 1, \dots, N$  within trial  $r$ . Finally, let  $\bar{\mathbf{X}} := [\mathbf{X}_1', \mathbf{X}_2', \dots, \mathbf{X}_R']'$

represent the matrix of all covariates with standardized columns (i.e., zero-mean columns with unit norm), capturing the covariates  $\mathbf{X}_r$  for all the trials  $r = 1, \dots, R$

the VAR model can then be expressed as:

$$\overline{\mathbf{y}}^{(c)} = \overline{\mathbf{X}}\boldsymbol{\omega}^{(c)} + \overline{\boldsymbol{\varepsilon}}^{(c)}, \quad (2)$$

where  $\overline{\boldsymbol{\varepsilon}}^{(c)} := [\boldsymbol{\varepsilon}_1^{(c)'}, \boldsymbol{\varepsilon}_2^{(c)'}, \dots, \boldsymbol{\varepsilon}_R^{(c)'}]' \sim \mathcal{N}(\mathbf{0}, \sigma^{(c)2}\mathbf{I})$  is a zero-mean Gaussian noise vector of size  $RN$  with variance  $\sigma^{(c)2}$ , and  $\boldsymbol{\omega}^{(c)}$  is a parameter vector accounting for the interactions in the network, for  $c = 1, 2, \dots, C$ .

In agreement with the parsing of the covariates in the matrix  $\overline{\mathbf{X}}$ , the parameter vector  $\boldsymbol{\omega}^{(c)} := [\boldsymbol{\omega}^{(c,1)'}, \boldsymbol{\omega}^{(c,2)'}, \dots, \boldsymbol{\omega}^{(c,C)'}]'$  in Eq. (2) is composed of a collection of cross-history dependence vectors  $\{\boldsymbol{\omega}^{(c,\tilde{c})}\}_{\tilde{c}=1:C}$ , where  $\boldsymbol{\omega}^{(c,\tilde{c})}$  represents the contribution of the history of neuron ( $\tilde{c}$ ) to the activity of neuron ( $c$ ) via the corresponding covariate vector  $\mathbf{h}_{r,n}^{(c)}$  encoded in matrix  $\overline{\mathbf{X}}$ . In particular the component  $\boldsymbol{\omega}^{(c,c)}$  is important in capturing the slow calcium florescence decay in an autoregressive fashion, and thereby excluding the transient effects of florescence decay from the GC analysis.

Next, we invoke the hypothesis of sparsity in the interactions among the neurons in the ensemble. In our model, the sparsity of the interactions can be captured through the sparsity of the parameter vector  $\boldsymbol{\omega}^{(c)}$ : when only very few components of  $\boldsymbol{\omega}^{(c)}$  are non-zero, neuron ( $c$ ) is only affected by the activity history of a few neurons in the ensemble. In addition, as the dimension of the parameter vector given by  $MC$  scales with the network size  $C$ , the hypothesis of sparsity enables the detection of

salient interactions within a large network, and thereby mitigates overfitting, especially when the observations are noisy and trials are limited in number.

Parameter Estimation: In order to define a framework for inferring a possible GC link ( $\tilde{c} \mapsto c$ ), two nested models are taken into account: 1) the VAR model in Eq. (2), where the contributing covariates from all the neurons are taken into account, referred to as the full model, and 2) the same model in which the covariates and parameters of a single neuron ( $\tilde{c}$ ) on neuron ( $c$ ),  $\tilde{c} \neq c$  are excluded, to which we refer as the reduced model. The parameters and covariates associated with the reduced model are denoted by  $\boldsymbol{\omega}^{(c \setminus \tilde{c})}$  and  $\bar{\mathbf{X}}^{\setminus \tilde{c}}$ , respectively.

The sparse parameter vector associated with either of the two models can be estimated by solving an  $\ell_1$ -regularized maximum likelihood (ML) problem for each neuron as follows:

$$\hat{\boldsymbol{\omega}} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \left( \frac{1}{2} \|\bar{\mathbf{y}}^{(c)} - \mathbf{X}\boldsymbol{\omega}\|_2^2 + \gamma \|\boldsymbol{\omega}\|_1 \right), \quad (3)$$

where  $\mathbf{X}$  takes the two values of  $\bar{\mathbf{X}}$  and  $\bar{\mathbf{X}}^{\setminus \tilde{c}}$  for the full and reduced models, respectively, the  $\ell_1$ -norm is defined as  $\|\boldsymbol{\omega}\|_1 := \sum_{m=1}^M |\omega_m|$ , and  $\gamma \geq 0$  is a regularization parameter tuning the sparsity level, which can be selected based on analytical results on  $\ell_1$ -regularized ML problems or via cross-validation. Given the parameter estimate  $\hat{\boldsymbol{\omega}}$ , the corresponding variance associated with the model can be computed as  $\hat{\sigma}^2 = \frac{1}{NR} \|\bar{\mathbf{y}} - \bar{\mathbf{X}}\hat{\boldsymbol{\omega}}\|_2^2$ .

Inference: The conventional measures of GC are based on ML estimates of the VAR parameters, and not the regularized ML as in our case. Hence, we need to

modify the GC measure and the corresponding deviance statistics, to account for the estimation bias incurred due to  $\ell_1$ -regularization. This new measure is the static VAR-based counterpart of a similar measure presented in our earlier studies in the context of dynamic sparse point process models (Sheikhhattar and Babadi, 2016) To this end, we modify the deviance difference statistic corresponding to the full and reduced models to compensate for the bias incurred due to sparse regularization. The bias can be computed for the full model as  $B^{(c)} := \mathbf{g}^{(c)'} \mathbf{H}^{(c)-1} \mathbf{g}^{(c)}$ , where  $\mathbf{g}^{(c)} := \bar{\mathbf{X}}' (\bar{\mathbf{y}}^{(c)} - \bar{\mathbf{X}} \hat{\boldsymbol{\omega}}^{(c)}) / \hat{\sigma}^{(c)2}$  and  $\mathbf{H}^{(c)} := -\bar{\mathbf{X}}' \bar{\mathbf{X}} / \hat{\sigma}^{(c)2}$  are the gradient and Hessian of the log-likelihood function for the Gaussian VAR model of Eq. (2), respectively. Similarly, the bias  $B^{(c \setminus \tilde{c})}$  for the reduced model can be computed by replacing the matrix of covariates and parameter estimate by  $\bar{\mathbf{X}}^{\setminus \tilde{c}}$  and  $\hat{\boldsymbol{\omega}}^{(c \setminus \tilde{c})}$ , respectively.

The deviance difference statistic associated with the two nested full and reduced models can be expressed as:

$$\overline{D^{(\tilde{c} \mapsto c)}} := NR \log \frac{\hat{\sigma}^{(c \setminus \tilde{c})2}}{\hat{\sigma}^{(c)2}} - B^{(\tilde{c} \mapsto c)}, \quad (4)$$

where  $B^{(\tilde{c} \mapsto c)} := B^{(c)} - B^{(c \setminus \tilde{c})}$  denotes the difference of bias terms corresponding to the full and reduced models.

We finally employ the inference framework presented in (Kim et al., 2011; Sheikhhattar and Babadi, 2016) to simultaneously test the statistical significance of all possible GC interactions and to control the FDR at a given significance level  $\alpha$ . This inference framework integrates an extension of classical results on analysis of deviance, and a multiple hypothesis testing procedure based on the Benjamini-

Yekutieli FDR control (Benjamini and Yekutieli, 2001). The weights of the detected links are further characterized using the Youden's J-statistic, which is a summary statistic for quantifying the strength of hypothesis tests. The excitatory or suppressive nature of GC links are determined by the effective sign of estimated cross-history parameters associated with shorter latencies.

To quantify the spread of the distribution of GC-link directions (Figure 2.12F), we first constructed a circular histogram of the GC-link angles which were computed from MATLAB built-in function `atan2`. Based on this histogram we used PCA to extract the long and short axes of the eclipse like distributions. Then all the original angles were projected onto the short axis and the resultant dot products (taking absolute values) were compared between ACX fields. The more the values are shifted towards 1, the larger the spread in the short axis, indicating a less 'pointy' distribution.

### Pupillometry

During 2P imaging, the arousal state of the animal was monitored through pupillometry (McGinley et al., 2015). In short, a camera was positioned around 20cm away from and towards the right eye of the head-fixed mouse. An ultraviolet LED was placed near the camera to restrict the pupil dilation to around 1/2 of the maximum dilation. The exposure time of the camera was set to 26ms and each frame was triggered by 2P "Frame Out" triggers and thus synchronized to 2P images.

To extract pupil size, each image was first cropped around the eye and MATLAB built-in function “imfindcircles” was used to determine pupil location and diameter. The pupil size over time was further smoothed with a time window of ~150ms. The onset of micro-dilation was determined by first inverting the trace (flip sign) and using MATLAB built-in function “findpeaks” with a minimum peak prominence of 10 $\mu$ m. Next we quantified the occurrence of micro-dilation before, during and after tone onset using 1-second windows, to investigate whether micro-dilation is more likely to occur following tone offset. We established confidence interval by shuffling tone onset time and counting the micro-dilation occurrence in reference to the shuffled stimulus onset. We performed such analysis for 10 sets of experiments (n=9 mice). If micro-dilation is more likely to occur during any specific time window, then the actual counts should exceed upper bound of the confidence interval. If the counts are within the confidence interval, then the occurrence of micro-dilation is equally likely to occur before, during or after tone presentation.

### Extracellular electrophysiology

We performed extracellular electrophysiology in CBA/CaJ and Thy1-GCaMP6s F1 crosses by either acutely recording from A1 neurons or chronically implanting electrodes. We used 16-channel linear arrays with 50 $\mu$ m spacing between adjacent contacts (A1x16-3mm-100-177-CM16, NeuroNexus) and a Neuralynx Cheetah system (32 channels). The acute surgery or implant surgery was similar to the cranial window implantations. In both cases, we first identified the location of A1 through

widefield imaging of GCaMP6s and we advanced the electrode at a depth of around 900 $\mu$ m, which was read out from the manipulator. Figure 2.11A-E used data from chronic implantation while Figure 2.11F-G used single unit data pooling from both acute and chronic recordings. LFP signals and single units were acquired as previously described (Petrus et al., 2014). Briefly LFPs were acquired at 30kHz (filtered between 1 and 6000Hz) and down-sampled by a factor of 100 (using MATLAB built-in function 'decimate') before analysis. To calculate local field potential (LFP) responses, we took the difference of mean LFP amplitude within a 50ms time window before and after tone onset/offset. To determine the significance, we used a paired t-test separately for each frequency and onset/offset and a significant change above baseline was considered a significant response. For spike extraction, raw headstage signal was filtered from 300Hz to 6000Hz and detected online with a threshold of 30 $\mu$ v. Single units were sorted offline using MClust-3.5 package (A. D. Redish et al.) and KluastaKwik algorithm (K. Harris). For single unit analysis, we calculated responses as the spike count change within a 500ms window before or after tone onset/offset and used paired t-test to determine the response significance for each frequency.

## **Results**

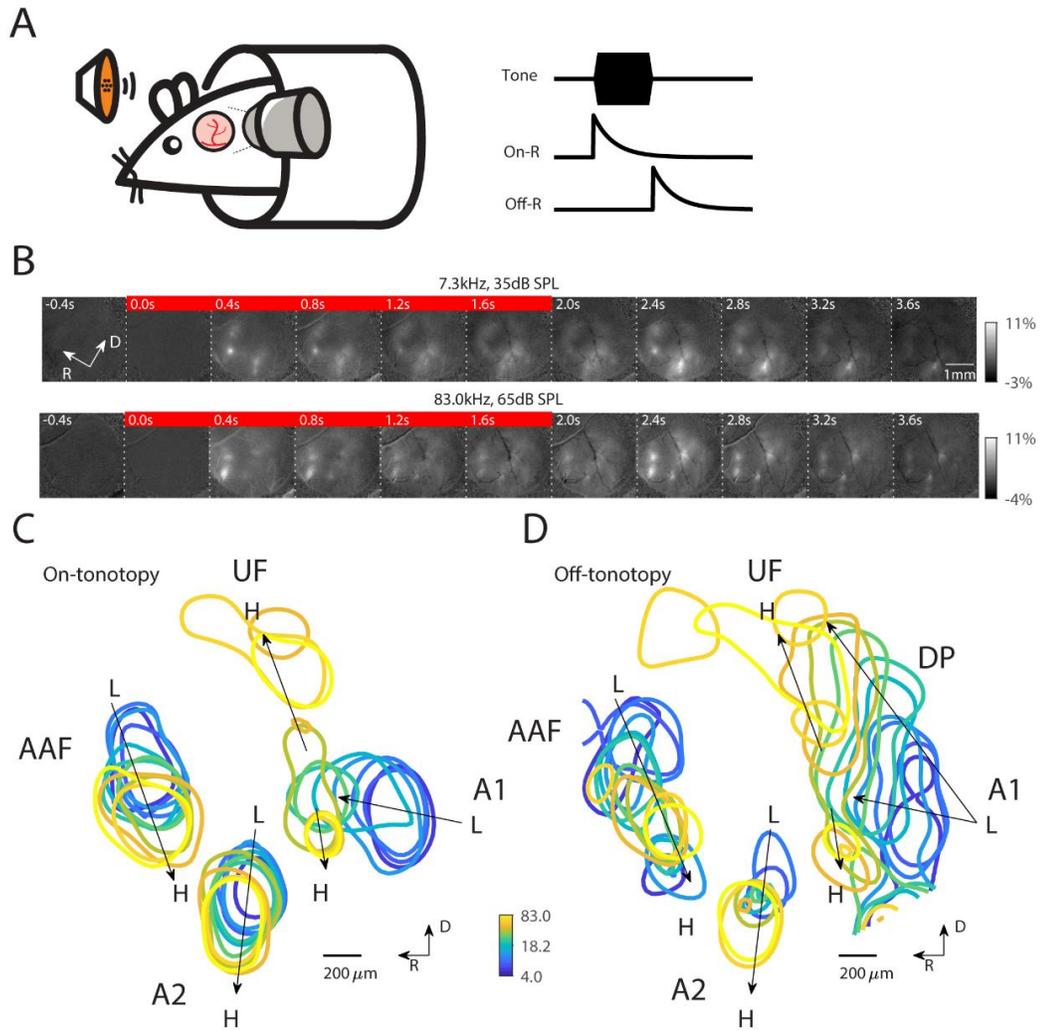
We set out to investigate the spatial organization of temporal sensitivity in mouse ACX on multiple spatial scales. Since the temporal sensitivity of ACX responses, especially Off-Rs, are sensitive to anesthesia (Recanzone, 2000; Qin et al.,

2007; Fishman and Steinschneider, 2009; Joachimsthaler et al., 2014), we performed our studies in ACX of awake animals. We used F1's of CBA/CaJ (JAX #000654) and Thy1-GCaMP6s (C57BL/6 background JAX #024275) crosses (Dana et al., 2014), which show normal adult hearing (Frisina et al., 2011) while having widespread cortical expression of GCaMP6s.

We first investigated the spatial distribution of On-R's and Off-R's on the mesoscale using widefield (WF) imaging. We imaged through a cranial window of 3mm radius over the left ACX of awake adult F1 animals (n=13) while the mice were presented with 2-second pure tones (Figure 2.1A). Tone onset resulted in spatially restricted fluorescence increases at several locations in the imaging field (Figure 2.1B, see 0.4s following tone onset, Figure 2.2A). Fluorescence increases were present in discrete locations corresponding to activations of primary as well as higher order ACX areas, putatively A1, AAF and A2 respectively. Following tone offset, we observed an additional increase of fluorescence (at 2.4s, or 0.4s after tone offset), which corresponded to an offset-response (Off-R) (Figure 2.1B). Off-R's are not due to changes in animal state after tone cessation (Figure 2.3). On-R and Off-R were also present in response to ultrasonic frequencies such as 83.0 kHz (Figure 2.1B). In both examples, the spatial pattern of On-R qualitatively matches previous results (Issa et al., 2014; Tsukano et al., 2015; Baba et al., 2016).

Varying sound frequency and level showed that both On-R and Off-R changed their response location with respect to tone frequency (Figure 2.2). We overlaid the contour of activation (see Methods) across ACX for each frequency at the respective threshold of On-R (Figure 2.1C) and Off-R (Figure 2.1D) and clear

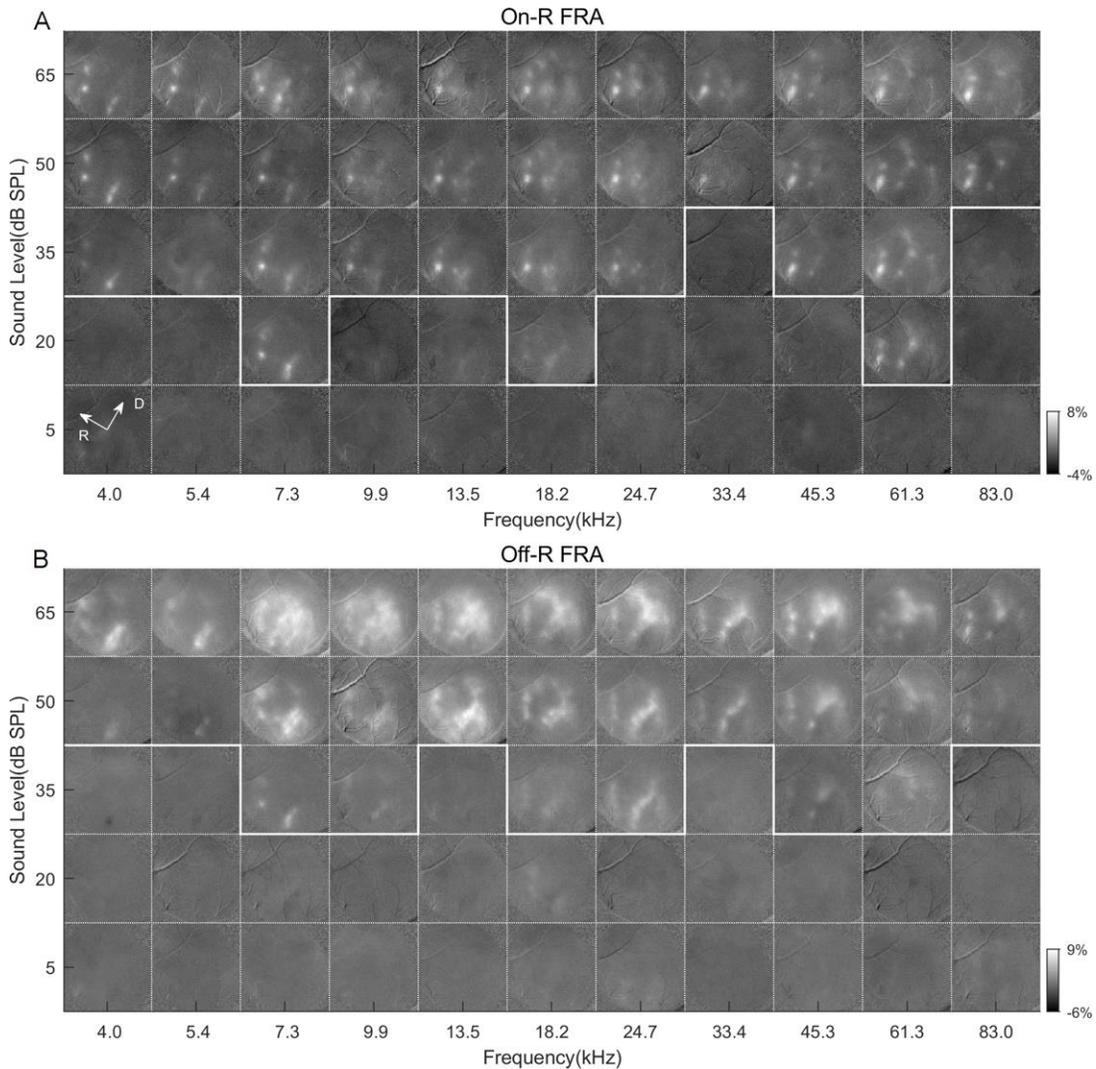
systematic changes of activated areas can be seen in multiple locations. Based on the relative positions of these gradients in the On-R we labeled areas as primary ACX (A1), Anterior Auditory Field (AAF) and A2. The detected gradients were consistent across all animals (Figure 2.4). A1 shows dual tonotopic axes: one from the caudal area towards the dorsomedial area (Ultrasonic Field or UF) and the other one reaching towards ventrolateral side (Figure 2.1C), largely consistent with prior reports (Polley et al., 2007; Issa et al., 2014; Tsukano et al., 2015) with the subtle difference that two on-tonotopic gradients in primary ACX share the low to mid frequency axis before splitting dorsally and ventrally. In addition to the expected On-R tonotopy, we observed that a tonotopic gradient is present for Off-Rs in A1, AAF and A2 in all animals (Figure 2.1D, Figure 2.4B). The Off-tonotopy gradient from A1 to UF overlapped with the On-tonotopy gradient. However, the Off-tonotopy gradient also extends dorsoposteriorly and thus covers more area dorsally than the On-tonotopy gradient. These results show that Off-Rs are present in multiple ACX areas and that Off-Rs are tonotopically organized. The differences in the tonotopic gradients between On-Rs and Off-Rs suggest that different microcircuits might underlie onset/offset processing.



**Figure 2.1 Both On-R and Off-R show global tonotopy**

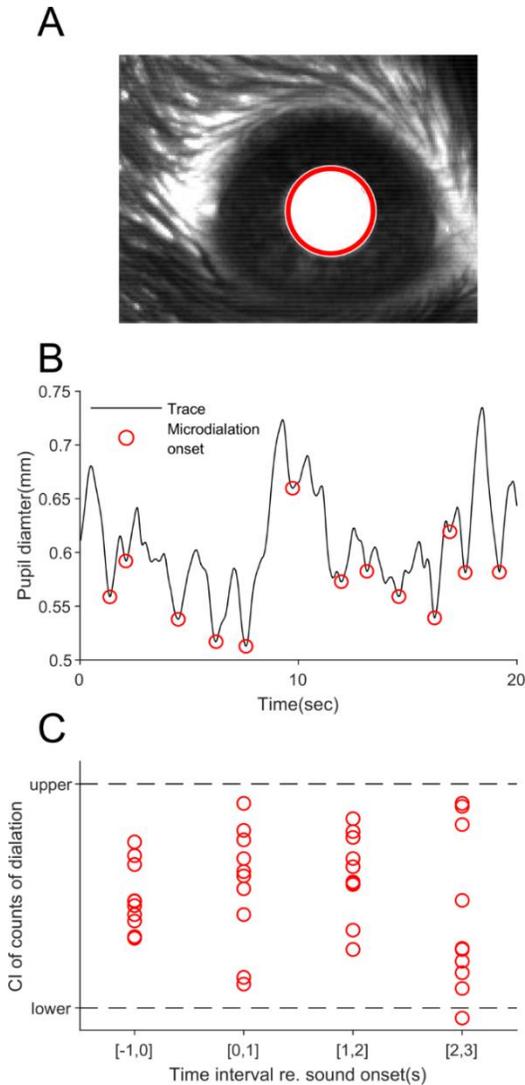
(A) Experiment paradigm: head fixed mouse passively listened to tones while left ACX was imaged for GCaMP6s. On- and Off-R are defined as increases in fluorescence following tone onset and offset, respectively. (B) Sequence of widefield images showing response to 7.3kHz tone at 35dB SPL and to 83.0kHz tone at 65dB SPL. The red bar indicates the images collected during tone presentation (0-2sec). (C) On-tonotopy showing the contour of 95 percentile of the response following tone

onset. Systematic shift of maximum activation location can be seen in A1, AAF and A2. (D) Same as in (C) but for Off-tonotopy.



**Figure 2.2 Tone onset and offset differentially activate large areas of ACX**  
On-R frequency response area (FRA). Baseline subtracted average images within 200-500ms after tone onset are plotted as a function of frequency and sound level. White solid lines show threshold at each frequency. (B) Off-R FRA. Average images

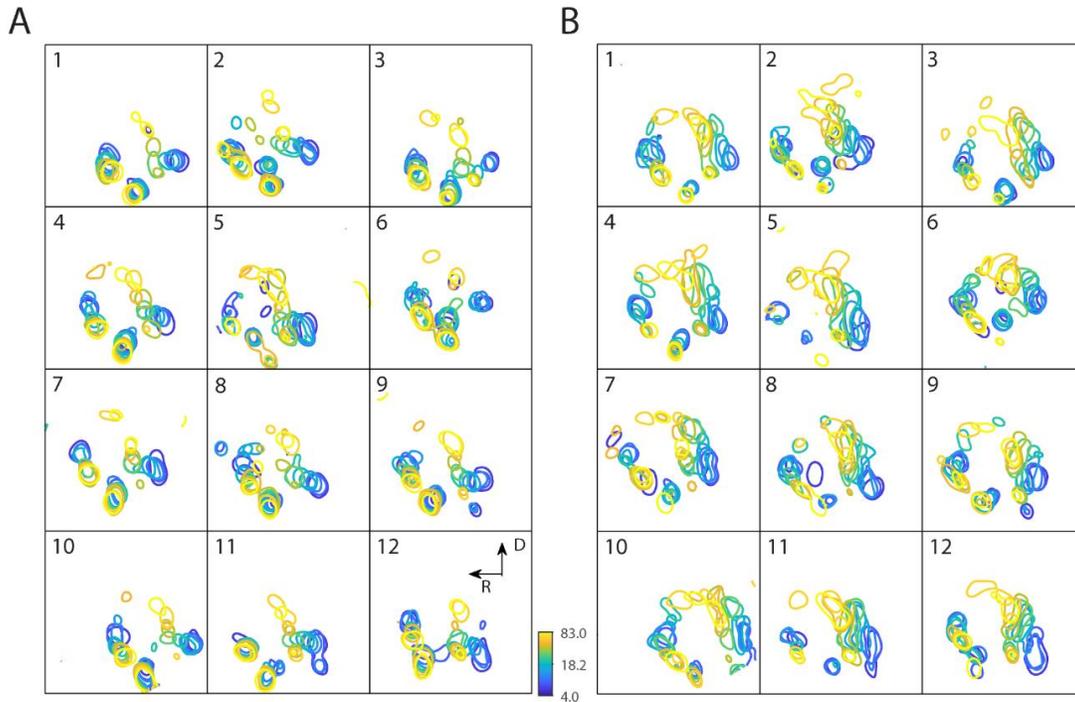
within 200-500ms after tone offset are plotted with images 0-200ms before tone offset used as baseline. Typically, Off-R had a higher threshold than on-response.



**Figure 2.3 On/Off-R is not related to the animal's arousal state**

We presented the tones to passively listening animals, and it is possible that the sudden termination of the sound could change the arousal state of the animal and that this caused Off-Rs. To investigate this possibility, we monitored the state of the animal through pupillometry (McGinley et al., 2015) and quantified occurrence of

pupil micro-dilation before, during and after tone presentation. Pupil dilation typically indicates increase in arousal state (Reimer et al., 2014). Example pupillometry image and pupil size detection (red circle). (B) Example pupil diameter trace and micro-dilation onset detected with MATLAB built-in function 'findpeak' with a minimum prominence of 10um. (C) Pupil dilation occurrence was quantified in four 1-second windows covering time before, during and after the tone presentation. The plot is normalized to the confidence interval computed through shuffling stimulus onset time. Most occurrence was well within the confidence interval, suggesting that pupil dilation happens equally likely before, during and after tone presentation and thus is unlikely to bias On-R or Off-R.



**Figure 2.4 ACX organization is stereotypical across mice**

On-tonotopy (A) and Off-tonotopy (B) of 12 other animals imaged. The colorbar indicates frequency of tones.

Distinct ACX areas show selectivity to temporal features

So far, we identified functional ACX areas based on their separate On-R and Off-R at threshold. Since these areas showed overlap, we sought to determine if ACX contained distinct functional areas based on the combination of their differential selectivity for On-R and Off-R throughout stimulus space (frequency and sound level) and if such an ACX segmentation could lead to the identification of unique ACX areas. We thus developed a method to perform unsupervised and unbiased image segmentation taking the entire temporal response of each imaged pixel into

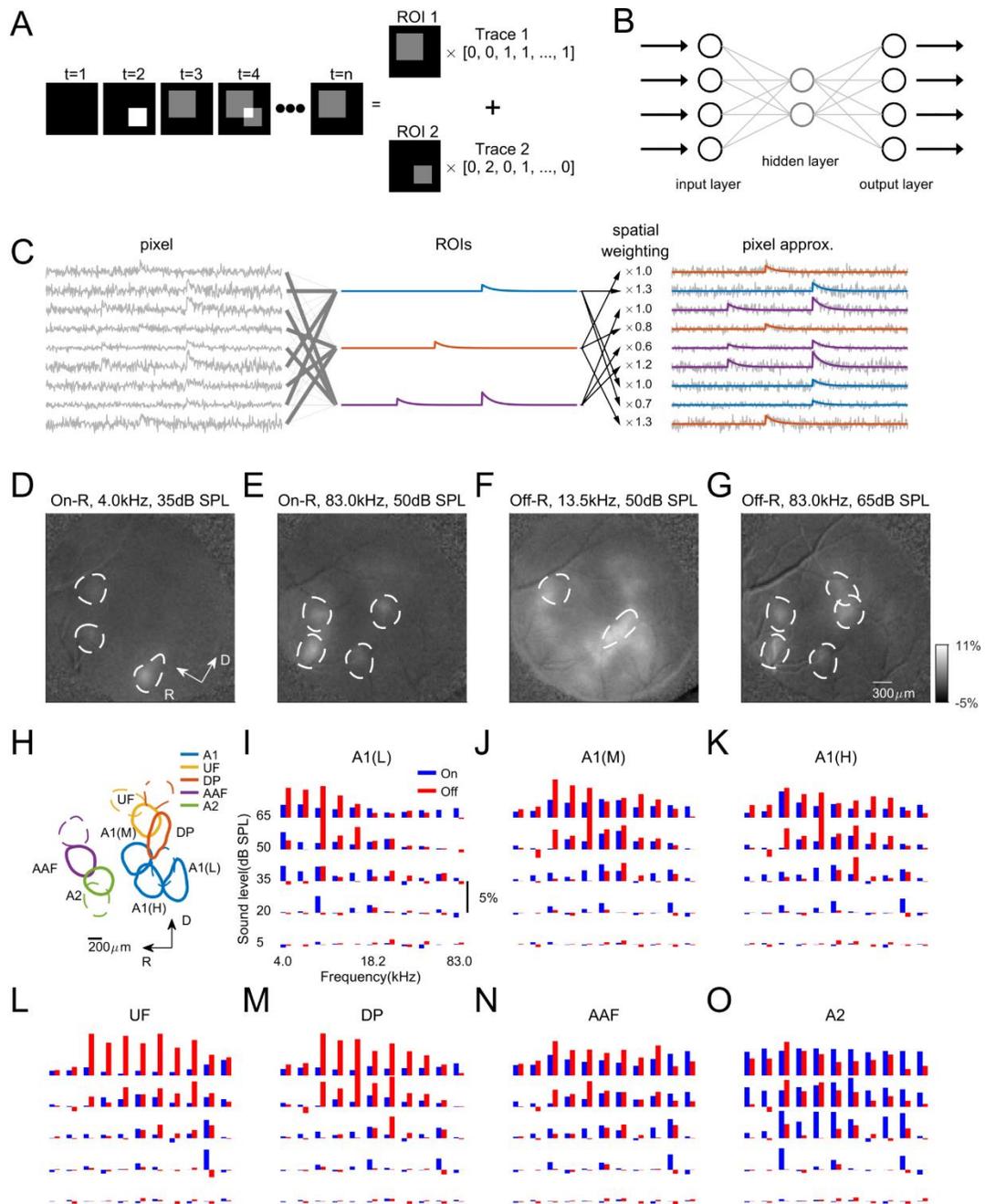
account. We expressed the sequence of activity of all imaged pixels as a linear combination of spatially distinct regions of interest (ROIs) weighted by temporal modulations (Figure 2.5A). To do so, we used an autoencoder neural network with non-negativity constraints on the weights (Whiteway and Butts, 2017). An autoencoder is a neural network with one or more hidden layers (Figure 2.5B). While the input and output layers have the same number of nodes, the autoencoder reduces the dimension in the image sequence by expressing the intensity of each pixel as the weighted sum of the activity of the hidden layer, which has a smaller dimensionality. These weights are interpreted as distinct spatial patterns of activity (or region of interest, ROI) and the activity of the hidden layer is interpreted as the temporal modulation (Figure 2.5C).

Typically, an autoencoder with around 50 ROIs achieved a good approximation of the acquired image sequence (Figure 2.6A). The resulting ROIs densely tiled the imaged area (Figure 2.6B, D) with minimum spatial overlap (Figure 2.6C). The lack of spatial overlap reflects the distinct selectivity of On/Off-R of different auditory fields, which also makes parsing auditory fields unambiguous. Additionally, the minimal overlap is likely due to our choice of the minimum number of ROIs to the desired degree of goodness of fit (Figure 2.6A). Adding ROIs would increase overlap but does not dramatically increase goodness of fit (Figure 2.6A). Next to verify our method, we compared the location of the ROIs with the evoked fluorescence increases. We found that the placement of ROIs agreed with the location of activation for both On-R (Figure 2.5D, E) and Off-R (Figure 2.5F, G), and their shapes reflected the contour of the fluorescence increases. Thus, our method can reliably identify the

regions of common activation and extract their temporal activation without prior knowledge of the spatial distribution of activity. This approach provides advantage over the traditional square/hexagonal grid segmentation as the choice of the size of the grid could be arbitrary and is likely to obscure the temporal selectivity of ROIs by grouping functionally separate cortical fields together. Moreover, while we apply this method here to segment ACX into functional fields, this method can be applied to arbitrary WF datasets for spatiotemporal analysis and image segmentation.

Identified ACX fields show distinct On/Off-R frequency response areas (FRAs) (Figure 2.5H-O) indicating that differences in the sensitivity to temporal features is a major determinant of ACX organization. The low-frequency selective A1 ROI (Figure 2.5I) shows predominant On-R and Off-R for tones of 4.0 to 7.3kHz while the mid-frequency selective A1 ROI (Figure 2.5J) shows responses mostly for frequencies of around 18.2kHz. The high-frequency selective A1 ROI (ventrolateral gradient of A1) typically have On/Off-R very similar to mid-frequency A1 due to their spatial proximity and the diffuse nature of the WF signal. However, the average On-R of high-frequency A1 ROI to 61.3kHz is larger than that of mid-frequency A1 ROI at threshold (Figure 2.5K). In contrast, UF ROI shows much higher selectivity to high frequencies (Figure 2.5L), consistent with the proposed role in processing of conspecific ultrasonic vocalizations (Stiebler et al., 1997). Figure 2.5M shows one Dorsoposterior (DP) ROI where very strong Off-R can be found. AAF ROIs (Figure 2.5N) show comparable On/Off-R while A2 ROIs (Figure 2.5O) show weaker Off-R. The spatial layout of these ROIs was also consistent across mice due to stereotypical WF response profile (Figure 2.4). These results demonstrate that ACX contains

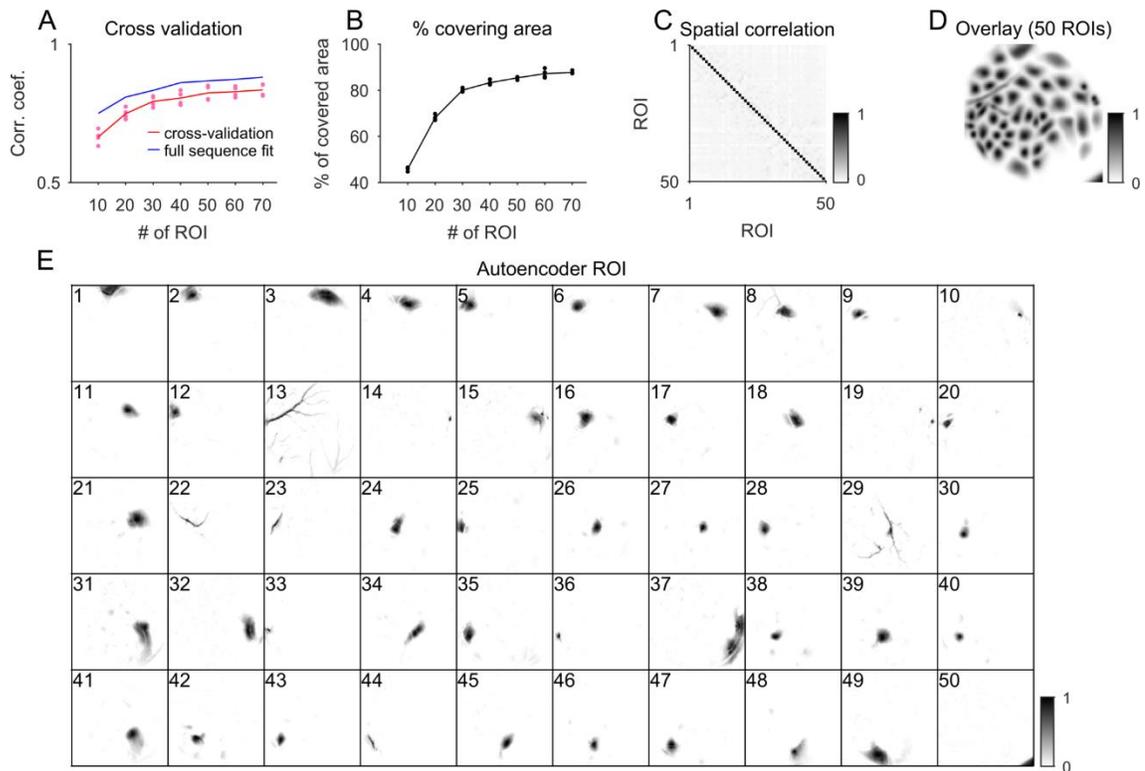
functional areas with distinct sensitivity to temporal features. Our image segmentation approach can better separate ACX into functional fields because it captures the different temporal dynamics of ACX fields.



**Figure 2.5 Widefield image segmentation using an Autoencoder reveals ACX areas with distinct On/Off selectivity.**

(A) Cartoon showing image segmentation. The example image sequence at any time

point can be expressed as the weighted summation of ROI 1 and ROI 2 by respective activity level. Our goal of image segmentation is thus to retrieve activated areas as well as their temporal activation traces. (B) Autoencoder is a neural network with one or more hidden layers between input and output layers, which have the same number of nodes. The weights between input/output layer and hidden layer are adjusted such that the output matches the input as closely as possible. The hidden layer typically has much fewer nodes than input/output layer to achieve dimension reduction. (C) Principle of fitting autoencoder ROIs. Original pixels (left) are linearly combined to produce ROIs (middle) such that each pixel in turn can be approximated (right) by the linear combination of these ROIs, while the weights are interpreted as spatial profile of the ROIs. (D-G) On- and Off-R spatial profiles overlaid with selected autoencoder ROIs to validate ROI placement. (D-G) share the same color scale. (H) Parcellation of ROIs into ACX fields. ROIs outlined in solid lines have the On/Off frequency response areas (FRAs) shown in (I-O). (I-O) On/Off-R amplitude is plotted as a function of frequency and sound level. Adjacent blue and red bars represent On/Off-R to the same frequency/sound level combination.

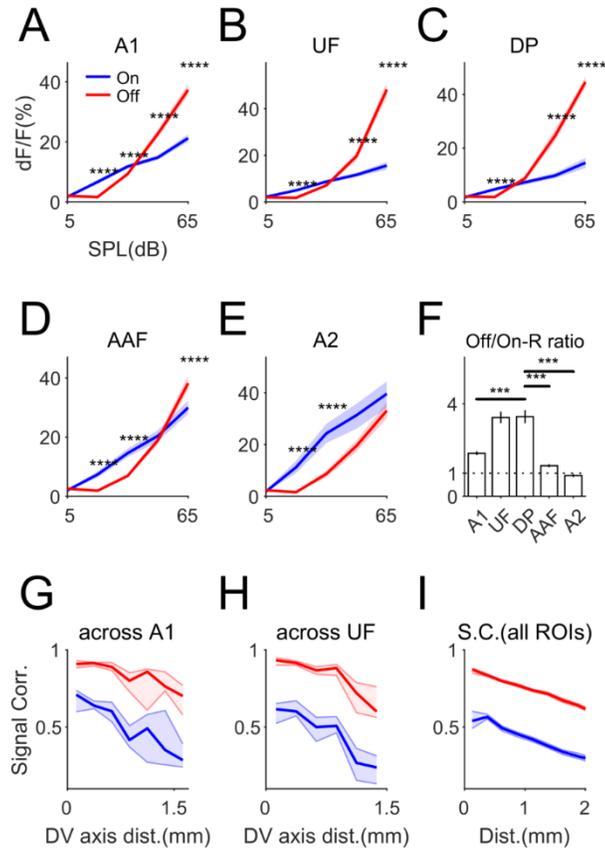


**Figure 2.6 Behavior of autoencoder fitting**

(A) Correlation with original image sequence obtained from cross-validation procedure and from full sequence fit as a function of ROI number. With 50 ROIs, the correlation value is around 0.8. (B) Areas covered by ROIs as a function of ROI number. (C) Spatial correlation of the 50 ROIs shown in (E). (D). Overlay of 50 ROIs. Note dense covering of cranial window. (E) Example of 50 ROIs shown individually.

ACX fields differ in thresholds and sound level dependence of On- and Off-Rs

So far, we observed differences between ACX areas in the selectivity to temporal stimulus dynamics. We next sought to characterize the response properties in these different ACX areas. Plotting the response amplitude in each field as sound level was varied revealed that Off-Rs have a higher threshold than On-Rs (Figure 2.7A-E). Moreover, Off-Rs can have higher amplitude than On-Rs (e.g. at 50 and 65dB SPL). Quantifying the preference of Off-R versus On-R across ACX fields at 65dB SPL showed that UF and DP are among the ACX fields showing the largest selectivity for Off-R (Figure 2.7F). These results suggest that while core ACX fields (e.g. A1, AAF) show robust responses to both tone onset and offset, higher areas away from the core fields can show dominant Off-Rs, especially at higher sound levels.



**Figure 2.7 On/Off-R show areal differences in amplitude and spatial distribution**

(A-E) Differential On-R and Off-R profile as a function of both sound level and auditory fields. On-R and Off-R profiles with respect to sound level for different auditory fields were obtained by summing over frequency in On/Off-FRAs. ‘\*\*\*’ indicates  $p < 0.001$ ; ‘\*\*\*\*’ indicates  $p < 0.0001$ . Dashed lines show 95% confidence interval around the mean. (F) Off- and on-response ratio at 65dB SPL as a function of auditory fields. Error bars show SEM. (G) On/Off-SC as function of distance along the dorsal-ventral axis, calculated among ROIs dorsal to A1 ROIs (see Methods). Off-R show higher SCs than On-Rs. (H) On/Off-SC calculated among ROIs dorsal to UF ROIs. (I) On/Off-SC among all ROIs.

Off-responsive areas are more spatially extensive than On-responsive areas

The different degrees of On-R and Off-R in the various ACX subfields suggest a different underlying circuit topology. To quantify the spatial topology on the large scale, we computed the signal correlation (SC) between individual ROIs among a dorsal-ventral slice in each ACX area (see Methods). In A1 and UF, the SCs computed from Off-Rs were significantly higher than SCs computed from On-Rs (Figure 2.7G, H). This relationship was maintained over distance suggesting that Off-Rs are more spatially extensive in the dorsal direction (Figure 2.7G, H). This pattern was also true across ACX (Figure 2.7I), suggesting that Off-Rs are more diffusively represented in all ACX areas. These results are consistent with our observation that dorsal ACX areas especially UF and DP are dominantly involved in encoding tone

offset (Figure 2.7F). Together, the areal differences in the tonotopic gradients (Figure 2.1) and the differences in SC between On-Rs and Off-Rs suggest that different intrinsic and ascending microcircuits within each area underlie the regional differences in onset/offset processing.

#### Neural populations in ACX areas differ in their selectivity to sound onset or offset

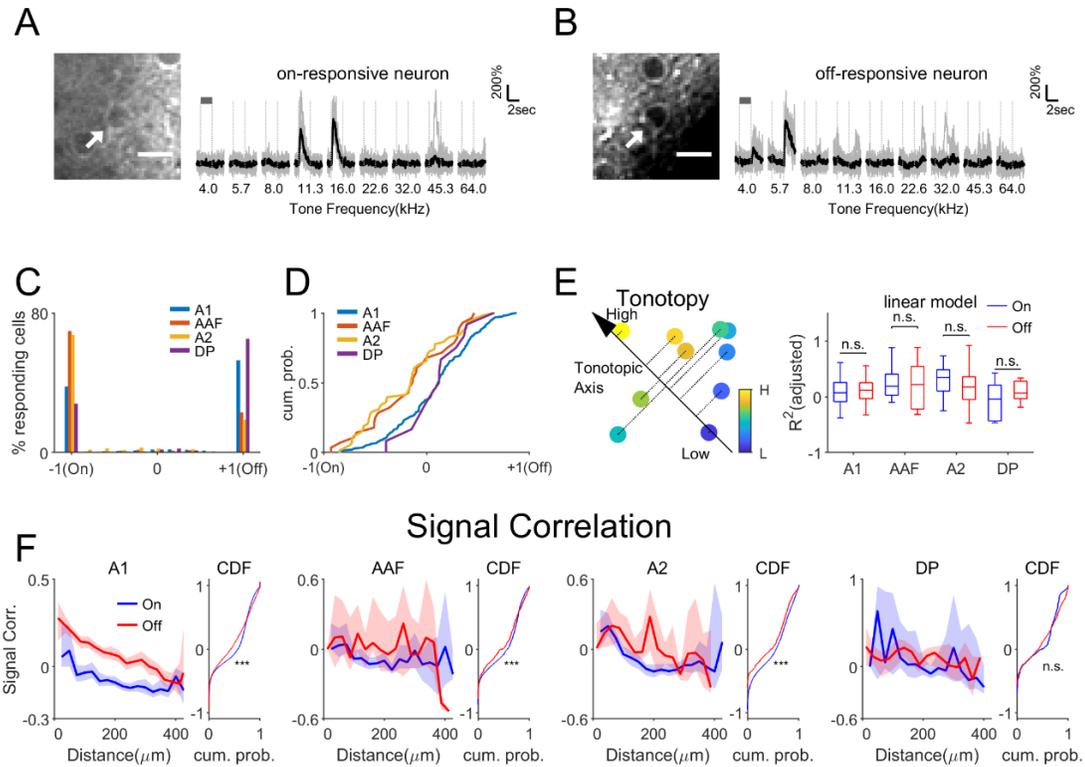
To investigate areal differences in the processing of tone onset and offset, we sought to analyze local microcircuits on the cellular level. We thus assessed the temporal stimulus preferences of single neurons in the different ACX areas using in vivo 2-photon imaging. We imaged cells in four auditory fields showing distinct temporal signatures (n=32 mice; A1: 67 field of views (FOVs), 19366 cells; AAF: 24 FOVs, 5425 cells; A2: 20 FOVs, 5918 cells; DP: 8 FOVs, 2573 cells). Cells in all ACX areas could show responses that are locked to either tone onset and offset respectively (Figure 2.8A, B, Figure 2.9). The cells showing On-R in ACX were sparse (A1, 5.05%  $\pm$  2.89%; AAF, 5.36%  $\pm$  2.58%; A2, 5.83%  $\pm$  4.53%; DP, 2.23%  $\pm$  1.29%, among all neurons imaged), while the same was true of Off-R (A1, 6.62%  $\pm$  4.34%; AAF, 2.14%  $\pm$  1.83%; A2, 2.28%  $\pm$  2.24%; DP, 4.64%  $\pm$  2.42%, among all neurons imaged). These results are consistent with electrophysiological studies reporting a sparse representation of sound in ACX (Hromádka et al., 2008). A limited number of neurons showed both On-Rs and Off-Rs (A1, 0.98%  $\pm$  0.90%; AAF, 0.54%  $\pm$  0.54%; A2, 0.95%  $\pm$  1.31%; DP 0.43%  $\pm$  0.57%, Figure 2.10A, among all neurons imaged), suggesting that most L2/3 neurons are either On- or Off-responsive.

We quantified the selectivity of On/Off-Rs by computing the On/Off-R Bias Index ( $OBI = \frac{Off - On}{Off + On}$ ) (Figure 2.8C). Most of the OBI values were -1 or 1, consistent with most neurons being exclusively on-responsive (On-only) or off-responsive (Off-only). In A1 and DP, Off-only neurons (53% and 65%, among all responding neurons) outnumber On-only neurons (38% and 28%, among all responding neurons), while in A2 and AAF reverse is true (67% and 70% vs. 23% and 19%, among all responding neurons). Neurons showing both On-R and Off-R constitute around 10% of all responding neurons and are more biased to Off-R in A1 and DP than in AAF and A2 (Figure 2.8D). We also confirmed these results in a separate analysis (Figure 2.10). Together these results show that ACX areas differ in both the number of On/Off-only neurons as well as in the On/Off-selectivity of individual neurons. Thus, the ACX areas are defined by the underlying cellular and population representation of tone onset/offset and the cellular response amplitudes. We next investigated the functional representation of tone onset and offset in these neuronal populations.

To further confirm our results and to get a view across layers, we implanted 16-channel linear multielectrode arrays into A1 of CBA $\times$ Thy1-GCaMP6s mice, spanning a cortical depth of 800 $\mu$ m. We first analyzed the local field potential (LFP) which reflects the combination of local neuronal activity and afferent input into A1 (Katzner et al., 2009; Liu et al., 2015; Herreras, 2016). First, we observed that more tone frequencies evoke Off-R compared to On-R (Figure 2.11A-C), consistent with the widespread nature of Off-R (Figure 2.7). Secondly, we calculated the OBI for all electrode contacts and the distribution of OBI was significantly shifted towards Off-R

(Figure 2.11D). These results confirm that Off-R evokes a wide activation in A1 and that A1 responses are biased towards Off-R.

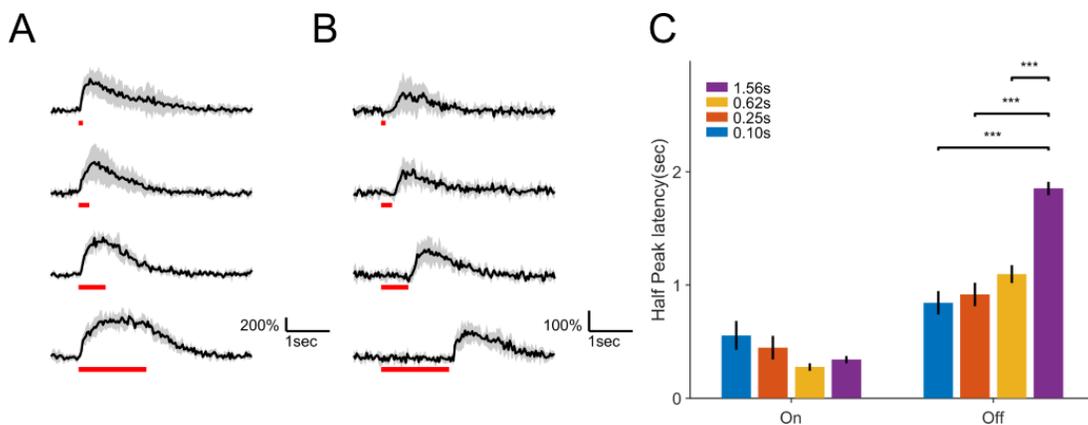
Previous electrophysiology studies have reported a higher proportion of neurons showing both On- and Off-R than our imaging results (Qin et al., 2007; Tian et al., 2013; Joachimsthaler et al., 2014). To identify potential sources for this discrepancy we recorded single units (n=220) from A1 of awake CBA $\times$ Thy1-GCaMP6s animals and analyzed their On/Off-R (Figure 2.11F-H). 200/220 units (91%) were deemed responsive to either tone onset or offset. Among the responsive units, 26% had only On-R, and 57% had both On- and Off-R, and 7% had only Off-R. We classified neurons based on their spike shape (wide vs. narrow) reflecting putative excitatory and inhibitory units and analyzed their OBI. We found that both classes showed similar OBI distribution (Figure 2.11G). Since our linear arrays allowed sampling of neurons across a large depth range we next compared OBI across depth. We found that OBI was dependent on the depth of the single units, with deeper layer units more biased to On-R (Figure 2.11H). Together these results suggest that A1 contains both On/Off-only neurons and that there is a depth-dependent distribution of these neurons consistent with sublaminal circuit differences in L2/3 (Meng et al., 2017).



**Figure 2.8 L2/3 neurons show distinct On-R and Off-R and are differentially distributed across ACX areas**

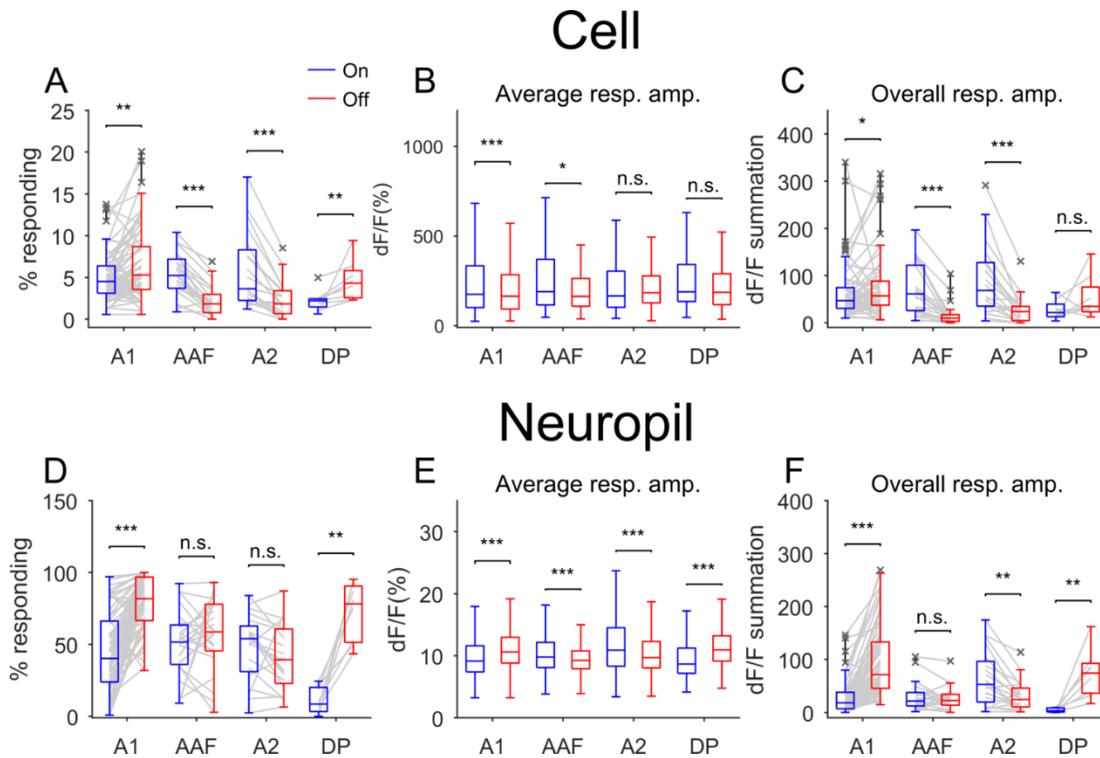
(A) An example on-responsive neuron (arrow). Vertical dotted lines indicate tone onset and offset respectively. Gray horizontal bar indicates tone duration. (B) An example off-responsive neuron. Scalebar: 10 $\mu$ m. (C) Histogram of cellular OBI values as a function of ACX fields.  $OBI = (Off-R - On-R) / (Off-R + On-R)$  while (D) shows cumulative distribution function (CDF) of values other than -1 and 1. Wilcoxon rank sum test, A1 vs AAF,  $z=2.77$ ,  $p=0.0056$ ; A1 vs A2,  $z=4.41$ ,  $p=1.02 \times 10^{-5}$ ; DP vs AAF,  $z=1.93$ ,  $p=0.053$ ; DP vs A2,  $z=2.49$ ,  $p=0.013$ . (E) Left: cartoon showing a linear model to predict the BF of On-R and Off-R with respect to the cells' spatial locations across all auditory fields. A direction is searched onto which the projection of the cell's coordinate best explains the cell's BF. Right: Goodness of fit of On-R and Off-R in cells of different auditory fields. (F)

Relationship between On- and Off-SCs and pairwise distance on the neuronal level for different auditory fields. Solid lines show median while the shading indicates the 95% confidence interval. The flanking panel shows CDF of on-SC and off-SC not regarding distance. ‘\*\*\*’ indicates  $p < 0.001$ . ‘\*\*’ indicates  $p < 0.01$ . A1, rank sum test,  $z = -13.6$ ,  $p = 4.33 \times 10^{-42}$ ; AAF,  $z = -3.52$ ,  $p = 4.30 \times 10^{-4}$ ; A2,  $z = -8.73$ ,  $p = 2.07 \times 10^{-18}$ ; DP,  $z = -2.93$ ,  $p = 3.4 \times 10^{-3}$ .



**Figure 2.9 Neuronal Off-R is time locked to tone offset.**

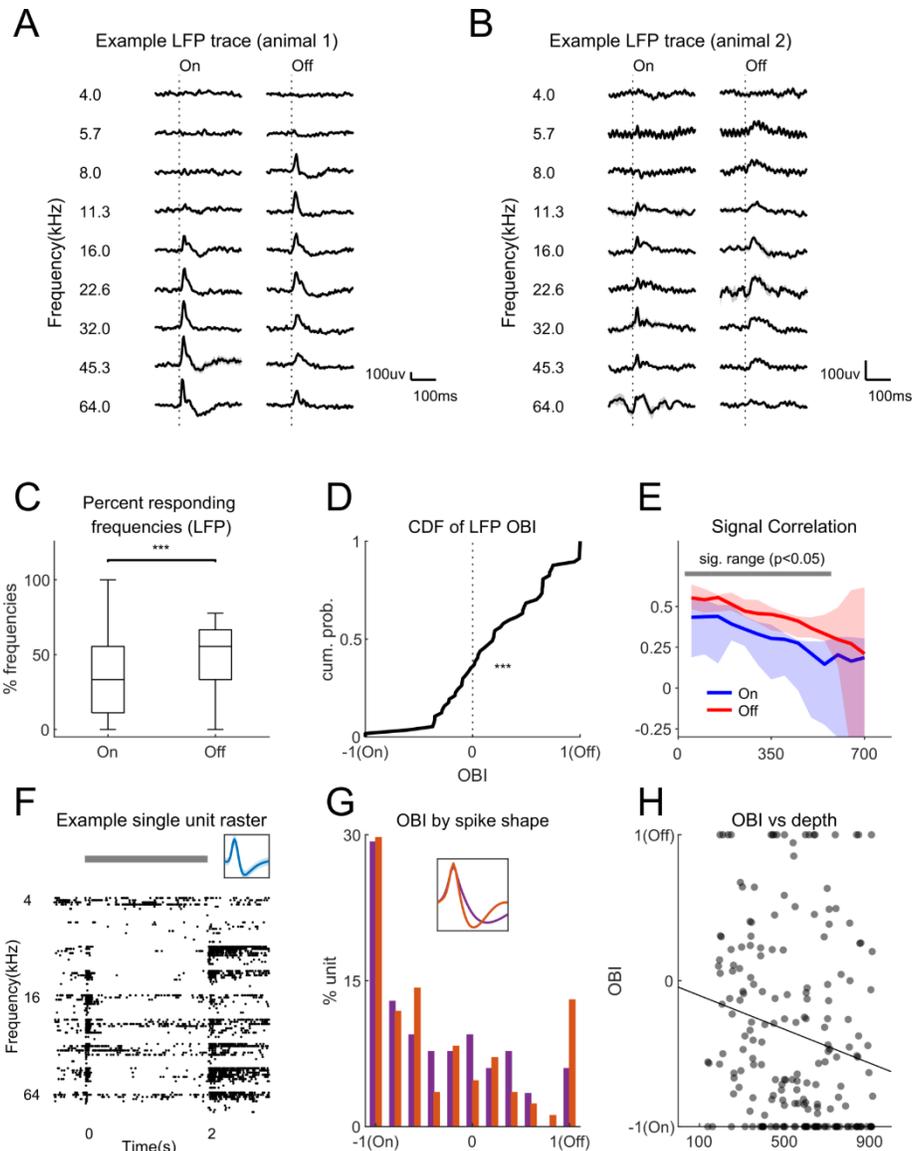
Tones of different durations were presented to the same group of neurons and the offset-responses were time locked to tone offset and thus were not ‘delayed’ response happening at specific time point after tone onset. (A) Example On-responsive neuron responding to 0.10, 0.25, 0.62, 1.56 second tones. The response is time locked to tone onset. Red lines represent tone durations. (B) Example Off-responsive neuron whose responses were time locked to tone offset. (C) Time to reach half of maximum activation was quantified for On- and Off-R respectively. For On-R, the latency is not a function of tone duration while Off-R latency shifts systematically as a function of tone duration. ‘\*\*\*’ indicates  $p < 0.001$ .



**Figure 2.10 Cellular and neuropil signals from different ACX fields show similar On/Off-R selectivity to that seen in widefield imaging**

(A) Percentage of on-responding and off-responding neurons as a function auditory field. Each gray line represents a field of view. In A1 and DP, more cells responded to tone offset than to onset while the opposite was seen in AAF and A2. Paired t-test: A1,  $6.62\% \pm 4.34\%$  vs.  $5.05\% \pm 2.89\%$ ,  $t(66)=-3.58$ ,  $p=0.0007$ ; AAF,  $2.14\% \pm 1.83\%$  vs.  $5.36\% \pm 2.58$ ,  $t(23) = 8.57$ ,  $p=1.29 \times 10^{-8}$ ; A2,  $2.28\% \pm 2.24\%$  vs.  $5.83\% \pm 4.53\%$ ,  $t(19)=5.70$ ,  $p=1.7 \times 10^{-5}$ ; DP:  $4.64\% \pm 2.42\%$  vs.  $2.23\% \pm 1.29\%$ ,  $t(7)=-2.50$ ,  $p=0.041$ . Few neurons showed both On-Rs and Off-Rs (A1,  $0.98\% \pm 0.90\%$ ; AAF,  $0.54\% \pm 0.54\%$ ; A2,  $0.95\% \pm 1.31\%$ ; DP  $0.43\% \pm 0.57\%$ ). (B) Average On/Off-R amplitude of individual neurons was pooled over field of views and plotted as a function of auditory fields. The relative response amplitude for On-R and Off-R could

vary across areas. Wilcoxon rank sum test, A1,  $z=3.34$ ,  $p=8\times 10^{-4}$ ; AAF,  $z=2.34$ ,  $p=0.0174$ ; A2,  $z=-1.21$ ,  $p=0.23$ ; DP,  $z=1.14$ ,  $p=0.26$ ). (C) Overall cellular On/Off-R amplitude (summing over all significant On-R or Off-Rs from all frequencies and all cells within a field of view) as a function of auditory field. The relative population composition and response amplitude well predicted the areal On/Off preferences on the large-scale (widefield). Wilcoxon signed rank test: A1,  $z=-2.30$ ,  $p=0.0214$ ; AAF,  $z=4.07$ ,  $p=4.6\times 10^{-5}$ ; A2,  $z=3.73$ ,  $p=1.89\times 10^{-4}$ ; DP,  $p=0.19$ . (D-F) Same as in (A-C) but plotted for neuropil. (D) Paired t-test: A1,  $t(66)=-10.4$ ,  $p=1.23\times 10^{-15}$ ; AAF:  $t(23)=-1.12$ ,  $p=0.27$ ; A2:  $t(19)=1.51$ ,  $p=0.15$ ; DP,  $t(7)=-6.99$ ,  $p=2.14\times 10^{-4}$ . (E) Wilcoxon rank sum test, A1,  $z=-33$ ,  $p=1.62\times 10^{-239}$ ; AAF:  $z=9.55$ ,  $p=1.32\times 10^{-21}$ ; A2:  $z=9.67$ ,  $p=4.21\times 10^{-22}$ ; DP,  $z=-10.1$ ,  $p=5.09\times 10^{-24}$ . (F) Wilcoxon rank sum test, A1,  $z=-7.61$ ,  $p=2.66\times 10^{-14}$ ; AAF:  $z=0.01$ ,  $p=0.99$ ; A2:  $z=2.07$ ,  $p=0.038$ ; DP,  $p=1.55\times 10^{-4}$ .



**Figure 2.11 Local field potential (LFP) in A1 shows bias towards Off-R in strength and spatial spread.**

(A, B) Two example electrode contacts showing On- and Off-R to different tone frequencies. Dotted vertical lines represent tone onset and offset respectively. (C) Percentage of frequencies evoking significant On- or Off-R shows an Off-R bias in LFP responses (Wilcoxon signed-rank test,  $p=9.5 \times 10^{-6}$ ). (D) LFP Off-R Bias Index (OBI) distribution is significantly shifted towards 1 (t-test,  $p=3.6 \times 10^{-6}$ ), suggesting an overall larger Off-R. (E) Signal correlation (SC) was computed for On- and Off-R

separately among all electrode contacts and Off-R SC was larger than On-R SC over distance (Wilcoxon rank sum test, 50 $\mu$ m: p=0.040; 100 $\mu$ m: p=0.036; 150 $\mu$ m: p=0.038; 200 $\mu$ m: p=0.041; 250 $\mu$ m: p=0.023; 300 $\mu$ m: p=0.011; 350 $\mu$ m: p=0.008; 400 $\mu$ m: p=0.003; 450 $\mu$ m: p=0.003; 500 $\mu$ m: p=0.004; 550 $\mu$ m: p=0.034). (F) Example raster plot of one single unit responding to both tone onset and offset. The inset shows the spike waveform. (G) Histogram showing OBI separately for wide and narrow spike waveforms (putative excitatory and inhibitory neurons). Both groups show prominent On-R and the majority of the units had both On- and Off-R. (H) OBI of single units depended on cortical depth. A linear model was used to quantify the depth dependence of single unit OBI and deeper cortical layers were more biased towards On-R (linear fit:  $y=5.8\times 10^{-4}x-0.043$ ,  $p=7.3\times 10^{-3}$ , adjusted  $R^2=0.031$ ).

#### Local tonotopy is heterogeneous for both On-R and Off-R in all areas

On the large-scale both On-R and Off-R show tonotopic organization (Figure 2.1, 2.5), while on the cellular scale tonotopy is heterogeneous in A1 of anesthetized mice (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Kanold et al., 2014). We tested if Off-R exhibited local tonotopy and if On-R and Off-R cells in ACX areas differed in local heterogeneity of frequency preference. We compared the degree to which On-R and Off-R are locally tonotopically organized by analyzing separate linear models between best frequency and spatial location of the cell (Figure 2.8E). Similar to prior studies in anesthetized mice (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Maor et al., 2016) we here find a lack of local tonotopy of frequency

selectivity in awake mice as the goodness of fit ( $R^2$ ) is low in general. Moreover, the models showed a similar  $R^2$  for On-R or Off-R across ACX areas, suggesting that the local heterogeneity of frequency selectivity between On-R and Off-R is similar within and across fields in mouse ACX.

#### ACX areas differ in the spatial pattern of neuronal correlated On-R and Off-R activity

Our results so far indicate regional differences in cellular selectivity. To gain insight into the spatial distribution of such circuits we calculated pairwise SCs which are reflective of shared inputs (Shadlen and Newsome, 1998). The dependence of the pairwise SCs on the spatial relationship between neurons is reflective of the underlying circuit topology. In A1 pairwise SCs for On-Rs are highest for nearby neurons and decrease with distance, consistent with prior results in anesthetized mice (Figure 2.8F) (Winkowski and Kanold, 2013). Such a decrease is also present in A2 while DP shows a patchy distribution of SCs for On-Rs. A weak SC gradient is present in AAF. Off-SC are larger than On-SC in most areas except for DP. In A1 these differences between On-SC and Off-SC are widespread, while such differences are present in patchy areas in AAF (~150-175 $\mu$ m) and A2 (~50-275 $\mu$ m). These results show that Off-R neuronal populations are more widespread, which could be due to a difference in the underlying intrinsic circuits or due to the spatial distribution of ascending input. We also validated this result by computing the SC among chronically implanted linear electrode contacts and a similar correlation structure was seen where Off-R SC was higher than On-R SC over distance (Figure 2.11E). This

result further suggests that Off-R is not only more widespread among different cortical columns but also along cortical depth.

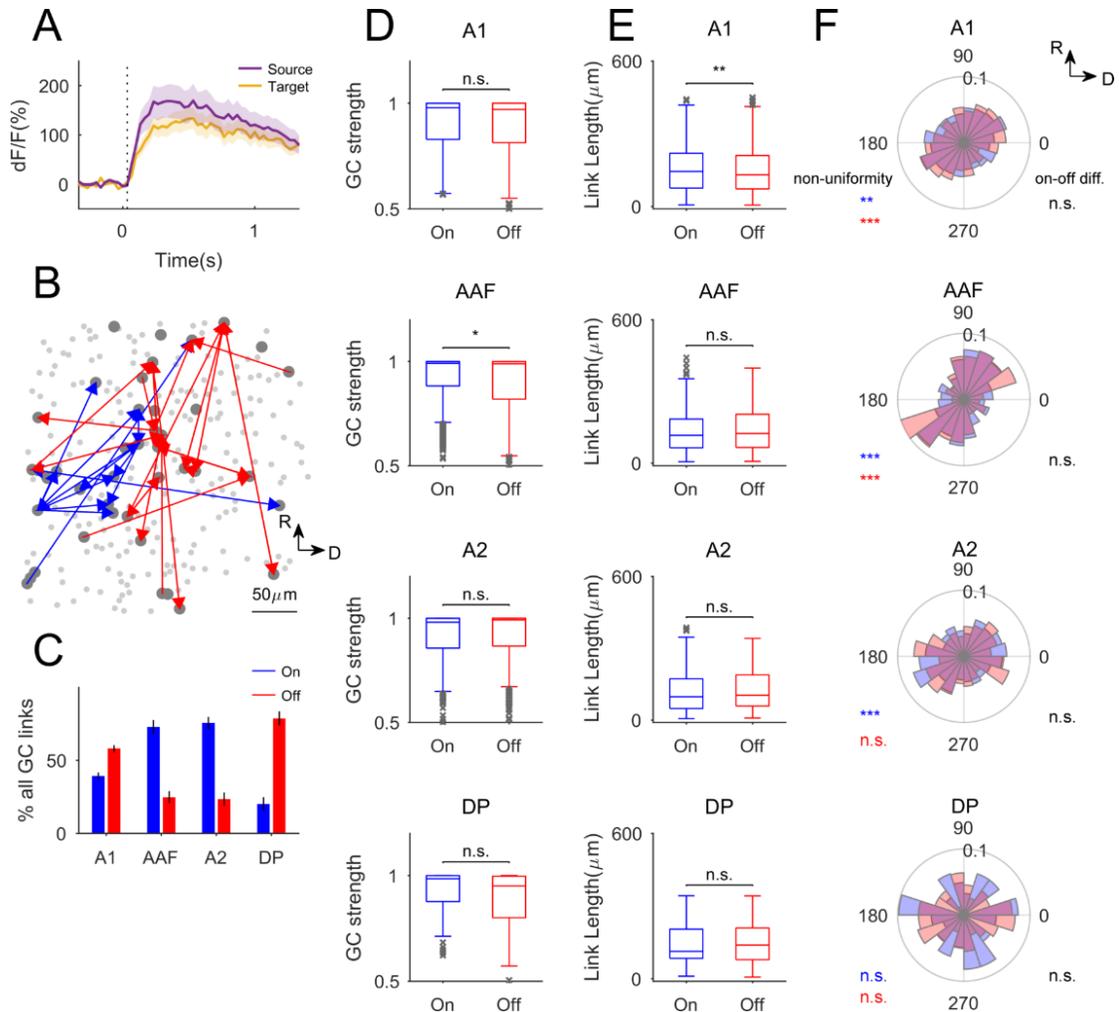
### Granger Causality analysis reveals areal differences in functional On/Off networks

The areal differences in pairwise SCs suggest different underlying neuronal networks. We thus sought to identify the functional networks in the different ACX areas. To do so, we performed Granger Causality (GC) analysis separately among On- and Off-responsive populations (Granger, 1969; Oya et al., 2007; Friston et al., 2013; Sheikhattar and Babadi, 2016; Francis et al., 2018; Sheikhattar et al., 2018). GC analysis provides a statistical/data-driven framework for inferring causal interactions between neurons from the neural data, by statistically testing if a neuron's activity can be predicted by the recent activity history of other neurons, and thus allows us to uncover functional networks (Granger, 1969; Sheikhattar and Babadi, 2016; Francis et al., 2018). The causal interactions (GC-links) can take effective positive or negative signs reflecting neuronal activities that are both increasing/decreasing or are of opposite directions (Francis et al., 2018). Our calcium indicator is expressed in excitatory neurons and thus we focused on positive GC links. An example of two GC linked neurons is shown in Figure 2.12A. Note that that the source trace preceded the target trace. Figure 2.12B shows one example field of view with most significant GC links labeled. We next proceeded to quantify the number, strength, length and directionality of these GC links. In A1 and DP, Off-GC-links outnumbered On-GC-links, while the opposite was true in AAF and A2 (Figure

2.12C). These areal differences in GC link number indicate higher respective interconnectivity and are consistent with the areal differences in the relative numbers of On-R and Off-R neurons. In contrast to these areal differences in link number, GC link strength (J-statistics) largely showed no difference except for AAF (Figure 2.12D), suggesting both On-R and Off-R networks are strongly functionally connected. Since the majority of cells are either On-R or Off-R these results indicate that ACX areas contain separate interdigitated On-R and Off-R networks.

Since 2-photon imaging allows us to identify the spatial location of imaged cells we next extracted the relative spatial positions of GC-linked cells and calculated the spatial properties of GC linked On and Off networks. First, we compared the length of GC-links in different ACX fields. Off-GC-links tend to have a larger number of shorter links in A1 (Figure 2.12E), suggesting that Off GC networks more densely cover the neural populations in A1 and are more spatially clustered. However, no difference is seen in other ACX fields (Figure 2.12E). Since ACX areas show tonotopic maps on a global scale, we next investigated if GC links also show a direction preference. We thus compared the distribution of On/Off GC-link directions. Except for DP and A2 Off-R, the distributions of the GC-link directions significantly deviate from uniform distributions (Figure 2.12F). In A1, AAF and A2, the ellipse like distributions have the long axis, reflecting a spatial bias of the cell pair interactions, roughly in parallel to the tonotopic axes. Thus, here we show that although local cellular populations lack precise tonotopic organizations when probed based on their frequency preference, there are nevertheless regularities in their functional spatial connectivity whose patterns are closely related to the tonotopic axis.

Moreover, we found no difference in the On/Off GC-link direction distribution (Figure 2.12F). This indicates that although On/Off-R populations are largely non-overlapping, they are spatially intermingled and parallel, consistent with the ‘salt-and-pepper’ structure in L2/3 of mouse ACX (Bandyopadhyay et al., 2010; Rothschild et al., 2010). We did notice that the distribution of GC-link directions in AAF tends to be more ‘pointy’ or narrower than in A1 or A2. We combined both On and Off GC-links and compared the spread in the direction of the short axis of the eclipse like distributions (see Methods). Indeed, AAF GC links were more narrowly distributed than in A1 ( $p=0.033$ ) and the difference between AAF and A2 was close to significance ( $p=0.058$ ). Thus, the spatial topology of the intrinsic functional architecture of L2/3 in different ACX fields differs. Moreover, our observation of these patterns indicates that spatially specific On/Off sub-circuits exist and that they are superimposed on the locally heterogeneous tonotopic map.



**Figure 2.12 GC analysis reveals distinct On/Off sub-networks**

(A) Fluorescence time course of GC-linked cells. (B) GC links in an example field.

Blue: On GC-links. Red: Off GC-links. Only GC links with J-statistics > 0.95 are

shown for clarity. (C) Proportion of GC-links as a function of ACX fields (false

discovery rate: 0.001). More Off GC-links are observed in A1 and DP (Wilcoxon

rank sum test, A1 on vs off:  $p=2.53 \times 10^{-7}$ ,  $z=-5.16$ ; DP,  $p=1.55 \times 10^{-4}$ ). In AAF and A2

On GC-links are more abundant (AAF  $p=5.44 \times 10^{-7}$ ,  $z=4.55$ ; A2  $p=3.32 \times 10^{-6}$ ,  $z=4.65$ ).

(D) J-statistics, a measure of GC-link strength as a function of ACX fields. Only AAF

shows a slight higher On GC-link strength (Wilcoxon rank sum test,  $p=0.0175$ ,  $z=2.38$ ). (E) GC-link length as a function of ACX fields. A1 contains shorter Off GC-links (Wilcoxon rank sum test,  $p=0.0022$ ,  $z=3.06$ ) (F) Distribution of direction of GC-links as a function of ACX fields. The non-uniformity of the distributions was tested using Chi-square goodness-of-fit test. A1, on:  $p=0.043$ , off:  $p=1.08 \times 10^{-22}$ ; AAF, on:  $p=1.48 \times 10^{-7}$ , off:  $p=7.77 \times 10^{-4}$ ; A2, on:  $8.15 \times 10^{-4}$ , off:  $p=0.42$ ; DP, on:  $p=0.89$ , off:  $p=0.17$ . On/Off distribution difference: Two-sample Kolmogorov-Smirnov test, A1:  $p=0.065$ ; AAF:  $p=0.82$ ; A2:  $p=0.$ ; DP:  $p=0.85$ .

#### The On/Off responsivity of MGB terminals determines areal responses

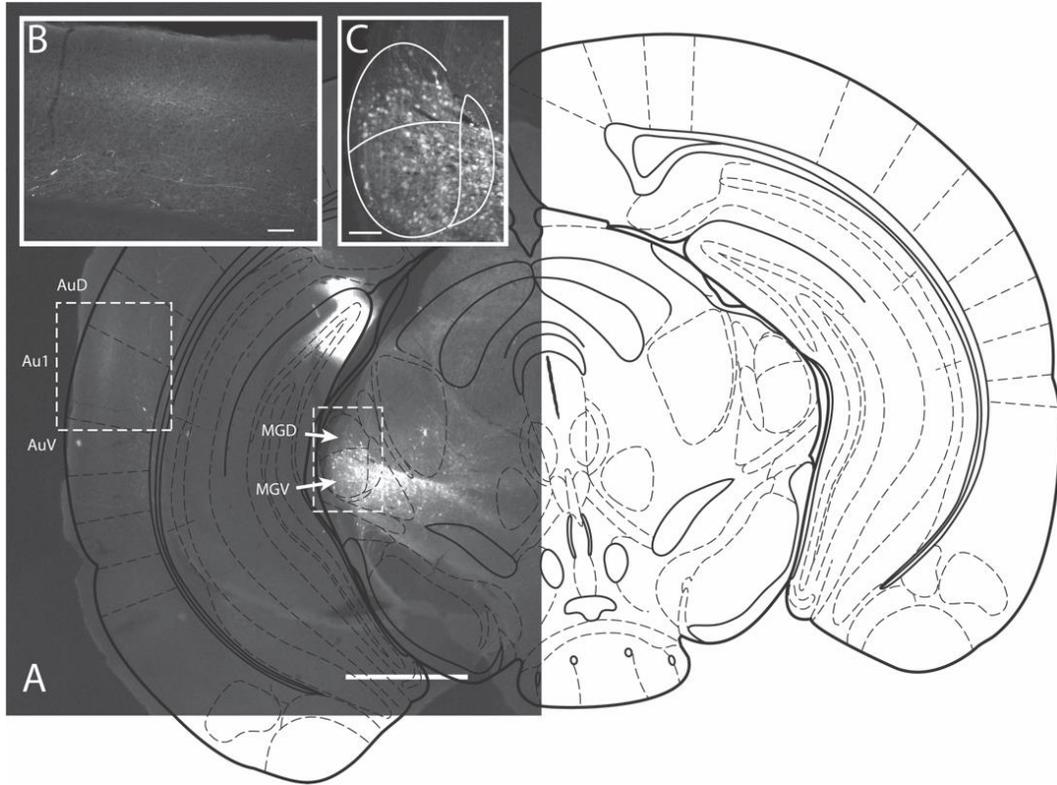
So far, our results indicate that ACX contains distinct functional areas that are defined by differing cellular selectivity and intrinsic connectivity. Since ascending inputs to ACX neurons determine the initial cellular selectivity to sound dynamics we next examined how the cellular on/off selectivity emerged from the inputs to ACX. The main ascending inputs to ACX are provided by medial geniculate body (MGB) axons which terminate on excitatory neurons ranging from L2/3 to L6 with the strongest input in L4 (Ji et al., 2015). In vivo patch clamp recordings in A1 have shown that non-overlapping sets of synapses can give rise to On- and Off-Rs (Scholl et al., 2010). Since different ACX areas receive dominant input from different subdivisions of the MGB we speculate that these sets of synapses reflect separate pathways from the MGB. To test this hypothesis, we injected virus expressing GCaMP6s into the MGB (see methods,  $n=7$  mice) and imaged the labelled terminals

in A1 (20 FOVs) after 3-4 weeks to allow expression (Figure 2.13). We focused on A1 because of its distinct difference in On-R and Off-Rs and because prior in vivo patch clamp recordings showed that in A1 On- and Off-R are carried by non-overlapping sets of synapses (Scholl et al., 2010). MGB terminals showed prominent On-R or Off-R (Figure 2.14A, B). The proportion of MGB terminals showing both On-R and Off-R is low ( $0.88\% \pm 1.06\%$ ). The proportion of MGB terminals showing either On-R or Off-R was similar (Figure 2.14C). OBI values for MGB terminals also indicated that most terminals were either on-only or off-only (Figure 2.14D). Thus, the majority of MGB terminals either relay On-R or Off-R suggesting the existence of distinct parallel pathways from the thalamus to A1. Terminals showing both On-R and Off-R had a more negative OBI compared to the distribution of OBI of the cellular response (Figure 2.14D inset), suggesting that there exists a transformation of On- and Off-R selectivity from MGB terminals to A1 cellular responses, which are more biased to Off-R. Moreover, given the prevalence of Off-R A1 neurons, this suggests a differential amplification of Off-Rs from MGB inputs to yield a larger fraction of Off-R neurons.

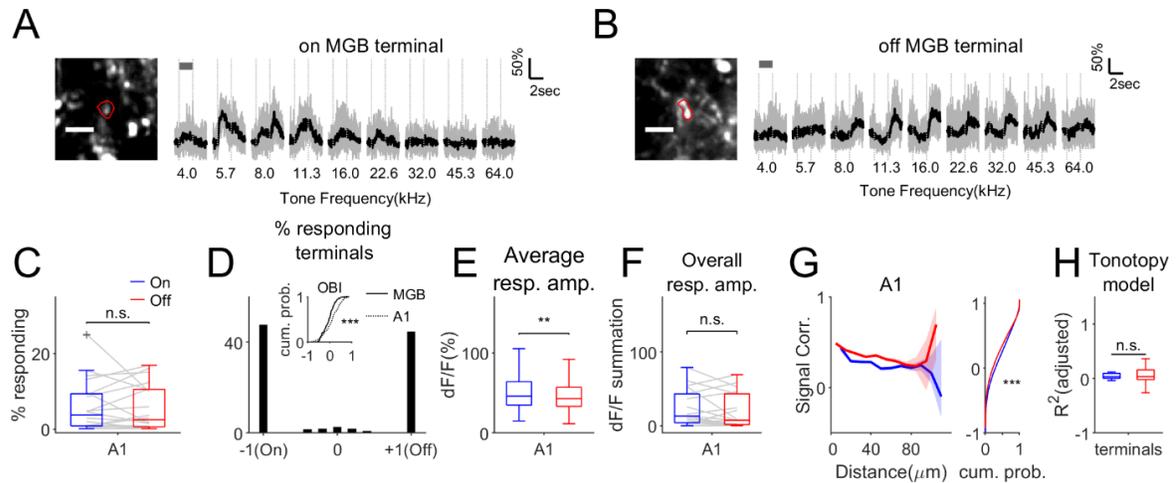
To gain insight into the transformation, we compared the average strength of On-R and Off-R pooling across terminals. Terminal On-Rs were larger than Off-Rs (Figure 2.14E) which is similar to the cellular responses (Figure 2.10B). However, unlike cellular response in A1, the On-R and Off-R MGB terminals have similar overall response amplitude (Figure 2.14F). This suggests that the larger Off-Rs in A1 cells are not generated by stronger or more numerous Off-R MGB afferents.

Local convergence of MGB inputs could lead to stronger cellular responses. Indeed convergence and temporal synchrony of thalamic inputs can strongly influence cortical neurons (Bruno and Sakmann, 2006). We have observed a distinct spatial SC structure in mesoscale (Figure 2.7G-I) as well as in cellular responses (Figure 2.8F) and we speculate that these properties could result from spatially structured MGB input. We thus calculated the SC of MGB terminals. Indeed, we find that off-MGB terminals had higher SC (Figure 2.14G) consistent with the majority cellular Off-Rs. Given that SCs on the cellular scale show a distance dependence, we investigated the distance dependence of On/Off-SC of MGB terminals. We find that Off-SC is higher than On-SC over a distance of 0-70 $\mu$ m indicating a higher degree of spatial spread of terminal Off-R. These results suggest that although individual MGB terminals do not respond to tone offset more strongly than to tone onset, local MGB terminals respond more similarly to tone offset, suggesting that the spatial correlation structure of MGB inputs is transformed into cellular tuning in A1, resulting in a more spatially extensive representation of tone offset.

Lastly, we investigated whether there is tonotopic structure in MGB terminal response. Using a linear model, we did not find a tonotopic structure in either On-R or Off-R in MGB terminals (Figure 2.14H), consistent with the observation that local On-R MGB projections in A1 are spatially heterogeneous in their tuning (Vasquez-Lopez et al., 2017). Together, our results suggest that the spatial meso-scale distribution of On- or Off-responsive A1 neurons is largely inherited from the spatial distribution of On/Off-R MGB terminals.



**Figure 2.13 MGB injections of AAV-GCaMP6s label terminals in A1**  
 Viral injection of AAV1.hSyn1.mRuby2.GSG.P2A.GCaMP6s.WPRE.SV40 in MGB.mRuby signal was imaged. (A) Overall image of injection site and cortical terminals (2x). Scale bar shows 1mm. Labeling in hippocampus is due to injection tract. (B) 10x view of auditory cortex. Terminals can be seen in layer 4 and deeper layers, consistent with typical MGBv projection pattern. Scale bar shows 100 $\mu$ m. (C) 10x view of MGB. Labeled cells can be seen in all divisions. Scale bar shows 100 $\mu$ m.



**Figure 2.14 MGB terminals in A1 largely show either On-R or Off-R and Off-R terminals show higher local signal correlations**

(A) Example on-responsive terminal. The image shows the contour of the terminal in red. Scalebar: 5 $\mu$ m (B) Same as in (A) but shows an off-responsive terminal. (C) No difference is observed in proportion of on- or off-responding terminals. On: 5.99%  $\pm$  6.72%; Off: 5.62%  $\pm$  6.00%; paired t-test,  $t(20)=0.34$ ,  $p=0.74$ . (D) Histogram of OBI values of MGB terminals in A1. Inset shows CDFs of OBI values other than -1 and 1 from MGB terminals and A1 L2/3 neurons (Wilcoxon rank sum test,  $z=3.64$ ,  $p=2.71 \times 10^{-4}$ ). (E) Individual MGB terminals in A1 show significant larger On-Rs (Wilcoxon rank sum test,  $z=2.91$ ,  $p=0.0036$ ). (F) No difference is observed in the overall On/Off-R amplitude (see Method, Wilcoxon rank sum test,  $z=0.85$ ,  $p=0.39$ ). (G) Off-Rs show higher off-SC over distance (0-70 $\mu$ m). Dashed lines show 95% confidence interval around the median. The right panel shows cumulative distribution function of On- and Off-SC not regarding distance. ‘\*\*\*\*’ indicate  $p<0.001$ . (H) No difference was found between the goodness of fit of linear On-R and Off-R tonotopy model in MGB terminals in A1.

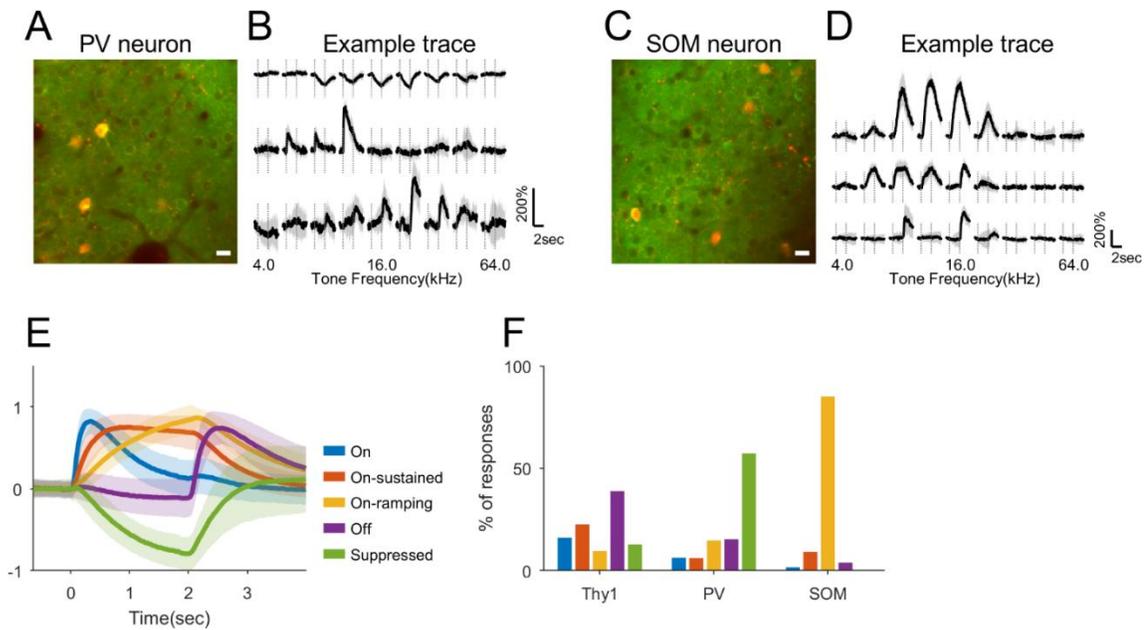
## Cortical inhibitory networks can amplify Off-R through disinhibition

The activity of cortical neurons is influenced by inhibition and we hypothesized that the On/Off-R selectivity of ACX L2/3 excitatory neurons is actively shaped by the local inhibitory network. To investigate this question, we focused on PV and SOM positive interneurons which are thought to control the activity of L2/3 neurons via a disinhibitory circuit (Pfeffer et al., 2013). We crossed Thy1-GCaMP6s mice with either PV-cre or SOM-cre mice and injected AAV-virus (see methods) expressing mRuby and GCaMP6s under control of FLEX switch sequence into ACX of F1 animals. We then imaged these animals (Thy1xPV-cre: n=8, 427 PV neurons; Thy1xSOM: n=6, 288 neurons). With this approach, local PV or SOM interneurons could be identified based on nuclear red fluorescence signal while allowing simultaneous imaging of (Thy1 positive) excitatory neurons and PV/SOM neural populations (Figure 2.15A, C). We presented 2 second tones to these F1 animals and found that although some PV and SOM interneurons displayed typical On/Off-R similar to those seen in excitatory neurons, a significant portion of these interneurons displayed a much slower temporal dynamics in their response to prolonged tones (Figure 2.15A, B). Most notably, PV neurons showed a slow decrease in fluorescence signal following tone onset (Figure 2.15B), which recovers following tone offset. SOM neurons showed similarly slow temporal responses albeit positive in sign (Figure 2.15D). To classify different response types, we performed k-means clustering on significant responses averaged across trials, pooling responses from both Thy1 (including traces from F1's of CBA/CaJ and Thy1-GCaMP6s crosses) and PV/SOM cells. We used correlation as the distance measurement

between averaged traces such that responses of similar temporal dynamics will be grouped together irrespective of the amplitude. We could identify 5 clusters with distinct temporal dynamics. The first cluster shows sharp increase in fluorescence signal following tone onset, which immediately decays (Figure 2.15E, 'On'). The second cluster shows less sharp fluorescence increase than the first cluster, but the signal sustains during the tone presentation, suggesting sustained spiking activities (Figure 2.15E, 'On-sustained'). The third cluster shows even slower dynamics with little plateau, which decays following tone offset (Figure 2.15E, 'On-ramping'). The fourth cluster shows sharp increase after tone offset and thus is categorized as typical Off-Rs (Figure 2.15E, 'Off'). The last cluster has similar dynamics as 'On-ramping' while opposite in sign (Figure 2.15E, 'Suppressed'). We next quantified the proportion of responses assigned to each cluster as a function of cell type. Distinct patterns are found among Thy1 and PV/SOM cells (Figure 2.15F). Thy1 cells, responses are predominantly assigned to 'Off', 'On' and 'On-sustained'. However, PV interneurons show predominant 'Suppressed' responses while SOM interneurons show mostly 'On-ramping'. These two response clusters show no difference in latency reaching half of peak amplitude ( $0.95 \pm 0.36$ s vs.  $0.85 \pm 0.26$ s,  $p=0.21$ , Wilcoxon rank sum test). However, the opposite sign suggests that SOM neurons suppress PV neurons during prolonged tone activation, which also consistent with previously discovered disinhibition circuit scheme (Pfeffer et al., 2013). These results indicate that the suppression of PV neurons by SOM neurons potentially allows a decrease of PV inhibitory input onto local excitatory populations, which in turn could amplify Off-R as the tail of inhibitory post-synaptic current from SOM to PV

interneurons could well last until after tone offset despite the cessation of firing of SOM.

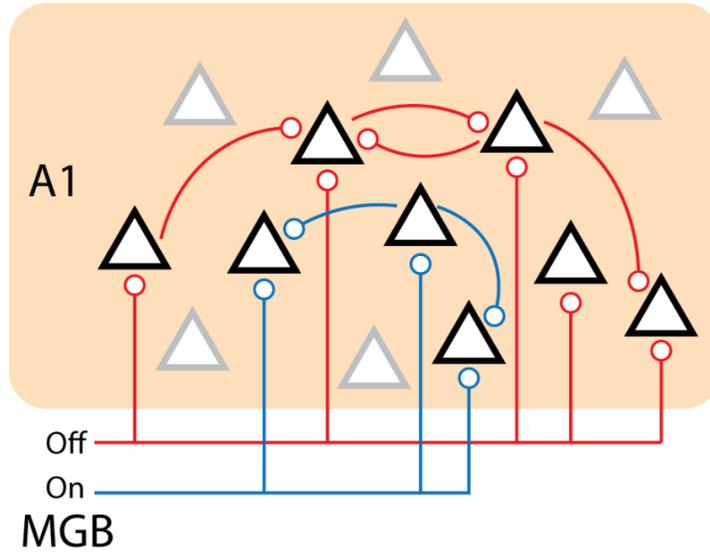
In summary, the above results suggest that the spatial distribution of On/Off-R MGB terminals determines the spatial distribution of On- or Off- responsive A1 neurons and that Off-Rs are amplified compared to On-Rs due to disinhibition through suppression of PV interneurons by SOM interneurons (Figure 2.15) as well as to increased local spatial clustering of Off-R MGB afferents (Figure 2.16).



**Figure 2.15 PV and SOM neurons show distinct temporal dynamics in response to prolonged tones**

(A) Example field of view showing both Thy1-GCaMP6s cells and PV positive interneurons virally expressing GCaMP6s and mRuby. Scalebar: 10 $\mu$ m (B) Three example PV interneurons showing suppressed response (top), On-R (middle) and Off-R (bottom). (C) Example field of view showing both Thy1-GCaMP6s cells and SOM positive interneurons virally expressing GCaMP6s and mRuby. (D) Three example SOM interneurons showing slow ramping responses following tone onset (top, middle) and Off-R (bottom). (E) K-means clustering on responses by Thy1, PV and SOM cells reveals distinct types of temporal dynamics in response to prolonged tones. All traces were normalized to maximum absolute amplitude before averaged within each cluster. Shaded regions show standard deviation. (F) Thy1, PV and SOM

cells show distinct proportion of response types. PV cells show predominantly “Suppressed” response while SOM cells show predominantly “On-ramping” response.



**Figure 2.16 Circuit model of separate On/Off processing stream from auditory thalamus to A1**

We propose that cortical On/Off-R are resultant from largely segregated On/Off thalamic input and the spatial pattern of these input determine the spatial layout of On/Off-R selective neurons. Further, Off-R cortical neurons have more recurrent connections that amplify the thalamic input compared to On-R circuitry. Black triangles represent On/Off-R neurons while gray triangles represent unresponsive neurons.

## Discussion

We show that the ACX encodes tone offset in a locally parallel, spatially extensive and yet globally tonotopically organized manner. We find that the ACX contains distinct functional areas that are characterized by distinct selectivity for tone onset or offset on the population level. Thus, the cortical representation of spectral information, i.e., tone frequency, is influenced by temporal dynamics of the spectrally static tones. Individual ACX areas contain populations of cells with distinct selectivity for On-R or Off-Rs. Moreover, we utilized GC analysis to reveal that ACX areas contain intermingled On/Off networks within L2/3. Therefore, areal specification is due to different relative numbers of On-R and Off-R neurons as well as due to distinct intracortical circuits. Imaging responses of PV and SOM interneurons in A1 and their distinct temporal dynamics point to disinhibition as one mechanism that amplifies Off-R. Moreover, by imaging MGB terminals we show that the areal and cellular On/Off-R specialization may arise from differences in MGB input which could be further enhanced by local spatially correlated activity of MGB terminals. Together, our results suggest that the differential dynamic responses originate from differential feedforward input from MGB that is further elaborated on by different intrinsic excitatory and inhibitory circuits in the different ACX regions. This indicates that ACX areas operate in parallel to extract differing temporal stimulus qualities. Our results also demonstrate that Off-Rs are tonotopically organized on the mesoscale. The lack of Off-R tonotopy in prior anesthetized studies (Baba et al., 2016) is likely due to Off-R being most prominent in awake animals

(Recanzone, 2000; Qin et al., 2007; Fishman and Steinschneider, 2009; Joachimsthaler et al., 2014).

We here develop and deploy a novel method to define functional ACX areas based on their temporal responses instead of the largest stationary response. This image segmentation technique is based on temporal coactivation of pixels (Whiteway and Butts, 2017). This method is unbiased and unsupervised and requires no prior assumptions on the distribution of cortical fields and can be applied to arbitrary WF datasets.

Besides tone onset and offset, ACX neurons can also be sensitive to other dynamic aspects of sound such as amplitude modulation, frequency modulation, sound duration and frequency sweep rate (Schreiner and Urbas, 1986; Heil et al., 1992; He et al., 1997; Baumann et al., 2015; Issa et al., 2017). While frequency sweep rate is topographically organized in mouse ACX (Issa et al., 2017), our results show that Off-R are also topographically represented.

We found an extensive representation of tone offset in A1 and DP neurons. A1 neurons receive On-Rs and Off-Rs synaptic inputs which are thought to be mediated by non-overlapping set of synapses (Scholl et al., 2010). We find that MGB terminals are mostly either exclusively On-responsive or Off-responsive suggesting that A1 neurons receive convergent input from such On- or Off-responsive MGB terminals. Further, we found that MGB terminals have on average weaker Off-R, suggesting that enhanced Off-R in A1 is indeed resultant from different On/Off-R input topology, or the spatial distribution of connections. While in A1 Off-R neurons outnumber On-R neurons, Off-R MGB terminals do not outnumber On-R MGB

terminals. Moreover, the On/Off-R bias differs between MGB terminals and A1 neurons. No evidence so far suggests that On- and Off-circuit has different quantal synaptic strength and thus the On/Off-R bias is more likely to result from differential convergence of connections. Together, these observations suggest the presence of local A1 circuits to amplify Off-R. Our results also suggest that a disinhibitory circuit formed by SOM and PV neurons could play this role. Nevertheless, MGB terminals share the same On/Off-SC structure as A1 L2/3 neurons, suggesting that the spatial pattern of cellular on/off-selectivity is largely determined by the spatial pattern of MGB projections. Although Off-Rs are of smaller amplitude than On-Rs in MGB terminals, the spatially more extensively distributed input could potentially involve more recurrent connections or convergence such that cellular Off-Rs are preferentially amplified especially at high sound levels. Indeed, a multilayer nonlinear neural network has been proposed to underlie the wide variety of On/Off-Rs observed in A1 (Deneux et al., 2016). Our work suggests that the MGB-A1 circuit could underlie this transformation. Ideally, our conclusion would be strengthened by simultaneously imaging MGB terminals and ACX postsynaptic neurons. However, such approach would be still be limited as corresponding terminals and postsynaptic neurons would not be necessarily localized in the same imaging plane and with the spatial resolution of 2P imaging it is still difficult to determine unequivocally presynaptic terminal and postsynaptic cell pairs.

MGB terminals showing On-Rs or Off-Rs likely originate from different MGB subdivisions. ACX receives MGB inputs via lemniscal and non-lemniscal pathways. The lemniscal pathway arises from the ventral MGB (MGBv) which shows

On-Rs (Aitkin and Webster, 1972; Imig and Morel, 1983; Redies and Brandner, 1991; Hackett et al., 2011). Multiple lines of evidence suggest that Off-Rs originate in non-lemniscal pathways. Off-Rs are predominantly observed in a sheet partially surrounding MGBv (He, 2001). Off-Rs can also originate from the dorsal and medial MGB (MGBd and MGBm). Indeed, we found that A2 and DP which receive MGBd input (Lee and Sherman, 2008; Llano and Sherman, 2008) show Off-Rs. Moreover, the spatial extensiveness of Off-Rs is consistent with the broad projection from MGBm to ACX through L1 (Huang and Winer, 2000; Lee and Winer, 2008). Thus, non-lemniscal pathways likely provide tone offset information to ACX. We imaged the terminals at roughly the same depth as the cellular population imaged (~150 $\mu$ m), and thus these terminal might reflect a mixture of lemniscal and non-lemniscal pathways as terminals from both MGBv and MGBd are present in L2 in A1 (Saldeitis et al., 2014). Our results show overlapping tonotopy of On-R and Off-R albeit areal differences, suggesting that lemniscal and non-lemniscal pathways are coarsely aligned but show distinct spatial patterning. The factors that determine these patterns are unknown.

The majority of responding A1 L2/3 neurons have either On-R or Off-R. Thus the spatial heterogeneity of tonal responses in A1 L2/3 might be due to intermingled cells receiving differing thalamic input. In S1, functionally different thalamic inputs from ventral posterior medial nucleus (VPM) and posterior medial nucleus (POm) are relayed to barrels and septa (Koralek et al., 1988; Lu and Lin, 1993) which are spatially separated and carry whisking-touch information (Yu et al., 2006) and temporal information on whisker movement respectively (Ahissar et al., 2000). Our

results suggest that in contrast to S1 functionally different thalamic inputs to A1 are spatially interdispersed. A1 L2/3 contains cells with distinct functional intracortical-circuits and shows a sublaminal organization (Meng et al., 2017). It is possible that the distinct On/Off subnetworks we identified might relate to these distinct subpopulations. And because recurrent inputs from subgranular layers are thought to be able to amplify thalamic inputs (Somers et al., 1995; Miller et al., 2001; Li et al., 2013), we speculate that Off-R cells receive stronger or more extensive inputs from subgranular layers. Prior electrophysiology studies have identified a larger proportion of neurons responding to both tone onset and offset (Qin et al., 2007; Joachimsthaler et al., 2014). The discrepancy most likely results from the difference in recording depth and the inclusion of multiunit activity, given the intermingled spatial distribution of On- and Off-R (Figure 2.12B) which could bias electrophysiological studies. Indeed, our electrophysiological recordings show that Off R's are more prevalent in superficial layers where we performed imaging (~150  $\mu\text{m}$  depth). Such a depth dependence of On/Off-R is consistent with reports showing that L5/6 neurons were less likely to generate Off-R (Volkov and Galazjuk, 1991). Thus, given that 2-photon imaging has much higher spatial resolution and lacks electrode bias our imaging results most likely revealed a highly specific On/Off-R selectivity in upper L2/3. Nevertheless, we find that among all responding neurons in A1, 53% were Off-responsive, close to 59% reported in awake cat by Qin et al (2007). Therefore, our imaging data revealed an Off-R bias and a separation of On/Off-R network in superficial layers possibly via distinct sublaminal circuits (Meng et al., 2017). Future imaging studies linking functional circuits to functional responses and focusing on

deeper cortical layers are required to address how On/Off-R networks are transformed across cortical layers.

We found that around 5% of neurons in A1 respond to tone onset/offset, which is consistent with a sparse representation of sound in A1 (Hromádka et al., 2008). However, previous imaging studies of A1 have reported around 20-30% response rate (Issa et al., 2014; Kato et al., 2015). This discrepancy likely arises from sampling of different populations of neurons. Issa et al. (2014) used cre-dependent GCaMP3 driven under syn1-cre or emx-cre, while such labeled populations had fewer visual responses compared OGB-1 labeled neurons in V1 (Zariwala et al., 2012), suggesting a non-uniform labeling of neuronal population. Kato et al. (2015) used viral expression of GCaMP6s under syn1 promotor, which densely labeled local populations close to the injection site. Our current study used the GP4.2 line, which labels about 70% of pyramidal cells in L2/3 in a relatively uniform fashion (Dana et al., 2014). It is likely that the difference in response rate between our study and prior imaging studies is due to the labeling of different but potentially overlapping populations, the difference in calcium indicator (GCaMP3 vs GCaMP6s), expression profile (transgenic vs viral expression), and cell selection criteria.

We find that L2/3 PV and SOM positive interneurons show distinct temporal dynamics from each other as well as from Thy1-expressing excitatory neurons. In particular, PV and SOM interneurons show predominantly opposite signs of responses. The suppression of the fluorescence signal in PV neurons corresponded to a reduction of the firing rate (Forli et al., 2018). Thus our data suggests a suppression of PV interneurons' firing by SOM interneurons. This is consistent with a

predominantly suppressed responses of L2/3 PV neurons to prolonged tones (Kato et al. (2015)). The opposing direction in the response of PV and SOM neurons is consistent with the proposed cortical processing scheme that PV interneurons receive inhibitory input from SOM interneurons (Pfeffer et al., 2013), and that such inhibition shows significant synaptic facilitation (Karnani et al., 2016). Finally, SOM neurons more readily inhibit PV neurons than local excitatory neurons (Cottam et al., 2013). We speculate that such inhibition could facilitate detection of changes in auditory streams, such as tone offset given that the duration of inhibitory post-synaptic current could outlast firing of SOM interneurons, which creates a window for elevated excitability in local pyramidal neurons before PV interneurons return to baseline firing. Although such inhibition of PV interneurons by SOM interneurons has not been experimentally confirmed in ACX, it is possible that this circuit is involved in processing prolonged auditory stream. Kato et al (2015) have shown that SOM interneurons potentially suppress responses to behaviorally irrelevant and prolonged tones by inhibiting PV interneurons and pyramidal cells. Furthermore, our results show that SOM neurons are active throughout tone presentation, in contrast to previous findings that SOM cells fire transiently during tone presentation, although such difference could be due to the animal's state (awake vs. anesthesia) (Li et al., 2014a; Chen et al., 2015). Thus, SOM interneurons are potentially important for auditory stream analysis and their interactions with PV neurons could facilitate change detection.

In conclusion, we have demonstrated by multiscale imaging a distinctly extensive parallel spatial representation of sound dynamics in ACX at multiple levels

and we propose that this spatial pattern is determined by the meso- and micro-scale spatial layout of thalamic input and by distinct intracortical circuits.

# **Chapter 3. Diversity of receptive fields and sideband inhibition with complex thalamocortical and intracortical origin in L2/3 of mouse primary auditory cortex**

*This chapter has been published in Journal of Neuroscience (Liu and Kanold, 2021)*

## **Abstract**

Receptive fields of primary auditory cortex (A1) neurons show excitatory neuronal frequency preference and diverse inhibitory sidebands. While the frequency preferences of excitatory neurons in local A1 areas can be heterogeneous, those of inhibitory neurons are more homogenous. To date, the diversity and the origin of inhibitory sidebands in local neuronal populations and the relation between local cellular frequency preference and inhibitory sidebands are unknown. To reveal both excitatory and inhibitory subfields we presented two-tone and pure tone stimuli while imaging excitatory (Thy1) neurons and two types of inhibitory neurons (PV and SST) in L2/3 of mice A1. We classified neurons into 6 classes based on frequency response area (FRA) shapes and sideband inhibition depended both on FRA shapes and cell types. Sideband inhibition showed higher local heterogeneity than frequency tuning, suggesting that sideband inhibition originates from diverse sources of local and distant neurons. Two-tone interactions depended on neuron subclasses with excitatory neurons showing the most nonlinearity. Onset and offset neurons showed dissimilar

spectral integration, suggesting differing circuits processing sound onset and offset. These results suggest that excitatory neurons integrate complex and nonuniform inhibitory input. Thalamocortical terminals also exhibited sideband inhibition, but with different properties from that of cortical neurons. Thus, some components of sideband inhibition are inherited from thalamocortical inputs and are further modified by converging intracortical circuits. The combined heterogeneity of frequency tuning and diverse sideband inhibition facilitates complex spectral shape encoding and allows for rapid and extensive plasticity.

## **Introduction**

One of the fundamental functions of sensory systems is to differentiate between distinct stimuli. Such stimulus selectivity requires that neural circuits possess selectivity for certain attributes of the sensory stimulus. Starting at the peripheral sensory epithelium, stimulus selectivity is achieved through functional so-called lateral, or sideband, inhibition. The visual system achieves this by the activation of neurons that reduce the activity in other neurons that have slightly differing receptive field properties, e.g. being sensitive to stimuli at a different spatial location. For example, ON/OFF receptive fields of the retinal ganglion cells signal size selectivity (Kuffler, 1953; Famiglietti and Kolb, 1976) and are shaped by such lateral interactions (Cook and McReynolds, 1998). In the auditory system, the mechanical properties of the basilar membrane creates frequency selectivity as traveling waves reach maximum amplitude at specific locations within cochlea (Von Békésy and Wever, 1960), which is further amplified by the movement of outer hair cells

(Fettiplace, 2020). Presentation of two tones causes nonlinear mechanical interactions in the cochlea which in turn causes suppression in inner hair cells (Ruggero et al., 1992). Thus, a tone of a different frequency than the primary tone can alter the responses to the primary tone and such frequencies constitute the inhibitory sideband, which sharpens stimulus selectivity even at the very first stage of sensory information encoding.

Lateral inhibition or sideband inhibition can also be found along the ascending auditory pathway, including in the cochlear nucleus (Greenwood et al., 1976; Nelken and Young, 1994; Davis and Young, 2000), inferior colliculus (Brimijoin and O'Neill, 2005; Mayko et al., 2012), medial geniculate body (Schreiner, 1981) and the auditory cortex (ACX) (Sutter and Schreiner, 1991; Nelken et al., 1994; Sutter et al., 1999; Li et al., 2014b; Kato et al., 2017). In the ACX, thalamic inputs are amplified (Li et al., 2013) and further processed by local microcircuits, resulting in the refinement in the frequency tuning where L2/3 neurons showed narrower tuning than L4 neurons (Winkowski and Kanold, 2013; Li et al., 2014b) despite the fact that frequency tuning tends to get broadened along auditory ascending pathway (Bartlett et al., 2011). While the relative contribution of different classes of inhibitory neurons to this tuning refinement is unclear (Li et al., 2014b; Kato et al., 2017), Parvalbumin (PV) or Somatostatin (SST) positive neurons are thought to mediate sideband inhibition in the primary auditory cortex (A1) (Li et al., 2014b; Kato et al., 2017; Lakunina et al., 2020). Moreover, it is unclear whether thalamocortical inputs also contribute to sideband inhibition in A1.

Excitatory A1 L2/3 neurons in a local area can show diverse tuning preferences (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Winkowski and Kanold, 2013; Kanold et al., 2014; Maor et al., 2016) and integrate excitatory and inhibitory inputs from a large region of the tonotopic map (Meng et al., 2017). In contrast inhibitory PV cells in a local area show a high degree of similarity (Maor et al., 2016). This raises the question whether sideband inhibition varies among local populations of L2/3 neurons and if any spatial patterns exist relative to local frequency tuning. To investigate inhibitory sidebands of A1 neurons and the relationships of these sidebands between neurons, we performed 2-photon imaging and probed neural responses to both pure tones (PT) and two-tone (TT) stimuli in excitatory (Thy1) and inhibitory (PV and SST) populations. We classified neurons based on the shape of their frequency response areas (FRAs) and found a differential degree of sideband inhibition and nonlinear frequency interactions among FRA types and cell types. Inhibitory sidebands of local neural populations showed high variability and heterogeneity indicating that a variety of inhibitory sources contributed to them. Imaging the activity of MGB terminals showed that inhibitory sidebands were present in MGB terminals, but that the tuning and sideband properties differed from those of cortical neurons. Our results thus suggest that inhibitory sidebands are created by non-uniform mechanisms between neurons, reflecting a complex interplay between existing inhibitory sideband structures in the feedforward MGB input and additional contribution of cortical inhibition. The combined heterogeneity in frequency tuning and sideband inhibition could further render

neurons' selective for spectral features and provide a rich local substrate for extensive and rapid plasticity.

## Methods

### Animal Procedures

All procedures were approved by the University of Maryland's Animal Care and Use Committee. To produce mice with normal hearing, all animals used in this study were F1 generations from the crosses between CBA/CaJ mice and other transgenic lines including Thy1-GCaMP6s (JAX#024275, GP4.3 (Dana et al., 2014)), PV-cre (JAX#017320) and SST-cre (JAX#013044). C57BL/6 are homozygous for the mutant *Cdh23* allele *ahl* that causes age related hearing loss while CBA/CaJ mice are homozygous for the wildtype *Ahl+* (Kane et al., 2012). Such crosses ensured that F1 offspring had one wildtype allele such that they had normal hearings (Frisina et al., 2011; Bowen et al., 2020). To express GCaMP6s in PV or SST neurons, we injected AAV1.Syn.Flex.mRuby2.GSG.P2A.GCaMP6s.WPRE.SV40 (gift from Tobias Bonhoeffer & Mark Huebener & Tobias Rose, Addgene viral prep # 68720-AAV1; <http://n2t.net/addgene:68720>; RRID: Addgene\_68720, ~30nl per site, 3-4 sites of injections) into the left auditory cortex of the F1 animals expressing PV-cre or SST-cre. We waited  $14.6 \pm 2.3$  days before starting imaging viral injected animals. We used 6 CBA/CaJxThy1-GCaMP6s mice (4 males, 2 females, 11 to 24 weeks old), 4 CBA/CaJxPV-cre mice (2 males, 2 females, 13 to 15 weeks old) and 8

CBA/CaJxSST-cre mice (4 males, 4 females, among which 6 of them were 13 to 27 weeks old and the rest 2 were 48 weeks old).

### Cranial window implant

We implanted cranial windows to perform imaging over the left A1 following the procedure outlined in Liu et al. (2019). First to prevent brain swelling during the cranial window implant, 0.1cc dexamethasone (2mg/ml, VetOne) was injected subcutaneously 2-3 hours prior to the start of the surgery. All surgery tools were sterilized with a bead sterilizer (18000-45, Fine Science Tools). The animals were anesthetized with isoflurane (Fluriso, VetOne) using a calibrated vaporizer (Matrix VIP 3000) with 4% for induction and 1.5-2% for maintenance. During surgery the body temperature of the animal was maintained at 36.0 degrees Celsius. After the head fixation, the hair on top of the head was removed by applying Hair Remover Face Cream (Nair). Application of betadine (Purdue Products) followed by a 70% ethanol rinse was repeated 3 times before the skin was removed. The surface of the skull was gently scrapped with a scalpel blade to remove the soft tissue. Muscles covering the left temporal bone were subsequently removed. After cleaning the skull, a custom 3D printed stainless steel headplate was mounted and secured using C&B-Metabond (Parkell Inc.). A circular craniotomy was then performed over the left auditory cortex with a diameter of ~3.5mm using a dental drill. Viral injections were performed at this point. Then a custom-made cranial window was placed over the exposed brain. The window consisted of 2 layers of 3mm round coverslips (64-0720,

CS-3R, Warner Instruments) stacked at the center of a 4mm round coverslip (64-0724, CS-4R, Warner Instruments) and secured with optic glue (NOA71, Norland Products). The edge of the cranial window was then sealed with Kwik-sil (World Precision Instruments). More Metabond was then applied to secure the window to the skull. After the surgery, 0.05cc Cefazolin (1 gram/vial, West Ward Pharmaceuticals) was injected subcutaneously and the animal recovered under a heat lamp for 30 minutes before being returned to the home cage. Medicated water (Sulfamethoxazole and Trimethoprim Oral Suspension, USP 200mg/40mg per 5ml, Aurobindo Pharms USA; 6ml solution diluted in 100ml water) substituted normal drinking water for 7 days before any imaging was performed.

### Viral injection into MGB

To label axon terminals of MGB in A1, we injected AAV.CamKII.GCaMP6s.WPRE.SV40 (a gift from James M. Wilson, Addgene viral prep # 107790-AAV9; <http://n2t.net/addgene:107790>; RRID: Addgene\_107790) into the MGB. Specifically, we used the coordinate AP -3.2mm, ML 2.1mm relative to Bregma to target the left MGB. We injected ~100nl of the virus at a depth of 3.0mm below pia. The cranial window was implanted using the same procedure as outlined above. For this experiment, we used 3 CBA/CaJ mice (1 female, 2 males) around 11-12 weeks old. Imaging was performed  $17.1 \pm 1.1$  days after viral injections.

### Widefield imaging and image processing

To identify the location of A1 we performed widefield imaging as previously described (Liu et al., 2019). The animal was head-fixed in a custom holder and the cranial window was illuminated with 470nm LED light (M470L3, Thorlabs Inc.) while the green fluorescence was collected using a PCO Edge 4.2 camera. The frame rate was 30 Hz and images had dimensions of 400 by 400 pixels, which were downsampled by a factor of 4 for analysis.

For image analysis, the 10 frames before sound onset were used as the baseline, whose average was subtracted from each of the 30 frames following the sound onset. These 30 frames were then averaged to reveal the location of fluorescence increase. Then we manually identified the location of A1 based on known tonotopy in the mouse ACX (Liu et al., 2019).

### Two-photon imaging and image processing

The animal was first head-fixed in a custom holder. Then the field of interest was determined by comparing the widefield map with the blood vessel patterns to ensure A1 was imaged. We imaged L2/3 neurons at a depth of  $250.6 \mu\text{m} \pm 48.4 \mu\text{m}$  and MGB terminals at a depth of  $117 \mu\text{m} \pm 19.5 \mu\text{m}$ . The size of the field of view was  $369 \mu\text{m}$  by  $369 \mu\text{m}$  for cellular imaging and  $92 \mu\text{m}$  by  $92 \mu\text{m}$  for MGB terminal imaging and we used the B-SCOPE (Thorlabs Inc.) with the microscope body tilted at 45 degrees such that the mouse head could be held upright. The excitation wavelength was 920 nm and images were collected with ThorImage software (Thorlabs Inc.) at a

frame rate of ~30 Hz. A 16x Nikon objective was used (NA 0.80) and the optic zoom was set to 2X for cellular imaging and 8X for MGB terminal imaging.

To extract cellular fluorescence, we manually placed circular regions of interest (ROIs) on identified cells with the contour of the ROIs roughly aligned with the shape of the cell's soma. To extract neuropil traces we used a ROI spanning 20 $\mu$ m from the cell center while excluding any soma ROIs within the distance. To calculate  $\Delta F/F$  traces, we followed the same procedure detailed in (Liu et al., 2019). Briefly, we first obtained the neuropil corrected traces for each cell using the following equation:

$$F_{corrected}(t) = F_{cell}(t) - 0.8 \times F_{neuropil}(t)$$

To determine the baseline we constructed a histogram of the corrected fluorescence trace and found the fluorescence value with the maximum count, which corresponded to the most frequently occurring fluorescence value for each cell. This value was chosen as the baseline of each cell. We then used the following equation to obtain  $\Delta F/F$  over time:

$$\frac{\Delta F}{F}(t) = \frac{F_{corrected}(t) - baseline}{baseline}$$

To extract fluorescence trace from MGB terminals, we first used an automated program to define ROIs. Specifically, we used a 2D image peak finder algorithm (Natan (2020). Fast 2D peak finder (<https://www.mathworks.com/matlabcentral/fileexchange/37388-fast-2d-peak-finder>), MATLAB Central File Exchange. Retrieved November 4, 2020) to localize the center of each bouton in the average image of the entire image sequence. For the parameters

of this algorithm, we used 15% maximum intensity of the image as the threshold and a gaussian filter with a sigma of  $\sim 0.54 \mu\text{m}$  (3 pixels). This algorithm finds local maxima and thus accounts for the uneven brightness across boutons. With the location of boutons defined, we proceeded to use a circular ROI with a diameter of  $\sim 1.4 \mu\text{m}$  (8 pixels) to extract raw fluorescence trace. For neuropil traces, we used a circular ROI with a radius of  $5 \mu\text{m}$  while excluding all bouton ROIs within the radius. The  $\Delta F/F$  for each bouton was then calculated with the same procedure as outlined above.

### Acoustic stimuli

We presented two sets of stimuli to obtain the FRAs and the inhibitory sidebands of the neural population, respectively. The first set consisted of 16 tones logarithmically spaced from 4kHz to 53.8kHz. The amplitudes of the tones were calibrated to 70dB SPL and attenuated from 0dB to 30dB SPL with a step of 15dB SPL. Each tone was 100ms in duration and had a 10ms linear ramp at the onset and the offset of the tone. The second set of stimuli consisted of both PTs and TT combinations. The PTs within this set were the same 16 tones in the first set with their individual amplitude calibrated to 60dB SPL. The TT combinations were constructed by drawing two distinct tones and obtaining the linear summation of the waveforms over time. Thus, the TTs were 63 dB SPL as they were the linear summation of two different frequencies. For each TT stimulus, the phases of the TTs were independently and randomly selected. The second set of stimuli were also

100ms in duration and had the same 10ms ramping. All sounds were presented using a custom-written MATLAB GUI that communicated with RX6 and PA5 (TDT Inc.) for actual waveform generation and sound attenuation. The sound was delivered with one ES1 speaker (TDT Inc.) driven by ED1 speaker driver (TDT Inc.). The speaker was situated 10cm away from the animal's head and at a 45-degree angle relative to the midline.

### Response significance

We determined the significance of the responses using a similar approach outlined in (Liu et al., 2019). First, a window of 10 frames before stimulus onset was chosen for measuring baseline activities, while a window of 20 frames spanning 0.2 sec to 0.83 sec after stimulus onset was chosen to measure evoked activity. We chose to start at 0.2 sec to ensure the maximum separation of response amplitude from baseline given actual neural activities as the latency to reach peak fluorescence change is around 160 ms for GCaMP6s (Chen et al., 2013). This choice also accounts for offset responses, which are also more delayed. Time-varying  $\Delta F/F$  values were obtained within both windows across trials (10 frames  $\times$  5 trials = 50 data points before sound onset, 20 frames  $\times$  5 trials = 100 data points after sound onset), and the 99.9% confidence interval (CI) of the mean of each sets of data points were obtained. A response was deemed significant if the lower bound of the post-stimulus CI was higher than the upper bound of the pre-stimulus CI.

## Classification of FRA shape

In order to classify the FRAs, we first resorted to unsupervised algorithm that helped to identify recurring shapes, which ultimately guided our manual classification. For unsupervised classification, we first aligned the FRAs of all responsive cells (pooling from all cell types) at the geometric center, calculated through weighted average of frequencies by significant responses:

$$F_c = \frac{\sum_{i,j} r(F_{i,j}) * F_i}{\sum_{i,j} r(F_{i,j})}$$

where  $i$  and  $j$  denote index of frequency and sound level respectively while  $F$  and  $r$  denote actual frequency and the response amplitude respectively. Non-significant responses were set to zero to ensure the validity of the average. This method of alignment was preferred over using best frequency or characteristic frequency because their measurement could be noisy and thus less robust. Next, a principal component analysis (PCA) was performed for dimensionality-reduction such that remaining number of coefficients account for 95% of the total variances in the aligned FRA. Then, we performed K-means clustering on the kept coefficients using a range of number of clusters (from 2 to 20) and used the T-distributed Stochastic Neighbor Embedding (t-SNE) algorithm to visualize the clustering results. We used correlation distance in the K-means algorithm in order to capture the similarity between FRA shapes, regardless of the response amplitude. Upon plotting the average aligned FRAs and inspecting the t-SNE plot (Figure 3.4A) we chose the number of clusters to be 6, among which some roughly corresponded to classic V and I shaped FRAs while some were sparsely responding to PTs. We then manually inspected the

FRAAs guided by the unsupervised clustering to generate the final classification result (Figure 3.4A). The final clusters differed in their general shapes, tuning properties at different levels and sparseness of the responses.

### Width of tuning curve or inhibitory sideband

Frequency selectivity is measured by the width of the tuning curve. To account for the variability of tuning across cells with different FRA shapes as well as to use a similar measure to quantify the broadness of both tuning curve and inhibitory sideband, with the latter not necessarily single-peaked, we used a sparseness measure as a surrogate to estimate tuning curve width. Specifically, we first calculate the sparseness of the tuning curve and inhibitory sideband with the following equation:

$$Sparseness = \frac{\sqrt{k} - \|x\|_1}{\|x\|_2 \times (\sqrt{k} - 1)}$$

where  $\|x\|_1$  and  $\|x\|_2$  are the L1 and L2 norm of vector  $x$ .  $k$  is the length of vector  $x$ , while  $x = r \odot s$ , which is the element wise product between  $r$ , a vector representing the amplitude of either tuning curve or inhibitory sideband, and  $s$ , a vector consists of 0s and 1s to indicate the significance at each frequency for either the tuning curve or the inhibitory sideband. The sparseness measure has values between 0 and 1, with 0 achieved by a vector with a uniform non-zero amplitude and 1 achieved by a vector with a single non-zero element. Thus, a widely tuned neuron will have sparseness close to 0 while a narrowly tuned neuron will have sparseness close to 1. We then used 1-Sparseness as a measure for tuning width.

### Signal correlations

We used the neural responses to the second stimulus set (16 PTs+120 TTs) to compare the signal correlations with or without the addition of the second tone.

However, such comparison would not be valid unless the number of PTs or TTs was the same. To achieve this, we first picked a frequency ( $F_i$ ) from all frequencies and gather average responses over all trials to all frequencies but  $F_i$ :

$$[\bar{r}(F_1), \dots, \bar{r}(F_{i-1}), \bar{r}(F_{i+1}), \dots, \bar{r}(F_n)]$$

where  $n$  is the total number of distinct frequencies (16 in the current study). Such generated vector would be of length  $n-1$ . We proceeded by forming the second response vector as follows:

$$[\bar{r}(F_1 + F_i), \dots, \bar{r}(F_{i-1} + F_i), \bar{r}(F_{i+1} + F_i), \dots, \bar{r}(F_n + F_i)]$$

which consisted of average responses to all TT combinations containing  $F_i$ . Such generated response vector had the same length of  $n-1$  and was matched in frequency with the PT response vector except for the introduction of the second tone. We computed the signal correlations between cell pairs using the above response vectors through computing the correlation coefficients and pooled data across frequencies for statistical analyses.

### Sideband inhibition and nonlinear frequency interactions

To determine whether the responses to TT caused significant reductions in response amplitude compared to responses to the BF, which was chosen based on the frequency evoking the maximum response among the PTs presented at 60 dB SPL in

our second stimulus set, we first chose a window of 0.5sec (15 frames) after the stimulus onset and centered at the peak of  $\Delta F/F$  change and gathered all time-varying  $\Delta F/F$  traces within the window across trials (15 frames  $\times$  5 repeats, 75 data points) for both the TT and the BF stimulus. Next, we used a bootstrap procedure with 1000 repeats to determine the 95% confidence intervals (CI) of the mean of the two sets of data points. If the upper bound of the CI of the TT stimulus was less than the lower bound of the CI of the BF stimulus, then the TT stimulus was considered to have resulted in a significant reduction in response compared to that of the BF. Similarly, to determine nonlinear interactions between responses to PTs, we gathered three sets of data points, belonging to responses to two distinct PTs and their TT combination, respectively. We bootstrapped the 95% CI of the summation of the mean responses to the two PTs and compared the boundary to the 95% CI of the mean response to the TT combination. If the two CIs were nonoverlapping, the interaction was considered significant and depending on the sign of the difference, it was characterized as facilitative (positive) or suppressive (negative). To characterize the total amount of such interactions, we summed over all found facilitative or suppressive interactions for individual cells. To calculate Suppression Facilitation Index (SFI), we used the following equation:

$$SFI = \frac{S - F}{S + F}$$

where S and F denote the absolute value of total suppressive and facilitative interactions, respectively.

### IQR analysis

To quantify the heterogeneity of local frequency tuning and sideband inhibition (Figure 3.7), we computed the interquartile range (IQR) of either BF and BIhF within a 100 $\mu$ m radius of the cell in question and pooled this value based on cell types. Specifically, all responding cells within the 100  $\mu$ m radius were identified and the absolute differences in octaves of BF and BIhF relative to the cell in question were calculated respectively. Then the IQR value was computed for  $\Delta$ BF and  $\Delta$ BIhF respectively and pooled according to cell types.

### Experimental design and statistical analyses

To compare tuning width between different cell type and FRA type combinations (Figure 3.5B), we used a three-way ANOVA with main factors of cell type, FRA type and tuning versus sideband (MATLAB built-in function ‘anovan’, 2017b). To compare specific groups, we used Tukey–Kramer multiple-comparison test (MATLAB built-in function ‘multicompare’, 2017b). For comparison of IQR (Figure 3.7), we used a two-way ANOVA with main factors of cell type and BF versus BIhF. For comparison of signal correlations (Figure 3.8), we used a two-way ANOVA with main factors of cell type and PT versus TT. For SFI comparison, we used one-way ANOVA with FRA types as the factor (Figure 3.9B). All bar graphs show mean $\pm$ SEM as indicated. Confidence intervals were constructed with MATLAB built-in function ‘bootci’ (Mathworks MATLAB, 2017b). For effect size, we computed Hedges’  $g$  using Measures of Effect Size (MES) toolbox (Harald

Hentschke (2020). [hhentschke/measures-of-effect-size-toolbox](https://github.com/hhentschke/measures-of-effect-size-toolbox)

(<https://github.com/hhentschke/measures-of-effect-size-toolbox>), GitHub. Retrieved November 12, 2020).

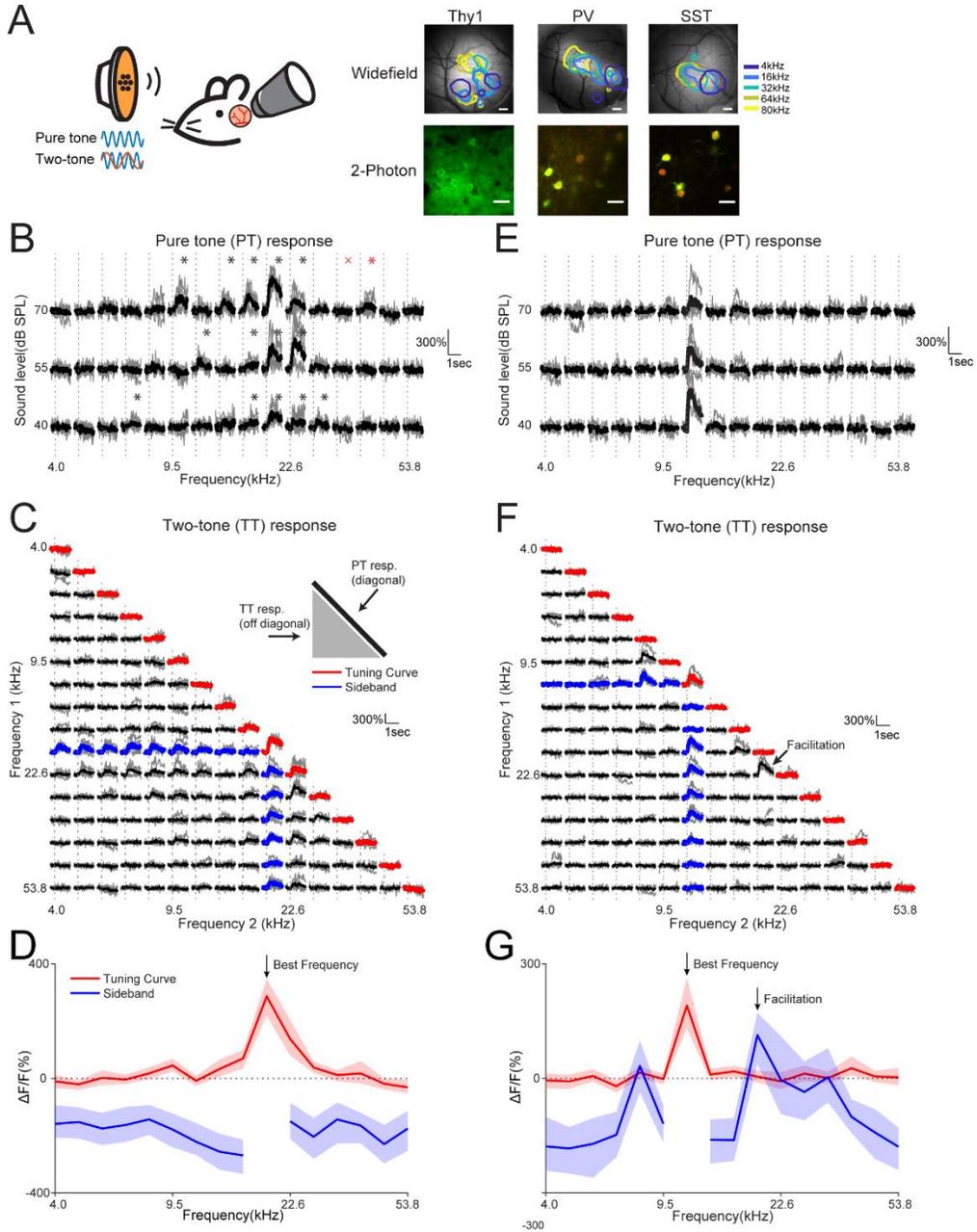
## Results

To characterize sideband inhibition in L2/3 of mouse primary auditory cortex (A1), we played both pure tones (PT) and two-tone (TT) combinations to passively listening awake male and female mice (Figure 3.1A). Conventionally, sideband inhibition is inferred by first choosing a reference tone typically at the best frequency (BF) of the neuron in question and then presenting other tones of varying frequencies and sound levels in combination with the chosen tone (Brosch and Schreiner, 1997). By definition, a tone at BF evokes the largest response among frequencies presented across all sound levels. If a tone at BF combined with other tones results in reduced responses, functional inhibition can be inferred at these frequencies (Sutter and Schreiner, 1991). However, this method is ill-suited for two-photon imaging where large neural populations are monitored simultaneously given the heterogeneity of local tuning (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Winkowski and Kanold, 2013) it would be impossible to use the same reference tone for every neuron. Therefore, we designed an alternative approach using a fixed sound level and presenting tones of all possible combinations given the chosen frequency range (3.75 octaves) and density (4 tones per octave). This strategy thus resulted in 120 distinct tone pairs given 16 different tones (4kHz to 53.8kHz, logarithmically spaced). We

presented two sets of stimuli. The first set of stimuli consisted of the same 16 tones at 3 sound levels (40/55/70dB SPL) in order to construct frequency response area (FRA). The second set of stimuli consisted of both individual PTs and all 120 tone pairs whose waveforms were linear summations of the PTs. In this second set, the PTs were presented at 60 dB SPL and thus the TTs were 63 dB SPL as they were the linear summation of two different frequencies. Figure 3.1 shows two examples of responses to the two sets of stimuli in excitatory neurons. Figure 3.1B shows the responses of an example neuron to pure tone (PT) stimuli. This neuron had a V-shaped FRA with its BF at 19kHz (Figure 3.1D, red curve). Presenting two tones (TT) showed that in the presence of a second tone other than 19kHz, the evoked change in the fluorescence signal ( $\Delta F/F$ ) was reduced compared to that evoked by the BF alone (Figure 3.1C). We thus inferred the inhibitory sideband as frequencies that, when presented together with BF, resulted in a reduction of responses to BF alone. This analysis revealed the presence of sideband inhibition flanking the BF (Figure 3.1D, blue curve). Figure 3.1E-G shows another Thy1 neuron with an I-shaped FRA. This neuron not only showed sideband inhibition, but also showed a facilitative TT effect (arrows in Figure 3.1F, G). As our goal is to infer the inhibitory sideband, we focused mainly on the suppressive effect.

We proceeded to record from both excitatory (Thy1-GCaMP6s (Dana et al., 2014)) and inhibitory (PV-cre and SST-cre animals with viral expression of GCaMP6s, Figure 3.1A) neurons. All mice used in this study were F1 generations from crosses with CBA/CaJ mice to ensure normal hearing throughout adulthood (Frisina et al., 2011). Table 1 lists the basic statistics of responding neurons such as

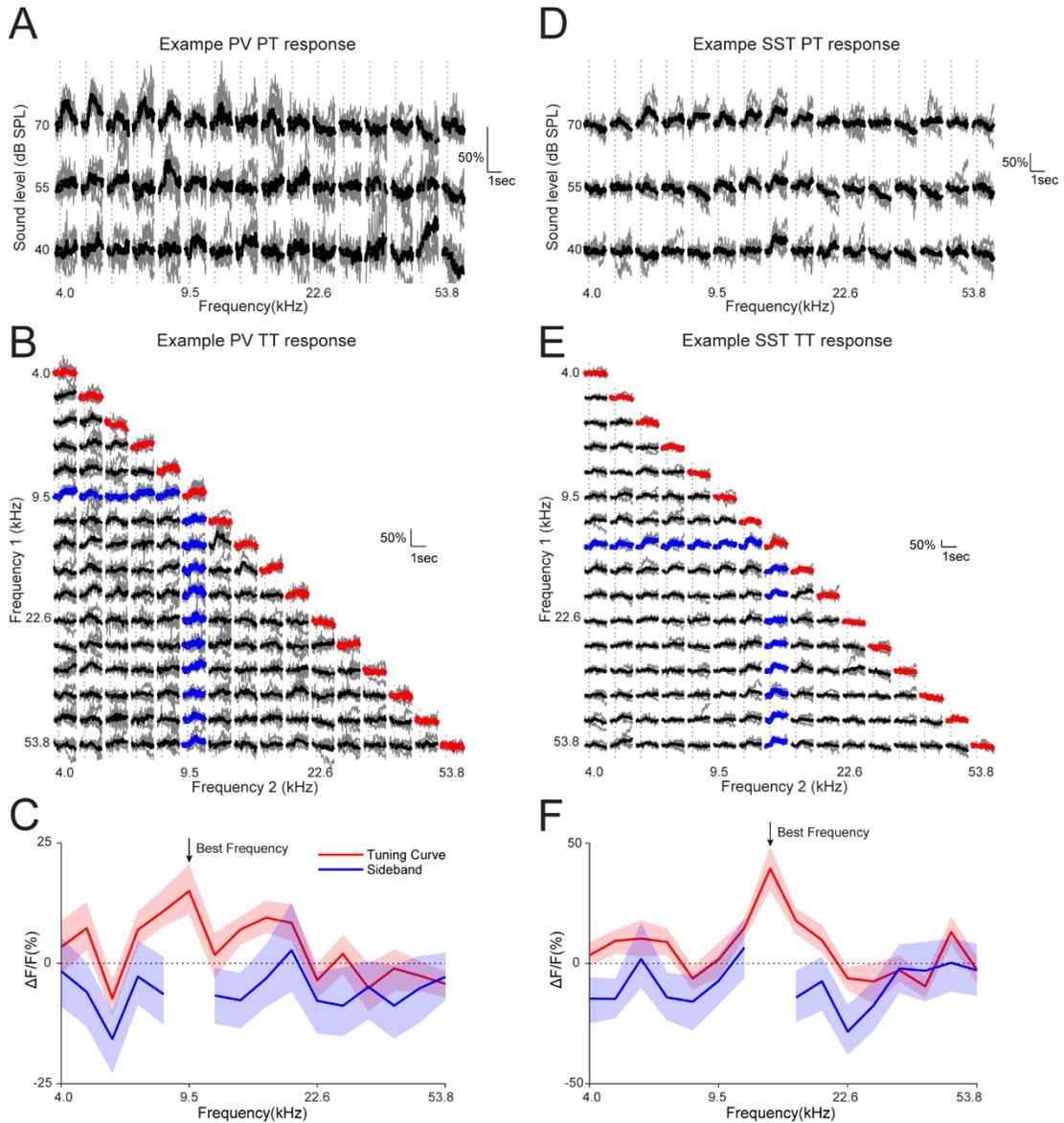
the percentage responding to PTs, to TTs or to both. We characterized a neuron as responding and analyzed it only if it responded significantly to at least one stimulus (see Materials and Methods). A total of 5576 Thy1 neurons, 1324 PV interneurons and 1451 SST interneurons passed this criterion and all subsequent analyses are focused on them. Figure 3.2 shows example responses from one PV and one SST interneuron, confirming that our approach is applicable to interneurons as well.



**Figure 3.1 Inhibitory sideband can be inferred using TT stimuli combined with 2-photon imaging of GCaMP6s across neural populations**

Example responses to PT and TT of two Thy1-GCaMP6s neurons within the same field of view and their respective tuning curves and inhibitory sidebands are shown here. (A) Experimental paradigm of the current study. Awake mice were passively listening to pure tone (PT) and two-tone (TT) stimuli while different cell classes in the left A1 were imaged. The location of A1 was determined by performing widefield imaging. Example widefield maps are shown on the right. The contour lines of different color indicate the region of fluorescence increase following the presentations of pure tones of corresponding frequencies. Scale bar = 300  $\mu\text{m}$ . The example 2-photon field of views show Thy1-GCaMP6s neurons and viral expression of GCaMP6s and mRuby in PV and SST interneurons. Scale bars = 20  $\mu\text{m}$ . (B) Example responses to PT of one Thy1 neuron. The gray traces represent responses in individual trials and average responses are plotted in black. Vertical dotted lines indicate the onset of the stimulus. The asterisks indicate significant responses determined by non-overlapping 99.9% confidence interval of  $\Delta F/F$  over pre- and post-stimulus period. For example, the pre- and post-stimulus CI was [-17%,19%] and [29%, 84%] respectively for the stimulus with the red asterisk, and thus this stimulus evoked a significant response. For the stimulus with the red cross, the pre- and post-stimulus CI was [-19%,18%] and [-22%, 11%] respectively and thus the response was not significant. (C) Example responses from the same neuron as in (B) to TT stimuli. Note that the traces on the diagonal were responses to PTs in the second stimulus set (see Materials and Methods) while off-diagonal traces were responses to TT combinations (figure inset). The traces with averages plotted in red or blue were used to construct tuning curve and sideband respectively in (D). (D)

Tuning curve and inhibitory sideband of the same neuron shown in (B) and (C). The solid lines indicate mean responses while the shaded regions show 99.9% confidence interval. For tuning curves, the  $\Delta F/F$  reflects the percentage of fluorescence change from baseline following the different PTs. For inhibitory sidebands,  $\Delta F/F$  following the BF tone was subtracted from the  $\Delta F/F$  following the TT stimuli containing the BF. A significantly negative value suggests a suppression of the response due to the presence of the second tone other than BF. (E-G) same as in (B-D) but for another Thy1-GCaMP6s neuron with an “I” shape FRA. Note that this neuron showed both TT suppression and facilitation. In (F) the arrow points to one example of TT facilitation where PTs presented alone failed to excite this neuron. (G) The sideband of this neuron showed mostly suppression except for one frequency (arrow) that evoked facilitation.



**Figure 3.2 Inhibitory sidebands can be inferred with TT stimuli in PV and SST interneurons**

(A) Example FRA of one PV interneuron. The gray traces show individual trials while the black traces show the trial average. Vertical dotted lines indicate the onset of the stimulus. (B) Example responses to TT from the same PV neuron as shown in (A). The diagonal responses were to PTs while the off-diagonal responses were to TTs that were the combinations of all PT pairs. (C) The tuning curve and inhibitory

sideband of the same PV neuron as shown in (A) and (B). The solid lines indicate mean responses while the shaded regions show 99.9% confidence intervals. (D-F) Same as in (A-C) but for an SST interneuron.

**Table 3-1 Basic response properties across cell types**

	Number of fields of view	Number of mice	Percentage responding to PTs (% , Mean±SD)	Percentage responding to TTs (% , Mean±SD)	Percentage responding to both (% , Mean±SD)
Thy1	57	6	35.5±10.9	40.9±11.5	25.4±10.3
PV	43	4	60.7±19.6	63.4±19.1	53.8±20.3
SST	65	8	79.5±13.1	81.6±13.3	74.7±15.0

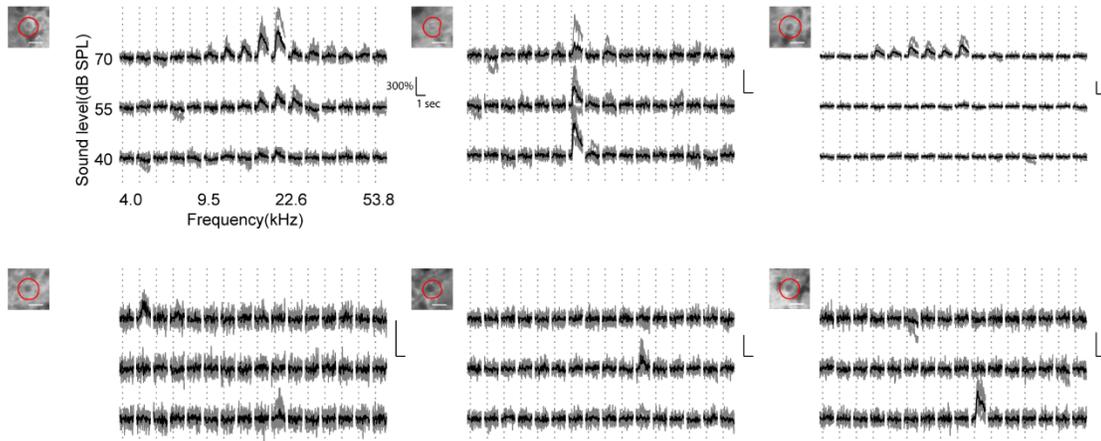
FRA shapes of A1 neurons form cell type dependent classes

We observed a high degree of variability in the shapes of FRAs of all responding neurons (Figure 3.3), similar to recordings in cats (Sutter and Schreiner, 1991; Sutter et al., 1999). We hypothesized that the shape of the FRAs could be linked to properties of sideband inhibition. We thus sought to first classify FRAs based on their shapes. In short, we aligned the FRAs at the geometric center by averaging frequencies weighted by responses across sound levels and performed K-means clustering based on PCA components that kept the 95% of the total variance in the aligned FRA. Figure 3.4A shows the t-distributed stochastic neighbor embedding (t-SNE) plot of the PCA scores, which embeds high-dimensional data such as the PCA scores for visualization in low-dimensional space, as well as generated labels and their corresponding average FRAs. These unsupervised classification results

suggest that there are at least 6 distinct types of FRAs given the stimulus set in the current study. Intuitively, cluster 1 and cluster 2 corresponded to typical V and I shaped FRAs, while the most distinct feature of cluster 3 was its wide tuning at the loudest sound level and sparser responses at lower levels. Clusters 4, 5 and 6 typically only responded to one frequency and sound level combination and thus had the sparsest responses. To further improve the accuracy of the clustering, we manually examined the labels assigned to each cell and corrected misclassifications. Among the total 8351 neurons, 4619 labels were corrected. The final labels were V, I, H, S1, S2 and S3, where “H” stands for “Horizontal” and “S” stands for “Sparse”. Figure 3.4A shows the average FRAs of the corrected labels. We quantified the variability within each cluster before and after manual correction by computing the interquartile range of the responses across neurons at each aligned frequency. The corrected clusters showed variability mainly within their respective shapes indicating that the manual classification better retains the consistency of shapes within each cluster (Figure 3.4B). We further quantified the proportion of misclassification by calculating for each final manual cluster the compositions of the original K-means result (Figure 3.4C), which shows that the misclassification happened majorly among putative V and I shaped neurons while H and S type neurons were mostly accurately classified. To further validate the distinctness of the FRA types, we plotted the response profiles either over frequency or over sound level as a function of cell types (Figure 3.4D). All FRA types had most of the frequency responses at the center of the FRA except for H type, which had the smallest slope in its cumulative curve over frequency, due to its wide tuning (Figure 3.4D, left). The differences between FRA types were more

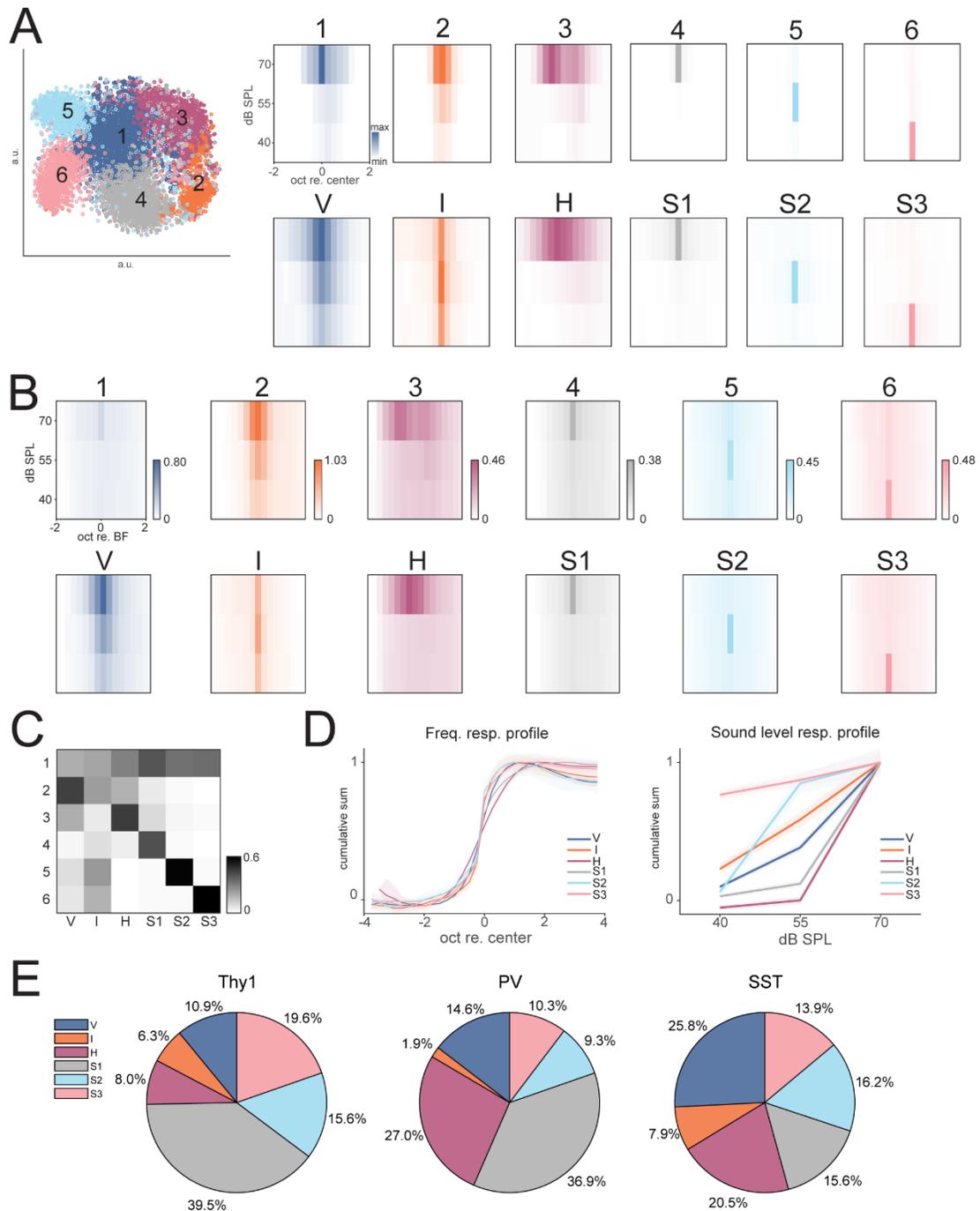
pronounced in the response profile over sound level (Figure 3.4D, right). All S types had most of the responses at the sound level to which they were most selective. I type had a rather linear profile while V and H showed supralinear profiles due to widening tuning at higher sound levels (Figure 3.4D, right). These results suggest that the labels generated by our semi-automatic classification reflect true differences in the neurons' selectivity to both frequency and sound levels.

We performed the classification on FRAs of all cell types and further quantified the proportion of FRA types within each cell type (Figure 3.4E). For Thy1 neurons, the vast majority of responding neurons belonged to the S types, consistent with the sparse coding of stimuli in sensory cortices (Hromádka et al., 2008). PV and SST neurons had a lower percentage of S type FRAs than Thy1 neurons. Most notably, both types of interneurons had higher proportion of H type FRAs than Thy1 neurons, suggesting a broadening of tuning as the sound level increases. Moreover, PV neurons were the least likely to have I shaped FRAs, consistent with their broad tuning (Li et al., 2014b). These results show a clear cell type dependent distribution of different FRA types in L2/3 of mouse A1.



**Figure 3.3 The FRAs of individual neurons show a large variability**

6 different FRAs from Thy1 neurons were plotted to show the variability of FRA shapes across neural populations. In each panel the image of the neuron was shown in the left upper corner with the red line indicating its contour. The white scale bar in the image shows 10  $\mu\text{m}$ . For FRA traces, all horizontal scale bars indicate 1 second while vertical scale bars indicate 300%  $\Delta F/F$ . All gray traces show individual trials while the black traces show the trial average. These examples differed in their sparseness of responses as well their selectivity for frequency and sound level. For example, in the second row these neurons responded only to one frequency and sound level combination and thus showed the highest level of sparseness in their FRAs.



**Figure 3.4 Classification of FRA shapes reveals distinct receptive field types**

(A) Left: t-SNE plot of PCA coefficients of aligned FRAs with data points color-coded with their corresponding K-means clusters. Right top row: average aligned FRAs of the 6 clusters as determined by K-means algorithm. Right bottom row:

average aligned FRAs of manual classification based on K-means result. The clusters 1-6 roughly correspond to the color-matched manual classification result in the bottom row. (B) The variability within the K-means cluster and that within the manually classified cluster were compared. We quantified the variability at each aligned frequency by computing the interquartile range of response amplitude at the aligned frequency across neurons. The color-matched plots share the same color axis. The manual classification results showed higher variability within their perspective shape, e.g. cluster 1 vs. V and cluster 2 vs. I. (C) To quantify the misclassification, for each final manual label the proportion of original K-means cluster is plotted. For example, the first column shows the proportion of the original clusters that constituted the final V-shape cells, i.e., 19.6% from cluster 1, 44.8% from cluster 2, 19.8% from cluster 3, 2.1% from cluster 4, 7.7% from cluster 5 and 6.1% from cluster 6. Thus, each column has the summation of one. The most misclassification happened among the putative V and I shape neurons (cluster 1 and 2) while H and S type neurons were more accurately assigned. (D) The FRA clusters differed in their frequency and sound level response profile. Each line represents the normalized cumulative summation of responses over either frequency (left) or sound level (right). For each cell the summation was normalized such that the maximum was one. The shaded regions show 95% confidence interval. (E) The proportion of FRAs types within each cell type.

### Sideband inhibition shows dependencies on FRA shape

We next sought to compare various properties of inhibitory sidebands across both cell types and FRA types. Despite the classification of FRAs, not all neurons responded to the PTs in the TT stimulus set, possibly due to the sparseness of responses or stimulus selectivity for particular frequency and sound level combinations. Table 2 shows the proportion of neurons from which we could infer inhibitory sidebands and we focused the following analysis on these subsets of neurons. First, we plotted the average tuning curves and inhibitory sidebands for each cell class (Figure 3.5A). We then quantified the width of both the tuning curve and the inhibitory sideband. Given that the shape of the tuning curves and sidebands among different neurons could be highly variable we resorted to a sparseness measure that could be applied to both the tuning curve and inhibitory sideband (see Material and Method). The fewer frequencies a neuron significantly responded to, the higher the sparseness of the tuning curve and given that the sparseness values are bound between 0 and 1, we used  $1 - \text{sparseness}$  as the width measure. We found that for all cell types, the width of the inhibitory sidebands was larger than the tuning curve width (Figure 3.5B, Table 3). Since inhibitory sidebands are thought to sharpen tuning curves (Li et al., 2014b), we hypothesized that the width of the tuning curve and the width of the inhibitory sideband would be negatively correlated. Indeed, a linear fit pooling all cell types and FRA types showed a significant negative slope between the width of tuning curve and that of the inhibitory sideband (Figure 3.5C,  $p=5.1 \times 10^{-13}$ ), suggesting a narrower tuning is associated with wider sideband

inhibition. These results are consistent with the notion that inhibitory sidebands in cortical neurons contribute to tuning curve sharpening (Li et al., 2014b).

We next investigated differences of tuning widths across cell types. The widths of the tuning curves were significantly different across cell types (ANOVA,  $p=9.9\times 10^{-35}$ ) and post-hoc multiple comparison test revealed that Thy1 neurons had narrower tuning width than both PV and SST neurons (Thy1 vs PV,  $p=9.9\times 10^{-10}$ , effect size measured by Hedges'  $g=-0.32$ ; Thy1 vs SST,  $p=9.6\times 10^{-10}$ , Hedges'  $g=-0.48$ ) while PV neurons had narrower tuning width than SST neurons ( $p=0.015$ ). However, both Thy1 and PV neurons showed wider inhibitory sideband than SST neurons (Thy1 vs SST,  $p=9.6\times 10^{-10}$ , Hedges'  $g=0.33$ ; PV vs SST,  $p=2.4\times 10^{-6}$ , Hedges'  $g=0.25$ ) while Thy1 and PV neurons did not differ ( $p=0.11$ ). Together, these results show that both tuning width and inhibitory sideband width depend on cell types. Specifically, Thy1 neurons had the narrowest tuning width among the three cell types while the inhibitory sideband width was comparable to PV neurons but wider than SST neurons. Most notably, the width of sidebands were much broader than those of the tuning curves across all cell types and FRA types, suggesting the highly selective frequency tuning which could be due to broad inhibitory synaptic inputs (Li et al., 2014b).

The differences between cell types in terms of tuning and sideband widths could be due to their BF. We thus compared the median BF across cell types and found that the median BFs were similar (Figure 3.6A). We next compared the tuning BF as a function of both cell types and FRA types (Figure 3.6B). Within specific cell types, some differences in BF exist across FRA types. Specifically, we found that

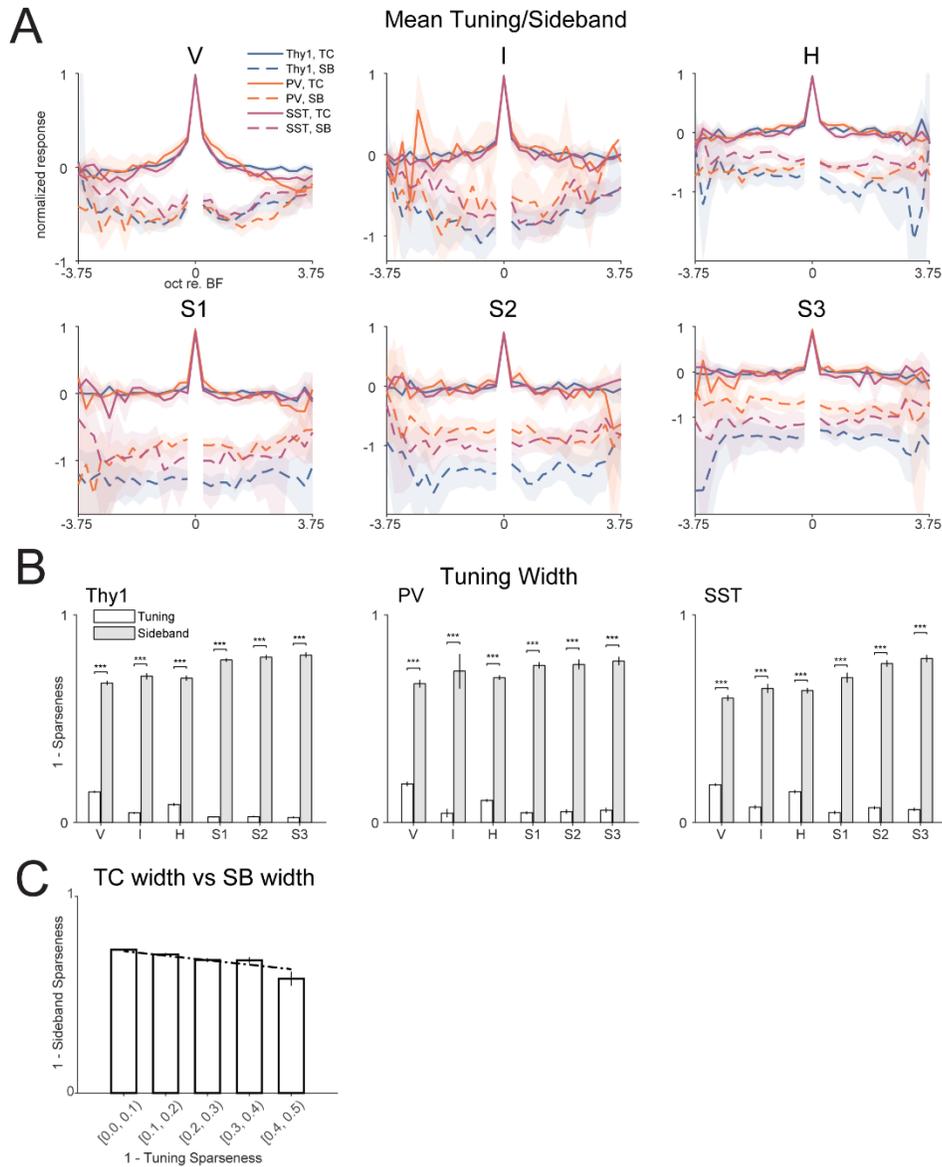
Thy1 I shape neurons had slightly higher BF median than V ( $p=0.025$ ), S1 ( $p=0.031$ ) and S2 ( $p=0.037$ ) shape neurons. The respective 95% CI of BF median difference in octaves, computed with bootstrap procedure, were [0, 0.5], [-0.25, 0.5] and [0, 0.6250]. Since these CIs contains 0, we conclude that the true difference between the BF distributions are relatively small. For PV neurons, only V and H shape neurons showed significantly different BF median ( $p=2.0\times 10^{-8}$ ) and 95% CI of the median difference was [-1, -0.5] octave. For SST neurons, V shape neurons also had a lower BF median than H shape neurons ( $p=0.0053$ ) with 95% CI of the median difference being [-1.25, -0.25] octave. Given that these differences only existed in specific FRA type pairs, they were not likely to significantly impact the results on tuning and sideband width. Similarly, we found that the tuning and sideband width were rather constant within the frequency range of the PT and TT stimuli set, while some differences exist for cells whose BFs were at the low or high end of the frequency range, which is likely due to the lack of data beyond the frequency extremes (Figure 3.6C). Together these results show that across the tonotopic axis of A1, the frequency selectivity of cortical neurons are similar and that they receive similar amount of sideband inhibition.

**Table 3-2 Proportion of neurons within each FRA types with sideband inferred**

	V	I	H	S1	S3	S3
Proportion (%)	88.8	69.4	68.1	24.5	32.9	25.4

**Table 3-3 P values, Wilcoxon signed-rank test, width of tuning curve vs. width of sideband inhibition and effect size measured by Hedges' g**

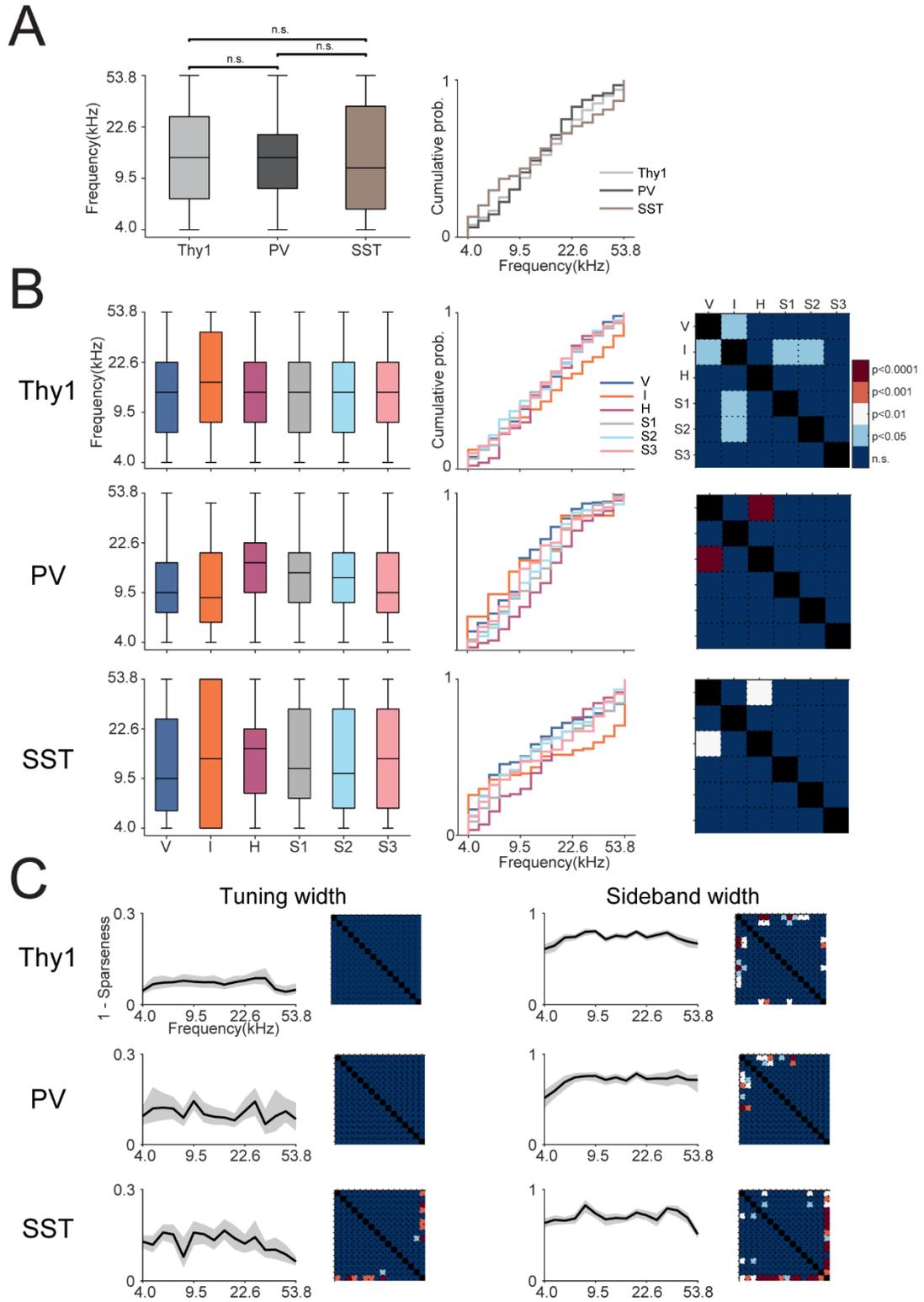
	V		I		H		S1		S3		S3	
	P value	Hedges' g										
Thy1	$3.3 \times 10^{-83}$	-2.79	$3.9 \times 10^{-39}$	-3.77	$3.6 \times 10^{-46}$	-3.64	$6.9 \times 10^{-78}$	-5.41	$1.1 \times 10^{-36}$	-5.71	$3.4 \times 10^{-37}$	-5.60
PV	$1.5 \times 10^{-26}$	-2.47	$9.8 \times 10^{-4}$	-3.04	$1.5 \times 10^{-40}$	-3.92	$2.3 \times 10^{-25}$	-4.95	$1.6 \times 10^{-10}$	-5.20	$2.4 \times 10^{-10}$	-5.86
SST	$6.4 \times 10^{-49}$	-2.13	$1.2 \times 10^{-17}$	-3.38	$5.2 \times 10^{-39}$	-2.76	$8.3 \times 10^{-18}$	-3.57	$6.5 \times 10^{-23}$	-4.92	$1.8 \times 10^{-15}$	-5.57



**Figure 3.5 All cell types and FRA types show broader inhibitory sidebands than tuning curves**

(A) The average tuning curve (solid lines) and inhibitory sideband (dash lines) were plotted as a function of both cell types and FRA types. TC: tuning curve; SB: sideband. For each cell, the amplitude of the tuning curve and that of the inhibitory sideband were normalized to the amplitude of the BF. Then the normalized tuning curve and inhibitory sideband were aligned at BF in order to construct both the mean

and the confidence interval using bootstrap procedures. (B) The widths of the tuning curves and the inhibitory sidebands as measured by 1-Sparseness were plotted as a function of cell types and FRA types. Inhibitory sidebands were considerably larger in width than those of the tuning curves. '\*\*' indicates  $p < 0.01$ . '\*\*\*' indicates  $p < 0.001$ . (C) Across all cells, the width of the tuning curve was negatively correlated with the width of the inhibitory sideband. The bar graph shows the inhibitory sideband width binned according to the width of the tuning curve. The dot-dash line shows the linear fit ( $y = 0.73 - 0.23x$ ,  $p = 5.1 \times 10^{-13}$ ).



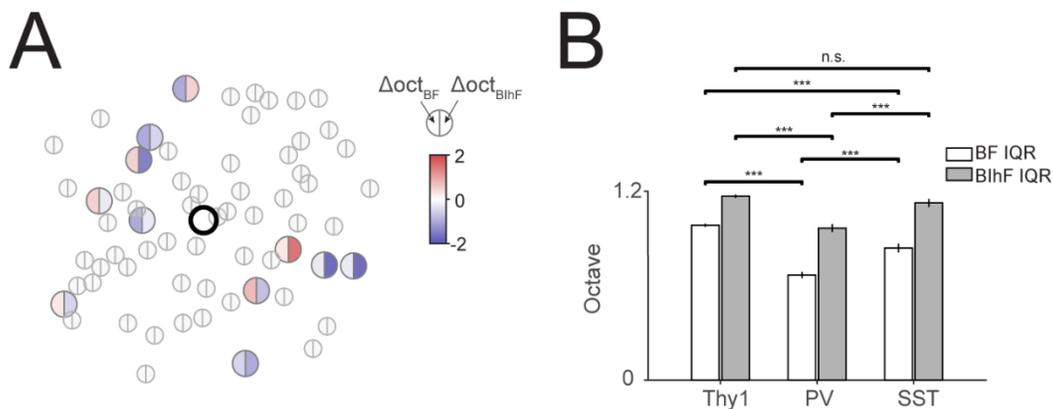
**Figure 3.6** BF of tuning curves were not different across cell types

(A) Boxplot and cumulative distribution function of tuning BF as a function of cell types. Median BF across different cell types were similar. (B) Boxplot and cumulative distribution function (left and mid column) of tuning BF as a function of both cell types and FRA types. The rightmost column shows the pairwise comparison significance of tuning BF as a function of FRA types. BF difference was only found in between specific FRA subtypes within each cell type. (C) The width of tuning curve and inhibitory sideband were largely similar across tuning BF in all cell types, except for at the low or high frequency end, which is likely due to the lack of data beyond the frequency range chosen for this study.

Inhibitory sidebands of local populations show higher degree of heterogeneity than frequency preference

Tonotopy on the mesoscale is a defining characteristic of A1, but on a finer spatial resolution such organization is largely lost as individual excitatory neurons in a local area can have heterogeneous frequency selectivity (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Winkowski and Kanold, 2013; Kanold et al., 2014; Maor et al., 2016; Liu et al., 2019). Since we here show the presence of inhibitory sidebands, we investigated whether such heterogeneity exists in the inhibitory sideband of different cell types in a local area. We quantified the heterogeneity of local tuning by computing the interquartile range (IQR) of the BF within a radius of 100 $\mu$ m. A large IQR would indicate a more diversely tuned local populations. Similarly, we defined the best inhibitory frequency (BIhF) as the frequency evoking

the strongest inhibition in the sideband and quantified the IQR of BIhF (Figure 3.7). A two-way ANOVA (cell type  $\times$  BF/BIhF) revealed a significant main effect of cell types on IQR ( $p=6.1\times 10^{-47}$ ). Specifically, Thy1 neurons had greater overall heterogeneity than PV neurons (post-hoc multiple comparison,  $p=9.6\times 10^{-10}$ , effect size as measured by Hedges'  $g=0.6050$ ) and SST neurons ( $p=1.1\times 10^{-9}$ , Hedges'  $g=0.2380$ ), while PV neurons showed less heterogeneity than SST neurons ( $p=9.6\times 10^{-10}$ , Hedges'  $g=-0.2540$ ). These results are consistent with in vivo patch clamp recordings showing a higher level of heterogeneity in excitatory than PV neurons (Maor et al., 2016). Second, the main effect of BF versus BIhF was also significant ( $p=6.5\times 10^{-57}$ ) with IQR of BIhF higher than IQR of BF across cell types ( $p=1.0\times 10^{-10}$ , Hedges'  $g$ : Thy1, -0.3488; PV, -0.4971; SST, -0.3379), suggesting that the heterogeneity of inhibitory sidebands was greater than that of tuning of local populations. This heterogeneity of inhibitory sidebands suggests that diverse sources of functional inhibition as an aggregate result in the inhibitory sideband. Lastly, the interaction term (cell type  $\times$  BF/BIhF) was also significant ( $p=2.8\times 10^{-3}$ ). Specifically, the difference between BF and BIhF IQR within Thy1 neurons were smaller than those within PV and SST neurons (ANOVA and multiple comparison, Thy1 vs PV,  $p=0.001$ , Hedges'  $g=-0.1722$ ; Thy1 vs SST,  $p=0.044$ , Hedges'  $g=-0.0931$ ). Together, these results suggest that the combined heterogeneity in the local populations' tuning and inhibitory sideband could further diversify a neuron's response to spectrally complex stimuli and thus makes its responses more selective.



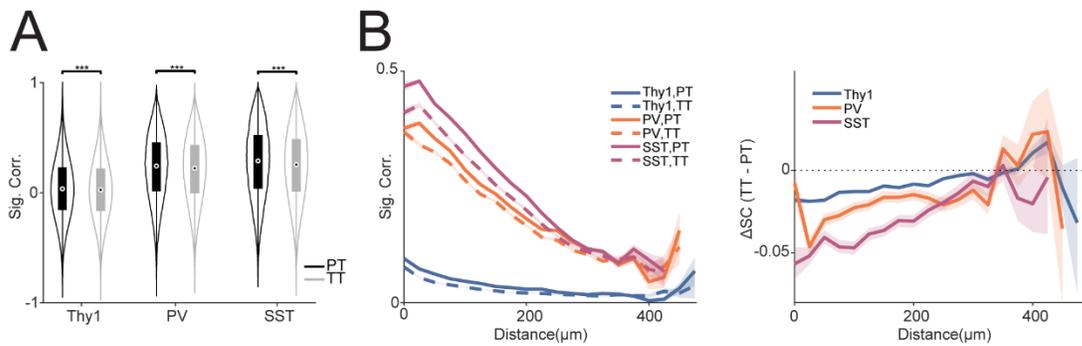
**Figure 3.7 Inhibitory sidebands show more local heterogeneity than tuning curves**

(A) Cartoon for IQR calculation. The cell in question is represented as the black circle at the center. Cells within a 100µm radius are plotted. The left and right half of the circle color code the difference of BF and BIhF in octave respectively with the center cell. Grey circles represent non-responding cells. The IQRs are then calculated taking the interquartile range of BF and BIhF  $\Delta$ oct. (B) The heterogeneity of the selectivity of the local tuning and inhibitory sideband was quantified by calculating interquartile range (IQR) of best frequency (BF) or best inhibitory frequency (BIhF) respectively. IQR of BIhF was larger than that of BF across all cell types. ‘\*\*\*’ indicates  $p < 0.001$ .

### The presence of the second tone decorrelates neuronal responses

The differences in the IQR of BF and BIhF suggest that the introduction of the second tone influences neuronal coding of the primary tone on the population level. Thus, to measure this influence we next quantified the signal correlations (SC) to the

primary tone with or without the second tone (Figure 3.8A). We performed a two-way ANOVA to determine the dependence of SC on both cell types and the addition of the second tone. The main effect of cell type was significant ( $p < 0.001$ ) and so was the main effect of adding the second tone ( $p < 0.001$ ). Specifically, across all cell types the SC of TT was lower compared to that of PT ( $p = 1.0 \times 10^{-10}$ ). Regardless of stimulus type, SCs of Thy1 cells were lower than both PV and SST neurons (Thy1 vs PV,  $p = 9.6 \times 10^{-10}$ , Hedges'  $g = -0.6534$ ; Thy1 vs SST,  $p = 9.6 \times 10^{-10}$ , Hedges'  $g = -0.7917$ ), while SCs of SST neurons were highest among the three cell types (PV vs SST,  $p = 9.6 \times 10^{-10}$ , Hedges'  $g = -0.1253$ ). This suggests that PV and SST neurons are more functionally homogeneous than Thy1 neurons, likely due to convergent local input, consistent with results by Maor et al. (2016). In all cell types, SC of TT were lower compared to SC of PT (Thy1,  $p = 2.1 \times 10^{-8}$ , Hedges'  $g = 0.036$ ; PV,  $p = 2.1 \times 10^{-8}$ , Hedges'  $g = 0.065$ ; SST,  $p = 2.1 \times 10^{-8}$ , Hedges'  $g = 0.095$ ). To investigate whether the addition of a second tone caused any change to the spatial pattern of neural correlations, we plotted the SCs against the distance of the cell pairs (Figure 3.8B). The SCs of PT and TT decreased over the distance in all cell types and the SCs of TTs were consistently below those of PTs (Figure 3.8B), suggesting a network level decorrelation of neural responses by the addition of the second tone. These results suggest that a spectrally complex stimulus would make neural responses sparser and effectively more selective to spectral features. These results also indicate that TT responses cannot be readily predicted from responses to PTs alone and that the linear and nonlinear frequency interactions in TT responses need to be characterized.



**Figure 3.8 The presence of the second tone decorrelates neuronal responses**

(A) Violin plot showing the distribution of signal correlations (SCs) as a function of both cell types and stimuli (PT vs TT). SCs of TT responses between pairs of neurons were significantly reduced compared to SCs of PT across all cell types. ‘\*\*\*\*’ indicates  $p < 0.001$ . (B) Left: SCs as function of distance, cell types and stimulus (PT vs. TT). SCs of TTs were smaller in value than SCs of PTs across distance. Solid lines correspond to SCs of PTs while dash lines correspond to SCs of TT. The shaded regions show 95% confidence interval. Right: the difference between SCs of TT and PTs as a function of distance and cell types. Shaded regions show 95% confidence interval.

### Nonlinear frequency interactions depend on both cell types and FRA types

Frequency interactions in the auditory system can be linear or nonlinear (Escabi and Schreiner, 2002). Our experimental design allowed us to investigate the degree of nonlinear interactions between frequencies beyond simple TT suppression. Specifically, if a neuron behaves like a linear filter, then its response to the TT stimuli would be the linear summation to the responses to each frequency presented in

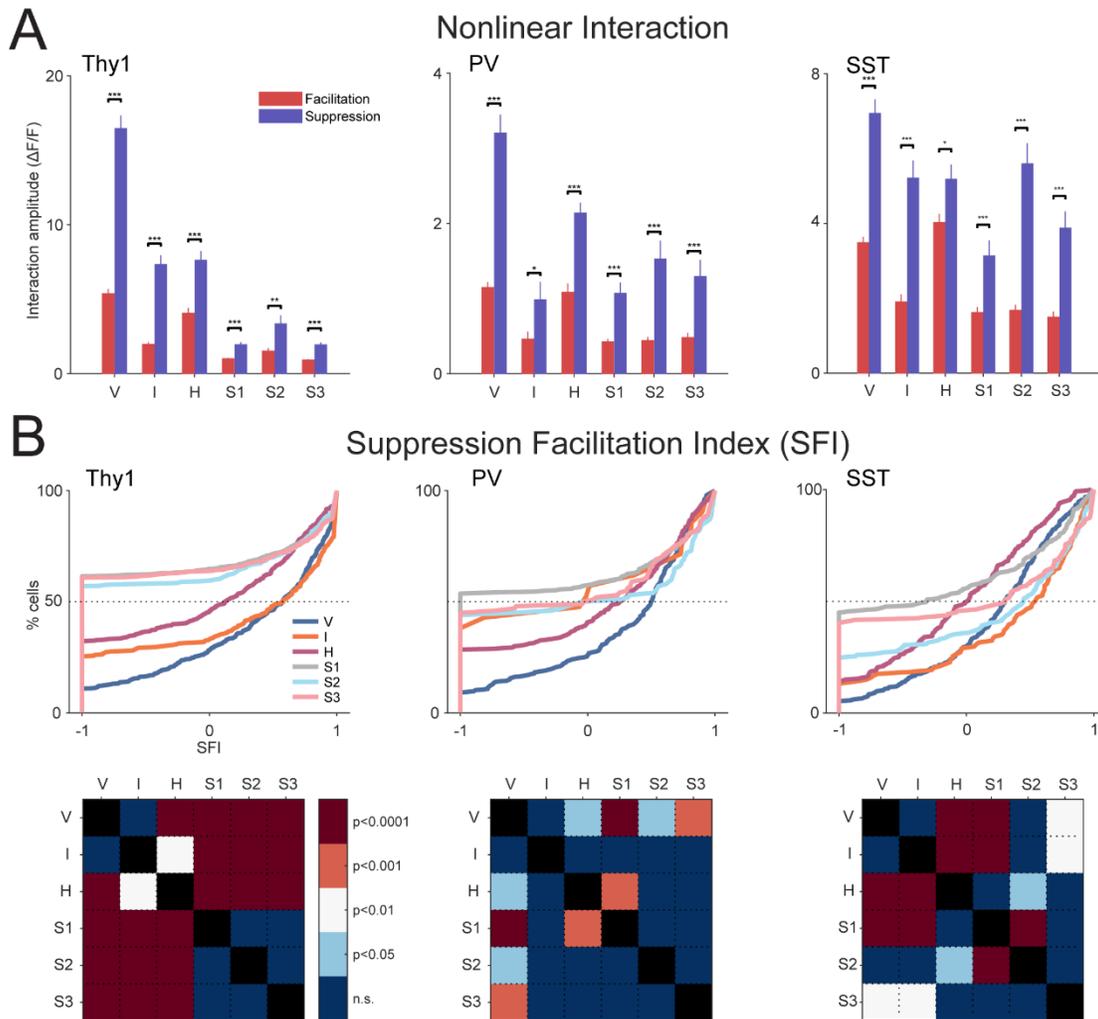
isolation:  $r(F_1 + F_2) = r(F_1) + r(F_2)$ . Any deviation from the linear assumption signals the presence of nonlinear interactions. We quantified the degree of nonlinear interactions across cell types and separated them based on whether the response to the TT stimulus was larger (facilitation) or smaller (suppression) than that predicted by the linear assumption (Figure 3.9A). All cell types showed nonlinear effects and significantly more suppression than facilitation (two-way ANOVA, main effect of facilitation vs. suppression: Thy1,  $p=1.6\times 10^{-76}$ ; PV,  $p=1.1\times 10^{-13}$ ; SST,  $p=2.4\times 10^{-35}$ ). Across all FRA types, except H type in SST neurons, all other FRA types showed the same dominance of suppressive nonlinear interactions (Wilcoxon rank sum test, see Table 4 for p values and effect size), which suggests the degree of nonlinear interactions could further depend on specific cell type and FRA type combinations. This analysis pooled the degrees of facilitation and suppression across cells and thus reflected the properties on a population basis. To investigate the bias of facilitation and suppression of individual cells, we calculated a Suppression Facilitation Index (SFI), which had values between -1 and 1, with 1 indicating only suppressive interactions and -1 indicating only facilitative interactions. The cumulative distributions of SFIs are shown in Figure 3.9B. For Thy1 cells, V, I and H type neurons had the most bias towards suppression while SFI distributions among all S types were not significantly different from each other (Figure 3.9B). While V and I type neurons had the same SFI distribution, both were higher than that of H type neurons (V vs H,  $p=2.1\times 10^{-8}$ , Hedges'  $g=0.5018$ ; I vs H,  $p=0.001$ , Hedges'  $g=0.3081$ ), suggesting a broader tuning width was associated with a lesser degree of suppressive frequency interactions. For PV neurons, SFI values were smaller in H

type neurons than in V type neurons ( $p=0.011$ , Hedges'  $g=0.3451$ ). For SST neurons, SFI values were also smaller in H type neurons than in both V and I type neurons (V vs H,  $p=3.4\times 10^{-5}$ , Hedges'  $g=0.4671$ ; I vs H,  $p=4.4\times 10^{-5}$ , Hedges'  $g=0.5799$ ).

Therefore, the shape of FRAs could serve as an indicator of the degree of nonlinear interactions and thus delineates functionally separate classes of cells. Specifically, the results above suggest that H type neurons might be more involved in integrating energy over frequency bands while V and I neurons serve as differentiators of the frequency content.

**Table 3-4 Wilcoxon rank sum test p values and Hedges' g, facilitation vs suppression across cell type and FRA types**

	V		I		H		S1		S2		S3	
	P value	Hedges' g	P value	Hedges' g	P value	Hedges' g	P value	Hedges' g	P value	Hedges' g	P value	Hedges' g
Thy1	$4.6 \times 10^{-31}$	-0.68	$4.7 \times 10^{-16}$	-0.64	$1.0 \times 10^{-6}$	-0.33	$9.7 \times 10^{-7}$	-0.15	$2.8 \times 10^{-3}$	-0.14	$1.1 \times 10^{-7}$	-0.23
PV	$1.7 \times 10^{-14}$	-0.82	$4.9 \times 10^{-2}$	-0.56	$9.2 \times 10^{-9}$	-0.44	$1.7 \times 10^{-5}$	-0.28	$2.4 \times 10^{-5}$	-0.55	$5.4 \times 10^{-4}$	-0.43
SST	$1.09 \times 10^{-16}$	-0.65	$4.4 \times 10^{-10}$	-0.84	$1.2 \times 10^{-2}$	-0.21	$8.2 \times 10^{-4}$	-0.32	$1.8 \times 10^{-11}$	-0.66	$7.5 \times 10^{-7}$	-0.52



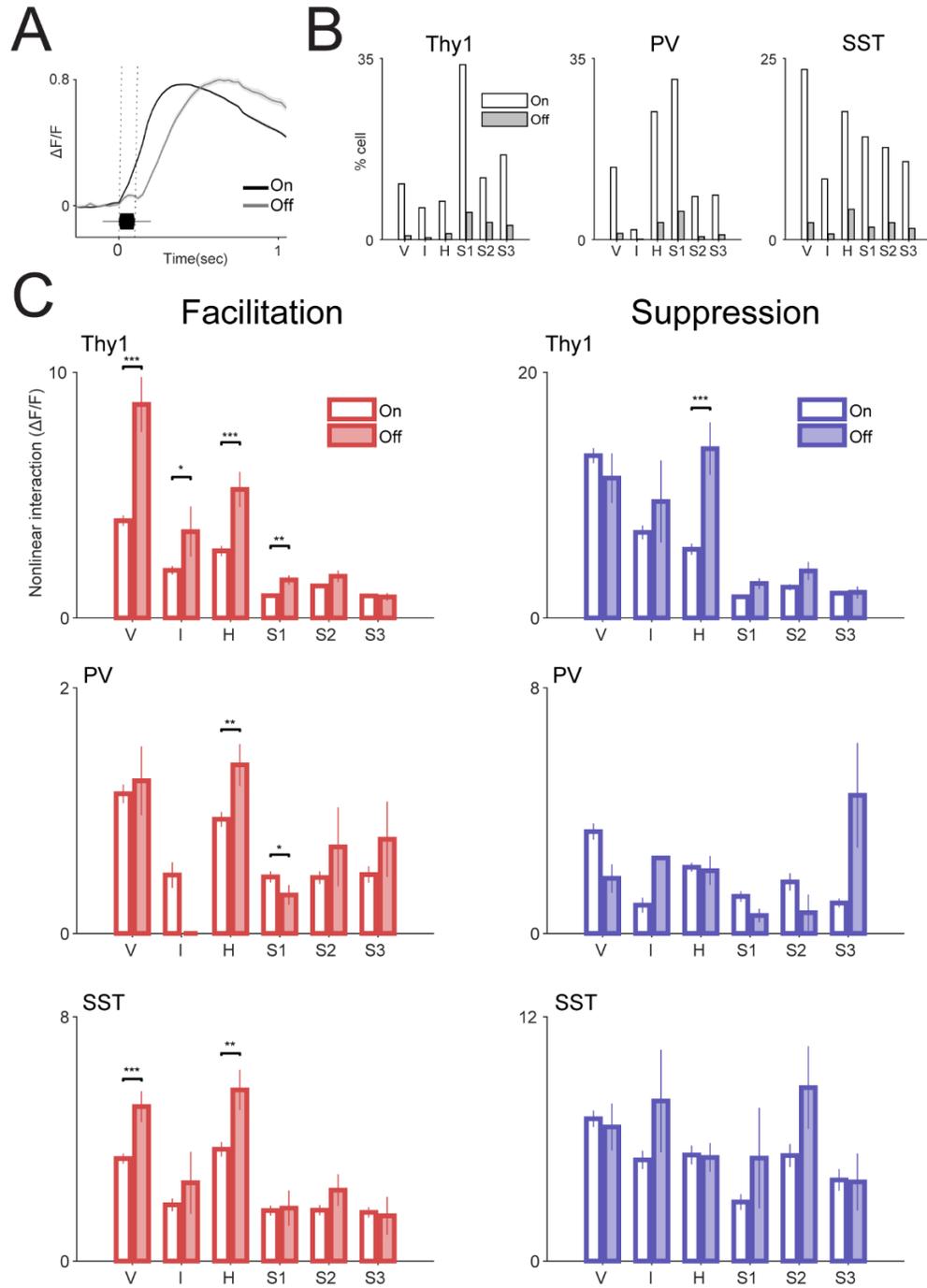
**Figure 3.9 Nonlinear frequency interactions show prominent suppression among all cell types while the relative facilitation/suppression strength depends specific cell/FRA type combination**

(A) The amplitude of facilitative and suppressive nonlinear frequency interactions as a function of FRA types and cell types. In all cell types, the average suppressive interactions were larger in amplitude than facilitative interactions. ‘\*’ indicates  $p < 0.05$ . ‘\*\*’ indicates  $p < 0.01$ . ‘\*\*\*’ indicates  $p < 0.001$ . (B) Upper row: Suppressive Facilitative Index (SFI, see Material and Method) as a function of FRA types and cell types. SFI measures the bias of suppressive or facilitative interactions within individual cells. 1 indicates pure suppressive interactions while -1 indicates pure facilitative interactions. Lower row: matrices showing p values of pairwise comparisons between different FRA types within each cell types.

#### Nonlinear frequency interactions as a function of onset and offset response

Our previous study demonstrated that dynamic sound features such as sound onset and offset are processed in parallel processing pathways in the auditory cortex (Liu et al., 2019). Thus, we sought to investigate whether these two pathways might also process spectral information differently and whether differences exist between these pathways in terms of the degree of nonlinear frequency interactions. We first quantified the latency of the cellular responses to identify onset and offset neurons (Figure 3.10A). We labeled neurons as either onset or offset neuron (see Materials and Methods) since they largely consisted of separate groups of neurons (Liu et al., 2019). The majority of neurons were onset neurons across cell types and FRA types (Figure 3.10B). The lower fraction of offset neurons compared to prior studies (Liu et al., 2019), is likely due to shorter stimulus duration (100 ms vs 2 seconds) and

intertrial intervals (1.5 seconds vs 5 to 5.5 seconds). We next compared the amplitude of facilitative or suppressive interactions between onset and offset neurons on a population basis (Figure 3.10C). In Thy1 neurons, offset V, I, H and S1 type neurons had larger facilitative interactions than onset neurons of the same FRA type (Figure 3.10C, Wilcoxon rank sum test, V:  $p=2.6\times 10^{-7}$ , Hedges'  $g=-0.93$ ; I:  $p=0.022$ , Hedges'  $g=-0.50$ ; H:  $p=1.5\times 10^{-5}$ , Hedges'  $g=-0.59$ ; S1:  $p=7.8\times 10^{-3}$ , Hedges'  $g=-0.32$ ). In contrast, only H type neurons showed larger suppressive interactions for offset neurons (Wilcoxon rank sum test,  $p=3.9\times 10^{-5}$ , Hedges'  $g=-0.82$ ). PV neurons generally showed the same degree of facilitative and suppressive interactions across onset and offset neurons, except that PV H type offset neurons showed larger facilitative interactions than corresponding onset neurons (Wilcoxon rank sum test,  $p=0.003$ , Hedges'  $g=-0.42$ ) while PV S1 offset neurons showed slightly smaller facilitative interactions (Wilcoxon rank sum test,  $p=0.026$ , Hedges'  $g=0.17$ ). SST neurons also showed similar degree of nonlinear interactions across onset and offset neurons except that V and H type offset neurons had larger facilitative interactions than V and H onset neurons (Wilcoxon rank sum test, V:  $p=2.2\times 10^{-7}$ , Hedges'  $g=-0.59$  ; H:  $p=1.5\times 10^{-5}$ , Hedges'  $g=-0.50$ ), which is similar to what seen in Thy1 neurons. These results show that a subset of offset neurons tended to have larger nonlinear facilitative interactions than their onset counterparts. This suggests that the offset pathway not only conveys temporal information, it also tends to integrate spectral information supralinearly to a larger degree than onset neurons, and thus could be more suited to encode the general energy level in the stimulus.



**Figure 3.10 Nonlinear frequency interactions as a function of response timing**

(A) The average traces of onset and offset responses pooled from all responding

neurons. Dotted vertical lines mark the onset and offset of the stimulus respectively.

(B) Proportion of onset and offset neurons as a function of cell type and FRA types.

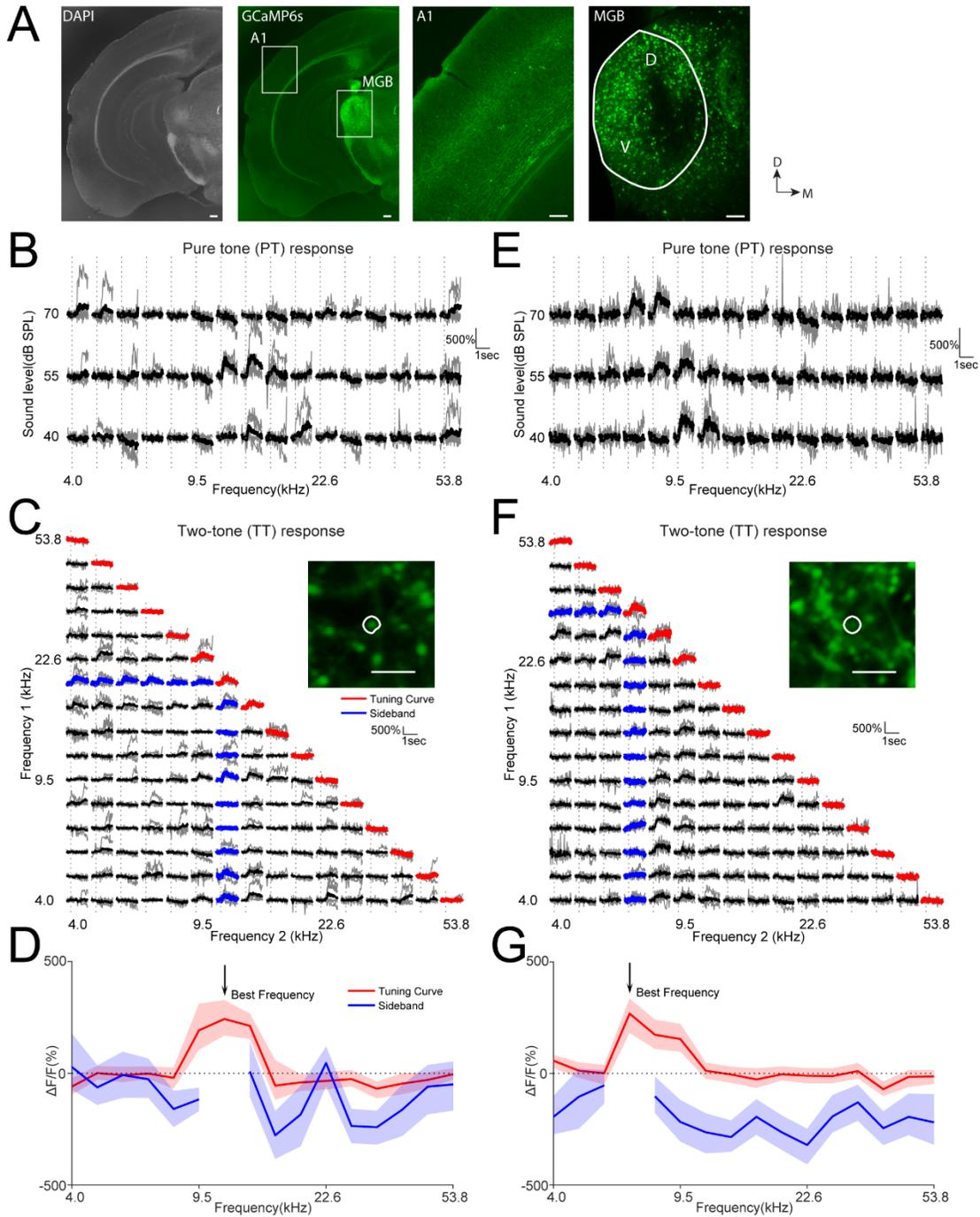
(C) Nonlinear frequency interactions as a function of response timing, FRA types and cell types. A subset of offset neurons showed a higher degree of facilitative interactions than their onset counterparts. ‘\*’ indicates  $p < 0.05$ . ‘\*\*’ indicates  $p < 0.01$ . ‘\*\*\*’ indicates  $p < 0.001$ .

### MGB responses exhibit sideband inhibition

A1 neurons receive feedforward input from the auditory thalamus or medial geniculate body (MGB) and it is thus possible that the sideband inhibition seen in cortical neurons is partially inherited from thalamocortical input. To investigate this question, we injected AAV.CamKII.GCaMP6s.WPRE.SV40 into the MGB and imaged MGB terminals in A1 (Liu et al., 2019) (n=3 mice, 17 FOVs, Figure 3.11A). Our labeling of MGB neurons included ventral and dorsal division and thus the imaged boutons likely consisted of both first and second order thalamic input (Figure 3.11A). The MGB terminals had denser labeling in L1 and L4, consistent with previous findings (Vasquez-Lopez et al., 2017). We imaged at a depth of  $117 \mu\text{m} \pm 19.5 \mu\text{m}$  and as L1 and L3b/4 MGB boutons show similar tuning (Vasquez-Lopez et al., 2017), the PT and TT responses were not likely affected by this choice of depth. Among all putative boutons recorded,  $47.1\% \pm 12.2\%$  responded to at least one PT stimulus and  $54.2\% \pm 12.8\%$  responded to at least one TT stimulus.  $40.3\% \pm 12.6\%$  responded to both PT and TT stimuli. Figure 3.11B-G show two example boutons

with respective FRAs and responses to TT and the corresponding sideband inhibition. These results show that sideband inhibition exists in thalamocortical input. Similar to our cellular data, we also observed considerable variability in the shapes of the MGB terminal FRAs and following the same approach as above we classified the FRA shapes (Figure 3.12A). However, the MGB clusters were not as separated as seen in cellular data (Figure 3.2B), which could be due to the limited signal to noise ratio of bouton imaging that resulted in failed detection of smaller responses. Nevertheless, the average FRAs of the clusters resembled those seen in cellular data (Figure 3.12A). Specifically, cluster 1 roughly corresponded to a combination of narrow ‘V’ and ‘I’ shapes while cluster 2 corresponded to putative ‘H’ shape. The average FRA of cluster 3 suggests a broad tuning, and yet the average tuning curve suggests otherwise (Figure 3.12A, B). Thus, this cluster has a large within cluster variation and might not have a well-defined FRA shape. We speculate that this cluster might represent responses from non-lemniscal pathway such as dorsal MGB (dMGB), as our injections were not restricted to particular divisions of MGB (Figure 3.12A) and dMGB responses are weakly tuned (Vasquez-Lopez et al., 2017). Clusters 4 and 5 likely corresponded to ‘S1’ shape and cluster 6 corresponded to ‘S2’ shape. However, ‘S3’ type was not recovered in MGB data. These results show that the responses of individual MGB boutons to tones also showed large FRA variability. We then proceeded to quantify the width of both tuning curves and inhibitory sidebands in these clusters. On average, all clusters showed much broader inhibitory sidebands than tuning curves (Figure 3.12B, C, Wilcoxon sign rank test, cluster 1:  $p=8.3\times 10^{-96}$ , Hedges’  $g=-3.20$ ; cluster 2:  $p=2.8\times 10^{-105}$ , Hedges’  $g=-3.97$ ; cluster 3:  $p=2.7\times 10^{-277}$ ,

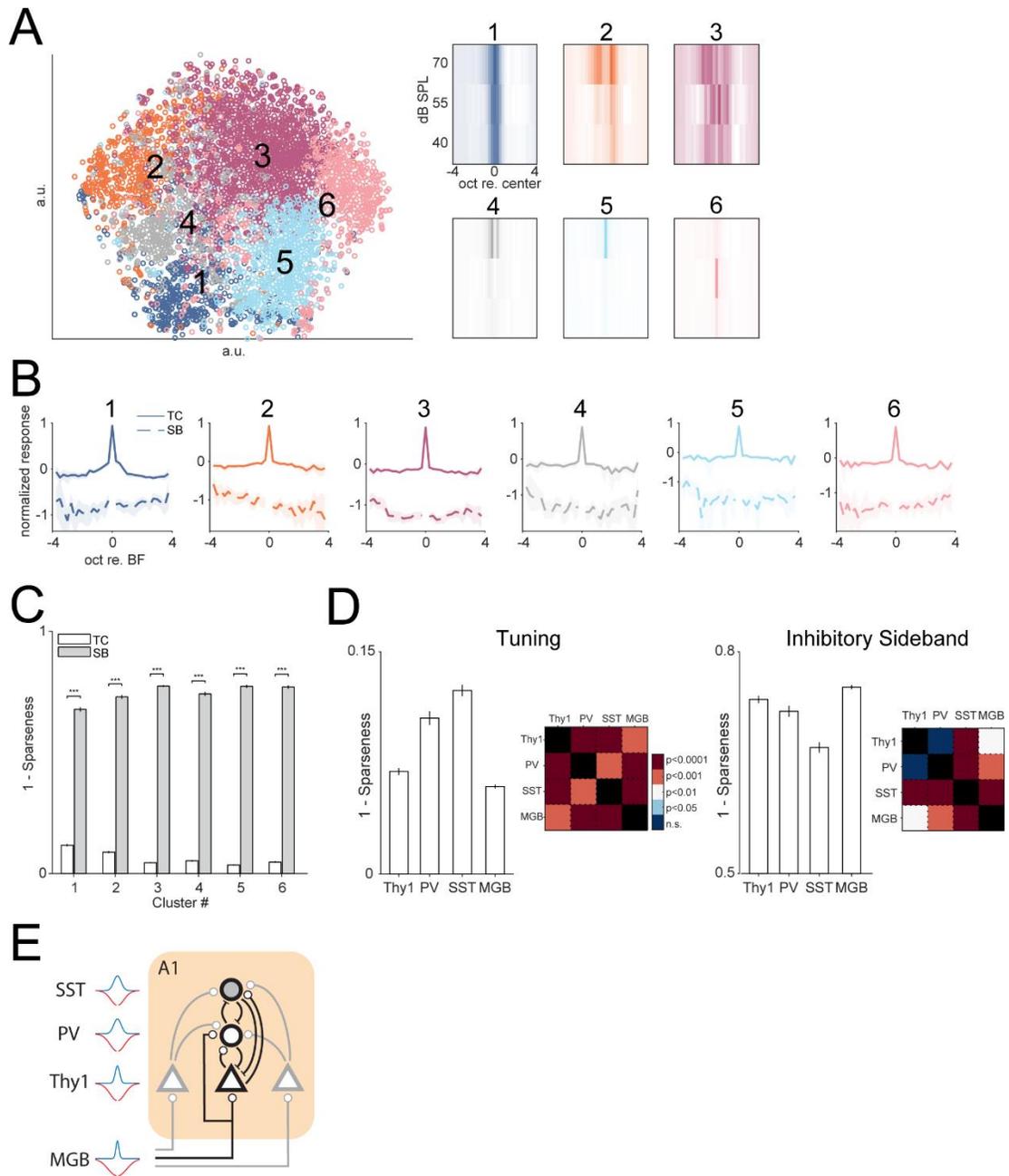
Hedges'  $g=-5.49$ ; cluster 4,  $p=4.7\times 10^{-85}$ , Hedges'  $g=-4.79$ ; cluster 5:  $p=8.1\times 10^{-129}$ , Hedges'  $g=-5.66$ ; cluster 6:  $p=1.7\times 10^{-104}$ , Hedges'  $g=-5.36$ ), also similar to our cellular data (Figure 3.5B). These data suggest that MGB input might contribute to sideband structures seen in cortical neurons. Finally, we compared the sparseness of tuning width and inhibitory sideband of MGB terminals with those of cellular data (Figure 3.12D). We found that MGB terminals showed significantly narrower tuning than all cortical cell types examined (ANOVA and multiple comparison, MGB vs Thy1,  $p=7.8\times 10^{-4}$ , Hedges'  $g=-0.11$ ; MGB vs PV,  $p=3.8\times 10^{-9}$ , Hedges'  $g=-0.49$ ; MGB vs SST,  $p=3.8\times 10^{-9}$ , Hedges'  $g=-0.67$ ). In contrast, MGB terminals showed the broader inhibitory sideband than those of cortical neurons (ANOVA and multiple comparison, MGB vs Thy1,  $p=0.0097$ , Hedges'  $g=0.086$ ; MGB vs PV,  $p=3.5\times 10^{-4}$ , Hedges'  $g=0.17$ ; MGB vs SST,  $p=3.8\times 10^{-9}$ , Hedges'  $g=0.42$ ). These results are consistent with narrower tuning of MGB neurons relative to A1 neurons in awake marmoset (Bartlett et al., 2011). Together, MGB terminals showed more pronounced sideband inhibition than the three cortical neuron types examined, which suggests that the narrowly tuned MGB feedforward input serves as the backbone for cortical inhibitory sidebands and that the wider tuning of cortical neurons reflect the differential patterns of convergence of connectivity onto different types of cortical neurons by both feedforward thalamocortical and intracortical input (Figure 3.12E).



**Figure 3.11 MGB terminals show inhibitory sideband structures**

(A) Brain slice showing GCaMP6s expression in MGB and MGB terminals in A1.

Left two images show DAPI staining and GCaMP6s epifluorescence at low magnification (2×). Scale bars represent 200 μm. Right two images show high magnification (10×) views of outlined areas. Scale bars represent 100 μm. Axon terminal GCaMP6s expressions can be seen in L1 and L4. (B-G) Two example MGB boutons' response to PT and TT are shown. (B) FRA of one example bouton. (C) Example responses to TT stimuli. Gray traces represent individual trials in both (B) and (C). The blue and red average traces indicate the responses used to construct the tuning curve and the inhibitory sideband in (D). The inset shows the image of the bouton, with the white contour line outlining the ROI mask. The scale bar represents 5 μm. (D) The tuning curve and inhibitory sideband of the bouton. (E-G) Same as in (A-C) but for another bouton.



**Figure 3.12 MGB terminals show narrow tuning and prominent sideband inhibition**

(A) t-SNE plot showing the distribution of center aligned FRAs of MGB terminal FRAs. The colors indicate the clusters from K-means algorithm. Right: average center aligned FRAs of the K-means clusters. (B) Average tuning curves and

inhibitory sidebands of clustered MGB terminals. All clusters showed narrow tuning and broad inhibitory sidebands. (C) Bar plot showing the width of tuning and inhibitory sideband as a function of clusters. In all clusters, the inhibitory sidebands were much broader than tuning curves. ‘\*\*\*’ indicates  $p < 0.001$ . (D) Bar plot comparing tuning width and sideband width among different cortical cell types and MGB terminals. MGB terminals show narrower tuning than cortical neurons while having broader inhibitory sidebands. (E) Cartoon showing a model for cortical tuning and sideband inhibition. The cortical neurons partially inherit inhibitory sideband structures from the thalamocortical MGB input and the width of tuning and sideband inhibition reflect the differential convergence of input within intracortical circuits, i.e. from neurons tuned to other frequencies (gray triangles and connections). Triangles represent Thy1 neurons while white and gray circles represent PV and SST neurons respectively.

## **Discussion**

We used a variant of the TT paradigm suitable for two-photon imaging and characterized the sideband inhibition of excitatory neurons (Thy1) and two classes of inhibitory neurons (PV and SST) in A1. Imaged cells clustered into subtypes based on the shapes of FRA and the degree of sideband inhibition and nonlinear frequency interactions. This suggests that sound information in A1 is locally processed in distinct neuronal population with varying degrees of complex integration of spectral information. The wide inhibitory sidebands suggest a tight control of frequency

tuning by inhibition. The mostly suppressive nonlinear frequency interactions also suggest inhibition as a key modulator of spectral integration and thus plasticity. Moreover, we found that cells showing onset and offset preferences showed different spectral integration, suggesting that these cells form different neuronal circuits. Finally, we found that MGB terminals show inhibitory sidebands, suggesting that feedforward inputs contribute to cortical sideband inhibition.

A1 neurons formed 6 distinct clusters based on their FRA. These clusters differed in their apparent shapes, namely in their frequency and sound level selectivity. In excitatory Thy1 neurons, the majority (~75%) of the responding cells belonged to the S types where their FRAs were patchy and sparse while ~25% of the responding cells belonged to V, I or H types. With an overall response rate of ~35%, this suggests about 9% all excitatory cells have “classic” V-shaped FRAs, consistent with a sparse coding regime with only about 5% auditory responsive neurons (Hromádka et al., 2008). Notably, our results suggest a negative correlation between tuning and inhibitory sideband width. Together, these results suggest that functional inhibition controls the sparseness and tuning width of responses in A1, consistent with the contribution of PV interneurons to sparse coding of local pyramidal cells (Liang et al., 2018). The wide range of inhibitory sidebands also implies that adjusting inhibitory sidebands could be a potential mechanisms to rapidly change cellular tuning during task performance (Fritz et al., 2010; Francis et al., 2018).

Pyramidal neurons in L2/3 of mouse A1 can be classified into 5 groups based on their cellular location and functional excitatory and inhibitory connectivity pattern (Meng et al., 2017), suggesting that these cells would likely show differential tuning

properties. Here the FRAs of L2/3 neurons could be classified into 6 different groups and we speculate that these clusters correspond to cells with differential connectivity patterns. L2 neurons are more likely to have recurrent lateral connections while deeper L3 neurons show more feedforward L4 input, which is correlated with a broader tuning at more superficial layers (Meng et al., 2017). Within our classification, H type neurons had the broadest tuning at 70dB SPL and PV and SST neurons are more likely to include such type of neurons. SST neurons in L3 of mouse A1 mostly receive excitatory input from superficial layers (Oviedo, 2017), suggesting that local connectivity patterns could bias tuning selectivity in vivo.

Our results suggest that sideband inhibition observed in cortical neurons originates both from MGB terminals and intracortical sources. The source of cortical sideband inhibition in A1 has been attributed either to PV neurons (Li et al., 2014b) or to SST neurons (Kato et al., 2017; Lakunina et al., 2020). One confounding factor is that these prior studies differed in their methods for inferring inhibitory sidebands. Li et al. (2014b) used the classic TT paradigm with one tone (100 ms in duration) of varying frequency and sound levels and the other held constant at the characteristic frequency of the cell. In contrast, Kato et al. (2017) used 1-second tones and inferred sideband inhibition from frequencies that resulted in below-baseline fluorescence deflection, which could underestimate inhibition. Lastly, Lakunina et al. (2020) used increasing bandwidth of band-passed noise (1s duration) centered at the BF of the cell and inferred sideband through suppression of the firing rate. The differences in stimulus duration can potentially bias the relative contribution of PV and SST neurons. PV neurons receive feedforward input as a part of the “driver” pathway that

typically exhibits synaptic depression (Lee and Sherman, 2010). In addition, inhibitory synapses onto pyramidal cells also show strong synaptic depression (Bridi et al., 2020). Together, these two mechanisms suggest that the inhibition from PV neurons onto pyramidal neurons weakens as stimulus duration is prolonged. In contrast, it is likely that SST neurons are specialized in sustained inhibition during long stimuli. Inhibitory neurons show within group cooperativity and cross-group inhibition (Karnani et al., 2016) and our previous work showed that SST and PV neurons show activity of opposite sign during prolonged stimuli (Liu et al., 2019). These results suggest that SST neurons mediate sustained inhibition to both local excitatory neurons and PV neurons. In contrast, PV neurons might be more involved in mediating disynaptic feedforward inhibition at the onset of the sensory stimuli (Li et al., 2014b). Nevertheless, the relative contributions of PV and SST towards sideband inhibition as a function of the duration of the sensory stimuli remain unclear.

SST neurons provide dense and non-specific inhibition onto local pyramidal cells (Fino and Yuste, 2011) and they have high connection probability with local pyramidal neurons (Levy and Reyes, 2012). Thus, if SST neurons play a major role in shaping inhibitory sidebands, neighboring pyramidal cells should have similar inhibitory sidebands. However, our IQR analysis of BIhF shows a higher degree of sideband heterogeneity than local frequency tuning. Therefore, our results suggest that sideband inhibition might not originate from a uniform source but reflects multiple sources. Given that we find prominent inhibitory sidebands in MGB terminal responses, our result suggests that inhibitory sidebands arise from the complex

interplay between feedforward input from the thalamus and the processing by local microcircuits. This view is consistent with a linear amplification of thalamocortical inputs by L4 (Li et al., 2013) which would include both excitation and disinaptic inhibition. The tuning of MGB terminals was narrower than the three types of cortical neurons, consistent with results in marmoset (Bartlett et al., 2011). Thus, it is likely that thalamocortical inputs contribute significantly to the observed cellular inhibitory sidebands. Moreover, as sideband inhibition originates from the auditory periphery (Ruggero et al., 1992) and can be observed at other stages in ascending auditory pathway (Konrad-Martin et al., 1998; Jen et al., 2002), we consider the inhibition seen in A1 as functionally diverse with aspects partially generated de novo in A1.

The relative contribution of MGB and intracortical inhibitory sources likely varies by relative frequency to BF. PV and SST neurons likely contribute to inhibitory sideband far from BF. Our data show that PV and SST neurons had higher SCs over distance than Thy1 neurons, suggesting that these interneurons are suited to convey inhibition of frequencies far from the CF. In addition, Natan et al. (2015) found no firing rate change to BF when either PV or SST neurons were silenced while responses to non-preferred frequencies were slightly disinhibited. In contrast, silencing PV neurons increased evoked firing rate yet produced no change in tuning width while silencing SST neurons increased tuning width but this effect was due to increased firing rate to spectral content far from BF (Lakunina et al., 2020). However, these results cannot explain the sideband structure close to BF that we observed in the current study. Such inhibition close to BF could result from feedforward thalamocortical or L4 input. Thus, PV and SST neurons might serve to temporally

truncate pyramidal neurons' responses. In A1 L4, the intracortical connections are thought to selectively amplify EPSCs evoked by frequencies close to BF as cotuned neurons tend to be interconnected (Liu et al., 2007). This circuit could also amplify the TT suppression close to BF. Specifically, as responses of TT and PT close to BF are amplified, their differences are effectively amplified as well, leading to a greater suppressive effect. Therefore we speculate that feedforward inputs contribute more strongly to sideband inhibition for frequencies close to BF.

We have shown here that the introduction of the second tone decreases neural correlations between cell pairs, which suggests that more spectral complex stimuli result in changes in population encoding. This potentially resulted from prominent suppressive nonlinear interactions between frequency components within the TT stimuli. Nevertheless, facilitative interactions do exist and the relative strength of suppression versus facilitation depend on specific FRA types. Typically, V and I type neurons are most biased towards suppression, while H type neurons show less such bias due to larger facilitative interactions. This is consistent with findings that L2/3 neurons of mouse A1 consist of subtypes that differentially integrate spectral information (Li et al., 2019). V and I type neurons behave similarly to band-pass filters while H type neurons behave as integrators of spectral energy. We speculate that these two types of processing could be more relevant to differentiation and detection of sound, respectively. Across cell types, Thy1 neurons showed larger nonlinear interactions overall than PV and SST neurons. However, a possible confound is that evoked fluorescence ( $\Delta F/F$ ) responses could potentially be smaller in PV and SST neurons due to a higher baseline fluorescence levels. As the suppressive

effect of TT cannot exceed the responses to PT, the detectable nonlinear interactions were bounded by the  $\Delta F/F$  evoked by PTs.

We found that both onset and offset responding neurons show nonlinear frequency interactions in their responses to TT stimuli. Notably, in some subtype of neurons (e.g. Thy1 H type), offset neurons show greater nonlinear interactions. To our knowledge this is the first time such phenomenon has been shown, suggesting that the pathway conveying offset information in mouse A1 might be engaged differently in integrating sound information and might consist of different neuronal circuits. Our previous study suggested that timing information is relayed from thalamus and further refined in cortical microcircuits (Liu et al., 2019). Our current result suggests that this pathway is also suited to convey complex spectral interactions, but it is unclear if the same cortical mechanisms involving PV and SST neurons contributing to spectral interactions are responsible for the same interactions in offset responses.

In summary, our modified TT paradigm has allowed us to reveal inhibitory sideband and nonlinear frequency interactions of large populations of neurons. Neurons form subclasses based on the shape of FRAs and the differential degree of nonlinear frequency interactions. Moreover, spectrally complex stimuli like TT resulted in decorrelation of neural responses, suggesting a sparse and more selective encoding of sound with rich spectral content. The local heterogeneity of sideband inhibition suggests that these sidebands might reflect integration of multiple sources. Lastly, we found prominent sideband inhibition in MGB terminals. Together, our results show that cellular tuning properties in a local region might not result from a

uniform set of sources but from feedforward input with existing sideband structures arising from subcortical processing stages and further diverse contributions from local cortical inhibitory sources.

# **Chapter 4. Interactive auditory task reveals complex sensory-action integration in mouse primary auditory cortex**

## **Abstract**

Sensory perceptions are not simply passive events but are active constructions by the brain. Predictive coding theory postulates that the brain actively makes predictions about the outside world and corrects them if any error arises. Thus, this process requires organisms to actively explore their environment and calibrate their expectations. In other words, the individual's neural system would be shaped in ways specific to their interactions with world. Where such predictions are formed is unclear, but the cerebral cortex is thought to play a key role. We here designed a novel interactive behavior paradigm where the animal actively modulated the carrier frequency of the sound stream for water reward. Given this paradigm, we identified neurons in primary auditory cortex (A1) that encoded complex interactions between sound and action, akin to representing the boundary conditions of our task. Furthermore, we found that layer 2/3 (L2/3) neurons encoded more action related information than L4. Lastly, a subset of L2/3 and L4 neurons responded to perturbations during active trials, constituting potential error signals. Together, our results show that complex interactions between action and sound already happen in A1 and that A1 neurons' responses potentially encode the violation of the general rule

of our task. Thus, primary sensory cortices do not simply encode sensory driven activity but represent the complex interaction of sensory input, expectations, errors, and behavioral outcomes.

## **Introduction**

Predictive coding theory has recently become a popular framework for explaining perception. The theory hypothesize that the brain perceives the outside world by actively making predictions according to its own model of the world, and constantly updating it via interactions between lower and higher order brain regions through error signals (Friston, 2010; Heilbron and Chait, 2017). However, such error signals would not arise unless individual organisms are actively engaged in testing such models by interacting with their environment. Thus, one can argue that predictions are an intrinsic part of performing actions, while the sensory outcome of actions provide new information to update the brain's hypothesis of the world. Therefore, testing predictive coding theory requires experimental paradigms that explicitly incorporate actions into the sensory process.

Recent studies in the visual system have provided a wealth of information on how action is integrated into sensory perceptions. For example, it has been robustly shown that locomotion increases visual responses in mouse visual cortex (Niell and Stryker, 2010; Mineault et al., 2016; Pagan et al., 2016). Furthermore, error signals were evoked in mouse primary visual cortex by the mismatch between expected visual flow and actual feedback (Keller et al., 2012), lending evidence for the predictive coding theory. In the auditory system, this framework has been proposed to

explain phenomena such as Stimulus Specific Adaptation (SSA) and Mismatch Negativity (MMN) (Carbajal and Malmierca, 2018), as these neuronal responses have been largely interpreted as error signals for violating auditory expectations. However, these paradigms are not interactive in nature and do not require the active engagement of the subject. In particular, SSA can be evoked in both awake and anesthetized animals (Ulanovsky et al., 2003; Malmierca et al., 2009). As such, these paradigms are insufficient for studying predictive coding. However, there have been more effort in using closed-loop designs in studying auditory perception (Nelson et al., 2013; Rummell et al., 2016; Schneider et al., 2018; Clayton et al., 2021). These studies have shown that ACX neurons' responses are modulated by the animal's action, and these actions can generate expectation specific effects in ACX. Nevertheless, unlike in visual closed-loop experiments, where the animal's action directly and continuously modulates the visual stimulus, e.g., the speed of visual flow being coupled to running, the auditory counterparts employed a largely simple relationship between the action and the sound, i.e., action triggers sound presentations. These discrete events might not form as tight a coupling between action and sensory feedback as in the continuous coupling between running and visual flow. Furthermore, as action modulation in sensory cortex depends on experience (Attinger et al., 2017), we speculate that by introducing a more complex action-sensory relationship in the auditory domain, we might observe distinct form of interactions between the two in ACX that are specific to this relationship.

Thus, to investigate such possibilities, we designed a novel behavior paradigm that allowed the mice to directly interact with the sound. Specifically, we trained mice

to turn a wheel, whose readout was coupled with the increase or decrease of the carrier frequency of the sound stream the mice were presented with. Mice needed to “steer” the carrier frequency either into the low or high end of the spectrum (target zones) for water reward. Unlike previous studies where the animals mostly had control over the timing of the sound, this paradigm allowed mice to continuously evaluate and control one attribute of the sound, i.e., frequency. Given this paradigm, we further introduced two forms of perturbations: Delay-Sound (DS) and Stop-Frequency (SF) where we briefly perturbed the default relationship between action and sound. To study neural responses in this paradigm, we used 2-Photon (2P) imaging of primary auditory cortex (A1) to investigate whether potential error signals evoked by the perturbations were present. Given that the cortex performs hierarchical computations we imaged both layer 4 (L4) and L2/3 to investigate if the error signal emerged in specific layers. Overall, we identified groups of neurons in L2/3 and L4 of A1 that were responsive to distinct features including sound or action with L2/3 showed more neurons responsive to actions than L4. Further, we found a group of neurons that responded to a particular form of interaction between action and sound, i.e., the decoupling (DCP) between the two either at the boundary conditions or due to the introduced perturbations. Our results suggest that A1 represents the complex relationship between action and sound in our paradigm as DCP responses signal the violation of the default rules.

## Methods

### Animals

All protocols and procedures are approved by the Johns Hopkins Institutional Care and Use Committee. In this study we trained and imaged 3 male and 3 female adult mice which were F1 generation of Thy1-GCaMP6s (JAX# 024275) crossed with CBA/CaJ mice (JAX# 000654). Such a strategy ensures that F1 offspring has both widespread genetic expressions of GCaMP6s in excitatory cells and minimum hearing loss throughout their lifespan (Frisina et al., 2011). The mice used in this study ranged from 4- to 8-month-old.

### Behavior Paradigm

The mice were trained on an interactive behavior paradigm. Specifically, the mice were presented with a frequency stream which started with either 10 kHz (low starting frequency) or 40 kHz (high starting frequency). By turning the wheel (63 mm in diameter) placed beneath their front paw, mice were able to modulate the carrier frequency by either turning to their right (carrier frequency increases) or turning to their left (carrier frequency decreases). The lower and upper bound of the carrier frequency was defined by the low and high starting frequency, respectively, beyond which left turning at lower frequency bound or right turning at upper frequency bound) produced no carrier frequency change. If in a given trial the mouse was presented with the low starting frequency (10 kHz), the mouse would need to increase

the carrier frequency by turning to its right and if the carrier frequency entered a target zone which was 0.25 octave wide and situated at the high frequency end (33.6 to 40 kHz) for a sufficient amount of time (0.2 sec), a hit was achieved. For trials with the high starting frequency (40 kHz), the target zone was located at the low frequency end (10 to 11.9 kHz) and thus, the mouse was required to turn left in order to achieve a hit. The animal was given a maximum of 6 sec to complete each trial. 22.5 degrees of rotation corresponded to 1 octave change in carrier frequency. After a hit was achieved, the frequency stream continued for 0.5 sec while its carrier frequency remained the same as the time of hit. During this period, a 10 Hz amplitude modulation was added to the frequency stream to reinforce the stimulus salience with the behavioral relevance. After the sound terminated, the servo-controlled waterspout was elevated and one drop of water (~5  $\mu$ l) was dispensed. The mouse had 2 sec for reward assumption before the waterspout was retracted by the servo. If the animal turned in the incorrect direction beyond 30 degrees, the trial was deemed incorrect and terminated. If animal failed to reach the target zone without turning incorrectly (beyond 30 degrees in the incorrect direction), the trial was deemed a miss. Both incorrect and miss trials were punished with an 8-sec timeout. There was a fixed 0.5 sec period between the end of reward or punishment period of the previous trial and the start of the next trial. At the beginning of each trial, the mouse was monitored for any spontaneous wheel turning activity and only after the mouse remained inactive for another 0.5 sec was the sound presented. The program waited indefinitely until this criterion was met. After the sound started, the first 0.1 sec was considered a 'grace period' where any wheel turning was discounted, and thus wheel turning

during this period did not produce carrier frequency change. The behavior program was implemented with LABVIEW 18.0 and was run on NI-USB 6343 (National Instrument).

During imaging, we controlled the onset of any sound such that it roughly aligned with the frame start trigger from the 2P imaging system. The jitter between the onset of the sound and the onset of one particular frame was around  $0.17 \pm 1.23$  millisecond. This approach allowed us to more precisely align neuronal responses with the animal's behavior. Furthermore, we employed a second device (NI PCI-6251) that used the same 2P frame start triggers as the acquisition clock to record a copy of the rotary activity. Thus, not only did we record the rotary activity within each trial at 100 Hz but we also recorded the rotary activity at 15 Hz (the 2P frame rate) throughout the entire imaging session.

#### Perturbation trials during behavior

To test the hypothesis that A1 encodes error signals induced by mismatch between the behavior and the sensory outcome, we introduced two forms of perturbations: Stop-Frequency (SF) and Delay-Sound (DS). In SF perturbations, we chose 4 frequencies outside the target zone (13.2, 17.4, 23.0, 30.3 kHz). These frequencies were symmetrical about the center of the spectral range (20 kHz) in the logarithmical space and equally spaced by 0.4 octave. During the SF perturbation trials, as the frequency trajectory reached one of the SFs, the carrier frequency would be frozen at that SF for 0.25 sec, artificially introducing a decoupling period. At the

end of this period, the paradigm resumed the normal coupling between the rotary and the carrier frequency. We presented SFs in blocks. Every 4 SFs form a block and within each block all 4 different SFs were presented. The order of the 4 different SFs within a block were randomized. We also balanced the starting frequencies by flipping them every time the same SF were encountered, i.e., if one particular SF was introduced in a low starting frequency trial in the previous block, it would be introduced in a high starting frequency trial in the current block and vice versa. We required that between SF trials, there would be a minimum of 2 hit trials and a maximum of 7 hit trials. Within those limits, it was determined with a probability of 0.5 whether SF would be introduced in each trial. However, if the animal failed to reach the SF in a given SF trial, in which case the perturbation effectively did not happen, the SF was carried over to the next trial with the same starting frequency. In all our imaging sessions with SF perturbations, we had a total of  $31.4 \pm 4.9$  (mean $\pm$ std) SF trials per session, which corresponded to  $15.9\% \pm 1.4\%$  of all trials. In a different set of imaging sessions, we introduced Delay-Sound (DS) perturbations. In such trials, we introduced a delay of 1 sec between the action and the sensory feedback, i.e., the carrier frequency change corresponding to any rotary activities was delayed by 1 sec. To balance trials with low and high starting frequencies, every DS trial had the opposite starting frequency of the previous DS trial. Similar to SF trials, we introduced DS with a probability of 0.5 if there were at minimum 3 hit trials but no more than 8 hit trials since the last DS trial. However, if the outcome of a particular DS trial was miss (the animal did not turn sufficiently in the correct direction) or incorrect (the animal turned in the wrong direction and effectively

produced no carrier frequency change), we deemed these trials as insufficiently perturbed and repeated the DS trial until the animal turned in the correct direction with a sufficient amount, which typically resulted in a hit. In our imaging sessions with DS perturbations, we had  $21.2 \pm 6.5$  (mean $\pm$ std) DS trials, which corresponded to  $11.6\% \pm 1.9\%$  of all trials. Both SF and DS perturbations were only introduced during imaging sessions.

### Behavioral training

Mice first received headplates implant and were allowed 5-7 days of recovery before the start of water deprivation. We restricted the water intake of mice to no more than 1 ml per day and the weight of mice dropped steadily within around 7 days to about 80% to 85% of the original weight. In the meantime, mice were introduced to the behavioral chamber and were accustomed to head-fixation before the training start. We devised 4 training stages for mice to systematically advance. The parameters of these stages were overall similar to those of the imaging sessions. The first 3 stages differed in the gain of the wheel, with each stage requiring the mouse to turn twice as much as the previous stage. Specifically, the gains were 5.625, 11.25, and 22.5 degrees per octave change in carrier frequency. In these stages, mice were rewarded as soon as a hit was achieved, and thus they consumed the reward during the period when the amplitude modulation was added to the frequency stream. In the 4th stage, the reward time was delayed to 0.5 sec after the end of the frequency stream in a hit trial and thus separated the reward consumption behavior from sensory

input in time. The rotary gain of the 4th stage was the same as in the 3rd stage. The mice advanced to the next stage if they completed at least 100 trials within a session and if the maximum hit rate over any consecutive 100 trials were above 60%. This relatively relaxed criterion was chosen such that mice were not overtrained on intermediate stages and could advance to the final stage more quickly, which shared the parameters of the actual experimental session. After mice reached a performance of about 70% to 80%, they were given cranial window implant and transferred to behaving under 2-photon imaging.

### Sound stimulus

During the behavior paradigm, the carrier frequency of the sound stream changed dynamically depending on the mouse's wheeling turning. The carrier frequency was bounded between 10 and 40 kHz. The wheel turning as readout by the rotary readings and the carrier frequency was updated 100 times per second. The carrier frequency changed with a step of 1/128 octave. Every 10 ms, the program translated the change in rotary reading ( $\Delta R$ ) into the change in the carrier frequency ( $\Delta CF$ ) and outputted a linear frequency sweep of 10-ms duration with its start at the current carrier frequency ( $CF$ ) and its end at the  $CF + \Delta CF$ . Therefore, the sound stream was essentially a piecewise linear frequency sweep that updated every 10 ms. At each time point, the rotary reading was smoothed over the previous 100 ms with a mean filter to produce a smoother frequency stream. The amplitude of the frequency stream at any time point was interpolated using an array of frequencies and their

calibrated amplitude at 60 dB SPL. Within the frequency bounds, 22.5 degrees of rotation mapped onto 1 octave change in carrier frequency.

Right after the conclusion of the active session, we presented the mice with pure tones that ranged from 7.1 to 56.5 kHz with 0.25 octave spacing at the same sound level as in the active session. These tones were used to construct tuning curves under the passive condition.

After the presentation of the pure tones, we presented mice with playbacks of the frequency streams selected from a subset of trials in the previous active session. Specifically, all perturbation trials were selected along with 10 hit trials, 6 incorrect trials and 6 miss trials which were evenly split between trials with low or high starting frequency. These selected trials were each repeated 4 times with a random order.

The sound waveform was generated by NI 6343 (National Instrument), which was used as input for ED1 speaker driver (Tucker-Davis Technologies) that drove an ES1 open field speaker (Tucker-Davis Technologies).

### Widefield imaging

To locate A1, we performed widefield imaging similar to previously described (Liu et al., 2019). In short, we used a blue LED of 470 nm wavelength (M470L3, Thorlabs Inc.) to illuminate the cranial window while imaging the excitation light with a PCO Edge 4.2 camera. We presented 5 tones ranging from 4 kHz to 64 kHz with 1 octave spacing and at 3 sound levels (30, 50 and 70 dB SPL). The images were

of size 330 by 330 and were sampled at 30 Hz. We identified A1 by identifying the low to high frequency gradient starting from the caudal side of the cranial window towards the rostromedial side.

### Two-photon imaging and analysis

We performed two photon imaging in A1 while the mouse was behaving under the microscope (Bruker Ultima 2Pplus). We imaged A1 with a 16X Nikon objective (NA 0.80) and at an optical zoom of 1X. The field of view was of size 1109.9 by 1109.9  $\mu\text{m}$ . The frame rate was 15 Hz. During the experiment, the head of the mouse was upright while the microscope nosepiece was rotated from the vertical position by about 50 degrees to match the angle of the cranial window surface. The imaging laser (Spectra-Physics InSight X3) was tuned to 920 nm wavelength. For analysis, we used Suite2P package to perform motion correction, automated ROI detection and raw cellular and neuropil fluorescence trace extraction (Pachitariu et al., 2016). We corrected the neuropil contamination using the following equation:

$$F_{corrected}(t) = F_{cell}(t) - 0.7 \times F_{neuropil}(t)$$

To convert the time-varying neuropil-corrected fluorescence trace into  $\Delta F/F$  traces, we computed the baseline using the same method as described before (Liu and Kanold, 2021). In short, as most excitatory neurons had sparse firing, the fluorescence would fluctuate around the baseline during most of the imaging session. Thus, for each neuron, we constructed the histogram of fluorescence values over time and identified the value that appeared most often at the resolution of the histogram

binning. We used this value for that neuron's baseline and then computed  $\Delta F/F$  over time using the following equation:

$$\frac{\Delta F}{F}(t) = \frac{F_{corrected}(t) - baseline}{baseline}$$

For a subset of our analyses, we used inferred spikes from the  $\Delta F/F$  traces. The spikes were extracted using OASIS package (Friedrich et al., 2017).

### Cranial Window Surgery

We followed a similar procedure for cranial window implant as previously described (Liu and Kanold, 2021). In short, 0.1 ml dexamethasone (2mg/ml, VetOne) was injected subcutaneously 2-3 hours before the surgery started in order to prevent brain swelling during surgery. All surgery tools were sterilized using a bead sterilizer (18000-45, Fine Science Tools). We anesthetized mice with isoflurane (Fluriso, VetOne) using a calibrated vaporizer (Matrix VIP 3000). We used 4% for induction and 1.5-2% for maintenance. During surgery the body temperature of the animal was maintained at 36.0 degrees Celsius. Next, we exposed the bone covering the auditory cortex by removing the skin and the muscles. A circular craniotomy was then performed over the left auditory cortex with a diameter of ~3.5mm using a dental drill. We then placed a custom-made cranial window over the exposed brain. The window consisted of 2 layers of 3mm round coverslips (64-0720, CS-3R, Warner Instruments) stacked at the center of a 4mm round coverslip (64-0724, CS-4R, Warner Instruments) and was secured with optic glue (NOA71, Norland Products). The edge of the cranial window was then sealed with Kwik-sil (World Precision

Instruments). Finally, C&B-Metabond (Parkell Inc.) was applied to secure the window. After the surgery, 0.05 ml Cefazolin (1 gram/vial, West Ward Pharmaceuticals) and 0.1 to 0.15 ml Carprofen (1 mg/ml) was injected subcutaneously, and the animal recovered under a heat lamp for 30 minutes before being returned to the home cage. For the two days after the surgery, the animal was given additional 0.05 ml dexamethasone (2mg/ml) and 0.1 to 0.15 ml Carprofen (1 mg/ml) daily. The animal was taken off training schedule for 3 to 5 days after the surgery and was given supplemental water.

#### Linear model construction and fitting

We constructed linear models that included several factors to explain the neuronal responses. In our first model (Figure 4.3A), the predictors were constructed from both trials and inter-trial periods. Each row of predictors was constructed from a 0.5 sec window, and each window was shifted by 0.25 sec. The predictors include frequency energy in different frequency bins (E), the carrier frequency change rate or sweep rate (S) and rotary activities (R). The predictor E could be thought of as a spectrogram with a low spectral resolution. Specifically, the boundaries of the 5 bins we used were 0, 0.25, 0.625, 1, 1.375, 1.75, 2, which were measured in octaves relative to the 10 kHz, the lower bound of the carrier frequency. The first ([0, 0.25]) and last bin ([1.75, 2]) corresponded to the low and high frequency target zone, respectively. The rest of the 4 bins were of equal size: 0.375 octave. For each of the time window the predictors were extracted from, we quantified the proportion of time

the carrier frequency resided in each of the 6 bins, in order to approximate the energy in each frequency bin. Next, we calculated the S predictor first by taking the derivative of the frequency trajectory represented in octave. We then constructed two histograms for positive and negative sweep rate respectively. For positive sweep rate, we used the following edges for the histogram: 0, 5, 10, 20 oct/sec. For negative sweep rate, we used the same edges with the opposite sign. In both cases, 0 oct/sec was not included in the histogram and thus only strictly non-zero sweep rate was considered. For each time window, such histograms were constructed, which measured the occurrence of different sweep rate. For the R predictors, we took the first order derivative of the rotary traces and summed positive and negative terms separately, which produced +R and -R, representing the absolute size of right and left turning events as measured in degrees. To model the effect of behavioral state, we introduced one “Task” term (T) that took value of 1 for predictor values constructed from ACT trials while taking value of -1 for predictor values constructed from PB trials. We then included the cross-product terms between sound encoding terms, i.e., E and S, and T to produce ET and ST predictors. If a particular neuron responded preferentially to E during ACT but not PB trials, then one would expect positive E coefficients and positive ET coefficient. Such interaction term was not constructed for R term as the animal showed little activity during the PB session. Next, to model the interaction between sound and action, we introduced the RE term, which was the R terms (separated according to left and right terms) times the summation of all frequency energies (summing over all E terms). Thus, the RE term was equivalent to the R terms windowed by the presence of sound. Finally, we also included a reward

term (Rw) that marked the presence of reward. For this model, the design matrix had 6 E terms, 6 ET terms, 6 S terms, 6 ST terms, 2 R terms, 2 RE terms, 1 Rw term and 1 T term. Each ACT and PB session pair would produce ~10,000 observations for the row of the design matrix. We used inferred spikes as the neuronal responses for the model due to the more accurate timing. Next, we fit our linear model using LASSO (MATLAB built-in function, `lasso`), which had the benefit of automated feature selection. For each neuron, we used a 10-fold cross-validation to determine the optimal regularization term value that achieved the minimum error, and subsequently used the same value for fitting all the observations.

To determine the sensitivity of neuronal responses to the different predictors, we calculated the  $\Delta R^2$ s by measuring the difference between full model  $R^2$  and the  $R^2$  calculated from shuffling a particular set of predictor values. For example, to determine the  $\Delta R^2$  of the E predictor, all 6 E terms were shuffled. Next, we grouped neurons based on which predictor had the highest  $\Delta R^2$ . We also required that the maximum  $\Delta R^2$  cross a threshold of 0.05.

For Delay-Sound (DS) perturbations, we extracted predictors from 4 non-overlapping time-windows of 0.5 sec duration (Figure 4.6A). The first window was aligned with the onset of the action. Due to the 1 sec delay between action and carrier frequency change, in the first second following action onset, the carrier frequency remained the same as the starting frequency. Thus, the first two time-windows correspond to this period, which we define as the action window, while the other two time-windows corresponded to the period when the delayed carrier frequency change occurred. We define this window as the sound window. Next, we extracted action

driven decoupling (DCP) in the action window and sound driven DCP in the sound window. Action driven DCP (DCPA) was defined as the period of time when the wheel turning failed to translate into carrier frequency change. Sound driven DCP (DCPS) was defined as the period of time when the frequency change did not correspond with the concurrent action (Figure 4.6B). We broke down DCPA by the starting frequency at which it occurred (low vs. high) and we broke down DCPS by the direction of the frequency sweep (up vs. down). All other predictors were constructed as in the first model. For the rows of the design matrix, we included observations from DS trials and the corresponding PB trials. Only DS trials with a delayed hit were included. We also included non-DS (NDS) trials where hits were achieved. These trials with hits were selected such that the responses due to the action could be separated from the genuine error responses. In total, the design matrix had 31 columns (6 E terms, 6 ET terms, 6 S terms, 6 ST terms, 2 R terms, 2 DCPA terms, 2 DCPS terms and 1 T term) and ~800 rows.

For our third model that focused on the Stop-Frequency (SF) perturbation, we used the same approach and constructed similar predictors. As the duration of the SF perturbation was 0.25 sec, we constructed our predictors from two time-windows of 0.25 sec. The first window spanned the 0.25 sec before SF onset, while the second window spanned the 0.25 sec after the SF onset. All predictors but DCP were constructed in the same fashion as in the first model. As SFs happened outside target zones, we computed the action driven DCP in the 4 middle frequency bins that we used for E predictors, where the 4 SFs were within each of the 4 bins. Thus, this model had 4 DCP terms, one for each of the 4 SFs. In total, the design matrix for the

SF perturbations had 31 columns (12 E and ET terms, 12 S and ST terms, 2 R terms, 4 DCP terms, 1 T term). For rows in the design matrix, all SF perturbation trials in the ACT sessions were included. A typical ACT session included ~30 SF trials. We next included the PB trials that corresponded to SF trials in the ACT sessions. As each SF trial was played back with 4 repeats, the PB sessions added ~120 observations to the design matrix. To further account for the contribution of pre-SF period to neuronal responses, we included so-called SIM trials, where the frequency trajectory was similar to that of the SF trial immediately before SF onset (Figure 4.7B). We hypothesize that if the responses during the post-SF period were due to pre-SF factors, then the responses to SF trials would be similar to those in SIM trials. For each of the 8 unique SF trials (4 SFs  $\times$  2 starting frequencies), we each selected the closest  $n$  matching SIM trials, such that the total number of SIM trials ( $8n$ ) were 5 times that of SF trials (~150 observations). In the end the design matrix had about 300 rows.

### Local Best Frequency

In Figure 4.6I, we investigate whether BSF depended on the Local Best Frequency (LBF), which measured the tuning preference of local population of neurons. To compute LBF, we first identified the neurons within a 100- $\mu$ m radius of the neuron in question. Next, we summed the tuning curves across individual neurons weighted by the respective significance, i.e., significant responses had a weight of 1 while non-significant responses had a weight of zero. Thus, we obtained a population

tuning curve whose peak represented the frequency that evoked the most responses in the local group of neurons, which we defined as LBF. The significance of tone evoked responses for each neuron was determined similar to described before (Liu and Kanold, 2021). In short, we constructed the 99.9% confidence interval of the pre-stimulus-onset and post-stimulus-onset  $\Delta F/F$  values, and one tone is considered significant if the corresponding confidence intervals were non-overlapping. For pre-stimulus-onset period, we chose a window of  $\sim 0.5$  sec duration immediately before the tone onset. For post-stimulus-onset period, we chose a window of  $\sim 0.5$  sec duration offset by  $\sim 0.27$  sec (4 frames) from the tone onset, in order to better capture the peak of the  $\Delta F/F$  trace.

### Spatial Clustering Analysis

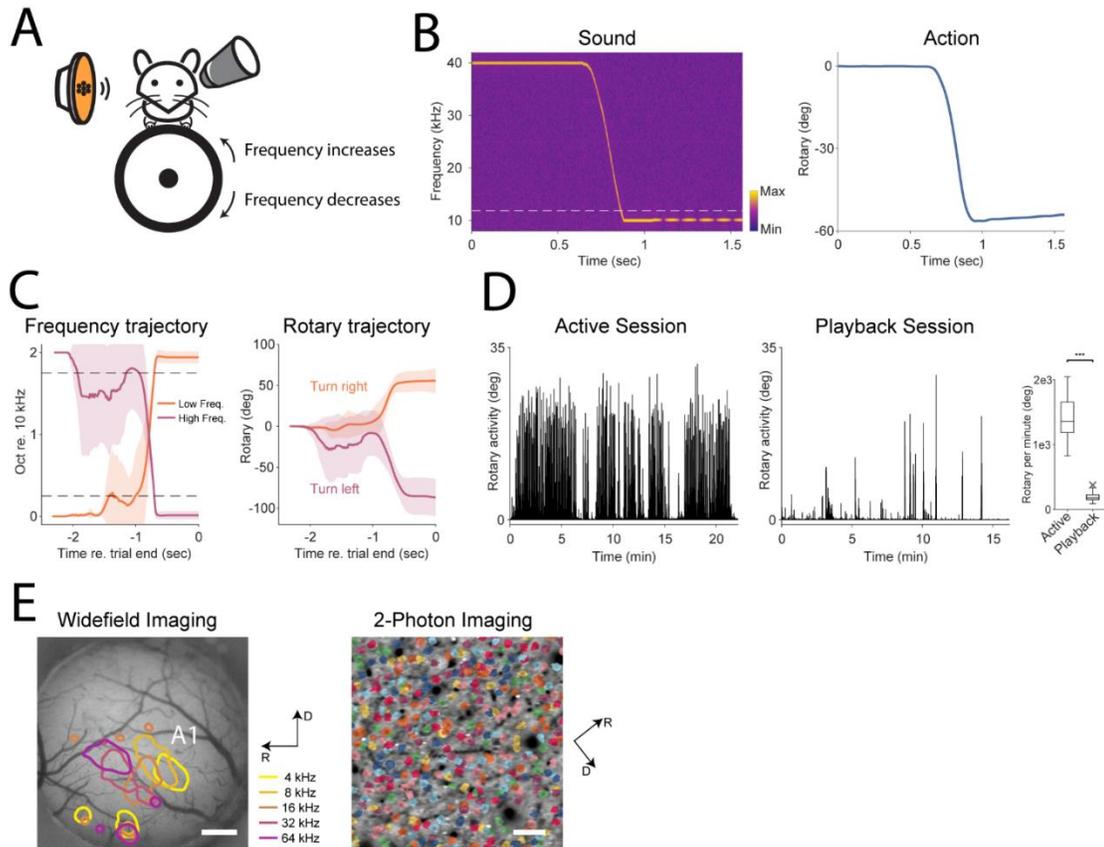
In Figure 4.8, we quantified the spatial clustering of different feature sensitive group following a similar strategy in Deneux et al. (2016). We calculated a homogeneity index for each neuron, which quantified the proportion of neurons within a  $100 \mu\text{m}$  radius that belonged to either the same group as the neuron at the center or belonged to the other feature sensitive groups. We determined the significance of the distribution of the homogeneity indices by comparing the actual distribution against the shuffled data, where we randomly assigned the feature sensitive identity to neurons in the FOV. The values of real and shuffled data were pooled from all FOVs and compared to determine significance.

## Results

### Mice learnt to perform an interactive auditory task

To investigate how the mouse A1 neurons encodes the interaction between action and sensory feedback and to investigate if such interactions emerge across the cortical layers, we imaged L2/3 and L4 of mouse A1 while the animal performed a novel interactive task (n=6 mice (3 male and 3 female); L2/3, 26 field of views (FOVs), 46328 neurons imaged; L4, 25 FOVs, 53211 neurons imaged). Our interactive task required the animal to change the frequency of a sound into a target frequency range. The mice were trained to control the frequency of the sound by turning a wheel placed beneath its front paw. The mice were presented with a frequency modulated sound stream and its carrier frequency would decrease or increase as the animal turned the wheel clockwise or counterclockwise (Figure 4.1A). The sound stream started either at the low end (10 kHz) or at the high end (40 kHz) of the predefined spectral range. The task required the animal to “steer” the low starting frequency (10 kHz) into the high target zone (33.6 kHz to 40 kHz, 0.25 octave in width), or “steer” the high starting frequency (40 kHz) into the low frequency target zone (10 kHz to 11.9 kHz, 0.25 octave in width). The carrier frequency would not decrease or increase beyond the low and high frequency boundary (10 and 40 kHz, respectively). Figure 4.1B shows the spectrogram of the microphone-recorded sound stimulus of one trial with the high starting frequency, along with the action of the animal as shown by the rotary reading (also see Video 1). Figure 4.1C shows the average frequency and rotary trajectories for hit trials with short hit latencies (within

2 sec). These results show that the animals turned the wheel robustly to perform the task. Immediately following each active (ACT) session, we presented animals with a playback (PB) session. In PB sessions, we first played pure tones to measure tuning curves of imaged neurons, following which we presented the animal with a subset of the frequency stream generated in the ACT session as a control condition. The playback session allowed us to measure the contribution of stimulus selectivity to neural responses with minimum confound from animal movement as previous studies suggest that auditory responses could be suppressed by the animals' action (Rummell et al., 2016; Schneider et al., 2018). In PB sessions, the animals remained relatively still and showed much less movement (Figure 4.1D, right). We identified A1 with widefield imaging and performed 2P imaging in A1 L2/3 and L4 (Figure 4.1E). The animals were able to perform the task well during imaging sessions. The overall hit rate of all behavioral sessions with simultaneous imaging was  $70.2\% \pm 8.9\%$  (Mean  $\pm$  STD). The maximum hit rate over any consecutive 100 trials during one session was  $75.6\% \pm 10.0\%$ . The average time to achieve a hit was  $1.05 \text{ sec} \pm 1.03 \text{ sec}$ .



**Figure 4.1 Mice were trained to perform a novel interactive behavior paradigm**

(A) Schematic of the behavior paradigm. The mouse was trained to modulate the carrier frequency by turning the wheel. (B) Example frequency stream and action from one hit trial. Left: the microphone-recorded spectrogram of the frequency stream. The white dash line indicates the boundary of the target zone for this trial. After the hit condition was met, the carrier frequency was frozen for 0.5 sec while a 10 Hz amplitude modulation was added to signal the trial outcome. Right: the rotary reading corresponding to the left turning action. (C) Average frequency and rotary trajectories of hit trials with short hit latencies ( $\leq 2$  sec) from one mouse in one active session. The solid lines represent the mean and the shaded regions represent standard deviation. (D) Left and middle: example rotary activities from one ACT session and

the subsequent PB session. Right: mice showed significantly less movement in PB sessions than in ACT sessions (Wilcoxon rank sum test,  $p=3.3\times 10^{-18}$ ). (E) Left: contour plots showing “hotspot” of fluorescence increase evoked by different tone frequencies during widefield imaging. A1 was identified through its caudolateral-mediodorsal tonotopic gradient. The scale bar represents 500  $\mu\text{m}$ . Right: A1 neurons under 2P microscopy. The example image shows a cropped region from the full FOV for clarity. Neurons were colored according to segmentation by Suite2P. The scale bar represents 50  $\mu\text{m}$ . The rotation of the 2P FOV was due to the rotation of the nosepiece relative to the scanning directions.

#### A1 neurons show both sensory and action driven activity during an interactive auditory task

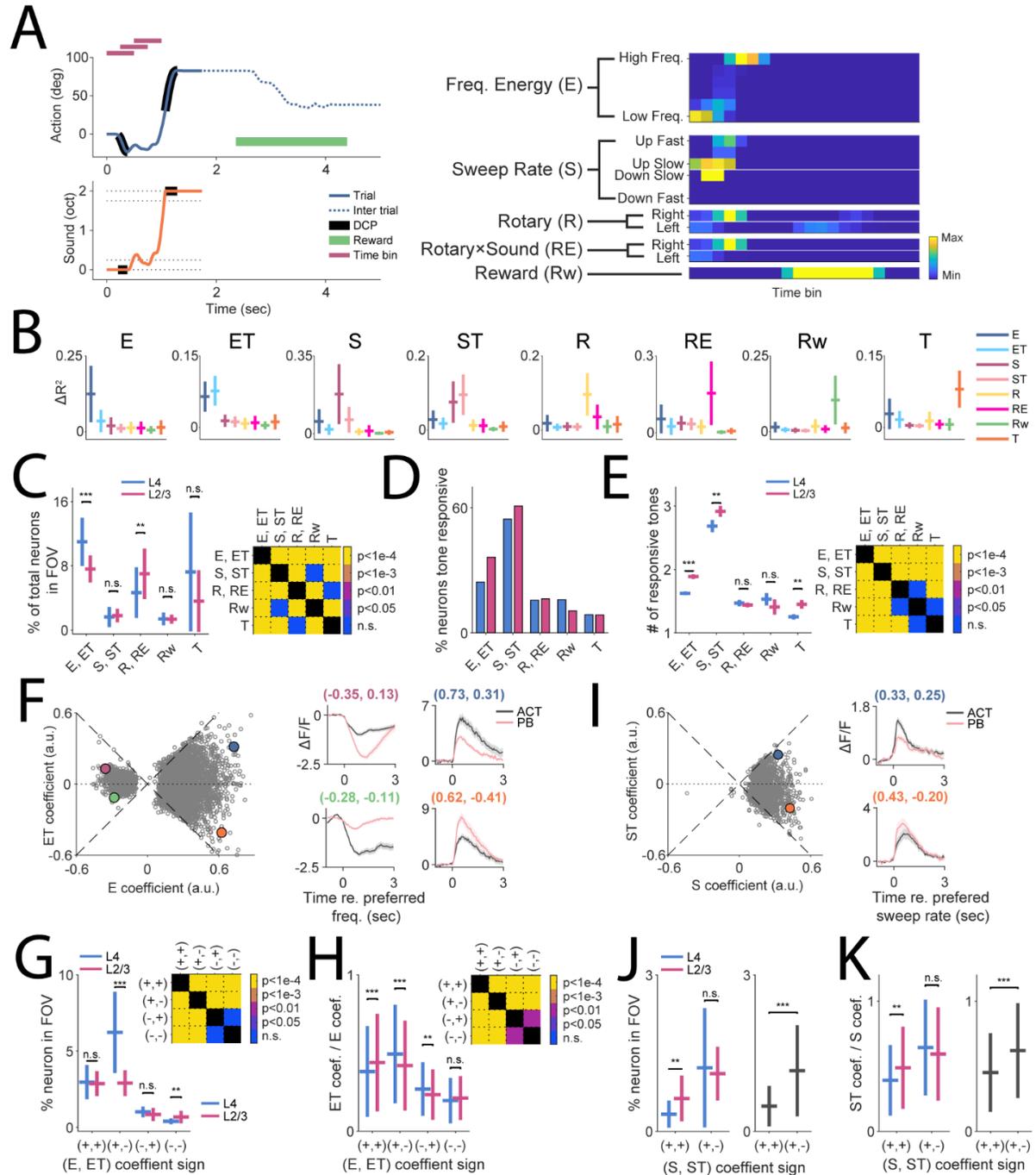
To begin to understand how neural responses encode action and the corresponding sensory outcome, we constructed a simple linear model whose predictors included energy in different frequency bins (E), the rate of frequency modulation (sweep rate, or S) and wheel turning (R) (Figure 4.3A). We added one interaction term between R and E as we hypothesize that neurons could show sensitivity to the synergy between the two (Figure 4.3A, see Methods). To account for the behavioral state dependence of the neuronal responses, we introduced one task term (T) that took value of 1 for predictor values constructed from ACT sessions and -1 for predictor values constructed from PB sessions. The T term alone can account for a shift of baseline activity levels across the behavioral states, while the interaction

terms between T and sound encoding terms, i.e., E and S, can account for sound driven responses that are behavioral state dependent (ET and ST term). We Finally, we added one term that accounts for the reward consumption (Figure 4.3A, Rw). Thus, this model has 8 groups of predictors that encode various features during the ACT and PB sessions (Figure 4.3B).

We fit our model with LASSO to automatically identify relevant factors influencing each neuron's responses. We clustered neurons into different feature sensitive groups by identifying the predictors with the maximum  $\Delta R^2$  within each neuron, calculated by shuffling predictor values (Figure 4.3B, see Method). We visualized the distribution of these groups of neurons in a lower dimension space projected by T-sne (Supplemental Figure 4.2A), which verified that such identified groups of neurons segregate in the projected space. These clusters further suggest that A1 neurons encode various features in addition to sound during the behavioral session. As a case in point, the Rw neuronal group was robustly and preferentially activated during reward consumption (Supplemental Figure 4.2B) while showing little sensitivity to other features (Figure 4.3B, Rw). The identification of this group further validated our approach to use regression to explain neuronal responses in A1.

Next, we quantified the fraction of neurons within each FOV that were assigned to each group as a function of layer (Figure 4.3C). Among all groups, E and ET neurons were the most abundant, while S and ST neurons were much fewer, suggesting that A1 neurons are more sensitive to the presence of sound energy within its receptive field, which is consistent with previous findings that A1 neurons were more likely to be driven by spectrally static stimuli than by frequency modulations

(Issa et al., 2017). Across cortical layers, L4 had more neurons assigned to E and ET group than L2/3 while the two layers showed similar proportions of S and ST neurons. In contrast, L2/3 showed more abundant R and RE neurons, suggesting more integration of action related information in L2/3. Next, we investigated the tuning properties of these feature sensitive groups of neurons (Figure 4.3D, E). Sound encoding neuronal groups, i.e., E, ET and S, ST, tend to show more neurons responsive to tones (Figure 4.3D) and showed broader tunings than other neuronal groups (Figure 4.3E). In addition, sound encoding L2/3 neurons showed broader tuning than L4 neurons, which is consistent with previous reports (Bowen et al., 2020). Together, these results suggest that there are functionally separate groups in A1 that preferentially encode sound or action related information. The differential proportion of sound and action sensitive neurons in L2/3 and L4 further suggests there is a shift towards encoding behaviorally relevant information from L4 to L2/3.

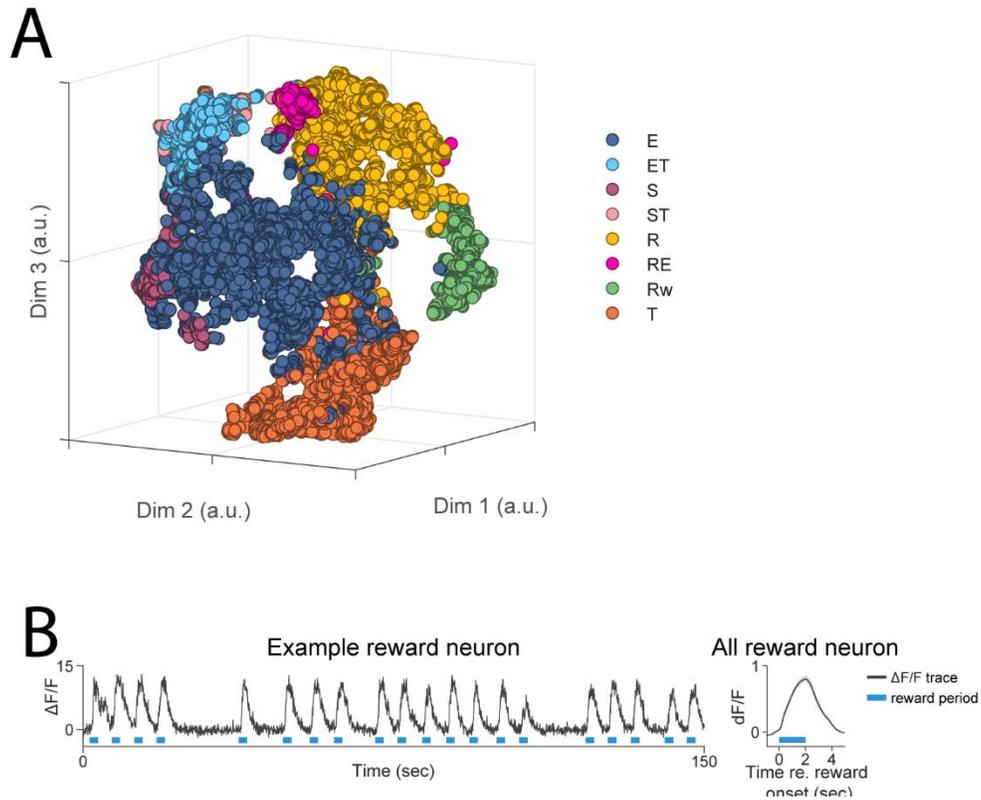


**Figure 4.2 A1 responses can be explained by various factors**

(A) Left: rotary trajectory (top) and frequency trajectory (bottom) of an example hit

trial. The part of action that was not translated into carrier frequency changes was marked in black (DCP). The green line marks the post-trial period when the animal consumes the water reward. The rotary activities were recorded both in trial (top, solid blue curve) and inter-trial (blue dotted curve). The purple lines represent the time windows (0.5 sec in duration) from which we constructed the predictors for the linear model. Each window was shifted by 0.25 sec from the previous one. Right: the predictors constructed from the same trial shown on the left. (B) The group of neurons sensitive to various features were identified through their maximum group  $\Delta R^2$  term. (C) The fraction of feature sensitive neurons as a function of cortical depth. L4 vs L2/3 Wilcoxon rank sum test (significance adjusted with Bonferroni correction): E, ET,  $p=1.8 \times 10^{-5}$ ; S, ST,  $p=0.22$ ; R,  $p=0.0086$ ; RE,  $p=1.8 \times 10^{-4}$ ; Rw,  $p=0.55$ ; T,  $p=0.015$ . (D) The fraction of neurons within each feature group that were tone responsive as a function of layer. (E) Left: within the tone responsive neurons of each feature group, the number of tones evoking significant responses in each neuron were plotted as a function of layer. L4 vs L2/3 Wilcoxon rank sum test: E, ET,  $p=1.2 \times 10^{-9}$ ; S, ST,  $p=0.0036$ ; R,  $p=0.40$ ; RE,  $p=0.11$ ; Rw,  $p=0.22$ ; T,  $p=0.0012$ . Right: multiple comparison between feature groups were plotted, pooling data from both layers. (F) Left: scatter plot the coefficients of the E and ET term from both E and ET group. The four quadrants of neurons showed different levels of behavioral state dependence in their frequency responses. The dashed lines were of slope 1 and -1, respectively. Right: the traces from four example neurons are shown. The colors of the titles match the four colored datapoints shown on the left. The traces are aligned to the onset of their preferred frequencies and vertically offset by their immediate pre-

onset baselines. (G) The fraction of E, ET neurons as a function of both the sign of the (E, ET) coefficient pair and cortical layer. L4 vs L2/3 Wilcoxon rank sum test: (+,+),  $p=0.77$ ; (+,-),  $p=9.2\times 10^{-8}$ ; (-,+),  $p=0.066$ ; (-,-),  $p=0.0058$ . The inset shows the multiple comparison between neurons with different (E, ET) coefficients signs regardless of the cortical layer. The (+,-) group had the most number of neurons. (H) The ratios between the absolute values of ET and E coefficients are plotted as a function of both the sign of the (E, ET) coefficient pair and cortical layer. L4 vs L2/3 Wilcoxon rank sum test, (+,+),  $p=7.6\times 10^{-7}$ ; (+,-),  $p=3.5\times 10^{-13}$ ; (-,+),  $p=0.012$ ; (-,-),  $p=0.12$ . The inset shows the multiple comparison between neurons with different (E, ET) coefficients signs regardless of the cortical layer. The (+,-) group showed the strongest modulation between ACT and PB session. (I) The same as in (F) but for S and ST group. Unlike E and ET group, the S coefficients were predominantly positive. (J) The fraction of S, ST neurons as a function of both the sign of the (S, ST) coefficient pair and cortical layer. L4 vs L2/3 Wilcoxon rank sum test: (+,+),  $p=0.0073$ ; (+,-),  $p=0.75$ . Right: comparison between (+,+) and (+,-) group regardless of the cortical layer, Wilcoxon rank sum test,  $p=1.2\times 10^{-9}$ . (K) The ratios between the absolute values of ST and S coefficients are plotted as a function of both the sign of the (S, ST) coefficient pair and cortical layer. L4 vs L2/3 Wilcoxon rank sum test: (+,+),  $p=0.0023$ ; (+,-),  $p=0.0285$ . Right: comparison between (+,+) and (+,-) group regardless of the cortical layer, Wilcoxon rank sum test,  $p=9.8\times 10^{-18}$ .



**Figure 4.3 Projection of feature sensitive neurons in a low dimensional space and reward neuron traces**

(A) 3D scatter plot showing the identified feature sensitive groups of neurons projected on the lower dimensional space generated by the T-sne algorithm. (B) Left: the fluorescence trace from an example reward sensitive neuron. The blue lines mark the reward consumption period. Right: the traces from all reward neurons aligned to the onset of the reward period. The solid line represents the mean and the shaded region represent the 95% confidence interval constructed through bootstrapping.

#### Sound driven responses in A1 were more suppressed during behavior

The interaction terms between sound encoding terms and the task term, i.e., ET and ST allow us to investigate the behavioral state dependence of sound evoked responses in A1. Depending on the sign of the coefficients of the sound encoding

term and its corresponding task interaction term, e.g., coefficients of E vs. coefficients of ET, one could determine first whether the sound feature (E and S) activates or suppresses the neuronal responses and second how the active state modulates such activation or suppression (ET and ST). For E and ET neurons, while the majority of neurons were activated by the presence of sound (positive E coefficients), a subset showed suppression (negative E coefficients, Figure 4.3F left). Both activation and suppression can be modulated by the behavioral state with a variety of strength, but the coefficients of the ET term were almost always smaller than the coefficients of the E term, confirming their modulatory roles (Figure 4.3F, left). Specifically, a neuron with positive E and ET coefficients would show boosted responses in the ACT session while a neuron with a positive E coefficient and a negative ET coefficient would show suppressed responses in the ACT session compared to the PB session. A neuron with negative E and ET coefficients would show less suppression in the PB session compared to the ACT session, while a neuron with a negative E coefficient and a positive ET coefficient would show more suppression in the PB session compared to the ACT session. Most E and ET neurons, regardless of cortical layers, showed positive E and negative ET coefficients (Figure 4.3G, inset). This suggests that the majority of A1 neurons showed more suppressed responses during the ACT session than during the PB session. However, these suppressed neurons were more abundant in L4 and in L2/3, where the neurons showing positive and negative ET coefficients were the same in number (Figure 4.3G). Furthermore, we quantified the strength of the task modulation by computing the ratio between the absolute values of ET and E coefficients (Figure 4.3H). This

analysis revealed that the neurons with positive E and negative ET coefficients showed the highest modulation strength, and such effect was also stronger in L4 than in L2/3. Together, these results suggest that although the majority of E and ET neurons showed activation in response to increase in sound energy, their responses were more suppressed during behavior and this suppressive modulation had the largest strength among all possible forms of modulation.

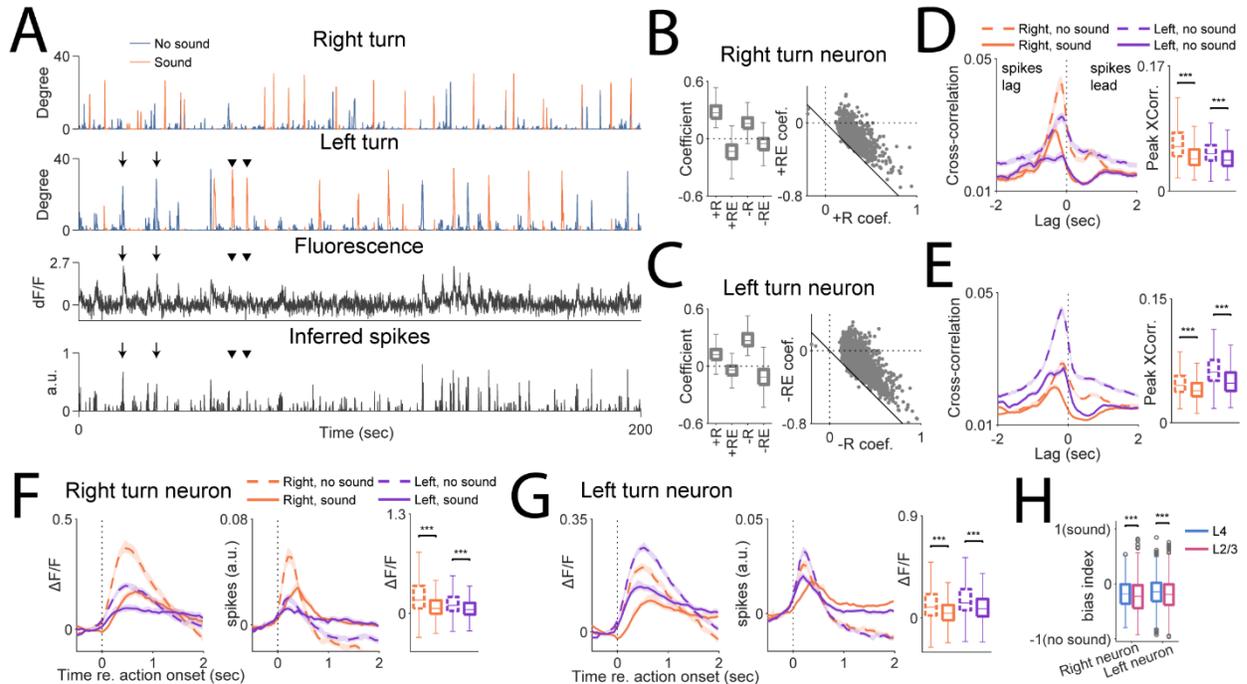
Next, we investigated whether such modulations also exist in S and ST neurons. Almost all S and ST neurons showed positive S coefficients, suggesting that their responses were activated by frequency sweeps. Similar to E and ET neurons, the majority of S and ST neurons showed a positive S and negative ST coefficient pair, although such neurons did not differ in numbers across L4 and L2/3 (Figure 4.3J). Furthermore, S and ST neurons also showed a similar stronger suppressive effect due to task engagement (Figure 4.3K). Overall, our results confirmed that A1 is in a more suppressed state during behavior, which is consistent with previous reports (Otazu et al., 2009; Kuchibhotla et al., 2017).

#### A1 neurons showed wheel-turning driven activities that were sound dependent

Having identified action sensitive neurons in A1, we proceed to examine their properties. Figure 4.4A shows an example R neuron identified as in Figure 4.3B. Examining this neuron's responses revealed that its activities were temporally correlated with leftward turning activities and that such correlation existed both within trials and during intertrial intervals. Furthermore, this neuron's responses tended to be smaller during trials than during intertrial intervals (Figure 4.4A, arrows and arrow heads). We separated R neurons into right turn and left turn preferring

neurons by identifying which term had a higher  $\Delta R^2$  and thus better explained the variability in the neuron's responses. Such identified neurons showed more prominent +R and +RE coefficients for right turn preferring neurons and more prominent -R and -RE coefficients for left turn preferring neurons (Figure 4.4B, C). However, for right turn preferring neurons, +R and +RE coefficients were significantly negatively correlated and the same was true for -R and -RE coefficients in left turn preferring neurons. This result suggests that similar to the example shown in Figure 4.4A, the majority of R neurons showed a more suppressed action driven responses during sound presentation. Examining the R and RE coefficient pairs revealed that almost all fell above the line with slope of -1 (Figure 4.4B, C), suggesting that RE coefficients were smaller than R coefficient in absolute value, and thus confirming that RE term negatively modulates action driven activities in R neurons. As our model lacks temporal resolution, we verified this result by constructing the correlograms between inferred spikes from R neurons with the onset of rotary activities broken down by turning directions and the presence of sound stimuli (Figure 4.4D, E). For both groups of neurons, the peaks of the correlogram had negative lags, suggesting that the onset of rotary activities preceded the spikes. For right turn preferring neurons, the correlogram with right turns during intertrial interval (no sound period) had the largest peak (Figure 4.4D, right). Similar result holds true for left turn preferring neurons (Figure 4.4E, right). We further examined the fluorescence traces and inferred spikes, temporally aligned to the onset of rotary activities broken down by turning directions and the presence of sound stimuli (Figure 4.4F, G). In both groups of neurons, turning in the preferred direction during no-sound intertrial period evoked

the largest responses. Lastly, we quantified the suppressive effect of sound presentation as a function of cortical layer, by computing the bias index using the response amplitude following action onset, i.e.,  $(\text{Resp}_{\text{sound}} - \text{Resp}_{\text{no sound}}) / (\text{Resp}_{\text{sound}} + \text{Resp}_{\text{no sound}})$ . This analysis revealed that L2/3 R neurons showed a stronger sound induced suppressive effect (Figure 4.4H). Together, these analyses showed that a subset of A1 neurons were selectively activated by the animal's specific action and thus potentially encodes the choice of the animal. Such neuronal responses are dynamically modulated by the presence of sound, which could provide a substrate of more complex interactions between sound and action within A1.



**Figure 4.4 Action sensitive (R) neurons showed sound dependent activities**

(A) Top two rows: example left- and right-turning rotary activities are plotted. Rotary activities during sound presentations are plotted in orange. Inter-trial rotary activities are plotted in blue. Bottom two rows: fluorescence trace and inferred spikes from one example action sensitive (R) neuron from the same behavioral session. This neuron was preferentially sensitive to left turns, and its response amplitude tended to be smaller during sound presentations (arrow heads) than during inter-trial intervals (no sound, arrows). (B) Left: boxplot showing the coefficients of right turn preferring R neurons. In these neurons, +R and +RE term coefficients are more prominent than the negative counterparts. Right: scatter plot showing the joint distribution of +R and +RE coefficients. +R and +RE coefficients had a significantly negative correlation ( $p=1.5 \times 10^{-228}$ ). The black solid line is of slope -1. (C) The same as in (B) but for left turn preferring R neurons. For these neurons, -R and -RE coefficients also had a

significantly negative correlation ( $p < 10^{-4}$ ). (D) Left: the correlogram between inferred spikes of right turn preferring R neurons and the onset of rotary activities broken down by both turn directions (right vs. left) and sound presence (sound vs. no sound). Right: boxplot showing the peak value of the correlogram under the four conditions. The peak correlogram values were higher during no sound period than during sound presentation for both right and left turns (Wilcoxon sign rank test, right, no sound vs. right, sound,  $p = 2.1 \times 10^{-74}$ ; left, no sound vs. left sound,  $p = 2.9 \times 10^{-32}$ ). (E) Same as in (D) but for left turn preferring R neurons. Similarly, the peak correlogram values were higher during no sound period than during sound presentation for both right and left turns (Wilcoxon sign rank test, right, no sound vs. right, sound,  $p = 1.6 \times 10^{-91}$ ; left, no sound vs. left sound,  $p = 1.7 \times 10^{-152}$ ). (F) Left and middle: the fluorescence and inferred spike traces of right turn preferring R neurons, temporally aligned to the onset of wheel turning and broken down by the turning directions and the presence of sound. Right: boxplot showing the  $\Delta F/F$  following the wheel turning onset, broken down by the four conditions.  $\Delta F/F$  values following wheel turning were higher during no sound than during sound presentation (Wilcoxon sign rank test, right, no sound vs. right, sound,  $p = 1.0 \times 10^{-68}$ ; left, no sound vs. left sound,  $p = 2.1 \times 10^{-33}$ ). (G) Same in (F) but for left turn preferring R neurons. Similarly,  $\Delta F/F$  values following wheel turning were higher during no sound than during sound presentation (Wilcoxon sign rank test, right, no sound vs. right, sound,  $p = 5.9 \times 10^{-107}$ ; left, no sound vs. left sound,  $p = 8.5 \times 10^{-96}$ ). (H) Boxplot showing the bias indices quantifying the strength of the suppressive effect of sound to neuronal responses evoked by wheel turning, broken down by neuron groups (left or right turn preference) and cortical layer. More negative values

suggest stronger suppressive effects. For both right and left turn preferring neurons, the suppression strength was higher in L2/3 than L4 (Wilcoxon sign rank test, right preferring neuron,  $p=4.3\times 10^{-4}$ ; left preferring neuron,  $p=6.2\times 10^{-8}$ ).

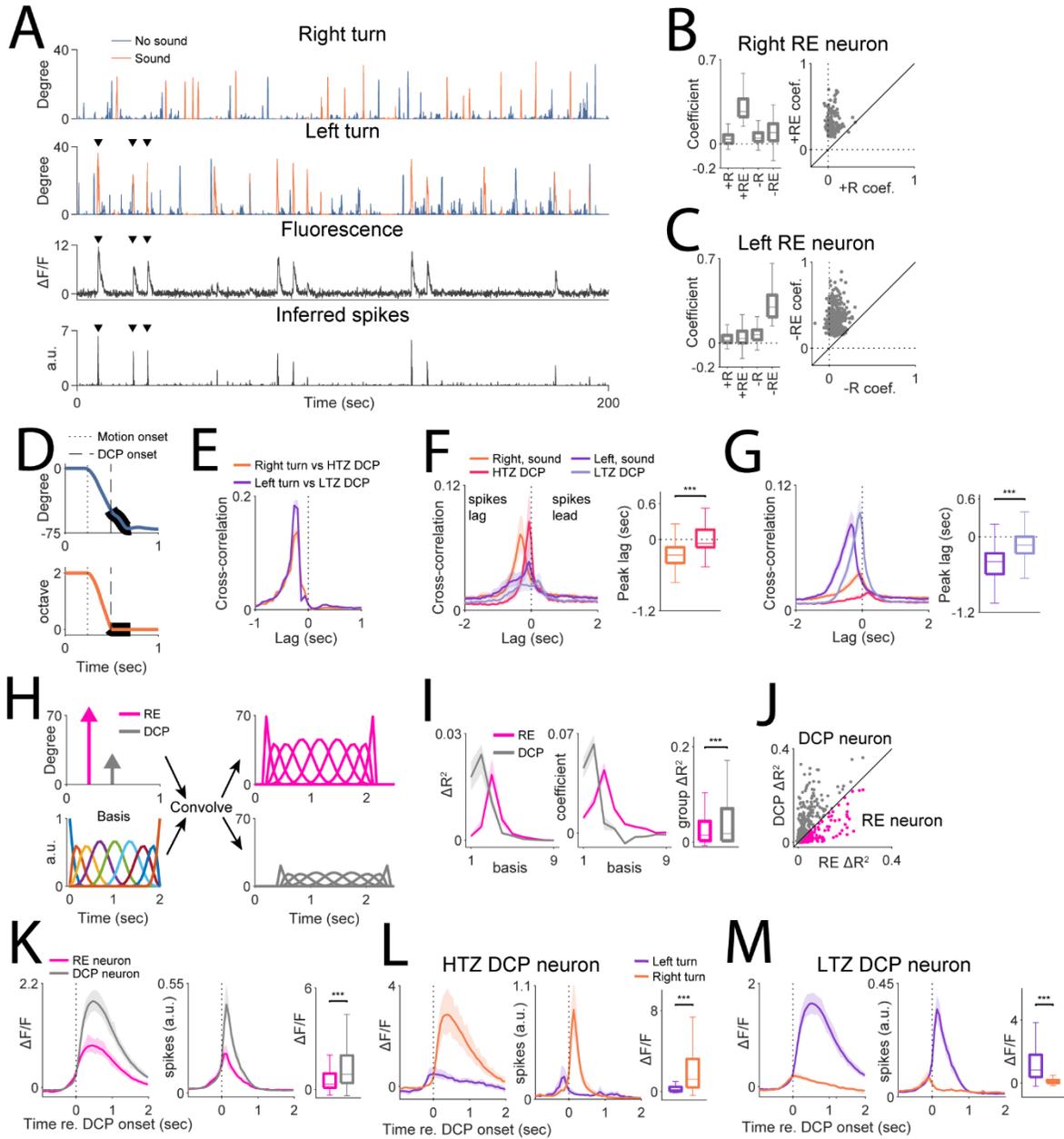
#### A subset of sound and action jointly sensitive neurons encode the spectrum boundary of the behavioral paradigm

We identified a group of neurons that were sensitive to actions specifically during sound presentation (RE, Figure 4.3B). Figure 4.5A shows one such neuron, which responded preferentially to left turns during trials but not during intertrial intervals. Similar to R neurons, RE neurons can be further separated depending on their sensitivity to left or right turns (Figure 4.5B, C). For right turn preferring RE neurons, their +RE coefficients were larger than +R coefficients, while left turn preferring RE neurons showed larger -RE coefficients than -R coefficients. These results confirm that RE neurons were driven by actions specifically during sound.

Due to the design of our interactive behavior paradigm, the carrier frequency cannot increase beyond the upper limit of the spectrum (40 kHz) or decrease below the lower limit of the spectrum (10 kHz). Thus, any wheel turning beyond these spectrum boundaries would essentially produce no carrier frequency change and result in a decoupling between the action and sound (DCP, Figure 4.5D). One goal of the current study is to introduce perturbations in the form of DCP (“Delay Sound” and “Stop Frequency” perturbations, see below) during the interactive task to induce error responses. However, DCP happens naturally at the spectral boundaries due to

the design of the experiments. We thus hypothesize that A1 neurons could be sensitive to this form of DCP intrinsic to the task. Nevertheless, as DCP happened during wheel turning, DCP and RE events were closely temporally related (Figure 4.5E) and thus difficult to distinguish with the coarse temporal resolution of our first model (0.5 sec window). Given this reason, true DCP sensitive neurons could be a subset of the RE neurons. We explored this possibility first by examining the correlogram between RE neurons' inferred spikes and motion onset and the correlogram between spikes and the onset of DCP events (Figure 4.5F, G). For both left and right preferring RE neurons, their inferred spikes had shorter lags from DCP event onset than from motion onset, which suggests that these responses could be more directly driven by DCP events. Next, we used a model with higher temporal resolution to address this question (Figure 4.5H). First, we expressed the wheel turning events and DCP events by shrinking these events to their onset time with corresponding event size (Figure 4.5H, left top). Next, we convolved these events with a set of 9 tent bases spanning 2 seconds (Figure 4.5H, left bottom) to get the predictors (Figure 4.5H, right). We fit the inferred spikes from RE neurons to these predictors and examined their explanatory powers (Figure 4.5I). For RE predictors, the 3<sup>rd</sup> basis had the highest  $\Delta R^2$  while for DCP predictors, the 1<sup>st</sup> and 2<sup>nd</sup> basis had the highest  $\Delta R^2$ , which further suggests that RE neurons' responses were more temporally aligned with DCP onset. In addition, DCP predictors had a larger overall  $\Delta R^2$  than RE predictors (Figure 4.5I), suggesting that DCP could better explain the responses. We assigned the subset of neurons that had higher DCP  $\Delta R^2$  as DCP neurons while the rest as RE neurons (Figure 4.5J). Next, we examined the

fluorescence and inferred spike traces temporally aligned according to the DCP onset in both DCP and RE neurons (Figure 4.5K). DCP neurons had a larger increase in  $\Delta F/F$  following the DCP onset than RE neurons (Figure 4.5K). The inferred spike trace showed that DCP neurons had a sharp increase in firing rate following DCP onset. We further separated DCP neurons based on their turn direction selectivity and both right and left turn preferring DCP neurons showed the most responses following the DCP onset (Figure 4.5L, M). Together, these results suggest that a subset of sound and action jointly sensitive neurons indeed responded robustly to the DCP of action and frequency at the spectrum boundary. Their responses could be interpreted as error signals, as the relationship between action and carrier frequency changes breaks down at the spectrum boundary. These responses could also guide the behavior of the animal as they signal reaching the target zone.



**Figure 4.5 A subset of action-sound jointly sensitive neurons encode the spectrum boundary of the task paradigm**

(A) Top two rows: example left- and right-turning rotary activities from one active

session. Rotary activities during sound presentations are plotted in orange. Inter-trial rotary activities are plotted in blue. Bottom two rows: fluorescence trace and inferred spikes from one example action-sound jointly sensitive (RE) neuron from the same behavioral session. This neuron was preferentially activated by left turns during trials (arrow heads). (B) Left: boxplot showing the coefficients of right turn preferring RE neurons. In these neurons, the +RE term had the most prominent coefficients. Right: scatter plot showing the joint distribution of +R and +RE coefficients. +RE coefficients were higher than +R coefficients (Wilcoxon rank sum test,  $p=9.6\times 10^{-25}$ ). The black solid line is of slope 1. (C) The same as in (B) but for left turn preferring RE neurons. Similarly -RE term coefficients were larger than -R coefficients (Wilcoxon rank sum test,  $p=1.7\times 10^{-69}$ ). (D) The rotary (top) and frequency (bottom) trajectory of an example trial. The decoupling (DCP) period where the rotary activities were not translated into carrier frequency changes due to the spectrum boundary was labeled in black in both plots. The dotted line marks the onset of the action while the dashed line marks the onset of the DCP period. (E) The correlogram between onset of left turn events and the onset of DCP at low frequency target zone (DCP LTZ, purple curve) and the correlogram between onset of right turn events and the onset of DCP within the high frequency target zone (DCP HTZ, orange curve) are shown. Both correlograms peaked at around -0.2 sec. The solid lines represent the mean, while the shaded regions represent standard deviation. (F) Left, the correlogram between inferred spikes of right turn preferring RE neurons and the onset of either motion (left or right turn) or DCP events (LTZ and HTZ). Spiking activities had a shorter lag from HTZ DCP events than from right turn onset (Wilcoxon sign

rank test,  $p=9.2\times 10^{-11}$ ). (G) Same as in (F) but for left turn preferring RE neurons. Spiking activities had a shorter lag from LTZ DCP events than from left turn onset (Wilcoxon sign rank test,  $p=2.6\times 10^{-12}$ ). (H) Left top: we expressed the rotary activities and DCP from the example in (D) by assigning the amplitude of these events (measured in degrees) to the respective onset time. In this example, the left turning event was of  $\sim 70$  degrees from the motion onset to the end of the DCP period. The DCP event was of size  $\sim 20$  degrees. These events were then convolved with a basis set spanning 2 seconds (left bottom) to construct predictors with frame rate temporal resolution (right, top and bottom). (I) Left and middle: the  $\Delta R^2$  and coefficients as a function of the 9 bases plotted separately for RE and DCP predictors. For RE predictors, the 3<sup>rd</sup> basis had the most explanatory power, while for DCP predictors, the 1<sup>st</sup> and 2<sup>nd</sup> basis had the most explanatory power. Right: boxplot showing the group  $\Delta R^2$  from all bases plotted as a function of RE and DCP. DCP predictors had higher  $\Delta R^2$  than RE predictors (Wilcoxon sign rank test,  $p=2.2\times 10^{-6}$ ). (J) The scatter plot showing the joint distribution of group  $\Delta R^2$  of RE and DCP term. We identified the putative DCP sensitive neurons as the subset of neurons with higher DCP group  $\Delta R^2$ . The solid line has a slope of 1. (K) Left and middle: the fluorescence and inferred spikes of RE and DCP neurons as grouped in (I) were temporally aligned to the onset of DCP and plotted. DCP neurons showed larger responses following DCP onset compared to RE neurons (right, Wilcoxon rank sum test,  $p=1.3\times 10^{-15}$ ). (L) Left and middle: fluorescence and inferred spike traces of HTZ DCP preferring neurons temporally aligned to the DCP onset. These neurons had larger  $\Delta F/F$  responses to right turn induced DCP (right plot, Wilcoxon sign rank test,  $p=1.0\times 10^{-9}$ ).

(M) Left and middle: same as (L) but for LTZ DCP preferring neurons. These neurons had larger  $\Delta F/F$  responses to left turn induced DCP (right plot, Wilcoxon sign rank test,  $p=2.0 \times 10^{-39}$ ).

### Delay Sound (DS) perturbations induced both action driven and sound driven error responses

The identification of DCP sensitive neurons indicate that A1 neurons can signal errors induced by the mismatch between action and carrier frequency change. However, the DCP at the boundary of the spectrum is intrinsic to the task design and thus we hypothesize that such DCP was expected by the animal. In order to actively violate the expectations of the animal, we introduced two forms of perturbation: “Delay Sound” (DS) and “Stop Frequency” (SF). DS perturbations aimed to perturbate the timing relationship between action and sound (perturbate “when”) while SF perturbations aimed to perturbate the sensory feedback (perturbate “what”).

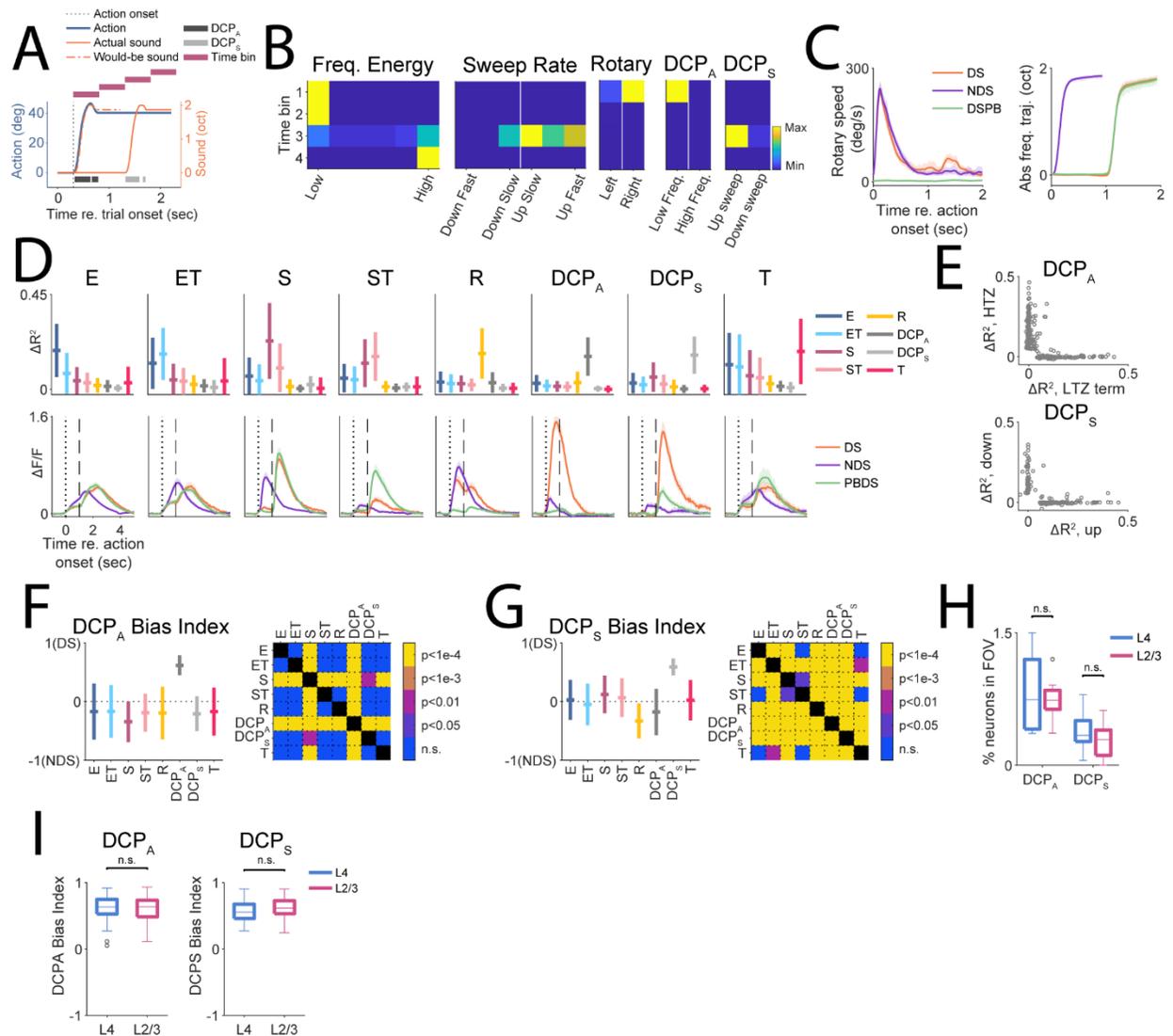
Figure 4.6A shows an example DS trial. We introduced a delay of 1 sec between the action and the carrier frequency change. Due to the delay, the carrier frequency remained same throughout the first 1 sec. Thus, any action during this period triggered an action induced DCP event ( $DCP_A$ ). After the 1 sec delay, carrier frequency started changing, but as it was not matched to the concurrent action, this triggers another DCP event but induced by sound ( $DCP_S$ ). In order to identify error responses triggered by  $DCP_A$  or  $DCP_S$ , we built a second linear model focused on the 2 sec after the action onset (Figure 4.6A). We constructed the predictors from 4 nonoverlapping time windows of 0.5 sec (Figure 4.6B). The first two windows fell

within the first 1 sec, which we denoted as the action window, while the other two windows fell within the second 1 sec, which we denoted as the sound window. Each trial thus contributed 4 data points to the model. In addition to DS trials, we also included non-DS (NDS) trials and playback of DS trials (DSPB). We only selected DS and NDS trials that resulted in hits where the animals showed a sufficient amount of turning, which was comparable across DS and NDS trials and resulted in similar frequency trajectories despite the delay in DS trials (Figure 4.6C). In the action window, both DS and NDS trials had wheel turning activities, while only NDS trials had corresponding frequency changes. DSPB trials shared the same frequency trajectories as DS trials but lacked wheel turning activities (Figure 4.6C, see also Figure 1D). These trials thus allowed us to distinguish the contribution from factors other than  $DCP_A$  or  $DCP_S$  to the neuronal responses. We proceeded to fit the model and identified different feature sensitive using a similar approach as in Figure 2B, and the average fluorescence traces from each group showed distinct patterns (Figure 4.6D). Most notably,  $DCP_A$  showed the most prominent responses in the action window during DS trials while the responses were much weaker during NDS trials in the action window. In contrast,  $DCP_S$  neurons showed prominent responses in the sound window in DS trials while the responses in the action window in NDS trials and the responses in the sound window in DSPB trials were also weaker. Other feature sensitive groups of neurons did not show such selectivity to DCP features. For example, the R neurons showed highly similar responses in the action window in both DS and NDS trials. The responses of  $DCP_A$  and  $DCP_S$  were also highly selective (Figure 4.6E). For  $DCP_A$  neurons, their responses were mostly explained by either

DCP at the low frequency target zone (resulted from right turns) or at the high frequency target zone (resulted from left turns). For DCP<sub>S</sub> neurons, their responses were evoked by either up frequency sweep or down frequency sweep. Thus, these error responses not only encoded the presence of DCP events, but they also encoded the specific content of the DCP events.

We proceeded to quantify the strength of these error signals. For DCP<sub>A</sub> induced errors, we computed the bias index between responses from the action windows in DS and NDS trials. For DCP<sub>S</sub> induced errors, we computed the bias index using the responses from the sound window in DS trials and responses from the action window in NDS trials. Among all feature sensitive neurons, DCP<sub>A</sub> and DCP<sub>S</sub> neurons showed the most positive bias indices and thus showed the strongest error responses (Figure 4.6F, G). These results suggest that our approach indeed identified putative error responses induced by DCP<sub>A</sub> and DCP<sub>S</sub>.

Next, we investigated whether such error signals had any dependence on cortical layers. First, we quantified the fraction of neurons in L4 and L2/3 that showed either DCP<sub>A</sub> or DCP<sub>S</sub> induced responses, but we found no difference. Similarly, L4 and L2/3 neurons showed no difference in the strength of either DCP<sub>A</sub> error responses or DCP<sub>S</sub> error responses (Figure 4.6I). Overall, these results suggest that these error signals were similarly represented in L4 and L2/3 of mouse A1.



**Figure 4.6 “Delay-Sound” (DS) perturbation induced error responses in a subset of A1 neurons**

(A) The frequency (blue solid curve) and rotary trajectories (orange solid curve) of one example DS trial are shown. The dotted dash curve shows the “would-be” frequency trajectory without the DS perturbation. Two forms of DCP (action induced, DCP<sub>A</sub> and sound induced DCP<sub>S</sub>) are marked in dark and light gray, respectively. Four windows of 0.5 sec duration were placed following the onset of the action to extract

the predictors as shown in (B). (B) The predictors extracted from the example trial in (A). (C) Left: average absolute rotary speed as a function of trial conditions. DS: delay-sound trials; NDS: non-delay-sound trials; DSPB, playback of delay-sound trials. Right: the corresponding absolute frequency trajectories of the three trial conditions. (D) Top: feature sensitive neurons were grouped according to their maximum  $\Delta R^2$  term. The horizontal lines represent the mean while the vertical lines represent standard deviation. Bottom: the average fluorescence traces from the corresponding feature sensitive neuronal groups as a function of trial conditions. The solid curves represent the mean while the shaded regions represent 95% confidence intervals. The dotted vertical lines mark the action onset while the dashed line mark the onset of the delayed frequency sweep. (E) Scatter plots showing the joint distribution of  $\Delta R^2$  of the LTZ and HTZ term for DCP<sub>A</sub> neurons (top) and the joint distribution of  $\Delta R^2$  of the up sweep and down sweep term for DCP<sub>S</sub> neurons. Both groups of DCP neurons showed a high selectivity for a single feature. (F) Left: DCP<sub>A</sub> bias index as a function of different feature sensitive groups shown in (D). The more positive the value of the bias index is, the higher the strength of the action induced error signal. Right: the multiple comparison between the different feature sensitive groups. (G) Left: DCP<sub>S</sub> bias index as a function of different feature sensitive groups shown in (D). The more positive the value of the bias index is, the higher the strength of the sound induced error signal. Right: the multiple comparison between the different feature sensitive groups. (H) Boxplot showing the percentage of neurons in a FOV that were assigned to either DCP<sub>A</sub> or DCP<sub>S</sub> groups as a function of cortical layer. There was no difference across layers (L4 vs. L2/3, Wilcoxon rank sum test,

DCP<sub>A</sub>, p=0.89; DCP<sub>S</sub>, p=0.26). (I) Boxplot showing the DCP<sub>A</sub> and DCP<sub>S</sub> bias indices as a function of cortical layer within DCP<sub>A</sub> and DCP<sub>S</sub> neurons, respectively. DCP<sub>A</sub> bias index showed no difference across layers (Wilcoxon rank sum test, p=0.42). DCP<sub>S</sub> bias index also showed no difference across layers (Wilcoxon rank sum test, significance adjusted with Bonferroni correction, p=0.032).

“Stop-Frequency” (SF) perturbations evoked frequency specific error responses in a subset of A1 neurons

The DS trials introduced action driven DCP events at the two starting frequencies. To further probe frequency dependency of DCP induced error responses, we introduced a second form of perturbation, i.e., “Stop-Frequency” (SF) perturbations. In SF trials, as the carrier frequency approached one of the SFs, randomly chosen for a given trial, the frequency stream would be ‘frozen’ for a brief period (0.25 sec) at the SF and thus introducing a frequency dependent DCP (Figure 4.7A). This brief action driven DCP created a deviation from the “would-be” frequency trajectory and we hypothesize that A1 also encodes DCP at these SFs and the amplitude of the DCP responses could be layer dependent.

To identify the neurons that were sensitive to DCP at these 4 SFs, we constructed a similar model to that of the DS perturbation but focused on the time immediately before and after the onset of the SF (see Method). Both SF trials and the corresponding PB trials were included, such that the contribution of sound feature selectivity to the responses could be accounted for. We also included non-perturbed ACT trials where the frequency trajectories were similar to the perturbed trials

immediately before the SF onset (Figure 4.7B, SIM trials). The similarity of the frequency trajectory and the presence of wheel turning activities help to account for the contributions from these factors to the neuronal responses. We hypothesize that true DCP responses are specific to SF trials and thus DCP sensitive neurons should show weaker responses to SIM and PB trials.

With this approach, we fitted a linear model using LASSO and we could identify neuronal groups that were sensitive to different features (Figure 4.7C). Next, we examined the response profile of these groups of neurons (Figure 4.7D). Except for DCP neurons, all other groups showed rather similar responses in SF and SIM trials, suggesting a lack of selectivity to DCP. In E, ET, S and ST neurons, PB trials tend to evoke larger responses, confirming a more suppressed state in ACT trials (see also Figure 2). R neurons showed similar responses in SF and SIM trials, with little activation in PB trials, confirming their sensitivity to actions. Moreover, these groups of neurons showed responses before DCP onset and thus were driven by factors other than DCP. In contrast, DCP neurons showed responses after DCP onset while showing much weaker responses in both SIM and PB trials. Thus, these neuron's responses cannot be accounted for by stimulus selectivity as well as the animal's action, and thus likely represent bona fide error responses. We quantified the strength of the error responses by computing a bias index between the response amplitude of the SF trials and the response amplitude of the SIM trials, i.e.,  $(\text{Resp}_{\text{SF}} - \text{Resp}_{\text{SIM}})/(\text{Resp}_{\text{SF}} + \text{Resp}_{\text{SIM}})$ . This analysis confirmed that among all groups of neurons, DCP neurons showed the highest selectivity towards SF trials (Figure 5E).

Together, these results showed that our approach could identify neurons sensitive to SF induced error responses.

Next, we investigated the frequency dependency of the error responses induced by SFs. We could separate DCP neurons into subgroups based on the particular SF that neurons were most responsive to (Figure 5F). These neurons showed a high degree of selectivity to SFs, as they responded primarily to one SF, which we defined as their Best SF (BSF). They also showed corresponding albeit weaker responses to the same SF in the PB trials, which suggests that these SF responses could be related to their tuning properties. Out of the total DCP neurons identified responding to SF (n=644, both layers included), 178 showed significant responses to pure tones in the passive setting. We plotted the BSFs against the BFs in these subset of tone responsive neurons and we found a positive relationship between BSFs and BFs within the spectral boundary (10 to 40 kHz, Figure 4.7G). This relationship holds both in L4 and L2/3. In addition, we quantified the distance between BSF and the closest frequency that evoked significant responses within the tuning curves of these neurons. The distribution of the distance centered around zero for both L4 and L2/3 (Figure 4.7G, right). These results suggest that in tone responsive DCP neurons, their SF selectivity is closely related to their tuning selectivity.

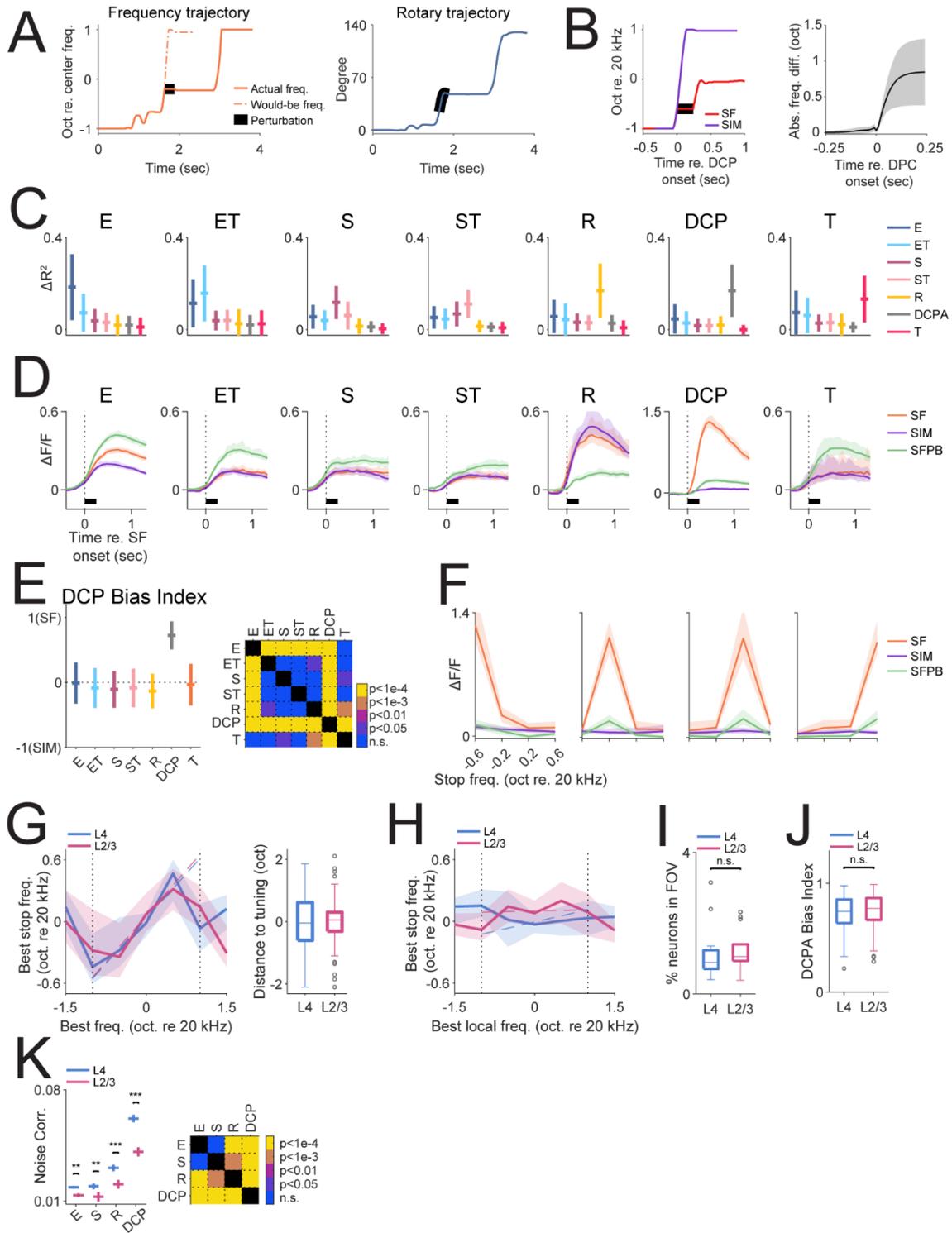
As the majority of these DCP neurons showed no tone responses, we investigated whether their SF selectivity could be predicted from the tuning properties of their local population. For this purpose, we defined local BF (LBF) as the frequency that most neurons responded to within a radius of 100  $\mu\text{m}$ . We plotted the

BSF against LBF, but we failed to find any relationship between the two measures (Figure 4.7H). This result shows that the SF selectivity is best predicted by the neuron's own tuning while the local population's frequency selectivity failed to predict SF selectivity, possibly due to the large variability in the local tuning preference (Bandyopadhyay et al., 2010; Rothschild et al., 2010).

Next, we investigated whether the encoding of SF induced error responses was layer dependent. We first quantified the fraction of SF responsive DCP neurons in L4 and L2/3, where we found no difference (Figure 4.7I). Next, we compared the strength of the error responses by computing the response bias index between SF and SIM trials defined as in Figure 4.7E and we found no difference in this measure (Figure 4.7J). These results suggest that L4 and L2/3 encode the SF induced error responses to a similar degree, mirroring the results from DS perturbations (Figure 5I).

The introduction of multiple SFs allowed us to investigate the correlational structure between neurons that were of different feature sensitivity, and we hypothesize that DCP neurons could have functional connectivity. To investigate this, we computed the noise correlation (NC), a measure that quantifies stimulus independent variability among neuronal pairs, either between DCP neurons and other feature groups of neurons (E, S and R, cross-group) or within DCP neurons (intra-group). We found that DCP neurons had the highest NC with other DCP neurons, while their NC with R neurons were higher than NC with E or S neurons (Figure 4.7K). We also found a consistently higher NC among neuronal pairs in L4 than in L2/3. These results suggest that DCP neurons form stronger functional networks within themselves while sharing stronger functional connectivity with action sensitive

neurons than sensory information sensitive neurons. These functional connections could underlie their highly nonlinear selectivity towards SF perturbations.



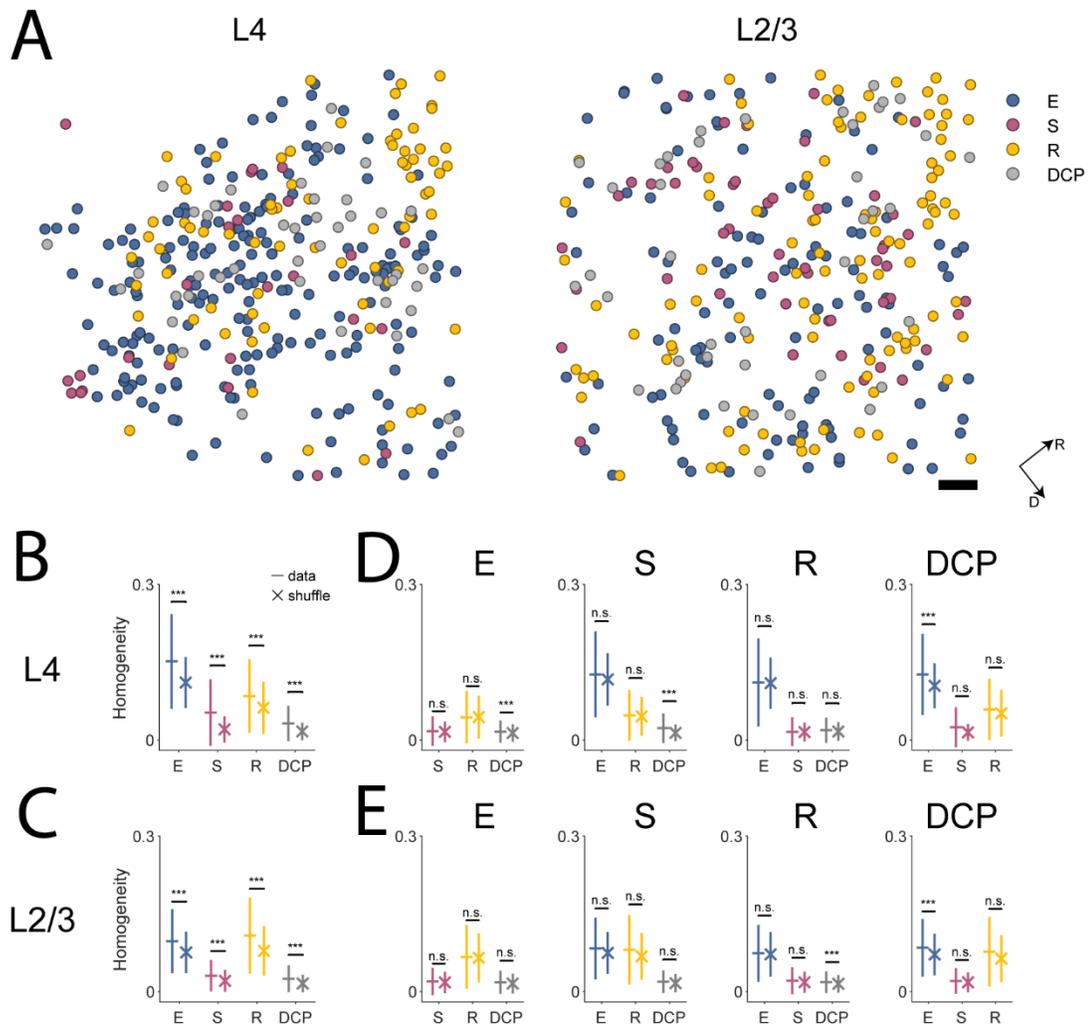
**Figure 4.7 “Stop-frequency” (SF) perturbations induced DCP responses at frequencies outside target zones**

(A) Example frequency and rotary trajectories of an example SF perturbation trial. Left: the orange solid line shows the frequency trajectory and the period marked in black indicates the occurrence of the perturbation, where the carrier frequency was frozen despite the rotary activity. The dash-dotted line shows the frequency trajectory if the perturbation had not been introduced. Right: the gray line shows the corresponding rotary activity of the trial, and the perturbation period was marked in black. (B) Left: the frequency trajectory of an example active trial with perturbation (red) and the frequency trajectory of an example SIM trial (purple) that shared a similar frequency trajectory to the SF trial immediately before DCP onset (marked in black). Right: the average absolute difference in frequency trajectories between SF trials and SIM trials. The difference was negligible before the perturbation onset and increased after perturbation onset. (C) The group  $\Delta R^2$  of identified feature sensitive neurons are shown. (D) Average  $\Delta F/F$  traces from SF, SIM and PB trials from the corresponding feature sensitive neurons as shown in (C). Solid lines show average across neurons while the shaded regions show 95% confidence interval. The vertical dotted lines indicate the onset of SF perturbation. The black bar indicates the duration of the SF perturbation. (E) Left: the bias indices quantifying the difference between SF and SIM trials were plotted as a function of feature sensitivity. Right: the multi-comparison table between the groups is shown. (F)  $DCP_A$  neurons showed distinct preference for SFs and neurons could be categorized by the SF that evoked the largest response. Each panel shows the average  $\Delta F/F$  responses over the subpopulation of

neurons with the same SF preference. (G) Left: for neurons showed passive tone responses, the SF that each DCP neuron was most sensitive to (Best Stop Frequency, BSF) was plotted against the neuron's best frequency (BF). BSF and BF were positively correlated within the carrier frequency boundary (10 and 40 kHz, marked by the dotted lines). The linear fit was plotted in dash lines (L4, slope=0.58,  $p=2.9 \times 10^{-5}$ ; L2/3, slope=0.59,  $p=3.0 \times 10^{-7}$ ). Right: for the same subset of tone responsive DCP neurons, the distance of BSF to the nearest frequency that evoked significant responses was shown as a function of cortical layer. Both distributions centered around zero. (H) Local BF (LBF), the frequency evoking the most responses from neurons within a radius of 100  $\mu\text{m}$  of DCP neurons, was plotted against the DCP neurons' BF. LBF was not correlated with BSF in DCP neurons (L4, slope=0.12,  $p=0.11$ ; L2/3, slope=0.01,  $p=0.87$ ). (I) Boxplot showing the bias indices from DCP<sub>A</sub> neurons as a function of cortical layer. No difference was found between L4 and L2/3 (Wilcoxon rank sum test,  $p=0.080$ ). (J) Boxplot showing the fraction of neurons per FOV identified as SF responsive as a function of cortical layer. No difference was found between L4 and L2/3 (Wilcoxon rank sum test,  $p=0.26$ ). (K) Left: cross-group noise correlation (NC) between DCP and E, S, R neurons (column 1-3) and intra-group noise correlation within DCP neurons (4<sup>th</sup> column). Vertical lines show the mean and the vertical lines show the SEM. L4 vs. L2/3, Wilcoxon rank sum test, NC DCP and E  $p=0.0012$ ; NC DP and S,  $p=0.0058$ ; NC DCP and R,  $p=2.6 \times 10^{-5}$ ; NC DCP vs DCP,  $p=4.6 \times 10^{-11}$ . Right: the multiple comparison table showing the difference of NC regardless of cortical layer.

### Feature sensitive neurons form spatial clusters

Lastly, we investigated the spatial distribution of different feature sensitive groups of neurons. Figure 4.8A shows two example FOVs from L4 and L2/3 respectively with identified feature sensitive neurons color coded. We hypothesize that neurons from the same feature sensitive group could form spatial clusters. To investigate this, we computed a homogeneity index (Deneux et al., 2016) that quantifies the fraction of neurons belonging to the same group within a 100  $\mu\text{m}$  radius and compared the distribution of values against shuffled data where the identity of neurons were assigned to randomly selected neurons in the FOV. This analysis revealed that all feature sensitive neurons were more likely to cluster with neurons of the same identity and thus were non-uniformly distributed in both L4 and L2/3 (Figure 4.8B, C). We further asked if this clustering also extended to neurons of different feature groups. For this purpose, we computed the fraction of neurons that belonged to other feature sensitive groups than the neuron at the center of the radius and compared the distribution of values against shuffled data (Figure 4.8D, E). The majority of the comparisons produced insignificant results, which suggest that neurons of different feature sensitive groups were largely randomly scattered across FOVs, despite their within-group spatial clustering, which might facilitate the interaction between neurons encoding different behaviorally relevant information by superimposing their spatial locations.



**Figure 4.8 Feature sensitive neurons form local clusters**

(A) Two example FOVs from L4 and L2/3 respectively showing the spatial distribution of feature sensitive neurons. DCP neurons contain all identified DCP sensitive neurons, i.e., spectral boundary, DS or SF sensitive neurons. E neurons contain both E and ET sensitive neurons from Figure 2. S neurons contain both S and ST sensitive neurons from Figure 2. R neurons contain both R and RE (excluding boundary sensitive neurons) from Figure 2. The black scale bar represents 100  $\mu\text{m}$ .

(B) and (C) The homogeneity indices that quantifies the fraction of local population

of neurons that belong to the same category are shown as a function of cortical layer. Data from all FOVs were pooled together. In both L4 and L2/3, each feature sensitive group showed significant within group clustering as actual data showed higher homogeneity indices than shuffled data (Wilcoxon rank sum test, data vs. shuffle, L4: E,  $p=2.2\times 10^{-128}$ ; S,  $p=4.9\times 10^{-29}$ ; R,  $p=8.2\times 10^{-22}$ ; DCP,  $9.4\times 10^{-17}$ ; L2/3: E,  $p=7.9\times 10^{-41}$ ; S,  $p=9.4\times 10^{-10}$ ; R,  $p=6.3\times 10^{-50}$ ; DCP,  $8.9\times 10^{-12}$ ). The markers show the mean while the vertical lines show the 95% confidence interval of the mean. (D) The cross-group homogeneity indices are shown as a function of the feature groups for L4. Each panel shows whether other feature groups spatially cluster in the vicinity of one particular feature group, e.g., the first panel shows the fraction of each group around E neurons. One tailed Wilcoxon rank sum test (significance adjusted with Bonferroni correction), E neurons: E vs. S,  $p=1$ ; E vs. R,  $p=1$ ; E vs DCP,  $p=4.1\times 10^{-5}$ ; S neurons: S vs. E,  $p=0.21$ ; S vs. R,  $p=0.96$ ; S vs. DCP,  $p=4.1\times 10^{-5}$ ; R neurons: R vs. E,  $p=1$ ; R vs. S,  $p=1$ ; R vs. DCP,  $p=0.56$ ; DCP neurons: DCP vs. E,  $p=2.0\times 10^{-5}$ ; DCP vs. S,  $p=0.0026$ ; DCP vs. R,  $p=0.29$ . (E) The same as in (D) but for L2/3 neurons. One tailed Wilcoxon rank sum test (significance adjusted with Bonferroni correction), E neurons: E vs. S,  $p=0.81$ ; E vs. R,  $p=1$ ; E vs DCP,  $p=0.022$ ; S neurons: S vs. E,  $p=0.12$ ; S vs. R,  $p=0.028$ ; S vs. DCP,  $p=0.061$ ; R neurons: R vs. E,  $p=0.97$ ; R vs. S,  $p=0.081$ ; R vs. DCP,  $p=5.5\times 10^{-6}$ ; DCP neurons: DCP vs. E,  $p=8.2\times 10^{-5}$ ; DCP vs. S,  $p=0.30$ ; DCP vs. R,  $p=0.022$ .

## Discussion

In this study, we trained mice to perform a novel and interactive auditory task while imaged large populations of neurons in A1. We identified separate groups of neurons sensitive to sound or action related information, and we showed that while L4 contained more sound encoding neurons, L2/3 contained more action driven neurons. For sound encoding neurons, their responses were typically more suppressed during active sessions than during passive sessions. Furthermore, we identified two groups of action sensitive neurons. One group was activated by actions both within trials and during intertrial intervals. However, their responses were more suppressed during trials where sound was presented. The second group of neurons were preferentially driven by the joint presence of action and sound, and we further show that a subset of these neurons encode the spectral boundary of our interactive task. Next, we introduced two forms of perturbations, i.e., “Delay-Sound” and “Stop-Frequency”, to evoke error responses in A1. We identified neurons in L4 and L2/3 that were selective to these perturbations, but we did not observe a difference in the strength of these error signals across the cortical layer. Finally, we show that the identified feature sensitive neurons were spatially clustered. Together, our results show that given our interactive behavioral task, a significant portion of A1 neurons were dedicated to encoding action related information, which could give rise to the highly nonlinear responses we observed either at the spectral boundary or evoked by the DS and SF perturbations.

Our interactive behavioral design has allowed us to systematically manipulate the coupling between action and sound feedback, such that error responses could be evoked. This approach is inspired by studies in the visual system where visual flow mismatch evoked error responses in the mouse visual cortex (Keller et al., 2012; Zmarz and Keller, 2016). The continuous coupling between wheel turning and the carrier frequency change is similar to the coupling of running behavior and the visual flow feedback. In such designs, the animal's action actively modulates the attribute of the sensory feedback, and we believe such tight coupling better engages the predictive mechanism in the brain. In the auditory system, similar efforts have been made to couple the animal's action, e.g., locomotion or lever pushing, with the auditory feedback (Rummell et al., 2016; Schneider et al., 2018; Audette et al., 2021). In these studies that employed a simple coupling, the sound was often only secondary to the task and might not necessarily require attention to the sound per se (Audette et al., 2021). Our task introduced a more complex mapping between sound and action while requiring the animal to pay attention to the sound presented, i.e., the direction of the turn depends on the initial frequency of the sound stream. Thus, the design of the current study is more suitable for the predictive coding framework.

In our task we used wheel turning to report the animal's behavior. We monitored the turning activities throughout the entire experiment session and together with the sound information, we were able to rule out responses due to sound and action while identifying putative bona fide error responses. Consistent with previous reports (Musall et al., 2019), we observed considerable action induced responses in A1 and identified action sensitive neurons through linear regression. What's more, we

found neurons responding specifically to reward consumption and these responses were robust and consistent. These findings further call for a disassociation of choice reporting behavior (wheel turning) and reward consuming behavior (licking) as the latter could produce responses that resemble sensory processing. In an interactive task design, it is impossible to temporally separate sensory processing signals from motor signals, and thus it requires the modeling approach, e.g., linear regression, to explain neuronal responses.

The above approach allowed us to identify neurons with specific selectivity to wheel turning directions. These neurons were activated by the animal's action, and it is possible that these neurons receive long range input from motor related areas. It has been shown that both primary (M1) and secondary motor cortex (M2) have projections to A1, and M2 innervate both pyramidal and parvalbumin positive neurons in A1, although the net effect of M2 activation in A1 could be feedforward inhibition (Nelson et al., 2013; Schneider et al., 2014). On the other hand, basal forebrain has also been shown to project to A1 and these cholinergic inputs in A1 are active during movement (Nelson and Mooney, 2016). It is thus possible that the action sensitive neurons identified in this study receive primarily input from basal forebrain given the positive sign of their responses. The selectivity of the turn directions in the action neurons also suggest that these neurons were not activated by the sound, if any, caused by the movement. In addition, the action neurons were also among the least tone responsive. Thus, these action neurons could serve as a functional network that broadcasts action specific signal to other A1 neurons, potentially giving rise to the action dependent error responses.

In this study, we identified both action driven and sound driven error responses in A1. The action driven error responses were evoked when the carrier frequency change failed to follow the wheel turning activities, which happened either due to the boundary of the frequency spectrum or due to the introduction of the perturbations. The error responses evoked by these decoupling events thus represent a highly nonlinear sensitivity to the combination of a stationary carrier frequency and concurrent wheel turning. These error responses could be interpreted as signaling the violation of the fundamental rule of the task, i.e., turning the wheel changes the carrier frequency. On the mechanism level, these responses could result from the nonlinear interactions between action encoding and sound encoding neurons. In DS trials, the action driven error responses were time locked to the onset of the action, while in SF trials, the error responses occurred after the SF onset. These phenomena further suggest that these responses can be explained in terms of low-level sound or action selectivity, but such selectivity was highly dynamic and dependent on previous action and sensory history.

One of the key hypotheses of the predictive coding theory is that cortical layers could form hierarchical structures in terms of error signaling (Heilbron and Chait, 2017). We thus investigated whether the strength of error signals differed across L4 and L2/3 in mouse A1. Despite the fact that more L2/3 neurons were action sensitive, we did not find a systematic shift in the error response strength across L4 and L2/3. This finding suggests that cortical layers might not be the minimum functional unit in terms of generating error responses, and it is possible that the hierarchy exists only across different brain regions.

In conclusion, we designed a novel interactive auditory task that allowed us to probe complex interactions between sound and action in mouse A1. We identified both sensory information and action sensitive neurons in A1, as well as neurons signaling error responses evoked by the decoupling between action and sound. Our approach potentially provides a new road map in terms of studying predictive coding in the auditory system.

# Chapter 5. Summary and Discussion

## Summary

In this dissertation, I used imaging techniques of various spatial scales to study the encoding of different aspects of sound in the mouse auditory cortex (ACX). 2-photon imaging allowed monitoring of large populations of neurons with high spatial resolutions while single photon or widefield imaging allowed monitoring activities of different auditory areas. With these techniques, I first studied the encoding of sound onset and offset in mouse ACX. I found that offset responses are tonotopically expressed over the surface of ACX on the scale of millimeters. With single photon imaging, I also found that offset responses dominate ACX, especially at high sound levels. At a finer spatial resolution, the ACX neuronal onset and offset responses mirror the responses seen at the mesoscale and the cortical offset could be resultant from the interplay between cortical microcircuit and thalamic feedforward input. Second, I used a modified version of the two-tone paradigm to study the inhibitory sideband structures in mouse primary ACX (A1). This paradigm allowed me to simultaneously infer the inhibitory sidebands of large populations of neurons. I classified excitatory neurons based on the shape of their receptive field (frequency response area, FRA) and revealed that both inhibitory sidebands and the amount of nonlinear spectral interactions depend on the particular FRA shape. Further, I showed that inhibitory neurons also showed inhibitory sidebands. Lastly, I studied the encoding of the interaction between sound and action in mouse A1 using a novel interactive behavioral paradigm. I found that the encoding of sound stimuli depended

on the behavioral state of the animal, and that the responses of the A1 neurons could be explained by various factors including their turning behaviors. Furthermore, I found that a subset of A1 neurons encode the decoupling between sound and action, and thus represent the general rule of the task. Together, these studies provided both new knowledge and new methodology on the studying of sound encoding in mouse ACX.

## **Discussion and Future Directions**

### Cortical processing of offset responses

In Chapter 2, I studied the encoding of sound offset in mouse ACX in not only excitatory neurons, but also in inhibitory neurons, as well as in thalamic input. We thus hypothesize that cortical offset responses are not generated *de novo* but rather inherited from thalamic feedforward input. However, cortical responses were more biased towards offset response than thalamic terminal responses. This result suggests that cortical microcircuit further processed sound offset responses and this processing could depend on the dynamics between excitatory and inhibitory neurons.

In the two types of common inhibitory neurons we studied, parvalbumin positive (PV) and Somatostatin positive (SST) showed mostly opposite sign of responses to prolonged presentations of tones. SST neurons typically showed a graduate increase in fluorescence while PV neurons showed a decrease of fluorescence. It is thus possible that SST neurons inhibit both PV and pyramidal cells

during the prolonged stimulus, and we hypothesize that because of the slow dynamics of the SST neuron's responses, their inhibition on local PV and pyramidal cells could extend for a brief period after the sound termination, causing a brief disinhibition that concur with the arrival of the feedforward offset responses relayed by the thalamic terminals. Thus, if this hypothesis is true, then inhibiting SST neurons during the sound presentation or transiently before the sound offset could reduce the offset responses in a local population of pyramidal cells.

To specifically test this hypothesis, one could employ optogenetic manipulations of SST neurons. Recent development in the generation more efficient opsins as well as in spatially and temporally precise manipulations of the neuronal populations have made this possible (Marshall et al., 2019). Specifically, one need to cross Thy1-GCaMP6s mice (Dana et al., 2014) with SST-cre mice, while injecting cre dependent red shifted inhibitory opsins such as Jaws (Chuong et al., 2014). Spatially precise targeting of SST neurons could be achieved with spatial light modulator (SLM), which is capable of generating arbitrary spatial patterns of stimulation (Nikolenko et al., 2008). Combining these approaches thus allows spatially and temporally accurate manipulations of the SST neurons, and I expect that inhibiting SST neurons either throughout the tone presentation or briefly before the tone offset would reduce offset responses in local pyramidal neurons.

### Spectral tuning in other areas than A1

In Chapter 3, I characterized the complex inhibitory sideband in mouse A1 and I show that the degree of nonlinear spectral interactions is associated with the FRA types. However, there had been very few studies focusing on other auditory cortical areas such as Anterior Auditory Field (AAF) and secondary auditory cortex (A2). Recently, it has been shown that spectrally dense sound stimuli resembling vocalizations preferentially active A2 (Kline et al., 2021). Specifically, A2 neurons responded more strongly than A1 neurons when the set of harmonics were temporally aligned. Therefore, it is possible that A2 neurons integrate spectral energy differently than A1 neurons. I have found that A1 neurons receive strong sideband inhibition, and A2 neurons could show the opposite effect with both broader spectral tuning as well as less sideband inhibition. In terms of AAF, a previous study has shown largely similar receptive field properties between A1 and AAF (Linden et al., 2003). However, anesthetized animals were used in this study and more recent imaging studies also suggest that the canonical tonotopic map used to identify AAF could be inaccurate (Issa et al., 2014; Romero et al., 2020). It thus calls for renewed effort to revisit spectral tuning in AAF, using imaging technique that allows better locating of AAF as well as the monitoring of large ensemble of auditory cortical neurons.

Therefore, one future direction to extend the result of Chapter 3 is to investigate the tuning properties of the AAF and A2 neurons with the same approach, i.e., using a similar two-tone paradigm to simultaneously infer the spectral tunings of large populations of neurons. In Chapter 3, efforts were made to classify auditory neurons with different FRA shapes. The validity of the classification could benefit

from the addition of the FRA shapes from both AAF and A2, which could further establish the spectral tuning preference of different auditory fields. Similarly, the same approach could extend to the common inhibitory neurons in AAF and A2.

### Interactive behavioral paradigm and cortical hierarchy

In Chapter 4, I designed a novel interactive behavioral paradigm with the goal to study the encoding of both action and sensory information in mouse A1. In this study, I observed significant action driven activities in mouse A1. Several questions remained to be addressed in this paradigm.

First of all, it is possible that the action driven activities could be the result of the form of action used to report animal's choice, i.e., wheel turning. To address this concern, one needs to perform a control experiment where the mouse is required to use the wheel for a simple tone discrimination task. This task would require the mouse to discriminate between two tones that share the same frequency content as the starting frequencies in our interactive behavioral task, i.e., 10 and 40 kHz. The mouse would be required to turn left for the 40 kHz and to turn right for the 10 kHz.

However, in this simple discrimination task, the carrier frequency is not coupled to the action, and thus this task is open-loop in nature. I hypothesize that the significant action driven activities in mouse A1 are specific to the interactive task, as these action driven responses provide a substrate for nonlinear interactions in A1 that gave rise to the decoupling induced error responses. Therefore, I would expect that in the control experiments with the open-loop design while the animal is required to turn the wheel

with the same amount, one would not observe such strong relay of action relevant information in A1.

Second, it remains to be shown which areas relay the action related information in A1. It has been shown that both primary (M1) and secondary motor (M2) cortex send long range input into A1 (Nelson et al., 2013). M2 projections in A1 both synapse onto pyramidal cells and inhibitory interneurons such as PV neurons (Nelson et al., 2013). However, it is also shown that the net effect of activating M2 projections is feedforward inhibitions in A1, which is the opposite of our observations in A1. Thus, M2 might not be the candidate region that activates action sensitive neurons in our study. Nevertheless, the artificial broad activation of M2 projections could mask the potential activations. In addition, extensive training with the interactive task could also modify the relative strength of the projection onto pyramidal neurons and inhibitory interneurons. Therefore, it remains to be tested whether M2 projections are involved in sending action related information to A1.

Another candidate for investigation is the basal forebrain (Nelson and Mooney, 2016). Although the basal forebrain also innervates both excitatory and inhibitory neurons in A1, the net effect of optogenetic stimulation is excitatory and thus is more in line with our observation of increased activity triggered by wheel turning. It has been shown that the basal forebrain terminals in A1 activate during locomotion and thus I expect that it also conveys information related to wheel turning to A1.

To specifically test the contribution of M2 and basal forebrain, one could express calcium indicators of different emission spectrum and simultaneously image

the terminal activities in A1 of trained mice. For example, one could use viral injection to target M2 for GCaMP6s expression, while targeting basal forebrain with RGECCO, a red shifted calcium indicator (Dana et al., 2016). By choosing an excitation wavelength suitable for both fluorophores, one could perform dual color imaging, and thus readout activities from both M2 and basal forebrain. Comparing the temporal coherence with wheel turning activities could shed light on which region is primarily responsible for driving the action related activities in A1.

Thirdly, it is unclear whether the decoupling induced error responses emerge in A1 or if it is present in earlier auditory processing stages. To answer this question, one would need to perform recordings from auditory subcortical nucleus, such as medial geniculate body (MGB) and inferior colliculus (IC). Indeed, it has also been shown that M2 also projects to MGB (Nelson et al., 2013). Thus, it is possible that the decoupling induced error responses could appear early in the auditory system.

Finally, in Chapter 4 I compared the strength of error responses across L4 and L2/3 in mouse A1 and I hypothesize that the magnitude of the error responses could depend on the cortical layer. However, there was little difference between the L4 and L2/3, and both showed similar error responses in terms of action driven decoupling and sound driven decoupling. Thus it is likely that the hierarchy, if any, of error signaling could exist between auditory processing stages but not necessarily across different cortical layers in the same area. Thus, to test this hypothesis, one would need to perform similar recording in higher order auditory areas such as A2.

According to the “Driver-Modulator” theory, A1 L5 neurons project to dorsal MGB, which projects to L4 of A2 (Lee and Sherman, 2010). Therefore, A2 presents as a

higher order processing stage, and it is possible that the error responses undergo further amplification through this pathway. In addition, ACX has also been shown to project to primary motor cortex (Zhang et al., 2016) and it remains to be investigated how these error signals interact with higher order brain areas and whether the neurons carrying error responses in A1 also have long projection targets. To answer this question would require a strategy that allows tagging neurons based on their activity. One candidate for this purpose would be CaMPARI, a calcium indicator that irreversibly changes color upon violet light illumination (Fosque et al., 2015). Simultaneous illumination with violet light during the introduction of decoupling could potentially label the neurons signaling error responses and their long-range projection target, if any, could be examined through histology.

Overall, the interactive behavior task design has the potential to provide rich information about the integration of action and sound along the auditory pathway, and I believe it will be a fruitful path to pursue.

# Bibliography

- Ahissar E, Sosnik R, Haidarliu S (2000) Transformation from temporal to rate coding in a somatosensory thalamocortical pathway. *Nature* 406:302.
- Aitkin LM, Webster WR (1972) Medial geniculate body of the cat: organization and responses to tonal stimuli of neurons in ventral division. *J Neurophysiol* 35:365-380.
- Atencio CA, Sharpee TO, Schreiner CE (2008) Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58:956-966.
- Atencio CA, Sharpee TO, Schreiner CE (2012) Receptive field dimensionality increases from the auditory midbrain to cortex. *Journal of neurophysiology* 107:2594-2603.
- Attinger A, Wang B, Keller GB (2017) Visuomotor coupling shapes the functional development of mouse visual cortex. *Cell* 169:1291-1302. e1214.
- Audette NJ, Zhou W, Schneider DM (2021) Temporally precise movement-based predictions in the mouse auditory cortex. *bioRxiv:2021.2012.2013.472457*.
- Baba H, Tsukano H, Hishida R, Takahashi K, Horii A, Takahashi S, Shibuki K (2016) Auditory cortical field coding long-lasting tonal offsets in mice. *Sci Rep* 6:34421.
- Bandyopadhyay S, Shamma SA, Kanold PO (2010) Dichotomy of functional organization in the mouse auditory cortex. *Nat Neurosci* 13:361-368.
- Bartlett EL, Sadagopan S, Wang X (2011) Fine frequency tuning in monkey auditory cortex and thalamus. *Journal of neurophysiology* 106:849-859.
- Baumann S, Griffiths TD, Sun L, Petkov CI, Thiele A, Rees A (2011) Orthogonal representation of sound dimensions in the primate midbrain. *Nature neuroscience* 14:423-425.
- Baumann S, Joly O, Rees A, Petkov CI, Sun L, Thiele A, Griffiths TD (2015) The topography of frequency and time representation in primate auditory cortices. *Elife* 4:e03256.
- Bayazitov IT, Westmoreland JJ, Zakharenko SS (2013) Forward Suppression in the auditory cortex is caused by the Cav3. 1 calcium channel-mediated switch from bursting to tonic firing at thalamocortical projections. *Journal of Neuroscience* 33:18940-18950.
- Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35:1798-1828.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*:1165-1188.
- Bowen Z, Winkowski DE, Kanold PO (2020) Functional organization of mouse primary auditory cortex in adult C57BL/6 and F1 (CBAXC57) mice. *Scientific reports* 10:1-14.
- Bregman AS (1994) *Auditory scene analysis: The perceptual organization of sound*: MIT press.
- Bridi MS, Shin SY, Huang S, Kirkwood A (2020) Dynamic recovery from depression enables rate encoding in inhibitory synapses. *Iscience*:100940.

- Brimijoin WO, O'Neill WE (2005) On the prediction of sweep rate and directional selectivity for FM sounds from two-tone interactions in the inferior colliculus. *Hearing research* 210:63-79.
- Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. *Journal of neurophysiology* 77:923-943.
- Bruno RM, Sakmann B (2006) Cortex is driven by weak but synchronously active thalamocortical synapses. *Science* 312:1622-1627.
- Carbajal GV, Malmierca MS (2018) The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends in Hearing* 22:2331216518784822.
- Chen I-W, Helmchen F, Lütcke H (2015) Specific early and late oddball-evoked responses in excitatory and inhibitory neurons of mouse auditory cortex. *Journal of Neuroscience* 35:12560-12573.
- Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* 499:295.
- Chuong AS, Miri ML, Busskamp V, Matthews GA, Acker LC, Sørensen AT, Young A, Klapoetke NC, Henninger MA, Kodandaramaiah SB (2014) Noninvasive optical inhibition with a red-shifted microbial rhodopsin. *Nature neuroscience* 17:1123-1129.
- Clayton KK, Williamson RS, Hancock KE, Tasaka G-i, Mizrahi A, Hackett TA, Polley DB (2021) Auditory corticothalamic neurons are recruited by motor preparatory inputs. *Current Biology* 31:310-321. e315.
- Cook PB, McReynolds JS (1998) Lateral inhibition in the inner retina is important for spatial tuning of ganglion cells. *Nature neuroscience* 1:714-719.
- Cottam JC, Smith SL, Häusser M (2013) Target-specific effects of somatostatin-expressing interneurons on neocortical visual processing. *Journal of Neuroscience* 33:19567-19578.
- Dana H, Chen TW, Hu A, Shields BC, Guo C, Looger LL, Kim DS, Svoboda K (2014) Thy1-GCaMP6 transgenic mice for neuronal population imaging in vivo. *PLoS One* 9:e108697.
- Dana H, Mohar B, Sun Y, Narayan S, Gordus A, Hasseman JP, Tsegaye G, Holt GT, Hu A, Walpita D (2016) Sensitive red protein calcium indicators for imaging neural activity. *Elife* 5:e12727.
- Davis KA, Young ED (2000) Pharmacological evidence of inhibitory and disinhibitory neuronal circuits in dorsal cochlear nucleus. *Journal of neurophysiology* 83:926-940.
- Decharms RC, Merzenich MM (1996) Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381:610-613.
- Deneux T, Kempf A, Daret A, Ponsot E, Bathellier B (2016) Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities. *Nature communications* 7.
- Escabi MA, Schreiner CE (2002) Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *Journal of Neuroscience* 22:4114-4131.

- Famiglietti E, Kolb H (1976) Structural basis for ON-and OFF-center responses in retinal ganglion cells. *Science* 194:193-195.
- Fettiplace R (2020) Diverse Mechanisms of Sound Frequency Discrimination in the Vertebrate Cochlea. *Trends in Neurosciences* 43:88-102.
- Fino E, Yuste R (2011) Dense inhibitory connectivity in neocortex. *Neuron* 69:1188-1203.
- Fishman YI, Steinschneider M (2009) Temporally dynamic frequency tuning of population responses in monkey primary auditory cortex. *Hear Res* 254:64-76.
- Forli A, Vecchia D, Binini N, Succol F, Bovetti S, Moretti C, Nespoli F, Mahn M, Baker CA, Bolton MM (2018) Two-photon bidirectional control and imaging of neuronal excitability with high spatial resolution in vivo. *Cell reports* 22:3087-3098.
- Fosque BF, Sun Y, Dana H, Yang C-T, Ohyama T, Tadross MR, Patel R, Zlatic M, Kim DS, Ahrens MB (2015) Labeling of active neural circuits in vivo with designed calcium integrators. *Science* 347:755-760.
- Francis NA, Winkowski DE, Sheikhattar A, Armengol K, Babadi B, Kanold PO (2018) Small Networks Encode Decision-Making in Primary Auditory Cortex. *Neuron* 97:885-897 e886.
- Friedrich J, Zhou P, Paninski L (2017) Fast online deconvolution of calcium imaging data. *PLoS computational biology* 13:e1005423.
- Frisina RD, Singh A, Bak M, Bozorg S, Seth R, Zhu X (2011) F1 (CBA× C57) mice show superior hearing in old age relative to their parental strains: Hybrid vigor or a new animal model for “Golden Ears”? *Neurobiology of aging* 32:1716-1724.
- Friston K (2005) A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences* 360:815-836.
- Friston K (2010) The free-energy principle: a unified brain theory? *Nature reviews neuroscience* 11:127-138.
- Friston K, Moran R, Seth AK (2013) Analysing connectivity with Granger causality and dynamic causal modelling. *Current opinion in neurobiology* 23:172-178.
- Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature neuroscience* 6:1216-1223.
- Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA (2010) Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nature neuroscience* 13:1011-1019.
- Froemke RC, Merzenich MM, Schreiner CE (2007) A synaptic memory trace for cortical receptive field plasticity. *Nature* 450:425-429.
- Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*:424-438.
- Greenwood D, Merzenich M, Roth G (1976) Some preliminary observations on the interrelations between two-tone suppression and combination-tone driving in the anteroventral cochlear nucleus of the cat. *The Journal of the Acoustical Society of America* 59:607-633.

- Guo W, Chambers AR, Darrow KN, Hancock KE, Shinn-Cunningham BG, Polley DB (2012) Robustness of cortical topography across fields, laminae, anesthetic states, and neurophysiological signal types. *J Neurosci* 32:9159-9172.
- Hackett TA, Barkat TR, O'Brien BM, Hensch TK, Polley DB (2011) Linking topography to tonotopy in the mouse auditory thalamocortical circuit. *J Neurosci* 31:2983-2995.
- He J (2001) On and off pathways segregated at the auditory thalamus of the guinea pig. *J Neurosci* 21:8672-8679.
- He J (2002) OFF responses in the auditory thalamus of the guinea pig. *Journal of Neurophysiology* 88:2377-2386.
- He J, Hashikawa T, Ojima H, Kinouchi Y (1997) Temporal integration and duration tuning in the dorsal zone of cat auditory cortex. *J Neurosci* 17:2615-2625.
- Heil P, Rajan R, Irvine DR (1992) Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. I: Effects of variation of stimulus parameters. *Hear Res* 63:108-134.
- Heilbron M, Chait M (2017) Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*.
- Henry KR (1985) Tuning of the auditory brainstem OFF responses is complementary to tuning of the auditory brainstem ON response. *Hearing research* 19:115-125.
- Herreras O (2016) Local field potentials: myths and misunderstandings. *Frontiers in neural circuits* 10:101.
- Hillyard SA, Picton TW (1978) On and off components in the auditory evoked potential. *Perception & Psychophysics* 24:391-398.
- Hillyard SA, Hink RF, Schwent VL, Picton TW (1973) Electrical signs of selective attention in the human brain. *Science* 182:177-180.
- Honma Y, Tsukano H, Horie M, Ohshima S, Tohmi M, Kubota Y, Takahashi K, Hishida R, Takahashi S, Shibuki K (2013) Auditory cortical areas activated by slow frequency-modulated sounds in mice. *PLoS One* 8:e68113.
- Hromádka T, Deweese MR, Zador AM (2008) Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol* 6:e16.
- Huang CL, Winer JA (2000) Auditory thalamocortical projections in the cat: laminar and areal patterns of input. *J Comp Neurol* 427:302-331.
- Imig TJ, Morel A (1983) Organization of the thalamocortical auditory system in the cat. *Annu Rev Neurosci* 6:95-120.
- Issa JB, Haefele BD, Young ED, Yue DT (2017) Multiscale mapping of frequency sweep rate in mouse auditory cortex. *Hear Res* 344:207-222.
- Issa JB, Haefele BD, Agarwal A, Bergles DE, Young ED, Yue DT (2014) Multiscale optical Ca<sup>2+</sup> imaging of tonal organization in mouse auditory cortex. *Neuron* 83:944-959.
- Jen PH-S, Wu FJ, Chen QC (2002) The effect of two-tone stimulation on responses of two simultaneously recorded neurons in the inferior colliculus of the big brown bat, *Eptesicus fuscus*. *Hearing research* 168:139-149.

- Ji X-y, Zingg B, Mesik L, Xiao Z, Zhang LI, Tao HW (2015) Thalamocortical innervation pattern in mouse auditory and visual cortex: laminar and cell-type specificity. *Cerebral Cortex* 26:2612-2625.
- Joachimsthaler B, Uhlmann M, Miller F, Ehret G, Kurt S (2014) Quantitative analysis of neuronal response properties in primary and higher-order auditory cortical fields of awake house mice (*Mus musculus*). *Eur J Neurosci* 39:904-918.
- Kane KL, Longo-Guess CM, Gagnon LH, Ding D, Salvi RJ, Johnson KR (2012) Genetic background effects on age-related hearing loss associated with *Cdh23* variants in mice. *Hear Res* 283:80-88.
- Kanold PO, Nelken I, Polley DB (2014) Local versus global scales of organization in auditory cortex. *Trends Neurosci* 37:502-510.
- Karnani MM, Jackson J, Ayzenshtat I, Tucciarone J, Manoocheri K, Snider WG, Yuste R (2016) Cooperative subnetworks of molecularly similar interneurons in mouse neocortex. *Neuron* 90:86-100.
- Kato HK, Gillet SN, Isaacson JS (2015) Flexible sensory representations in auditory cortex driven by behavioral relevance. *Neuron* 88:1027-1039.
- Kato HK, Asinof SK, Isaacson JS (2017) Network-level control of frequency tuning in auditory cortex. *Neuron* 95:412-423. e414.
- Katzner S, Nauhaus I, Benucci A, Bonin V, Ringach DL, Carandini M (2009) Local origin of field potentials in visual cortex. *Neuron* 61:35-41.
- Keller GB, Bonhoeffer T, Hübener M (2012) Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* 74:809-815.
- Kim S, Putrino D, Ghosh S, Brown EN (2011) A Granger causality measure for point process models of ensemble neural spiking activity. *PLoS computational biology* 7:e1001110.
- Kline AM, Aponte DA, Tsukano H, Giovannucci A, Kato HK (2021) Inhibitory gating of coincidence-dependent sensory binding in secondary auditory cortex. *Nature communications* 12:1-17.
- Konrad-Martin DL, Ruëbsamen R, Doërrscheidt GJ, Rubel EW (1998) Development of single- and two-tone responses of anteroventral cochlear nucleus neurons in gerbil. *Hearing research* 121:35-52.
- Kopp-Scheinflug C, Tozer AJ, Robinson SW, Tempel BL, Hennig MH, Forsythe ID (2011) The sound of silence: ionic mechanisms encoding sound termination. *Neuron* 71:911-925.
- Koralek K-A, Jensen KF, Killackey HP (1988) Evidence for two complementary patterns of thalamic input to the rat somatosensory cortex. *Brain research* 463:346-351.
- Kuchibhotla KV, Gill JV, Lindsay GW, Papadoyannis ES, Field RE, Sten TAH, Miller KD, Froemke RC (2017) Parallel processing by cortical inhibition enables context-dependent behavior. *Nature neuroscience* 20:62-71.
- Kuffler SW (1953) Discharge patterns and functional organization of mammalian retina. *Journal of neurophysiology* 16:37-68.
- Lakunina AA, Nardoci MB, Ahmadian Y, Jaramillo S (2020) Somatostatin-expressing interneurons in the auditory cortex mediate sustained suppression by spectral surround. *Journal of Neuroscience*.

- Langner G, Albert M, Briede T (2002) Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (*Chinchilla laniger*). *Hear Res* 168:110-130.
- Lee CC, Sherman SM (2008) Synaptic properties of thalamic and intracortical inputs to layer 4 of the first- and higher-order cortical areas in the auditory and somatosensory systems. *J Neurophysiol* 100:317-326.
- Lee CC, Winer JA (2008) Connections of cat auditory cortex: I. Thalamocortical system. *Journal of Comparative Neurology* 507:1879-1900.
- Lee CC, Sherman SM (2010) Drivers and modulators in the central auditory pathways. *Frontiers in neuroscience* 4:14.
- Levy RB, Reyes AD (2012) Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *Journal of Neuroscience* 32:5609-5619.
- Li H, Liang F, Zhong W, Yan L, Mesik L, Xiao Z, Tao HW, Zhang LI (2019) Synaptic mechanisms for bandwidth tuning in awake mouse primary auditory cortex. *Cerebral Cortex* 29:2998-3009.
- Li L-y, Li Y-t, Zhou M, Tao HW, Zhang LI (2013) Intracortical multiplication of thalamocortical signals in mouse auditory cortex. *Nature neuroscience* 16:1179.
- Li L-y, Xiong XR, Ibrahim LA, Yuan W, Tao HW, Zhang LI (2014a) Differential receptive field properties of parvalbumin and somatostatin inhibitory neurons in mouse auditory cortex. *Cerebral cortex* 25:1782-1791.
- Li L-y, Ji X-y, Liang F, Li Y-t, Xiao Z, Tao HW, Zhang LI (2014b) A feedforward inhibitory circuit mediates lateral refinement of sensory representation in upper layer 2/3 of mouse primary auditory cortex. *Journal of Neuroscience* 34:13670-13683.
- Liang F, Li H, Chou X-l, Zhou M, Zhang NK, Xiao Z, Zhang KK, Tao HW, Zhang LI (2018) Sparse Representation in Awake Auditory Cortex: Cell-type Dependence, Synaptic Mechanisms, Developmental Emergence, and Modulation. *Cerebral Cortex*.
- Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM (2003) Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J Neurophysiol* 90:2660-2675.
- Liu B-h, Wu GK, Arbuckle R, Tao HW, Zhang LI (2007) Defining cortical frequency tuning with recurrent excitatory circuitry. *Nature neuroscience* 10:1594-1600.
- Liu J, Kanold PO (2021) Diversity of receptive fields and sideband inhibition with complex thalamocortical and intracortical origin in L2/3 of mouse primary auditory cortex. *Journal of Neuroscience* 41:3142-3162.
- Liu J, Whiteway MR, Sheikhattar A, Butts DA, Babadi B, Kanold PO (2019) Parallel Processing of Sound Dynamics across Mouse Auditory Cortex via Spatially Patterned Thalamic Inputs and Distinct Areal Intracortical Circuits. *Cell Rep* 27:872-885 e877.
- Liu X, Zhou L, Ding F, Wang Y, Yan J (2015) Local field potentials are local events in the mouse auditory cortex. *European Journal of Neuroscience* 42:2289-2297.

- Llano DA, Sherman SM (2008) Evidence for nonreciprocal organization of the mouse auditory thalamocortical-corticothalamic projection systems. *J Comp Neurol* 507:1209-1227.
- Lu S-M, Lin RC-S (1993) Thalamic afferents of the rat barrel cortex: a light-and electron-microscopic study using Phaseolus vulgaris leucoagglutinin as an anterograde tracer. *Somatosensory & motor research* 10:1-16.
- Malmierca MS, Cristaudo S, Pérez-González D, Covey E (2009) Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *Journal of Neuroscience* 29:5483-5493.
- Maor I, Shalev A, Mizrahi A (2016) Distinct spatiotemporal response properties of excitatory versus inhibitory neurons in the mouse auditory cortex. *Cerebral Cortex* 26:4242-4252.
- Marshall JH, Kim YS, Machado TA, Quirin S, Benson B, Kadmon J, Raja C, Chibukhchyan A, Ramakrishnan C, Inoue M (2019) Cortical layer-specific critical dynamics triggering perception. *Science* 365:eaaw5202.
- Mayko Z, Roberts PD, Portfors C (2012) Inhibition shapes selectivity to vocalizations in the inferior colliculus of awake mice. *Frontiers in neural circuits* 6:73.
- McGinley MJ, David SV, McCormick DA (2015) Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron* 87:179-192.
- Mendelson J, Cynader M (1985) Sensitivity of cat primary auditory cortex (AI) neurons to the direction and rate of frequency modulation. *Brain research* 327:331-335.
- Meng X, Winkowski DE, Kao JP, Kanold PO (2017) Sublaminar subdivision of mouse auditory cortex layer 2/3 based on functional translaminar connections. *Journal of Neuroscience* 37:10200-10214.
- Merzenich MM, Knight PL, Roth GL (1975) Representation of cochlea within primary auditory cortex in the cat. *J Neurophysiol* 38:231-249.
- Miller KD, Pinto DJ, Simons DJ (2001) Processing in layer 4 of the neocortical circuit: new insights from visual and somatosensory cortex. *Current opinion in neurobiology* 11:488-497.
- Mineault PJ, Tring E, Trachtenberg JT, Ringach DL (2016) Enhanced spatial resolution during locomotion and heightened attention in mouse primary visual cortex. *Journal of Neuroscience* 36:6382-6392.
- Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK (2019) Single-trial neural dynamics are dominated by richly varied movements. *Nature neuroscience* 22:1677-1686.
- Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, Goldberg EM, Geffen MN (2015) Complementary control of sensory adaptation by two types of cortical interneurons. *Elife* 4.
- Nelken I, Young ED (1994) Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus type IV units to narrowband and wideband stimuli. *Journal of neurophysiology* 71:2446-2462.
- Nelken I, Prut Y, Vaadia E, Abeles M (1994) Population responses to multifrequency sounds in the cat auditory cortex: one-and two-parameter families of sounds. *Hearing research* 72:206-222.

- Nelson A, Mooney R (2016) The basal forebrain and motor cortex provide convergent yet distinct movement-related inputs to the auditory cortex. *Neuron* 90:635-648.
- Nelson A, Schneider DM, Takatoh J, Sakurai K, Wang F, Mooney R (2013) A circuit for motor cortical modulation of auditory cortical activity. *Journal of Neuroscience* 33:14342-14353.
- Niell CM, Stryker MP (2010) Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65:472-479.
- Nikolenko V, Watson BO, Araya R, Woodruff A, Peterka DS, Yuste R (2008) SLM microscopy: scanless two-photon imaging and photostimulation using spatial light modulators. *Frontiers in neural circuits* 2:5.
- Otazu GH, Tai L-H, Yang Y, Zador AM (2009) Engaging in an auditory task suppresses responses in auditory cortex. *Nature neuroscience* 12:646-654.
- Oviedo HV (2017) Connectivity motifs of inhibitory neurons in the mouse auditory cortex. *Scientific reports* 7:1-9.
- Oya H, Poon PW, Brugge JF, Reale RA, Kawasaki H, Volkov IO, Howard III MA (2007) Functional connections between auditory cortical fields in humans revealed by Granger causality analysis of intra-cranial evoked potentials to sounds: comparison of two methods. *Biosystems* 89:198-207.
- Pachitariu M, Stringer C, Schröder S, Dipoppa M, Rossi LF, Carandini M, Harris KD (2016) Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *BioRxiv:061507*.
- Pakan JM, Lowe SC, Dylida E, Keemink SW, Currie SP, Coutts CA, Rochefort NL (2016) Behavioral-state modulation of inhibition is context-dependent and cell type specific in mouse visual cortex. *Elife* 5:e14985.
- Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. *Nature neuroscience* 7:658-663.
- Petrus E, Isaiah A, Jones AP, Li D, Wang H, Lee HK, Kanold PO (2014) Crossmodal induction of thalamocortical potentiation leads to enhanced information processing in the auditory cortex. *Neuron* 81:664-673.
- Pfeffer CK, Xue M, He M, Huang ZJ, Scanziani M (2013) Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience* 16:1068.
- Polley DB, Read HL, Storace DA, Merzenich MM (2007) Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *J Neurophysiol* 97:3621-3638.
- Qin L, Chimoto S, Sakai M, Wang J, Sato Y (2007) Comparison between offset and onset responses of primary auditory cortex ON-OFF neurons in awake cats. *Journal of neurophysiology* 97:3421-3431.
- Recanzone GH (2000) Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. *Hear Res* 150:104-118.
- Recanzone GH, Guard DC, Phan ML (2000) Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J Neurophysiol* 83:2315-2331.

- Redies H, Brandner S (1991) Functional organization of the auditory thalamus in the guinea pig. *Exp Brain Res* 86:384-392.
- Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, Tolias AS (2014) Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *neuron* 84:355-362.
- Romero S, Hight AE, Clayton KK, Resnik J, Williamson RS, Hancock KE, Polley DB (2020) Cellular and widefield imaging of sound frequency organization in primary and higher order fields of the mouse auditory cortex. *Cerebral Cortex* 30:1603-1622.
- Rothschild G, Nelken I, Mizrahi A (2010) Functional organization and population dynamics in the mouse primary auditory cortex. *Nature neuroscience* 13:353-360.
- Ruggero MA, Robles L, Rich NC (1992) Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression. *Journal of neurophysiology* 68:1087-1099.
- Rummell BP, Klee JL, Sigurdsson T (2016) Attenuation of responses to self-generated sounds in auditory cortical neurons. *Journal of Neuroscience* 36:12010-12026.
- Russell I, Sellick P (1977) Tuning properties of cochlear hair cells. *Nature* 267:858-860.
- Saldeitis K, Happel MF, Ohl FW, Scheich H, Budinger E (2014) Anatomy of the auditory thalamocortical system in the mongolian gerbil: Nuclear origins and cortical field-, layer-, and frequency-specificities. *Journal of Comparative Neurology* 522:2397-2430.
- Schmidt M, Berg E, Friedlander M, Murphy K (2009) Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm. In: *Artificial Intelligence and Statistics*, pp 456-463.
- Schneider DM, Nelson A, Mooney R (2014) A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* 513:189-194.
- Schneider DM, Sundararajan J, Mooney R (2018) A cortical filter that learns to suppress the acoustic consequences of movement. *Nature* 561:391-395.
- Scholl B, Gao X, Wehr M (2010) Nonoverlapping sets of synapses drive on responses and off responses in auditory cortex. *Neuron* 65:412-421.
- Schreiner C (1981) Poststimulatory effects in the medial geniculate body of guinea pigs. In: *Neuronal mechanisms of hearing*, pp 191-196: Springer.
- Schreiner CE, Urbas JV (1986) Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). *Hear Res* 21:227-241.
- Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60:1823-1840.
- Schreiner CE, Urbas JV (1988) Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res* 32:49-63.

- Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience* 18:3870-3896.
- Sharpee TO, Atencio CA, Schreiner CE (2011) Hierarchical representations in the auditory cortex. *Current opinion in neurobiology* 21:761-767.
- Sheikhhattar A, Babadi B (2016) Dynamic estimation of causal influences in sparsely-interacting neuronal ensembles. In: *Information Science and Systems (CISS), 2016 Annual Conference on*, pp 551-556: IEEE.
- Sheikhhattar A, Miran S, Liu J, Fritz JB, Shamma SA, Kanold PO, Babadi B (2018) Extracting neuronal functional network dynamics via adaptive Granger causality analysis. *Proceedings of the National Academy of Sciences*:201718154.
- Somers DC, Nelson SB, Sur M (1995) An emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of Neuroscience* 15:5448-5465.
- Stiebler I, Neulist R, Fichtel I, Ehret G (1997) The auditory cortex of the house mouse: left-right differences, tonotopic organization and quantitative analysis of frequency representation. *J Comp Physiol A* 181:559-571.
- Sutter M, Schreiner C, McLean M, O'connor K, Loftus W (1999) Organization of inhibitory frequency receptive fields in cat primary auditory cortex. *Journal of Neurophysiology* 82:2358-2371.
- Sutter ML, Schreiner CE (1991) Physiology and topography of neurons with multi-peaked tuning curves in cat primary auditory cortex. *Journal of neurophysiology* 65:1207-1226.
- Tao C, Zhang G, Zhou C, Wang L, Yan S, Tao HW, Zhang LI, Zhou Y, Xiong Y (2017) Diversity in excitation-inhibition mismatch underlies local functional heterogeneity in the rat auditory cortex. *Cell reports* 19:521-531.
- Thevenaz P, Ruttimann UE, Unser M (1998) A pyramid approach to subpixel registration based on intensity. *IEEE transactions on image processing* 7:27-41.
- Tian B, Kuśmierk P, Rauschecker JP (2013) Analogues of simple and complex cells in rhesus monkey auditory cortex. *Proceedings of the National Academy of Sciences*:201221062.
- Trujillo M, Measor K, Carrasco MM, Razak KA (2011) Selectivity for the rate of frequency-modulated sweeps in the mouse auditory cortex. *Journal of Neurophysiology* 106:2825-2837.
- Tsukano H, Horie M, Bo T, Uchimura A, Hishida R, Kudoh M, Takahashi K, Takebayashi H, Shibuki K (2015) Delineation of a frequency-organized region isolated from the mouse primary auditory cortex. *J Neurophysiol* 113:2900-2920.
- Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. *Nature neuroscience* 6:391.
- Valdés-Sosa PA, Sánchez-Bornot JM, Lage-Castellanos A, Vega-Hernández M, Bosch-Bayard J, Melie-García L, Canales-Rodríguez E (2005) Estimating brain functional connectivity with sparse multivariate autoregression.

- Philosophical Transactions of the Royal Society of London B: Biological Sciences 360:969-981.
- Vasquez-Lopez SA, Weissenberger Y, Lohse M, Keating P, King AJ, Dahmen JC (2017) Thalamic input to auditory cortex is locally heterogeneous but globally tonotopic. *eLife* 6.
- Volkov I, Galazjuk A (1991) Formation of spike response to sound tones in cat auditory cortex neurons: interaction of excitatory and inhibitory effects. *Neuroscience* 43:307-321.
- Von Békésy G, Wever EG (1960) *Experiments in hearing*: McGraw-Hill New York.
- Whiteway MR, Butts DA (2017) Revealing unobserved factors underlying cortical activity with a rectified latent variable model applied to neural population recordings. *J Neurophysiol* 117:919-936.
- Williamson RS, Hancock KE, Shinn-Cunningham BG, Polley DB (2015) Locomotion and task demands differentially modulate thalamic audiovisual processing during active search. *Current Biology* 25:1885-1891.
- Winkowski DE, Kanold PO (2013) Laminar transformation of frequency organization in auditory cortex. *Journal of Neuroscience* 33:1498-1508.
- Yu C, Derdikman D, Haidarliu S, Ahissar E (2006) Parallel thalamic pathways for whisking and touch signals in the rat. *PLoS biology* 4:e124.
- Zariwala HA, Borghuis BG, Hoogland TM, Madisen L, Tian L, De Zeeuw CI, Zeng H, Looger LL, Svoboda K, Chen T-W (2012) A Cre-dependent GCaMP3 reporter mouse for neuronal imaging in vivo. *Journal of Neuroscience* 32:3131-3141.
- Zhang S, Xu M, Chang W-C, Ma C, Do JPH, Jeong D, Lei T, Fan JL, Dan Y (2016) Organization of long-range inputs and outputs of frontal cortex for top-down control. *Nature neuroscience* 19:1733-1742.
- Zhou M, Liang F, Xiong XR, Li L, Li H, Xiao Z, Tao HW, Zhang LI (2014) Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. *Nature neuroscience* 17:841-850.
- Zmarz P, Keller GB (2016) Mismatch receptive fields in mouse visual cortex. *Neuron* 92:766-772.