Encoding of Concurrent Sounds in the Monkey Inferior Colliculus

by

Shawn M. Willett

Department of Neurobiology
Duke University

Date:_____

Approved:

_____
Jennifer M. Groh, Supervisor

_____
Stephen G. Lisberger

_____
Lindsey L. Glickfeld

_____
Henry S. Greenside

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Neurobiology in the Graduate School
of Duke University

2020

ABSTRACT

Encoding of Concurrent Sounds in the Monkey Inferior Colliculus

by

Shawn M. Willett

Department of Neurobiology
Duke University

Date:_____
Approved:

_____
Jennifer M. Groh, Supervisor

_____
Stephen G. Lisberger

_____
Lindsey L. Glickfeld

_____
Henry S. Greenside

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Neurobiology in the Graduate School of
Duke University

2020

Abstract

The inferior colliculus (IC) is an auditory midbrain nucleus essential to the perception of sound frequency and the localization of sound source; yet it remains unclear how the firing rate of primate IC neurons contribute to the localization of concurrent sounds of variable sound frequencies. In this work, I extracellularly recorded the activity of 105 IC neurons while two adult macaque monkeys reported the location(s) of either a single bandpass filtered sound or two concurrent bandpass filtered sounds spatially separated by 24° and separated in sound frequency by 0.25 - 2 octaves. Monkeys performed this task well, with an accuracy of about 80% on single sound trials and about 90% on dual sound trials. The improvement in performance on dual sound trials was not explained by dual sound modulations of IC neural response functions. On dual sound trials, IC neuron receptive fields broadened, and sound frequency accounted for less variance in the dual sound response; and these changes decreased the performance of a maximum-likelihood decoder in correctly labeling the condition of a held out dual sound trial by about 20%. Overall, these results suggest that changes to the IC neural response functions elicited by the presence of a second, concurrent, sound should impair rather than facilitate the IC encoding of concurrent sounds and that an alternative explanation is required to account for monkey performance. I next investigated if recently discovered response alternations, suggested to underlie the

encoding of concurrent sounds, were present in the recorded populations. These response alternations occur when an IC neuron alternates its firing rate between the rate corresponding to each component sound of a dual sound pair. These response alternations were observed in about 60% of IC neurons and their contribution to the population response remained stable across the full, 2 octave, range of frequency separations tested. Thus, response alternations are a general mechanism used by the IC to potentially facilitate the encoding of multiple sounds and these results add to a growing body of work observing response alternations across brain areas. The measurements I performed clearly indicate that neurons in the primate IC are sensitive to not only sound frequency and location but also the number of sounds in the environment. Future empirical and theoretical work is needed to elucidate how exactly these response alternations arise and are read out by downstream neurons to allow for the perception of concurrent sounds.

Dedication

I dedicate this work to my parents Suzanne and Todd Willett, who filled my childhood with opportunities to engage with the wonders of the natural world. Their support made possible my pursuit of a doctoral degree. Without them I would not be here. Thank you.

Contents

List of Figures

Acknowledgements

This work was made possible by the unbridled positivity, enthusiasm, and support provided by my mentor Jennifer Groh. I am indebted to many of the Duke faculty, especially those on my committee, Henry Greenside, Lindsey Glickfeld and Stephen Lisberger and also Surya Tokdar, John Pearson, Nina Sherwood, Dale Purves, and Jeff Beck for feedback on my work and numerous invigorating conversations. I also owe much to the many members of the Duke neuroscience community for creating a stimulating and engrossing learning environment, particularly I would like to thank Nicolas Brunel and my close friends Charlie Giattino, Sam Brudner and Mike Young. I am grateful to the past and present members of the Groh lab for friendship, lively debate and encouragement; I am especially thankful to Jeff Mohl, Stephanie Schlebusch and Valeria Caruso. Lastly, I would like to thank the many staff who went above and beyond their station in helping me throughout my time at Duke, specifically I thank Yimin Wei, Edward McLaurin, Tanya Schreiber, and LaDonna Huseman.

1. Introduction

The natural world is cluttered with objects that generate sensations in nervous systems. To behave successfully in these noisy environments sensory systems must encode and maintain information about behaviorally relevant cues, even if these cues overlap in time, space, or content. Although done for rigor and interpretability of the results, most work in sensory neuroscience is done in contrived environments where animals are presented with a solitary stimulus; and, most of the work done with multiple stimuli probe how neural populations integrate cues from different senses (e.g. Stein & Stanford, 2008) or how extraneous cues are filtered out of neural signals (e.g.Maunsell, 2015). So, the field's understanding of how a population of neurons represents simultaneous stimuli, as is often faced in nature, is quite limited. The current work sheds light on this topic by investigating the representation of concurrent sounds in the monkey inferior colliculus (IC), an auditory midbrain nucleus.

In the chapters below, I present evidence that concurrent sounds reduce neural selectivity for sound frequency, degrading the representation of sounds in the IC (chapter 3). I then show that recently discovered response fluctuations (Caruso et al., 2018) occur in a large proportion of IC neurons and consistently contribute to the population response even for sound pairs with wide frequency separation, indicating these response fluctuations are a general mechanism IC neurons exploit to encode

concurrent sounds (chapter 4). The following sections provide background and motivation for this work.

1.1 The representation of sound in the auditory midbrain

The auditory system is responsible for encoding the spectral, temporal, and spatial properties of sounds. It does this with, arguably, the most elegant biology in the known universe. Sounds are initially generated by a vibrating source which creates a change in air pressure (air is the typical medium for primates). In the mammalian system, these sound waves are reflected and amplified by the external auditory meatus (i.e. pinna and ear canal) before it reaches the tympanic membrane (i.e. ear drum). The movement of the tympanic membrane results in vibrations of the middle ear ossicles initiating a traveling wave in the fluid filled cochlea (i.e. inner ear). Due to the unique and extraordinary properties of the inner ear (for review see: Hudspeth, 2014), the traveling wave propagates a specific distance along the cochlea dependent on the sound frequency, where it is then transduced into neural signals by inner hair cells residing on the basilar membrane. Therefore, there is a contiguous mapping of sound frequency along the basilar membrane, a property known as tonotopy. Downstream nuclei inherit this tonotopy making it the dominant organizational principle of the auditory system. The cochlea then projects to the cochlear nucleus which in turn projects auditory signals into a multitude of parallels pathways (Cant & Benson, 2003).

Crucially, the transduction of sound waves is reliant on only the amplitude and spectral content of a given sound and intrinsically contains no spatial information about the sound source. Indeed, localization of sounds must be computed from additional cues. Some of these cues, initially outlined around the turn of the 20th century (Rayleigh, 1907; Thompson, 1882), arise due to differences between sounds at the two ears and are referred to as binaural cues. The first cue arises due to a sound arriving at one ear earlier than the other generating an interaural time difference (ITD). The second cue arises due to the acoustic shadow of the head creating an interaural level difference (ILD). These cues are extracted through specialized connections between the cochlear nucleus and the superior olivary complex and are predominately thought to subserve localization of sounds in the horizontal plane (for review see:Grothe, Pecka, & McAlpine, 2010). A third cue, thought to underlie vertical localization, arises monoaurally and depends on spectral analysis of the reflection of different sound frequencies by the pinna (Blauert, 1997). Interestingly, these spectral cues are learned and can be altered through changes to the external ear throughout life (Hofman, Van Riswick, & Van Opstal, 1998). Nonetheless, a large aspect of the subcortical auditory system evolved to extract spatial information from the monaural and binaural cues.

Notably, invertebrates, birds and reptiles, amphibians, and mammals independently evolved tympanic membranes (Grothe et al., 2010). So, although there are well described and elegant systems for sound localization described in other species

(Knudsen & Konishi, 1978; Mason, Oshinsky, & Hoy, 2001) the implementation in the mammalian system could be quite different. The monkey auditory system, then, provides a number of advantages for studying the representation of concurrent stimuli. First, the monkey auditory system is quite similar to that of humans with largely overlapping acoustic ranges (Moore, 2012; Pfingst, Laycock, Flammino, Lonsbury-Martin, & Martin, 1978; Stebbins, Green, & Miller, 1966), a similar coarse tonotopic organization of the IC (Bulkin & Groh, 2011; Ress & Chandrasekaran, 2013), and similar patterns of ILDs and ITDs measured by the macaque head-related transfer function (Spezio, Keller, Marrocco, & Takahashi, 2000). Secondly, monkeys can be trained to perform many demanding behaviors. In the present work monkeys were trained to report the location of a single or each of two concurrent sound(s) by an eye movement to the sound source(s). Behavior is necessary to ground the interpretation of neural activity (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017), and in the case of the present work the behavioral task ensures neural activity recorded on dual sound trials underlies the encoding of multiple sounds.

The foremost reason to investigate the representation of concurrent sounds within the inferior colliculus (IC) is because it is a coding bottleneck. Nearly every ascending auditory pathway converges into the IC (Aitkin & Phillips, 1984) bestowing IC neurons with sensitivity to a myriad of sound parameters (for review see: *The Inferior Colliculus*, 2005). Among them it is known that the IC contributes to the perception of

sound frequency (Lim, Lenarz, & Lenarz, 2009; Pages et al., 2016) and sound localization (Jenkins & Masterton, 1982; Kelly & Kavanagh, 1994). Importantly, the representation of sound frequency and sound location in the primate IC are coarse (Bulkin & Groh, 2011; Groh, Kelly, & Underhill, 2003; A. Ryan & Miller, 1978; Versnel, Zwiers, & van Opstal, 2009; Zwiers, Versnel, & Van Opstal, 2004). For example, a given pure tone can activate 20-80% of frequency modulated sites (Bulkin & Groh, 2011). Spatial receptive fields are also broad, typically responding monotonically as a function of sound eccentricity (Groh et al., 2003; Zwiers et al., 2004). In short, the nearly complete convergence of auditory signals into the IC implies that if there are perceived sounds, information about each should be observable in the neural activity of IC neurons.

The rate-based nature of the code for sound location (McAlpine, Jiang, & Palmer, 2001) is particularly puzzling for the localization of concurrent sounds, since neurons cannot fire at two rates simultaneously. How, then, do humans and monkeys localize simultaneously presented sound?

A potential way to overcome this rate code constraint is to potentially segregate out the signal for each sound into segregated neural populations. The most likely mechanism to segregate neural populations in the monkey IC is through frequency selectivity. In fact, it is known that sound frequency can play an important role in concurrent sound localization. For example, if two sounds of the same center frequency are presented simultaneously the perceived sound source is a phantom directly between

the two actual sources, an illusion known as summing localization. However, as the frequency separation of the two sounds increases the illusion attenuates (Blauert, 1997). This was also seen in previous work in monkey behavior when performance on a dual sound localization task improved as the frequency distance between the two sounds increased (Caruso et al., 2018).

The following chapters investigate the role of frequency tuning and frequency separation in the encoding of concurrent sounds.

2. Methods

This section describes the methods shared between the following chapters.

2.1 General

All procedures were approved by the Duke Institutional Animal Care and Use Committee. Two adult, female, rhesus monkeys participated in this study. General procedures have been previously described in greater detail (Caruso et al., 2018; Groh, Trause, Underhill, Clark, & Inati, 2001). Briefly, under aseptic conditions a head post and a scleral search coil (Judge, Richmond, & Chu, 1980; Robinson, 1963), were surgically implanted. After recovery monkeys were trained to perform a sound localization task as described below. Once the task was learned, a recording chamber (Crist Instruments) was implanted over the left IC (in both monkeys) based on stereotaxic coordinates and was confirmed after implantation with structural MRI (Groh et al., 2001)

2.2 Training and behavior

During behavioral training and recording sessions monkeys performed single and dual sound localization tasks as shown in figure 1A. Monkeys were seated in a dark, sound attenuated room with head movements restrained. The illumination of a fixation light (LED) at 0° in both the horizontal and vertical dimensions signified the start of the trial. Monkeys began the trial by acquiring and holding fixation on the light. After a variable duration of fixation (600-700 ms) either a single sound or two simultaneous sounds would be presented while the monkey held fixation. After a

variable overlap of fixation and sound presentation (1000-1100 ms) the fixation light was turned off cueing the monkey to saccade to the origin of the sound or sounds. The monkey had a brief temporal window for their eyes to leave the fixation window and enter the first target window. The monkey then had to hold fixation in the target window briefly before initiating a second saccade to the second target window or until the single sound trial ended. The sound remained on until the end of the trial. If the monkeys performed the trial correctly, they were rewarded with a grape juice solution (30% grape juice, 70% water). Reward was delivered through a drinking tube and controlled via a solenoid.

The sound localization task used in this study was designed and implemented with proprietary software (Beethoven, RYKLIN Software Inc) and was adapted from a previous study (Caruso et al., 2018). There were two major adaptations. First, additional sound frequencies were used. Second, sounds in this study are only located at ±12°, meaning on dual sound trials the sounds are 24° apart compared to 30° in Caruso et al. (2018).

Initially, monkeys were trained to saccade to the origin of a sound (e.g. to the location of the speaker). Monkey Y was then trained to saccade to the source of sounds presented in a sequence. Over time the inter-sound-interval was decreased until eventually the sounds were presented simultaneously (Caruso et al., 2018). Monkey Y

was trained on all conditions except the control conditions. Monkey N was trained on the final task, including all conditions.

Eye position was acquired through an implanted scleral coil and Riverbend field system. Eye coil outputs were calibrated each day before behavioral and recording sessions. All behavioral data, including eye position data, were saved in Beethoven files for offline analysis.

2.3 Sound stimuli

All sounds consisted of bandpass filtered noise constructed by adding a series of tones. Sounds were frozen for each experimental session; that is for any given experimental session every instance of the sound stimulus of a given frequency involved the same waveform. The center frequencies were equally spaced in 0.25 octave intervals: 420, 500, 595, 707, 841, 1000, 1189, 1414, 1682 and 2000 Hz. These sounds were chosen due to the large proportion of IC neurons responsive at these frequencies (Bulkin & Groh, 2011), increasing the chance that more than one sound would significantly drive any recorded neuron.

All sounds were generated using proprietary software (Beethoven, RYKLIN Software Inc.) and passed through appropriate amplifiers to two speakers matched in their frequency response characteristics. Speakers were positioned at ±12° in the horizontal azimuth and at 0° in elevation at a distance approximately 1 meter from the monkey. Sounds were calibrated using a microphone at the position where the

monkey's head would be during the experiment. Sounds were calibrated to be 55 dB SPL when presented in isolation; these stimuli resulted in 60 dB SPL sound reaching the monkey on dual sound trials. The intensity difference is not expected to have substantive effects on neural activity or behavior (Caruso et al., 2018). We kept the sound source signals constant across conditions because in the real-world simultaneously occurring sounds are louder than when the component sounds are presented in isolation. All experiments were conducted in an anechoic chamber; the wall and ceilings were lined with echo attenuating foam (Sonex) and the floor was carpeted.

Across all experimental sessions there were 56 conditions, involving either one or two sounds in combination. For simplicity, we will refer to the sounds in two subsets, "A" and "B". The A sounds were used to assess the frequency tuning curves and consisted of the 8 middle sound frequencies (500 to 1682 Hz). The B sounds flanked the A sounds, at either 420 or 2000 Hz. On single sound trials any one of the 10 (8 A + 2 B) sounds could be presented in isolation at either ± 12°. On dual sound trials, an A sound was presented at either ± 12° and paired with a B sound at the opposite location. The dual and single sound trials accounted for 52 of the 56 conditions (8 A sounds x 3 paired conditions (none, 420 or 2000 Hz) x 2 locations = 48 + 2 B sounds presented in isolation x 2 locations = 56 conditions). To ensure monkeys were actively perceiving both sounds and not saccading to remembered locations in response to more complex spectral content (e.g. the dual sound trials have power in more bands than single sound trials),

an added control accounted for the last 4 conditions. The control trial presented a 1189 Hz sound with either B sound at either ±12°, that is to say the 1189 Hz and B sound were targeted to the same speaker location (2 sounds (1189+420 Hz or 1189+2000 Hz) x 2 locations = 4). Monkeys typically made single saccades on these control trials and reached high performance levels.

To acquire enough trials per condition and to accommodate cell holding times (typically 1-2 hours), the number of conditions per recording session were limited by presenting the A sounds at only one location (either ±12°) each day. Therefore, each recording session consisted of 32 conditions. The eight A sounds presented alone or in combination with either B sound accounted for 24 of the 32 conditions (8 sounds x 3 paired tones x 1 location = 24). To allow counterbalancing, the two B sounds were still presented at both locations (2 sounds x 2 locations = 4). The control conditions were not changed across days and accounted for the last 4 conditions. (2 sounds x 2 locations = 4). However, the control condition was added after monkey Y recordings were completed, therefore monkey Y recording sessions only consist of 28 conditions.

Conditions were randomly interleaved: any given trial had a 45% chance to be a single sound trial, a 45% chance to be a dual sound trial, or a 10% chance to be a control trial (or for monkey Y a 50% chance to be either a single- or dual-sound trial). Within the control and dual sound trial types the likelihood of any condition was equal. The likelihood of any single condition was equal except on 50% of single sound trials the B

sounds were presented at the opposite location of the A sounds to counterbalance the location of the single sounds.

2.4 Neural recordings

Neural activity was acquired through a Multichannel Acquisition Processor (MAP system, Plexon Inc). Single and multiunit activity signals were sent to an external speaker for auditory detection of neural activity. Once single units were isolated, using the box method (Plexon SortClient software), spike times were sent to Beethoven and saved for offline analysis.

Prior to each recording session the chamber was cleaned and a grid and turret (Crist Instrument, 1mm spaced grid) was inserted into the chamber. Prior to recordings, grid positions likely to intercept the IC were localized using structural MRI (Groh et al., 2001). A single tungsten micro-electrode (FHC, ~1-5 MΩ) was backloaded into a stainless-steel guide tube. The guide tube was then manually inserted approximately 1 cm into the brain. The electrode was then advanced with a microdrive (NAN Instruments) at 100-200 μm/s until the electrode tip was just outside of the guide tube. This was indicated by a mark on the electrode and confirmed by a change in the hash typical upon electrode entrance into the brain. The depth on the microdive was then zeroed and the electrode was advanced at roughly 8 μm/s while monkeys sat passively listening to sounds of variable frequencies presented every few seconds. Once the expected depth and/or sound related activity (determined by ear) was encountered in

the hash, electrode movement was stopped, and the electrode was allowed to settle for approximately 1 hour. The settling depth typically occurred 22 mm after exiting the guide tube. The electrode was then advanced in 1 μm increments until a cell was well isolated. If no cells were well isolated the electrode was retracted at roughly 8 μm/s.

A total of 105 neurons were recorded (monkey Y right IC, N=45; monkey N right IC, N=59; monkey N left IC, N=1).

3. Multiple sounds degrade the frequency representation in monkey inferior colliculus

3.1 Introduction

Natural environments contain numerous stimuli. A fundamental question is how the brain simultaneously encodes multiple items, particularly when more than one item falls within a given neuron's receptive field. Two possible (and potentially complementary) solutions to this multiplicity conundrum have been proposed: that receptive fields effectively shrink or shift to limit the number of stimuli within them, and/or that fluctuating activity patterns allow individual neurons to contribute to encoding only one particular stimulus at any given time. There is evidence in favor of the fluctuating activity hypothesis (Caruso et al., 2018), but whether the underlying tuning curves may also change to facilitate encoding multiple stimuli remains unknown. Here, we evaluated whether and how the presence of more than one sound alters the frequency tuning properties of neurons in the inferior colliculus (IC).

The monkey IC is an ideal structure to probe how neurons code concurrent stimuli for two reasons. First, nearly every auditory signal converges to the IC prior to ascending to the thalamus (Aitkin & Phillips, 1984), so information must be preserved at this stage in order to be available to higher areas (for review see *The Inferior Colliculus*, 2005). Causal studies have indicated a role in frequency perception (Pages et al., 2016) and sound localization (Jenkins & Masterton, 1982; Kelly & Kavanagh, 1994). Second, the representations of sound frequency and location in the monkey IC are coarse (Bulkin &

Groh, 2011; Groh et al., 2003; A. Ryan & Miller, 1978; Versnel et al., 2009; Zwiers et al., 2004). A given pure tone can activate anywhere from 20-80% of frequency modulated sites (Bulkin & Groh, 2011), implying that broadband natural stimuli are likely to be capable of activating an even higher percentage. Neurons tend not to have circumscribed spatial receptive fields but respond monotonically as a function of increasing sound eccentricity over a broad range of space (a hemifield or more, Groh et al., 2003; Zwiers et al., 2004).

In short, the almost complete convergence of ascending auditory signals in the IC, together with the broad frequency and spatial tuning in this structure, suggests that two sounds can potentially recruit similar populations of IC neurons. Nevertheless, humans (Best, Schaik, & Carlile, 2004; Perrott, 1984a, 1984b; Zhong & Yost, 2017) and monkeys (Caruso et al., 2018) are capable of localizing two concurrent sounds, implying that the IC population must overcome these coding constraints.

One possibility is that IC neurons sharpen or shift their frequency tuning curves to limit the degree to which any two stimuli fall within their receptive field. Tuning curves are not immutable (Weinberger, 1995). For example, spatial tuning in cat auditory cortex is modulated by behavioral state: units show broad selectivity under anesthesia (Imig, Irons, & Samson, 1990; Middlebrooks, Clock, Xu, & Green, 1994; Middlebrooks & Pettigrew, 1981) but become more selective in the awake cat (Mickey & Middlebrooks, 2003), particularly when performing a spatial auditory task (Lee & Middlebrooks, 2011).

Surround suppression in the visual system provides another example (Chettih & Harvey, 2019; Hartline & Ratliff, 1957) as do response profile modulations associated with experience (Freedman, Riesenhuber, Poggio, & Miller, 2005) or auditory attention (Connell, Barczak, Schroeder, & Lakatos, 2014).

Could the monkey IC also sharpen tuning curves to facilitate encoding of multiple stimuli? Previous work in the rabbit IC showed concurrent sound presentation broadens spatial tuning (Day, Koka, & Delgutte, 2012). However, it remains to be seen if frequency tuning is affected by the presence of a concurrent sound. The present study investigates how frequency tuning curves are modulated by the presence of an additional sound and if these changes are beneficial to decoding stimulus information from the population.

Single neuron activity was recorded from the IC while monkeys made saccades to one or two simultaneously presented sounds. Monkeys performed at least 86% correct on dual sound trials for all frequency separations. Frequency tuning curves were often altered when a second sound was presented but did not appear to shrink or shift. Rather, most neurons actually broadened their tuning curves. A maximum-likelihood decoder used to decode the stimulus on a held-out trial performed slightly worse when the held-out trial was a dual-sound trial. Thus, the frequency tuning curve modulations due to the presence of an additional sound seem to limit the available stimulus information encoded by the population of IC neurons. We conclude that changes to the

frequency tuning properties of IC neurons in the presence of multiple sounds are unlikely to solve the multiplicity problem.

3.2 Methods

3.2.1 Behavior

All behavioral analysis was done on attempted trials, defined as those in which the monkey successfully completed the fixation epoch and left the fixation window after the go cue. Proportion correct was defined as the # of Correct Trials / # of Attempted Trials

3.2.2 Neural activity

All analysis of neural activity was done on correct trials and during the first 500 ms after sound onset, i.e. during a period of time when monkeys were fixating (to reduce effects related to eye movements known to occur in the macaque IC (Bulkin & Groh, 2012a, 2012b; Groh et al., 2001; Porter, Metzger, & Groh, 2006, 2007). Neurons were classified as responsive if their spike count 500 ms after sound onset across all tested frequencies was significantly different from their spike count during baseline, the 500 ms prior to sound onset (t-test, $p < 0.05$). Ninety-three of the 105 neurons were classified as responsive. Neurons were classified as frequency selective, or tuned, if their spike counts 500 ms after sound onset were significantly modulated across all 10 sound frequencies (one-way ANOVA, $p < 0.05$). Sixty-three of 105 neurons were classified as frequency selective.

3.2.3 Point image

The point image refers to the proportion of recorded units that are responsive to a given condition (Capuano & McIlwain, 1981). Neurons were classified as responsive if, for a given condition, their spike count 500 ms after sound onset was significantly different than their spike count 500 ms prior to sound onset (t-test, $p < 0.05$).

3.2.4 d' (d-prime)

To investigate how differently a cell responded for A vs B sounds we calculated a sensitivity index or d':

$$d' = \frac{\mu_A - \mu_B}{\sqrt{\frac{1}{2}(\sigma_A^2 + \sigma_B^2)}}$$

Where μ and σ^2 corresponds to the mean spike count and spike count variance for sound A or B. This metric quantifies the separability of the sound A and sound B spike count distributions. We took the magnitude of the d' value (|d'|) since the order in the numerator is arbitrary (the A or B sound could be the first term). A |d'| value of 0 indicates the sound A and sound B spike count distributions are similar and higher values of |d'| indicate the sound A and sound B spike count distributions are more distinguishable.

3.2.4 Equivalent rectangular receptive fields (ERRFs)

ERRFs serve as a measure of the width of the frequency receptive field and were calculated in a similar fashion to previous work (Lee & Middlebrooks, 2011). The area under the (non-baseline subtracted) tuning curve was measured with the trapz()

function in MATLAB, which integrates the area using the trapezoid method. This area was then divided by the peak firing rate of the tuning curve, creating a rectangle with a width dependent upon tuning curve breadth and frequency modulation of peak firing rate.

3.2.5 Maximum-likelihood decoding

The decoding algorithm used here was adapted from previous work (Wallisch, 2014) and is similar to other implementations of maximum-likelihood decoding (Day & Delgutte, 2013; Jazayeri & Movshon, 2006). The goal of the decoder was to infer the stimulus that was presented on a set of held-out trials (1 held-out trial of a given condition per cell). The predicted condition was the condition (1 out of 26 possible conditions) that had highest population-log-likelihood. This computation was repeated 1000 times per condition. Since there were not 1000 unique trials per condition for each cell the particular held-out trial on a given repeat was sampled from all possible trials with replacement.

To determine the population-log-likelihood, first the likelihood for each cell was computed given the held-out trials spike count. Second, the weights were determined by finding the average spike count for each condition without the held-out data. Then likelihood for all 26 conditions given the spike count was computed:

$$\ln\left(Condition\ Likelihood\right) = \ \ln\left(P(spike\ count \mid frequency)\right)$$

This results in a vector with 26 elements for each cell. Each element corresponds to the log-likelihood of that condition for that held-out spike count. Essentially the log-likelihood is given by the difference between the held-out spike count and the mean firing rate across all conditions. These log-likelihood vectors are then summed across all cells to generate a population-log-likelihood:

$$\ln(Population\ Condition\ Likelihood) = \sum_{i=1}^{N} \ln\left(P(spike\ count_i \mid stimulus)\right)$$

Where spike count$_i$ corresponds to the *ith* neuron spike count, and N is the number of neurons in the population. Therefore, for each repeat of the decoding computation, the result is a 26-element vector where each element corresponds to the population-log-likelihood for each condition. The predicted condition for that repeat is the element with the largest population-log-likelihood.

Importantly, for any repeat the inferred condition could have been one of the 26 conditions (8 A sounds x 3 sound pairs = 24 conditions + 2 B sounds = 26 conditions). We used all responsive units (N = 93) for the decoding analysis.

3.3 Results

Monkeys could successfully report the locations of all sound combinations used in this study. Across all attempted dual sound trials, monkeys performed roughly 90% correct (Fig. 1B), compared to the 80% correct on single sound trials. This high degree of performance on the dual sound trials was only modestly affected by the frequency

separation between the two sounds: monkeys performed at about 87% correct for the smallest frequency separation of 0.25 octaves (Fig. 1C), rising slightly to 91% correct for the largest frequency separation of 2 octaves.

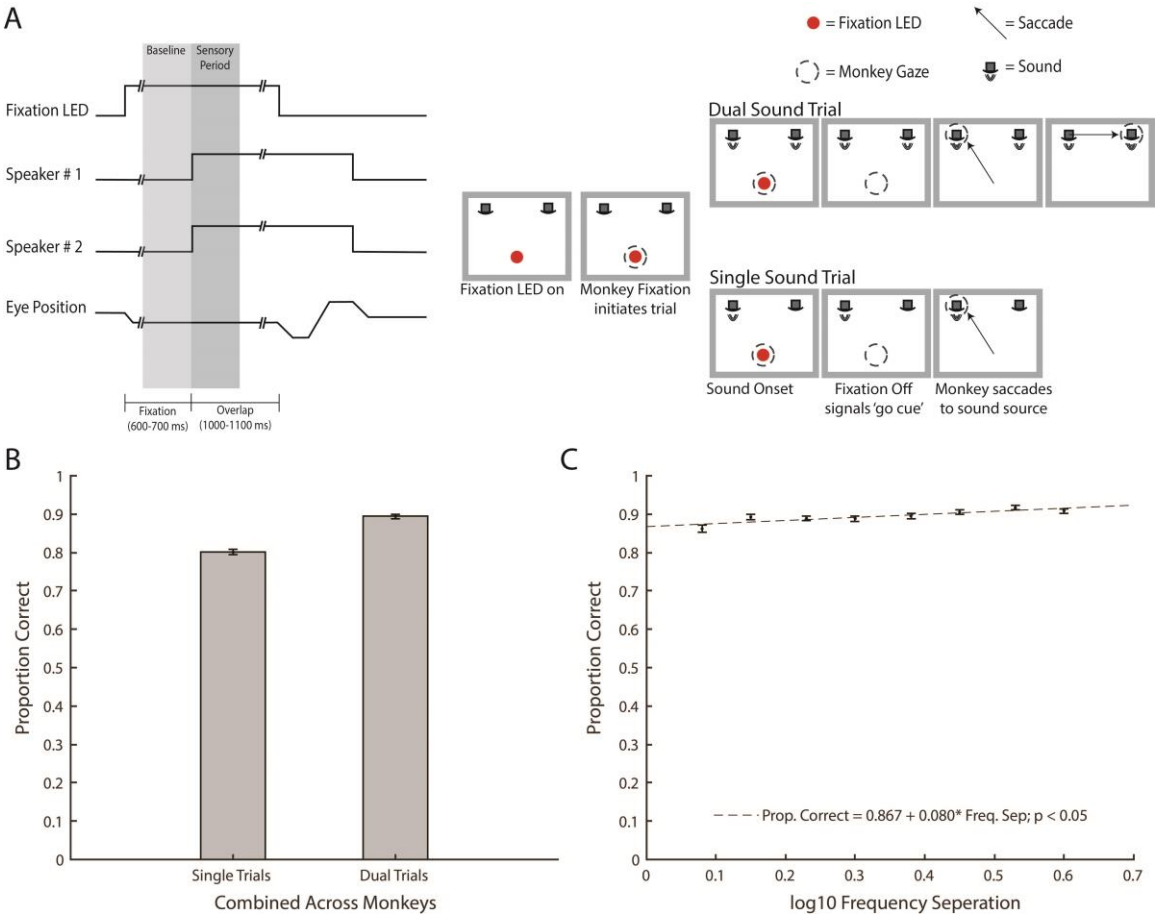


Figure 1: Task schematic and behavioral performance. (A) Monkeys were trained on a sound localization task. Left panel: A time course schematic of the task. Speaker 2 onset and the second saccade and fixation would only occur if it was a dual-sound trial. Right Panel: In the sound localization task the monkey would initiate a trial by fixating an LED. After variable interval fixation (600-700 ms), either 1, or 2 simultaneous, sound(s) were presented from one or two locations. After a variable interval of sound/fixation overlap (1000-1100 ms) the fixation light is turned off, which cued the monkey to report to the origin(s) of the sound(s) via saccade(s). On single sound trials monkeys made a single saccade. On dual-sound trials monkeys made a sequence of saccades to the origin of one sound and then the other. (B) The combined

proportion of correct single-sound trials (mean = 0.8011 ± 0.0069) and dual-sound trials (mean = 0.8941 ± 0.0060) across both monkey Y and monkey N for 105 behavioral sessions collected during cell recordings. (C) Change in performance across both monkeys as a function of the frequency separation between the two sounds. Performance increases as distance between sound frequencies increase (significantly fit by a line with a slope > 0; slope 0.867, $p < 0.05$). Error bars represented standard error of the mean (B & C).

This good performance does not appear to be the result of recruiting largely unique sets of neurons to encode each of the two stimuli. Of the N=93 sound-responsive neurons we recorded (identified by t-test, see Methods), 45-70% responded significantly to any given stimulus presented individually (Fig. 2, red curve). The large proportion of responsive units suggests that sounds presented in this experiment recruit overlapping populations of neurons. The low frequency bias (negative slope of the red curve in Fig. 2), is similar to previous macaque IC findings (Bulkin & Groh, 2011).

When a second sound was added, more units responded (Fig. 2, blue and black curves vs. the red curve). This increase in the number of responsive units at the population level should not occur if individual units were sharpening their tuning.

We next compared the sensitivity of individual IC neurons to the specific pairs of sounds used. This analysis focused on the roughly 67% (N = 63) of responsive neurons whose spike counts were significantly modulated by frequency (one-way ANOVA, $p < 0.05$). We quantified the separability of the spike count distributions of the two components of the dual sound conditions (e.g. 500 (A-sound) vs. 420 (B-sound) when either were presented in isolation) using a d' analysis (see Methods).

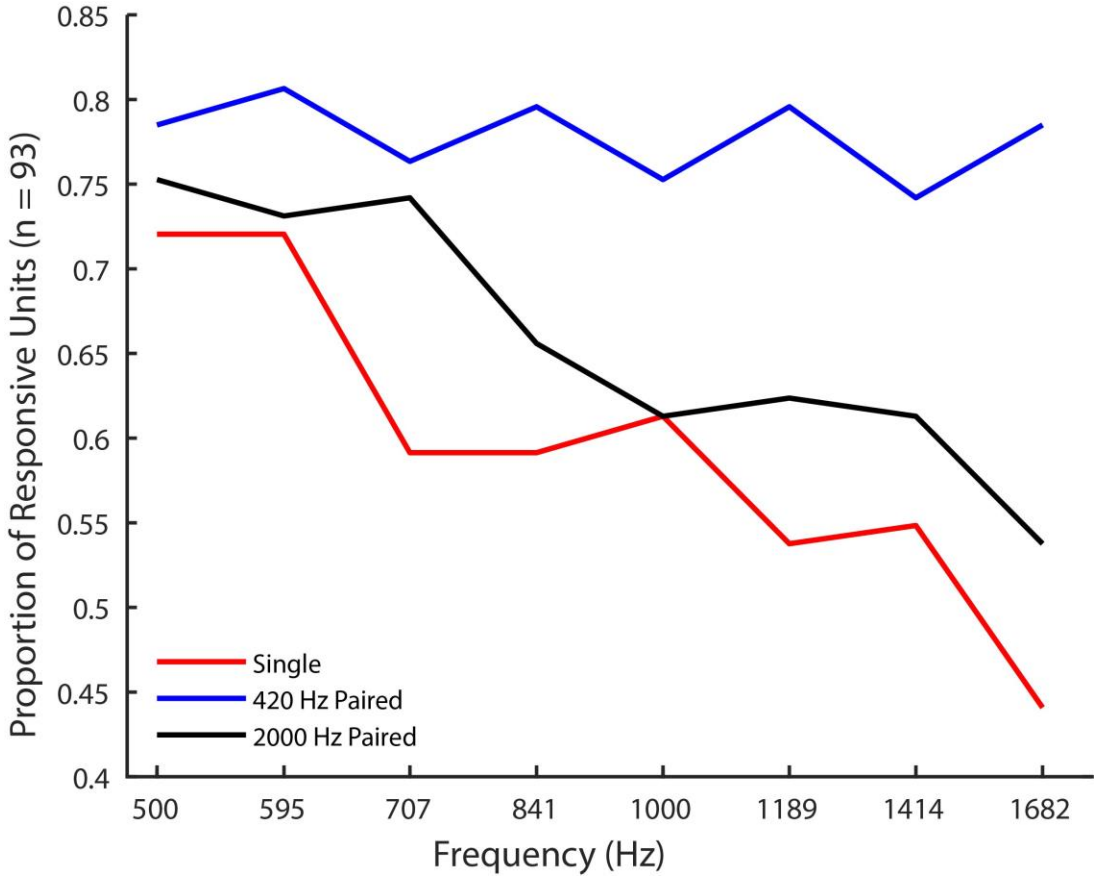


Figure 2: A point image plot, or the proportion of the responsive population that responds to a given frequency. Each point corresponds to the proportion of the responsive population that is significantly responsive compared to baseline for the given condition. A unit was counted as responsive if its firing rate in 0-500 ms after sound onset was significantly different than its firing rate from -500-0 ms before sound onset (t-test, $p<0.05$). Red corresponds to sounds presented in isolation, while blue and black correspond to dual-sound trials paired with 420 Hz or 2000 Hz, respectively.

Figure 3A shows the results for the comparison with a 420 Hz B sound. Units were sorted based on the sum of their |d'| values across conditions, with the units showing the highest aggregate |d'| on top. For the smallest frequency separation (left-

most column), few neurons exhibit large d' values: only about a third exhibit a d' value

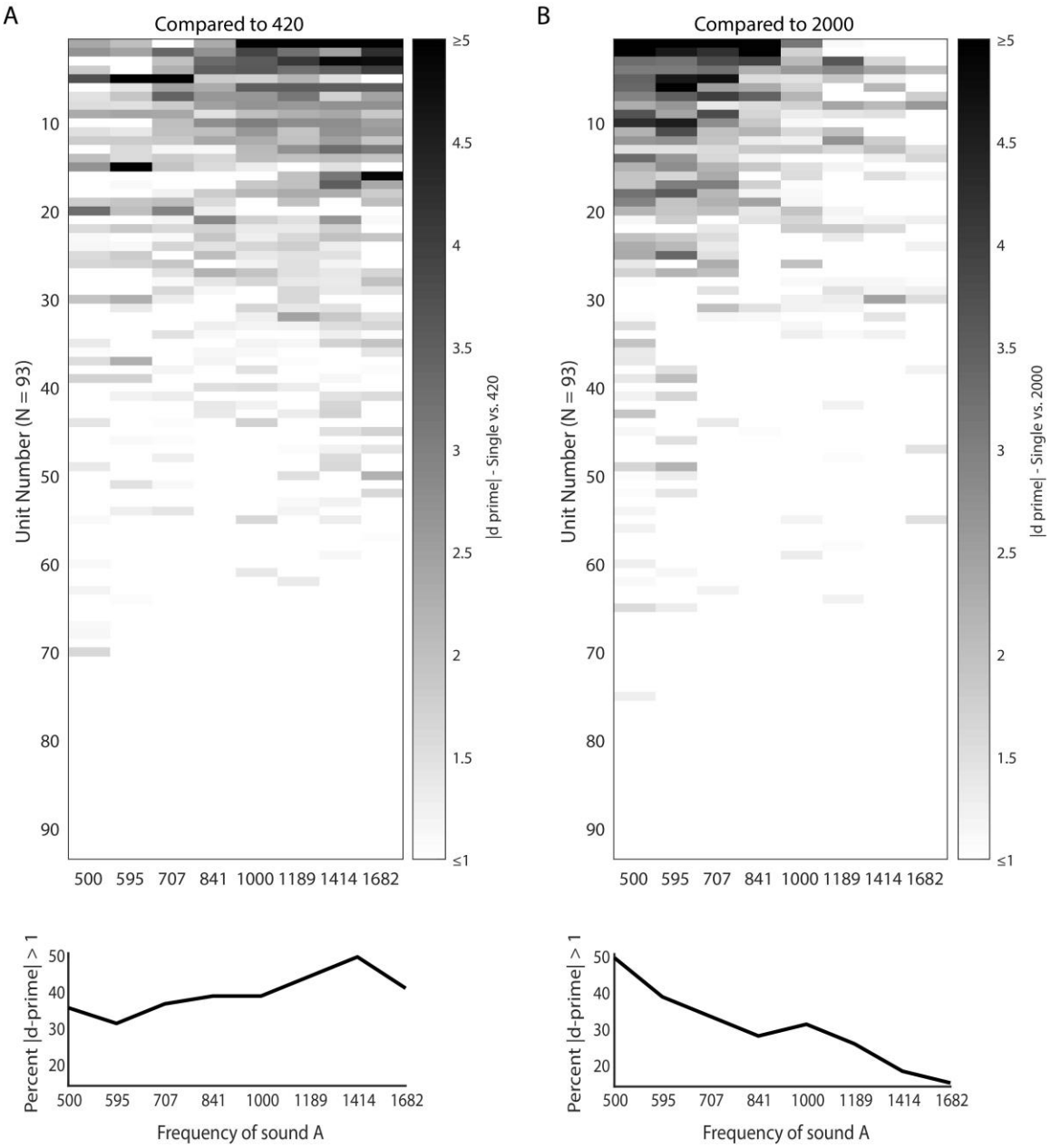


Figure 3: A sensitivity index, or d', between the A and B sounds when they are presented in isolation. (A) Top Panel: The magnitude of d' between each of the 8 A sounds and 420 Hz. A higher |d'| corresponds to a larger difference between the spike count distributions of the two sounds. Each row corresponds to a unit. Each column corresponds to a particular sound frequency. Units are sorted by their total |d'| across

all 8 A sounds compared to 420 Hz. Bottom Panel: The percent of responsive neurons (N = 93) that have a |d'| value > 1 for each frequency compared to 420 Hz. (B) The same as panel (A) but compared to 2000 Hz. Note these units are sorted by their total |d'| across all 8 A sounds compared to 2000 Hz. Therefore, unit n in A does not necessarily correspond to unit n in panel B.

greater than 1 (bottom panel). Figure 3B shows the complementary pattern for the 2000 Hz B sounds. Note that the unit order in two panels are different; both are sorted so the unit with the highest summed |d'| within a panel is at the top. For both panels, a large minority of cells barely differentiated between any of the A and B sounds, as evidenced by the white bands at the bottom of the figure (roughly unit 70 to 93). Overall, this analysis suggests that only a fraction of the recorded populations (roughly the top 20 units in both panels) possess sufficient sensitivity to underlie the improvement in performance due to frequency separation on dual sound trials (Fig. 1C).

We then evaluated whether neural response patterns change to facilitate the encoding of simultaneously presented sounds. Figure 4 compares the single and dual sound tuning profiles of six example cells. On single sound trials (red traces), neurons could prefer low frequencies (Fig. 4A, E, F), intermediate frequencies (Fig. 4B, D), or high frequencies (Fig. 4C). Concurrent sounds could change tuning in multiple ways. The neurons in figure 4A-C all become less responsive to sounds overall: the responses in the dual sound cases were generally lower than the responses to the "better" of the two single sounds (e.g. black curves fall between the red and blue curves). Some neurons appear to shift their preferred frequency (Fig. 4D), and some neurons appear

unaffected by the presence of an additional sound (Fig. 4E, F). Interestingly, some neurons displayed multi-peaked tuning curves that were robust across single and dual sound conditions (Fig. 4F). Qualitatively, it seems that the majority of neurons become less responsive in the presence of simultaneous sounds, but whether this suppression leads to receptive fields that sharpen or broaden requires quantification of tuning curves across the population.

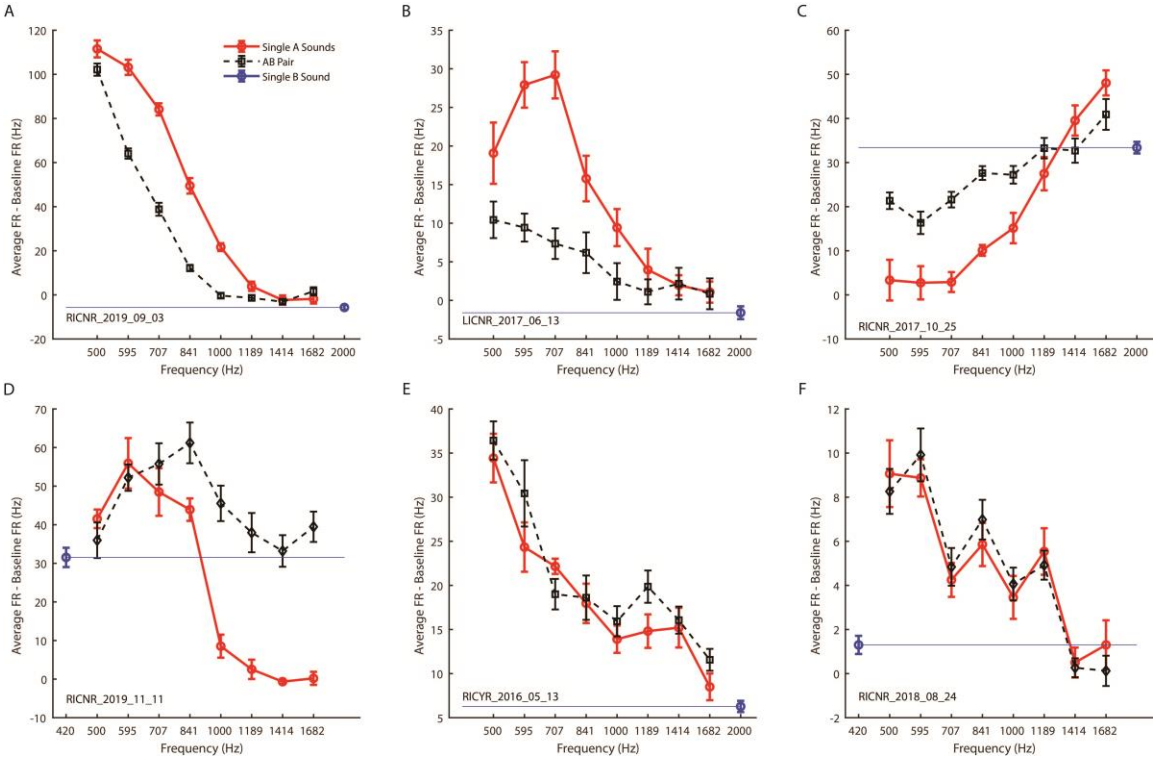


Figure 4: Example modulations of frequency tuning curves due to the presence of an additional sound in 6 different cells. Example modulations of frequency tuning curves due to the presence of an additional sound in 6 different cells. (A) The tuning curve across conditions for a cell that displays sharpening in the dual-sound condition. Each point corresponds to the baseline (-200 ms bin w.r.t sound onset) subtracted mean firing rate (500 ms bin w.r.t sound onset). Red points correspond to sounds presented in isolation. The blue point corresponds to the paired sound when presented in isolation. The blue line is a reference line, so the firing rate of the paired sound is visible across the

panel. The black points correspond to the dual-sound condition. (B and C) The same as panel (A) but for two different cells that show a decrease in response and selectivity. (D) The same as panel (A) but for a cell that shifts its preferred frequency away from the paired sound. (E & F) The same as panel (A) but for two cells that seem to be unaffected by the presence of an additional sound. Error bars represent standard error of the mean(A-F).

To do this, we selected units that were frequency selective to the A sounds when presented alone (N=57, one-way ANOVA, see Methods). The number of frequency-selective units dropped to N = 47 (82% of N=57) when the A sounds were paired with either 420 or 2000 Hz. To evaluate effect size, we compared the F-statistic (a metric of how much frequency modulates spike count) in the dual vs. single conditions (Fig. 5A, C). If the neurons were more selective in the dual sound conditions, the F statistic should be higher for these conditions and the data would tend to lie above the line of slope one, but they do not. Figure 5 B and D quantify this further by subtracting the single sound F statistic from the dual sound F statistic for a given neuron. If the neurons were more selective on dual sound than on single sound trials, these distributions would be positively skewed, but they are not. In fact, the negative skew, i.e. worsening frequency selectivity, is significant for the 420-Hz pairing (Fig. 5B, mean = -7.989 ±1.219 S.E.M, $p < 0.05$). Overall, then, neurons' frequency selectivity was either unaffected by a second sound or was reduced by it.

The preceding analysis considers how the means and variances in firing depend on sound frequency but doesn't directly investigate whether tuning curves change in systematic ways, such as changes in the width or peak (best frequency).

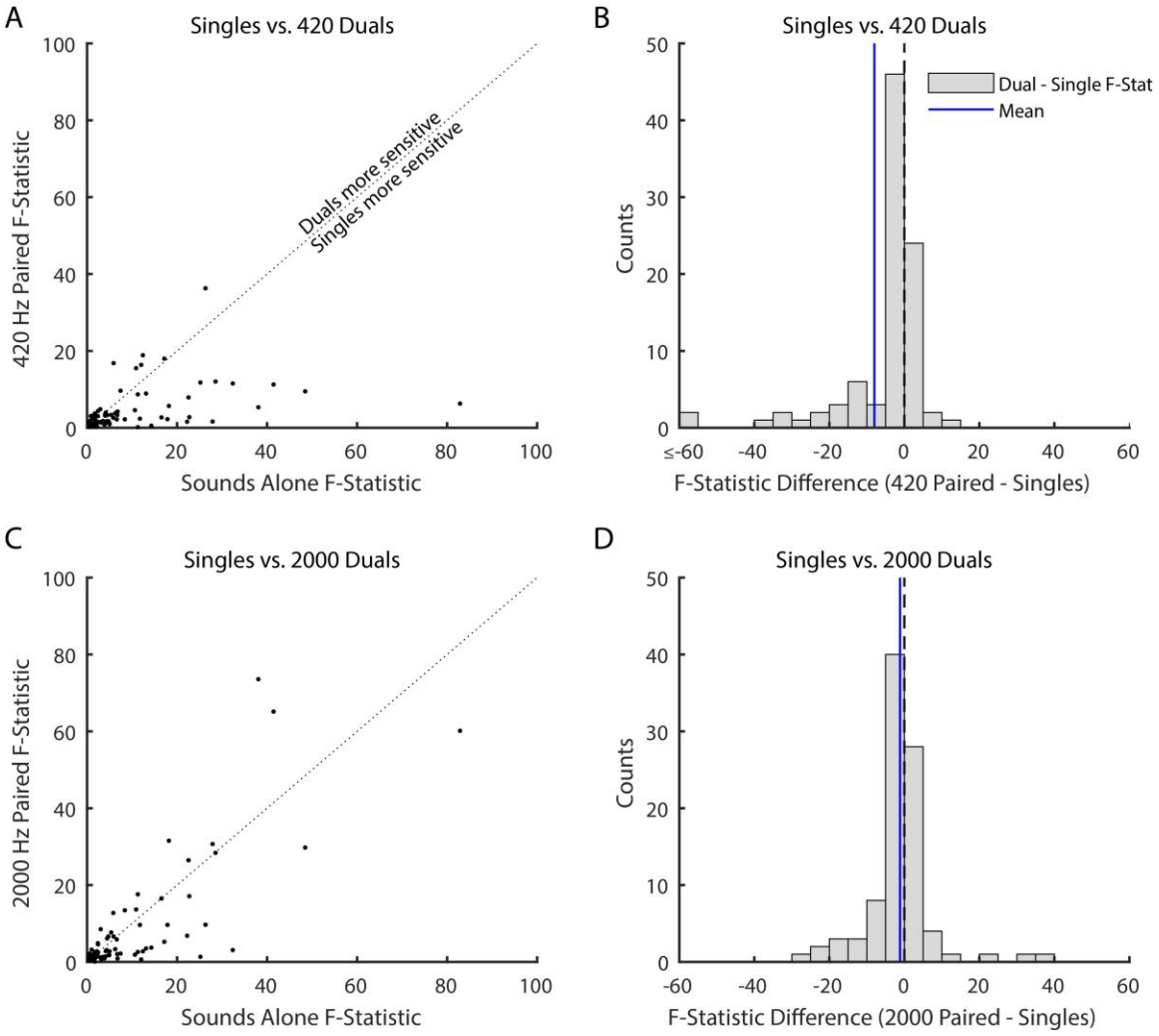


Figure 5: Frequency modulation of firing rate is affected by the presence of an additional sound. (A) The F-statistic measured from a one-way ANOVA investigating firing rate across different frequencies (all 8 A sounds). A high F-statistics indicates that there are large differences in firing rate across the different frequencies. The single-sound F-statistics are plotted against the 420 Hz paired F-statistics. Each point corresponds to one unit, out of 93 sound responsive units. Points above the unity line indicate that frequency modulates firing rate more in the 420 Hz paired condition. Points on the unity line indicate there is no change in the modulation of firing rate across the two conditions. Points below the unity line indicate that frequency modulates firing more in the single-sound condition. (B) The distribution of change in F-statistic per cell ($F\text{-Statistic}_{Single} - F\text{-statistic}_{Dual}$). A value of 0 indicates no change across the two conditions, a value greater than 0 indicates the 420 Hz paired condition was more sensitive to frequency and a value less than 0 indicates the single-sound condition was

more sensitive to frequency. The single-sound condition was significantly more sensitive to frequency (mean = -7.99 ± 1.22 S.E.M, $p < 0.05$). The blue vertical line indicates the mean of the distribution. The black vertical line is at 0. (C) The same as (A) but for compared to the 2000 Hz paired condition. (D) The same as (B) but compared to the 200 Hz paired condition. There was no difference in the modulation of firing rate across the two conditions (mean = -1.14 ± 0.9 S.E.M, $p = 0.21$). Note that in (A & C) an outlier data point was removed for visualization purposes. In panel (A) its position is x =364.5, y = 31.2. In panel (C) its position is x = 364.5, y = 396.3

To investigate possible changes in tuning curve width, we implemented the equivalent rectangular receptive field analysis (ERRF) from Lee and Middlebrooks (2011). Briefly, the ERRF divides the area under the tuning curve by the peak firing rate, creating a metric that incorporates the number of frequencies to which the unit responds based in part on the level of activity evoked by those stimuli. For example, many of the cells display a suppression in the dual sound condition (Fig. 4 A-C), yet these suppressions seem to be of different kinds. Generally, suppressions that compress the whole firing rate range (Fig. 4B & C) tend to increase ERRF width while suppressions that affect some conditions but leave the peak untouched (Fig. 4A) tend to decrease the ERRF width.

To compare tuning curve width between single and dual sound conditions, the distribution of each were plotted against each other (Fig. 6B & D). If the ERRF widths were similar across single and dual sound conditions the distribution would fall upon the unity line (black line in Fig. 6B, D). If the ERRF widths were narrower in the dual sound condition the distribution would be shifted below unity and if the ERRF widths were narrower in the single sound condition the distribution would be shifted above

unity. It is quite apparent that ERRF widths are broader in the dual sound conditions, for both high- and low-frequency paired sounds.

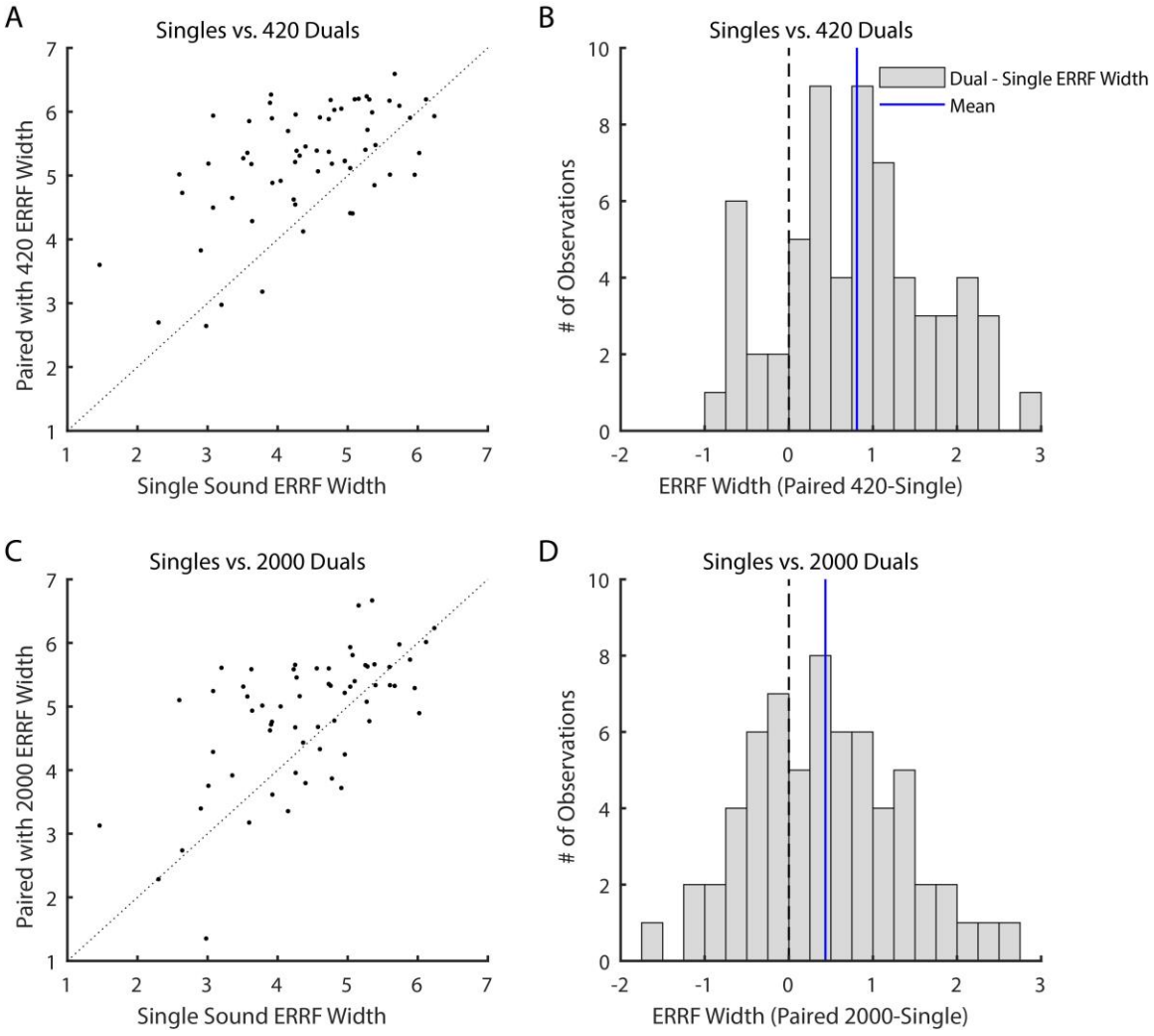


Figure 6: ERRF width is larger in dual-sound conditions. (A) The width of the ERRF for each unit in the single sound condition plotted against the width of the ERRF for each unit plotted in the 420 Hz paired condition. Each point represents a unit, out of 63 frequency modulated cells. The dashed line indicates unity, or ERRFs of equal width in the single sound and 420 Hz paired condition. Points above unity indicate ERRFs of greater width in the dual-sound condition. Points below unity indicate ERRFs of greater width in the single-sound condition. (B) The distribution of change in ERRF width per cell ($ERRF_{420} - ERRF_{Single}$). A value of 0 indicates the ERRF width did not change across conditions. Values greater than 0 indicate the ERRF is broader, while values less than 0

indicated the ERRF is narrower. The ERRF width in the 420 Hz dual sound condition is significantly broader (mean = 0.75±0.11 S.E.M, $p < 0.05$). The blue vertical line indicates the mean of the distribution. The black vertical line is at 0. (C) The same as (A) except for the 2000 Hz paired condition. (D) The same as (C) except for the 2000 Hz paired condition. The mean in the 2000 Hz paired condition is significantly broader (mean = 0.26 ± 0.11 S.E.M, $p < 0.05$).

To evaluate whether these width changes are significant, the difference between each neuron's dual sound ERRF width and single sound ERRF width was calculated. The 420 Hz paired ERRF widths were significantly broader than the single sound ERRF widths (Fig. 6C); the difference distribution was significantly greater than 0 (mean = 0.75±0.11 S.E.M, $p < 0.05$). Although the effect was smaller, the 2000 Hz paired ERRF widths were also significantly broader than the single sound EERF widths (Fig. 6D); the difference distribution was significantly greater than 0 (mean = 0.26 ± 0.11 S.E.M, $p < 0.05$). In summary, across the population the majority of neurons actually broaden their tuning curve in response to presentation of two simultaneous sounds.

Could some neurons shift their tuning curves, exhibiting a different best frequency in response to dual sounds (e.g. Fig. 4D)? To investigate if neurons shifted their tuning, we selected neurons with firing rates significantly modulated by frequency and with single sound best frequencies that fell within the middle six A sounds (595-1414 Hz), leaving out the edge cases for which a shift outside the tested range could not in principle be observed (N=17). We then compared their single sound best frequency to their best frequency in the dual sound conditions. Seventeen neurons met these criteria. Figure 7 shows how best frequency changes in the 420 Hz paired condition (Fig. 7A) or

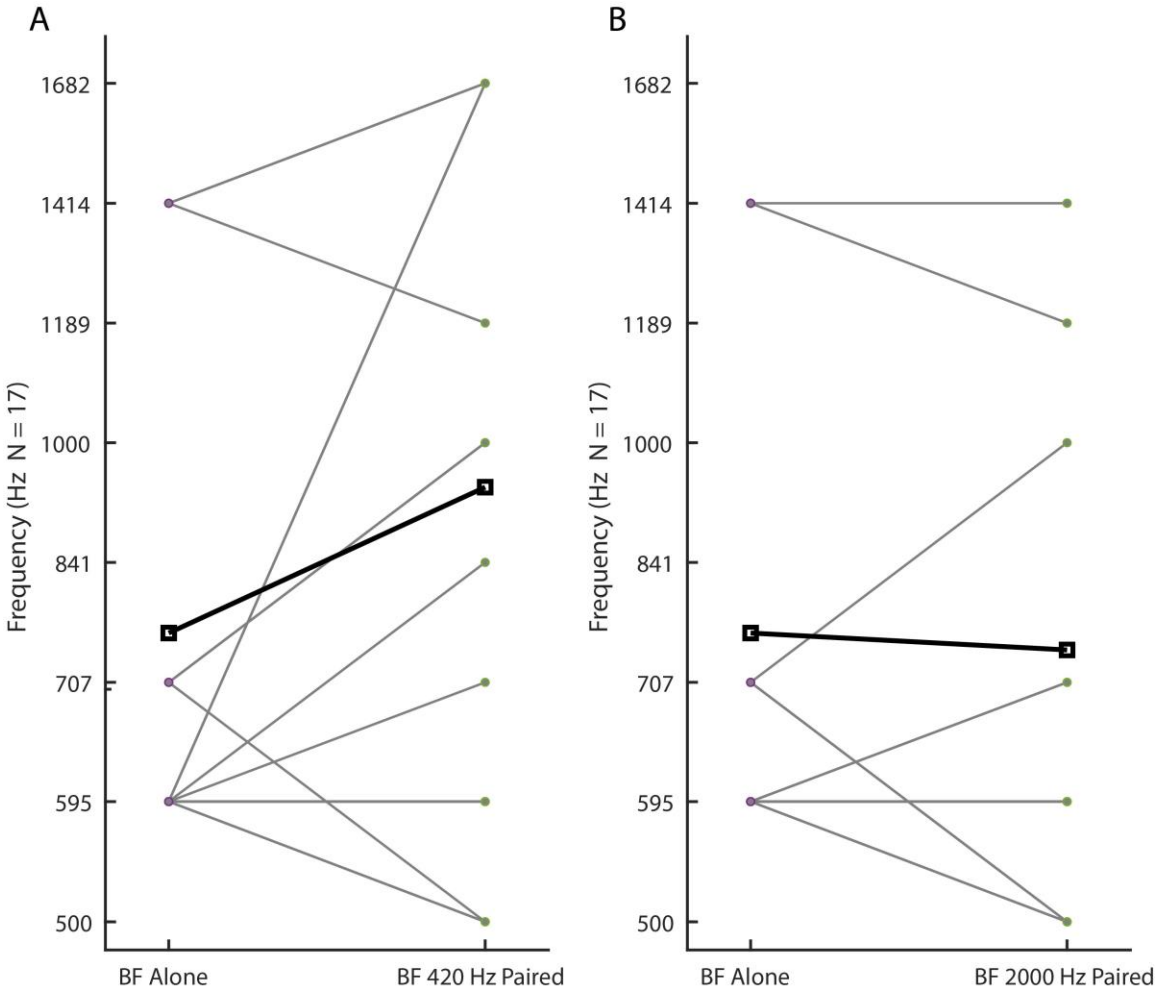


Figure 7: The presence of an additional sound does not cause a large shift in best frequency. (A) The single-sound best frequency plotted against the 420 Hz paired best frequency in the 17 cells that have best frequencies between 595-1414 Hz (grey lines). The bold line indicates the mean best frequency across the population. There is no significant shift in best frequency across the two conditions (single mean = 759.3 ± 76.48 S.E.M vs. dual mean = 937.6 ± 113.6 S.E.M, p = 0.072). (B) The same as (A) but for the 2000 Hz paired condition. There is no significant shift in best frequency across the two conditions (single mean = 759.3 ± 76.48 S.E.M vs. dual mean = 740.9 ± 75.67 S.E.M, p = 0.56).

the 2000 Hz paired condition (Fig. 7B). Although, there is a slight shift of the mean best frequency away from the paired sound, it is not significant for either the 420 Hz dual

sound (single mean = 759.3 ± 76.48 S.E.M vs. dual mean = 937.6 ± 113.6 S.E.M, p = 0.072) or the 2000 Hz dual condition (single mean = 759.3 ± 76.48 S.E.M vs. dual mean = 740.9 ± 75.67 S.E.M, p = 0.56). These findings suggest that shifting of receptive fields is not a ubiquitous computation within the population and is unlikely to substantially contribute to the encoding of simultaneous sounds.

In short, the modulations of frequency tuning on dual sound trials do not appear to be optimized for the preservation of information about either sound. To determine the impact of multiple sounds on the efficacy of coding stimulus related information, we evaluated how this code might be read out.

We first used a population-maximum-likelihood decoder to infer the condition from the observed spike count on a set of held-out trials (see Methods). Spike counts were assumed to be Poisson-distributed, and neurons were assumed to be independent of each other (a reasonable assumption given they were recorded at different times). For each held-out trial, the log-likelihood for each condition was summed across the population and the condition with the maximum-population-log-likelihood was predicted to be the held-out trials condition. This computation was repeated 1000 times per condition.

Overall, the decoder inferred every condition above chance (chance = 1/26). Figure 8A, C, and E shows the proportion of correctly predicted conditions across the 3 paired sound trial-types, sounds presented in isolation, sounds paired with 420 Hz, or

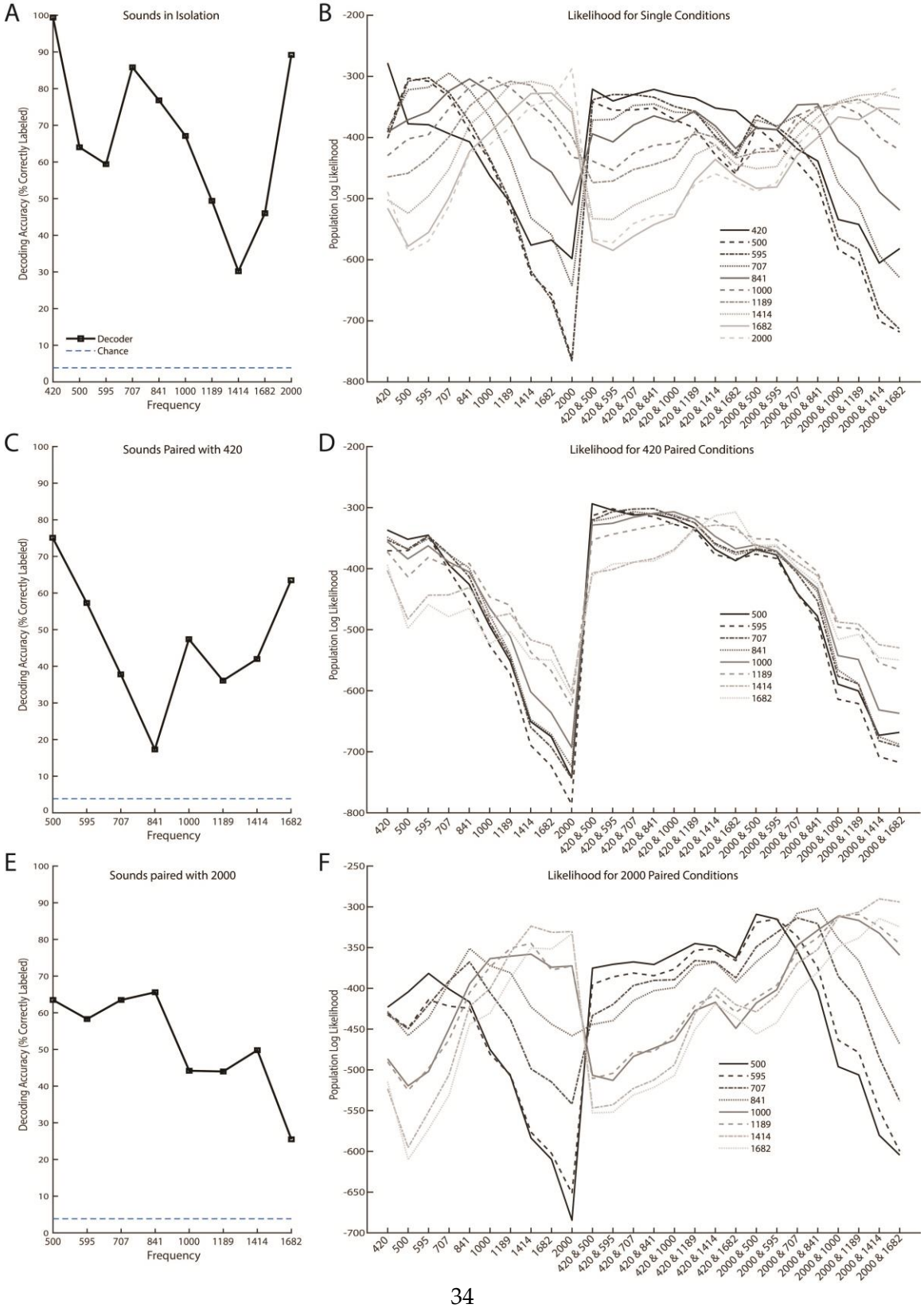
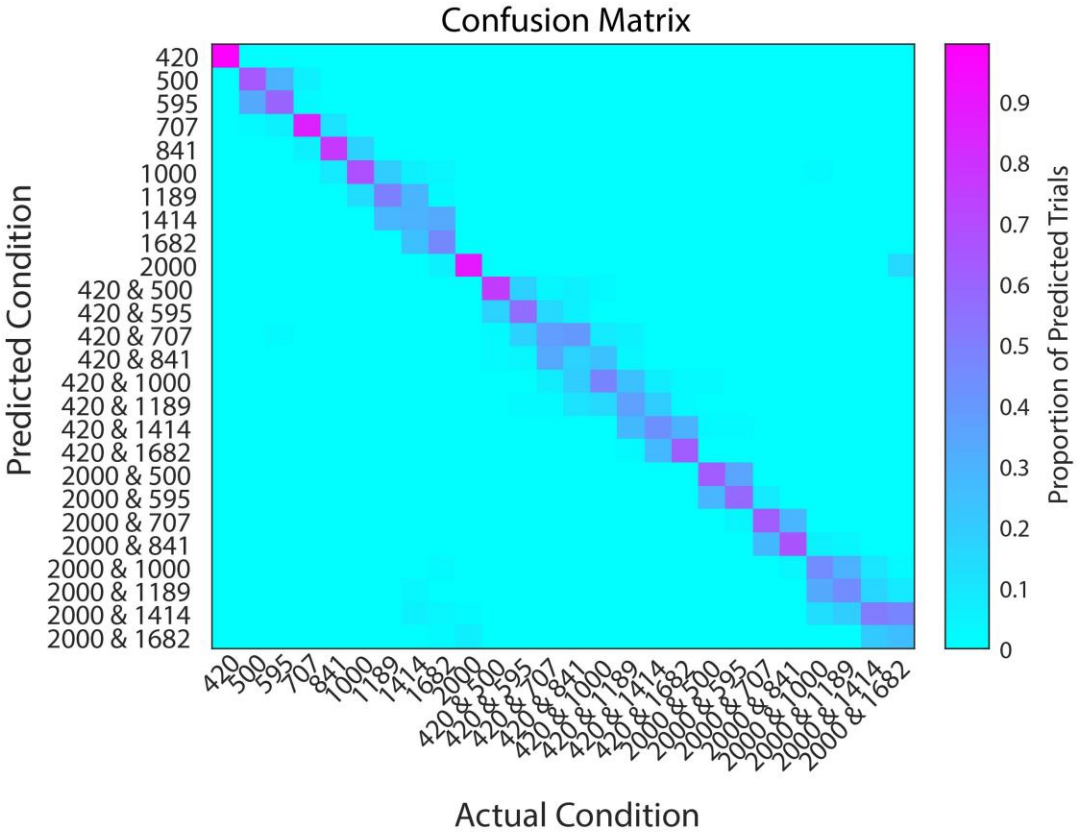

A
Sounds in Isolation
Decoding Accuracy (% Correctly Labeled)
Frequency
Decoder
Chance
B
Likelihood for Single Conditions
Population Log Likelihood
C
Sounds Paired with 420
D
Likelihood for 420 Paired Conditions
E
Sounds paired with 2000
F
Likelihood for 2000 Paired Conditions

Figure 8: Accuracy of population maximum-log-likelihood decoding and mean log-likelihood functions across conditions. (A) The accuracy of the population (N = 93) maximum-log-likelihood decoder across each single sound condition. Each point corresponds to the percent of correctly labeled repeats (out of 1000) for each condition. Chance level prediction is 1/26 or ~ 3.85% (B) The mean population log-likelihood function for each condition across all potential predicted conditions. Each condition had the chance of being labeled as any one of the 26 single or dual conditions. (C) The same as panel (A) but for 420 Hz paired conditions. (D) The same as panel (B) but for 420 Hz paired conditions. (E) The same as panel (A) but for 2000 Hz paired conditions. (F) The same as panel (B) but for 2000 Hz paired conditions.

sound paired with 2000 Hz, respectively. Qualitatively, the mean across all sounds presented in isolation (~66.7% correct) was higher than the mean across all 420 Hz paired conditions (~47.1% correct) or the mean across all 2000 Hz paired conditions(~51.8% correct). The likelihood functions show that the confusion in decoding tends to come from proximal conditions (Fig. 8B, D, F). This can be further seen in the confusion matrix (Fig. 9A) where the off diagonals are slightly darker in the dual compared to single sound conditions. This confusion arises due to the increase in likelihood when a low frequency sound is present (420 Hz paired and when 2000 Hz is paired with a low frequency: Fig. 9B), which may stem from the populations low frequency bias (Fig. 2). Ultimately, the changes in the frequency response functions seem to degrade the amount of information concerning each sound, resulting in worse decoding on dual sound trials.

Until now all the analyses presented probed how monkey IC encoding of sound changes in the presence of an additional sound. With the implementation of the maximum-likelihood decoder, we can now ask if single and dual sounds are encoded

A


Confusion Matrix
Predicted Condition
Actual Condition
Proportion of Predicted Trials
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
420
500
595
707
841
1000
1189
1414
1682
2000
420 & 500
420 & 595
420 & 707
420 & 841
420 & 1000
420 & 1189
420 & 1414
420 & 1682
2000 & 500
2000 & 595
2000 & 707
2000 & 841
2000 & 1000
2000 & 1189
2000 & 1414
2000 & 1682

B

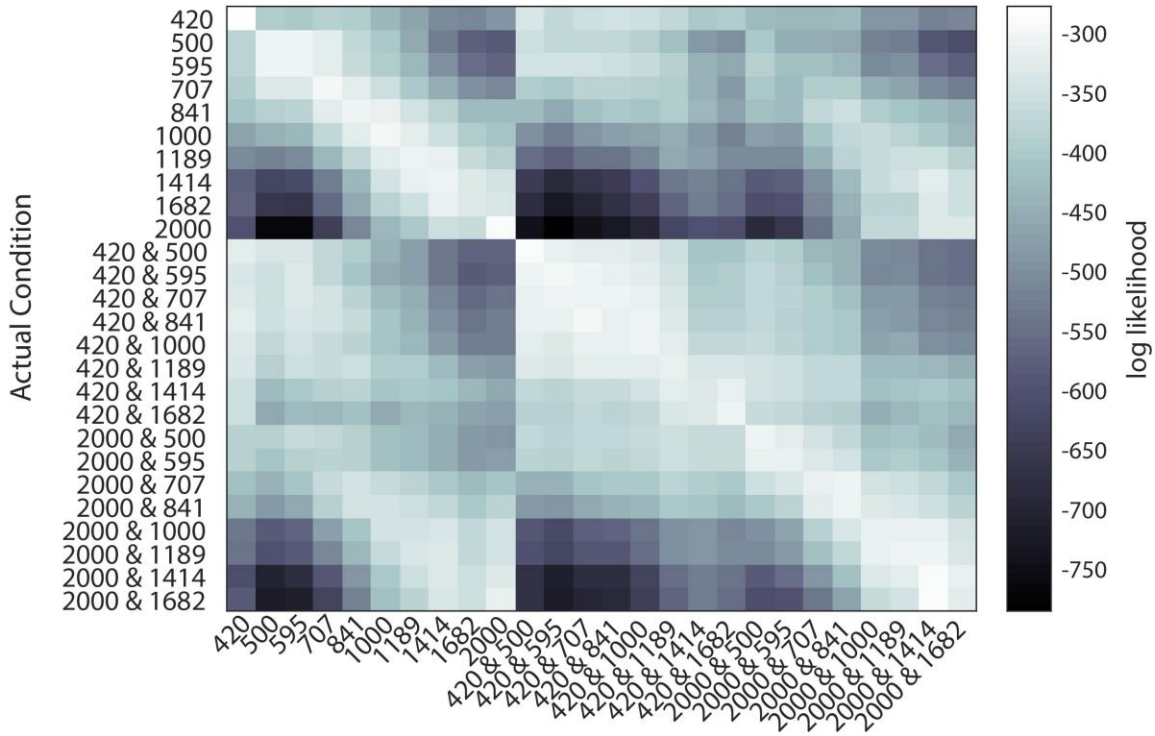

Actual Condition
log likelihood
-300
-350
-400
-450
-500
-550
-600
-650
-700
-750
420
500
595
707
841
1000
1189
1414
1682
2000
420 & 500
420 & 595
420 & 707
420 & 841
420 & 1000
420 & 1189
420 & 1414
420 & 1682
2000 & 500
2000 & 595
2000 & 707
2000 & 841
2000 & 1000
2000 & 1189
2000 & 1414
2000 & 1682

Figure 9: Confusion and mean-likelihood matrices for the decoder. (A) The confusion matrix for the decoder. Each column corresponds to the actual condition and each row corresponds to the predicted condition. The color of each square is the proportion of predicted trials (out of 1000) that fell into that bin. The diagonal corresponds to correct predictions. (B) A matrix of mean log-likelihood across conditions for the decoder. Each row corresponds to the likelihood function for that condition while each column corresponds to the mean log-likelihood for that condition across all actual conditions (e.g. leftmost column corresponds to the mean log-likelihood for 420 Hz across all conditions).

with similar neural codes. To investigate if the monkey IC uses a similar code for the single and dual sound conditions, we attempted to decode the dual sound condition using the decoding weights learned from the single sound trials. Each held-out dual sound trial could, therefore, be labeled as one of the 8 A sounds. Performance was markedly worse in the 420 Hz condition when decoded with the single sound weights (accuracy = 16.5%; chance 12.5%) compared to the dual sound weights (accuracy ~ 47.1%; chance ~ 3.846%). This is evident in the confusion matrix (Fig. 10A) which shows 595 Hz is the most predicted stimulus for seven of the eight conditions. The performance in the 2000 Hz condition when decoded with the single sound weights (accuracy = 40.65%; chance = 12.5%) is also worse compared to decoding with the dual sound weights (accuracy ~ 51.8% correct; chance ~3.846%). Unlike the 420 Hz paired condition, the 2000 Hz paired conditions seem to confuse proximal conditions (Fig. 10B). These results show that the IC would need to use different coding strategies in the single and dual sound conditions, implying an IC readout may need to implement a different set of weights depending on the number of sounds present.

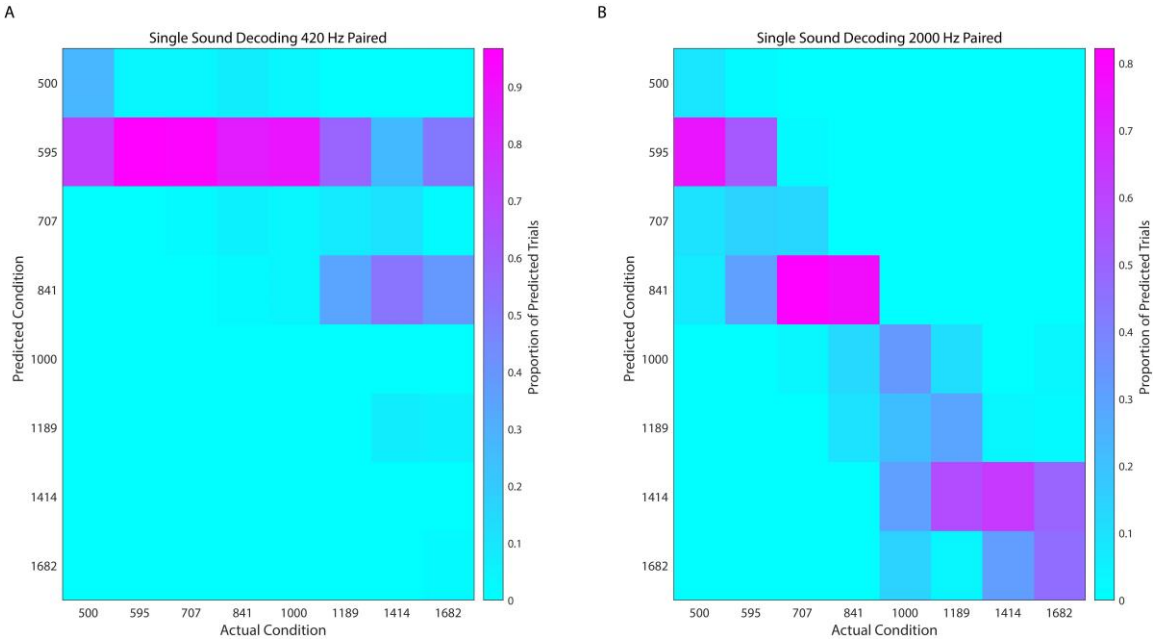


Figure 10: Decoding the paired condition with the single-sound weights impairs performance. (A) The confusion matrix for decoding 420 Hz paired conditions using the weights learned from single-sound trials. Each column corresponds to the actual condition and each row corresponds to the predictable condition. The color of each square is the proportion of predicted trials (out of 1000) that fell into that bin. The diagonal corresponds to correct predictions. (B) The same as (A) but for decoding the 2000 Hz paired conditions.

3.4 Discussion

How the brain simultaneously encodes multiple items, particularly when more than one item falls within a given neuron's receptive field, is an often-overlooked problem in neuroscience. It is assumed the population representation overcomes ambiguity in single unit encoding of stimuli, but the problem remains if the items recruit largely overlapping populations. To address this question, we first verified that monkeys could report the occurrence of two sounds whose center frequencies are

separated by as little as 0.25 octaves (and whose bandwidth overlaps). We then investigated how frequency response functions of are affected by the presence of that second, concurrently presented sound. We found that the sounds we presented recruited largely overlapping populations (Fig. 2 & 3), and that on dual sound trials the majority of these cells' firing rates were less modulated by sound frequency (Fig. 5) and had their tuning curves broadened (Fig. 6). There was little evidence of systematic shifts in best frequency (Fig. 7).

Overall, these changes in frequency response functions reduce the information available to the maximum-likelihood decoder, decreasing the decoding accuracy in the dual sound conditions (Fig. 8 & 9). The poor decoding performance when using the single sound weights to decode dual sound conditions (Fig. 10) suggests that a different read-out is needed depending on the number of sounds – a startling possibility given that the brain cannot have prior knowledge of how many sounds are in the environment except by virtue of what it detects via sensory input.

A limitation of this study is that it is not known how many neurons might be needed to support perception of any given stimulus. It is possible that the small proportion of neurons that responded to only one of a pair of sounds were responsible for the monkeys' perceptual abilities.

Our results seem to be in-line with prior work showing degradation of information when there are multiple stimuli presented (Day & Delgutte, 2013; Day et al.,

2012; Henry & Kohn, 2020). If this reduction in information scales with number of stimuli it could potentially explain the finding that the number of identifiable sounds saturates at three (Zhong & Yost, 2017).

Generally, these results refute the notion that the changes in frequency response functions could be used to overcome the multiplicity problem and support recent evidence of alternations in firing rates between those evoked by individual stimuli (Caruso et al., 2018) may be a more fruitful research direction.

In the present study, we made the simplifying assumption that the time-and-trial-pooled average response of a neuron to a combination of stimuli is reflective of the information that the neuron encodes. If the underlying activity on dual sound trials actually fluctuates between the A-like and B-like response patterns, the time-and-trial-pooled average will be a poor measure of the information present in neural signals. The fact that average responses to AB sounds were generally between the average responses to A and B sounds presented alone supports the possibility that fluctuations may underlie at least a portion of the results observed here. Future work will test this possibility.

In principle, changes in frequency tuning as investigated here and fluctuating activity patterns as investigated previously (Caruso et al., 2018) both have the potential to limit the degree to which a given neuron is faced with the task of encoding more than one stimulus at a time. It remains possible that information-preserving changes in

frequency tuning might be more evident when tested with a wider range of sound frequencies, allowing a fuller exploration of frequency tuning curves. However, the granularity of our testing, 0.25 octave spacing, was clearly coarser than monkeys' perceptual abilities, and the overall range, 2 octaves, should have been adequate to demonstrate an effect if changes in frequency tuning were a major contributor to these perceptual abilities.

4. Consistency of the monkey inferior colliculus response to concurrent sounds

4.1 Introduction

In the natural world, animals often behave in noisy environments where sounds overlap in frequency, space, and time. Survival in these noisy environments indicates that animals can segregate and encode behaviorally relevant sounds, even if sounds co-occur. In fact, recent work shows both humans (Best et al., 2004; Perrott, 1984a, 1984b; Zhong & Yost, 2017) and monkeys (Caruso et al., 2018) can localize concurrent sounds. However, the neural implementation underpinning the representation of simultaneous sounds remains ambiguous.

A structure involved in the representation of simultaneous sounds is the inferior colliculus (IC). The IC, a midbrain nucleus, affects the perception of both sound frequency (Lim et al., 2009; Pages et al., 2016) and sound location (Jenkins & Masterton, 1982; Kelly & Kavanagh, 1994) and is an obligatory site of convergence for the myriad of subcortical auditory pathways (Aitkin & Phillips, 1984). This convergence and known sensitivity to numerous features of sounds (for review see: *The Inferior Colliculus*, 2005) led to the notion that the IC may serve as a basis for the recognition of auditory objects (Griffiths & Warren, 2004; Grothe et al., 2010). Thus, if a mammal perceives simultaneous sounds then information concerning each sound must be encoded in IC neural activity.

Indeed, ideal observer decoders trained on IC population activity decode the number of sound sources (in rabbits:Day & Delgutte, 2013) as well as the number and frequency of sounds (in monkeys: Willett & Groh, 2020; chapter 3). However, the addition of a concurrent sound decreased decoder performance. This decrease in performance was likely due to the response suppression seen in individual IC neurons on multi-sound trials; generally, neurons respond somewhere between the two component sound responses on dual sound trials (Day et al., 2012; chapter 3). These results resemble findings in vision where multiple stimuli or objects normalize responses across visual cortex including V1 (Busse, Wade, & Carandini, 2009; Henry & Kohn, 2020) and IT Cortex (Bao & Tsao, 2018; Sheinberg & Logothetis, 2001; Zoccolan, Cox, & DiCarlo, 2005) Although recognized as a canonical computation (Carandini & Heeger, 2012), normalization is a poorly suited computation for the preservation of information about multiple stimuli since information about the individual stimuli is lost (Orhan & Ma, 2015). Conversely, neural responses invariant to the presence of other stimuli, a winner-take-all response, are well suited for the preservation of information about each stimulus, assuming some portion of the population is sensitive to each stimulus. Recently, Caruso et al. (2018) discovered macaque IC neurons that implement a time-varying winner-take-all strategy, alternating their firing rates, across or within trials, between the rates corresponding to each of two concurrently presented sounds. These response alternations could allow for the representation of multiple stimuli across

time within the same neural population (Caruso et al., 2018). Importantly, the pooling of neural activity across time and trials, a ubiquitous step in the processing of neural data, occludes the recognition of response alternations. In fact, averaging across time and trials would make such alternations appear as response normalization; potentially misleading the field in our understanding of how brains encode concurrent stimuli.

Although recent work decoded the number and frequency of sounds from the activity of the macaque IC, the decoder could not account for the high accuracy of the monkey's behavior on dual sound trials (chapter 3). A potential explanation for this discrepancy, is that these neurons alternate their responses across time or trials to encode concurrent sounds (Caruso et al., 2018) and a decoder with weights trained on the time and trial averaged responses of neurons fails to account for those response fluctuations. However, it remains to be seen if the fluctuating activity will be observed under different stimulus conditions. For example, Caruso et al. (2018) used stimuli proximal in center frequency (approximately ≤ 1 octave), which may push the IC population into a regime biased towards response alternations. Conversely, sounds disparate in center frequency may activate more segregated populations, exploiting the tonotopic structure of the macaque IC (Bulkin & Groh, 2011; A. Ryan & Miller, 1978; Versnel et al., 2009), and relaxing the need for response alternations potentially biasing the IC population towards some other response pattern. This raises an important question in the representation of simultaneous sounds: Are IC neurons consistent in

how they encode information concerning concurrent sounds or does their encoding of concurrent sounds depend on the particular sounds in the auditory scene?

To address this question, we recorded the activity of isolated neurons from the IC while monkeys localized two simultaneously presented sounds. We found the large diversity in IC dual sound responses was well captured by modeling each neuron as a linear combination of the two single sound responses, showing many, but not all, neurons normalize their dual sound response. However, upon closer inspection of the trial-by-trial structure in dual sound responses we found that much of the response normalization could be explained by alternations between the two component sound responses across time. These response alternations occurred in roughly 63% of IC neurons and for the full, 0.25-2 octave, range of frequency separations tested. Therefore, response alternations are a general mechanism IC neurons exploit to encode concurrent stimuli. These response fluctuations in conjunction with winner-take-all responses could allow for downstream decoders to remain stable independent of the number sounds in the auditory scene, alleviating the need for more complex read outs of the neural activity encoding multiple objects.

4.2 Methods

4.2.1 Neural activity

Unless otherwise noted, all analysis of neural activity was done during the first 500 ms after sound onset, i.e. during a period of time when monkeys were fixating (to

reduce effects related to eye movements known to occur in the macaque IC (Bulkin & Groh, 2012a, 2012b; Groh et al., 2001; Porter et al., 2006, 2007). Neurons were classified as responsive if their spike count 100 or 500 ms after sound onset, across all tested frequencies, was significantly different from their spike count during baseline, the 100 or 500 ms prior to sound onset (t-test, $p < 0.05$). These bins were chosen to include neurons with complex transient and/or sustained response profiles, typical of macaque IC neurons (Ryan & Miller, 1977). Ninety-three out of 105 neurons were responsive in the 500 ms bin and 95 out of 105 were responsive in the 100 ms bin. Together, 101 out of 105 (~96.2%) of the recorded neurons were responsive to sound.

4.2.2 Combination metric

To determine how the AB response related to the A and B response we calculated a combination metric:

$$Combination\ Metric = \frac{AB - \min(A, B)}{range(A, B)}$$

Where AB, A, and B correspond to the average response of the neuron to each of those conditions. This metric captures many different possible AB response combinations; values less than 0 indicate an inhibition below min(A,B) , values of 0 or 1 correspond to winner-take-all for the non-preferred or the preferred sound respectively, values between 0 and 1 correspond to a weighted average, and values greater than 1 signify response enhancement or summation. Importantly, for trials where A = B the

denominator would become 0. Therefore, we only included trials for which A ≠ B as defined by the separation factor described below (see Methods: Whole Trial).

4.2.3 Linear combination

We modeled the time and trial averaged mean AB response as a linear combination, a weighted sum, defined as:

$$FR_{AB} = \alpha FR_A + \beta FR_B$$

Where FRAB, FRA, and FRB are 16 element vectors where each element corresponds to the baseline subtracted average firing rate for a given AB, A, or B condition, respectively. The α and β weights were fit via the fitlm() function in MATLAB which also calculated the reported residuals and adjusted R2 values. Once the weights were found we predicted the FRAB from the right side of the weighted sum equation.

4.2.4 Whole trial

The statistical analysis of whole trial fluctuations was previously described (Caruso et al., 2018) and vetted (Mohl, Caruso, Tokdar, & Groh, 2020). Briefly, we classified each AB condition (8 A sounds x 2 B sounds = 16 AB conditions per neuron) based on the spike count from each single sound A and B trials and their corresponding AB dual sound trials. The spike count was measured from 0-500 ms after sound onset. Two assumptions were made concerning the spike count distributions.

The first was that the single sound spike count distributions followed Poisson statistics. To test the Poisson assumption on single sound trials A and B of a given AB

condition, we used an approximate chi-square goodness of fit test with Monte Carlo p-value calculation (Caruso et al 2018). The second was that the A and B spike count distributions were significantly separable, which was done by calculating the intrinsic Bayes factor of a model that the A and B rates were equal versus a model that the A and B rates were unequal (Caruso et al., 2018).

For AB conditions that passed these two assumptions, we modeled the single sound A and B spike counts as Poisson distributions with unknown rates λA, denoted Poi(λA), and λB, denoted Poi(λB). Four models were considered for the distribution of AB spike counts:

1. Mixture: a mixture distribution $\alpha \cdot$ Poi(λA) + (1-α) $\cdot$ Poi(λB) with α an unknown mixing weight
2. Intermediate: a single Poi(λAB) where some λAB is between λA and λB
3. Outside: a single Poi(λAB) where λAB is either larger or smaller than both λA and λB
4. Single: a single Poi(λAB) where λAB exactly equals one λA and λB

The relative likelihood of these competing models was determined by computing each models posterior probability with equal prior weights (1/4) assigned to each model, and with default Jeffreys' prior (Berger, 2006) on model specific Poisson rate parameters, and a uniform prior on the mixing weight parameter α. The Jeffreys' prior was truncated to appropriate ranges for the intermediate and outside models. Posterior model

probabilities were calculated by computation of relevant intrinsic Bayes factors (Berger, 2006). It is important to note that sensitivity of detection was not equal across the models, because the Single model is a limiting case of the 3 other models. Yet, this analysis correctly labels simulated data for each model, though with less confidence for the single model (Mohl et al., 2020).

4.2.5 DAPP

The DAPP model was previously described in detail (Caruso et al., 2018; Glynn et al., 2019). The DAPP model describes the dynamics of spike trains on dual sound trials as an admixture of those occurring on single sound trials. The DAPP model initially breaks each trial into time bins. We performed this analysis on the first 1000 ms after sound onset (still during fixation) with 50 ms time bins. The AB response for a given bin on a given trial was modeled as:

$$\lambda^{AB}(t) = (1 - \alpha(t))\,\lambda^{A}(t) + \alpha(t)\,\lambda^{B}(t)$$

Where the λA(t) and λB(t) were the component sound responses at time t. Importantly, the weight parameter α defines the contribution of either component sound to the AB response a time t. So, the dynamics of α(t) capture the dynamics of the dual sound response. We modeled the α(t) function as a smooth Gaussian process with three parameters corresponding to the mean, variance, and correlation of α across bins and trials. Although the α at a given time for a given AB trial could be distinct, it was enforced that the dual trials share dynamic patterns assuming these three parameters

were drawn from a common, unknown, distribution P, which we called a dynamic pattern generated and was a characteristic of the AB condition for a given neuron. P was estimated via a Markov chain Monte Carlo computation (see Glynn et al., 2019). Nevertheless, the estimation of P allowed for the generation of simulated α(t) curves that the given neuron would be likely to produce on future dual sound trials.

We characterized the simulated α(t) curves by the range of α values across all bins of a trial and the time aggregated average value of α (Caruso et al., 2018). The waviness index of the AB condition was computed as the odds of seeing an α(t) function with a change of α of at least 50% between its peak and trough:

$$r_w = \frac{P(|\alpha| > 0.5)}{P(|\alpha| < 0.5)}$$

Where P denotes the sampling proportion of the simulated α draws. The triplet's extremeness index was computed as the odds of seeing an α(t) function with its long-term average being closer to the mid-way mark of 50% than the extremes:

$$r_c = \frac{P(\bar{\alpha} \in (0.25, 0.75)}{P(\bar{\alpha} \notin (0.25, 0.75)}$$

The two indices were then threshold to generate a 2-way classification of each AB condition (Caruso et. al 2018). On waviness, an AB condition was categorized as Wavy or Flat, or ambiguous if rw > 1.3, rw > 0.77, or, 0.77 ≤ rw ≤ 1.3, respectively. On

extremeness, the categories were Central, Extreme, or ambiguous according to whether rc > 3.24, rc <1.68, or, 1.68 ≤ rc ≤ 3.24, respectively (Caruso et al., 2018).

4.3 Results

Two monkeys were trained to report the location of a single sound or two simultaneous sounds via eye movements to the sound source(s) (Fig. 1A). The sounds consisted of bandpass filtered noise that varied across 10 central frequencies, 420-2000 Hz. On single sound trials 1 of the 10 sounds was presented at ±12° azimuth. On dual sound trials 1 of the 8 middle sounds (referred to as A sounds: 500-1682 Hz) was presented at ±12° azimuth while a sound of either 420 or 2000 Hz (referred to as B sounds) was presented at the opposite speaker location. Monkeys performed this task with an average accuracy of 84.7% across all conditions (Fig. 1B), confirming, as previously reported (Caruso et al., 2018), that monkeys can not only perceive but also localize two concurrent sounds.

If monkeys perceive two separate sounds, then neural activity in the IC must contain information concerning each sound. To investigate the representation of simultaneous sounds, we recorded the activity of 105 units from the IC of monkeys while they performed the localization task. Approximately 96.2% (N = 101) of IC neurons were responsive to presented sounds. Importantly, all analyses were done on correct trials, ensuring that the neural activity recorded on dual sound trials reflected the encoding of multiple sounds.

The response of IC neurons to the presentation of two concurrent sounds was complex and difficult to predict from knowledge of the single sound responses. Figure 11 shows the average responses of 5 example neurons to all A (red curve), B (blue points), and AB (black curve) sound combinations, the left column is conditions of B = 420 Hz and the right column is conditions of B = 2000 Hz. Under certain conditions, some neurons appear to normalize, or average, the responses of the two component sounds (Fig. 11A & C-left, 11B-right), while some neurons appear to be invariant to the presence of a concurrent sound (Fig. 11D – left, A & E – right). Other neurons respond more than the maximum of either component sound, known as a response enhancement (Fig. 11C & D – right), and others seem to transition between normalization, winner-take-all, and enhancement depending on the specific sound combination (Fig. 11B & E – left). Knowledge of the AB response function under a specific paired B sound (420 or 2000 Hz) did not give an intuition for how the response function will change when paired with the other B sound (Fig. 11 left vs. right column). These 5 example units make apparent the heterogeneity of the IC response to two concurrent sounds, but whether the population is biased towards a particular combination rule needs quantification across all neuron's dual sound responses.

To determine the distribution of dual sound responses we defined a combination metric for each AB condition across all neurons (see Methods). Combination metric values less than 0 correspond to an AB response that is less than the response to either

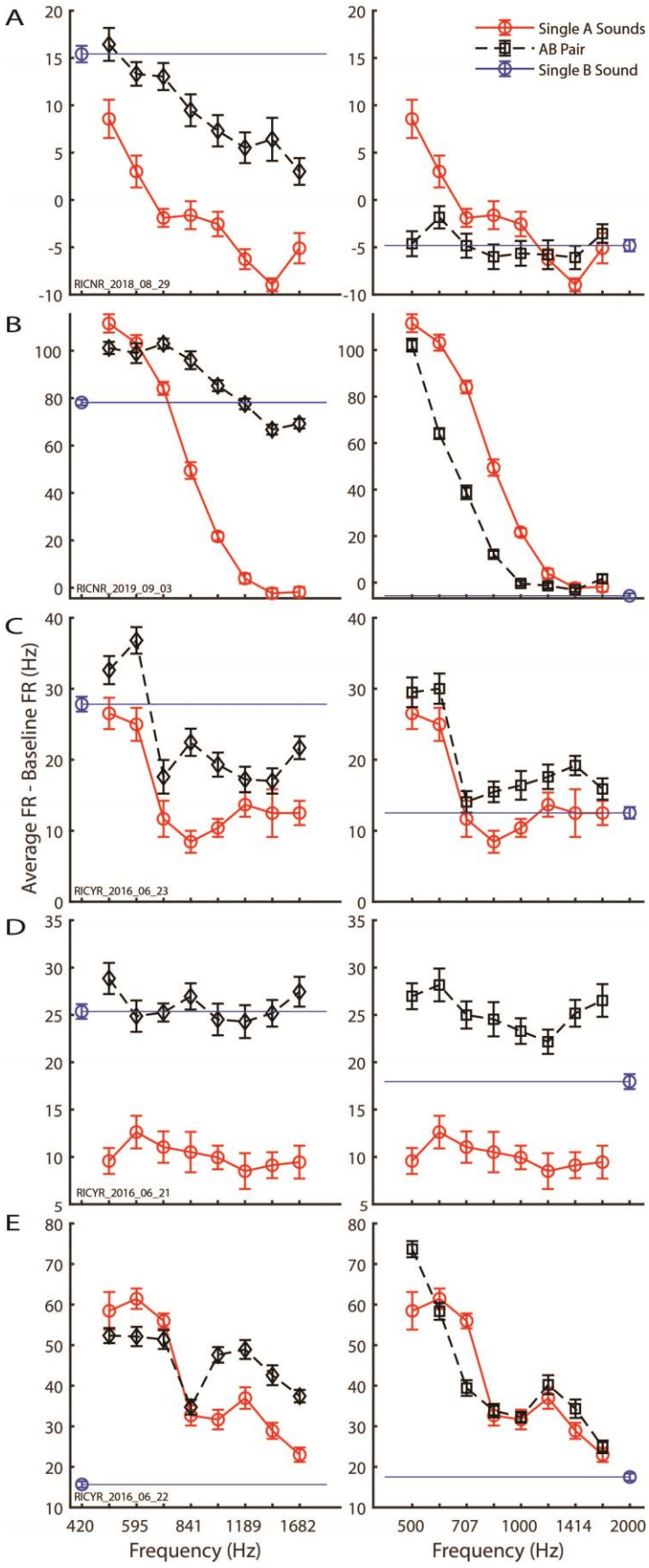

A
B
C
D
E
Single A Sounds
AB Pair
Single B Sound
RICNR_2018_08_29
RICNR_2019_09_03
RICYR_2016_06_23
RICYR_2016_06_21
RICYR_2016_06_22
Average FR - Baseline FR (Hz)
Frequency (Hz)
Frequency (Hz)
420
595
841
1189
1682
500
707
1000
1414
2000

Figure 11: Example units show the heterogeneity of dual sound responses in monkey IC. (A) The response function of an example neuron to the A sounds presented in isolation (red curve), the B sound presented in isolation (blue point), and the AB combination (black curve). The left column corresponds to B = 420 Hz and the right column corresponds to B = 2000 Hz. This neuron appears to normalize, or average, the two component sound responses when B = 420 Hz but appears to respond as if only B is present if B = 2000 Hz. (B) An example neuron that seems to enhance its response for some AB conditions if B = 420 Hz but to averages the two component sounds if B = 2000 Hz. (C). An example neuron that averages the two single sound responses if B = 420 Hz but enhances its response if B = 2000 Hz. (D). An example neuron that is not tuned to sound frequency and shows a winner-take-all response if B = 420 Hz but a summation of the component sound responses if B = 2000 Hz. (E) An example neuron that transition between a weighted average, enhancement and winner-take-all across both B conditions. Note the A and B sounds are played at different locations; A is played at ±12° and B is played at -1*A position. The responses shown here are the baseline subtracted (-200 to 0 ms), across trial average firing rate for the first 500 ms after sound onset (0 to 500 ms).

component sound (the A or B presented in isolation). Values of 0 correspond to an AB response that is equivalent to the response of the non-preferred component sound, while values of 1 correspond to an AB response that is equivalent to the response of the preferred component sound. Thereby, values of 0 or 1 signify an AB response invariant to the presence of the second sound, or a winner-take-all response captured by the non-preferred or preferred sound, respectively. Values greater than 1 correspond to an AB response that is greater than the response to either component sound, or a response enhancement. While values greater than 0 but less than 1 correspond to an AB response that is a between the two component sound responses, a response normalization or average. Note, winner-take-all and average responses are identical if the response to A and B are identical; therefore, we only included conditions in which a neuron's response to A and B differed (n = 655 conditions). The distribution of combination metric values

(Fig. 12) confirms the heterogeneity seen in the example units (Fig. 11). Approximately 71% (n = 470) of the AB conditions fall between 0.301 and 1.38 (the range within 1 standard deviation of the mean (mean = 0.835, standard deviation = 0.534). This indicates the bulk of the IC population implements a response normalization, though there are also many AB responses that show enhancement or winner-take-all responses. Although this metric provides an intuition for the type of computations the IC implements when faced with the need to encode two concurrent sounds, it is unclear if individual IC neurons are biased towards a particular dual sound response pattern (e.g. if one AB response is winner-take-all are the other AB responses also winner-take-all).

Describing a neuron's response to multiple stimuli as a linear combination of two consistently weighted component responses is one of the most successful, and interpretable, models in neuroscience (Carandini & Heeger, 2012). Indeed, averaging, winner-take-all and enhancement are all well by described by a linear combination, specifically a weighted sum. This approach can even capture the transition between these computations across conditions, though it requires additional parameters (Bao & Tsao, 2018).

Therefore, to determine if IC neurons time and trial averaged responses to concurrent sounds are consistent, we modeled the average AB response for each condition as a weighted sum of the average responses to the two component sounds (see Methods). This model adequately describes the AB responses of some neurons but

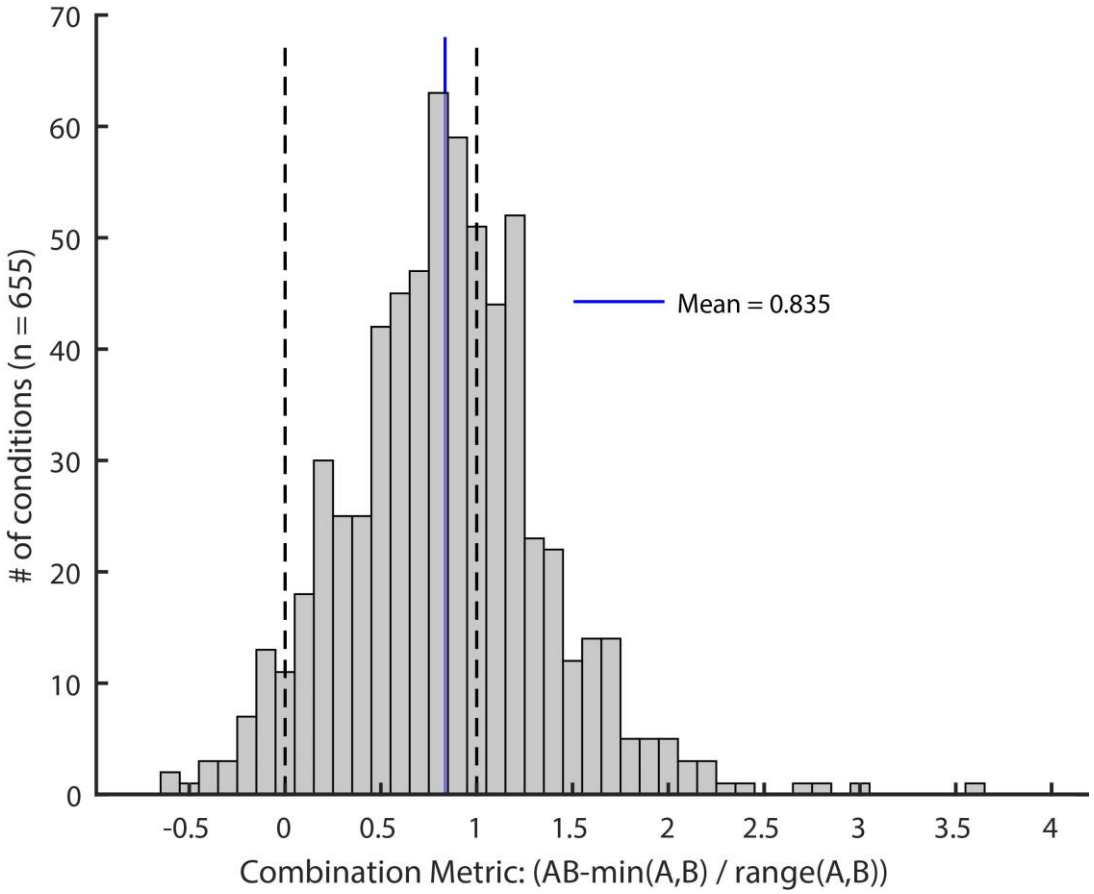


Figure 12: The bulk of the IC population falls within the weighted average, or normalization, regime. The combination metric for each AB condition with separable component sound responses (n = 655 conditions from 85 neurons) was included. The vertical dashed lines demarcate 0 and 1 or a winner-take-all to the non-preferred or preferred component sounds, respectively. Bins between 0 and 1 correspond to AB responses that average the two component sound responses. The blue line indicates the mean of the distribution (mean = 0.8316 ± 0.5337). Bin edges = [-0.75:0.1:3.95].

struggles under some conditions (Fig. 13). The predicted AB response functions, generally, fit the example neuron response functions well (Fig. 13A-E, black vs. grey curves). The distribution of weights (Fig. 13F) confirms that throughout the IC population the AB responses of many cells are well described as a consistent winner-take-all, average (normalization), or enhancement. Winner-take-all neurons are clustered

around the points of the shaded triangle region (weights of 0 and 1 or 1 and 0), while points falling within the shaded triangle indicate neurons that average the component sound responses or show a response enhancement. However, upon comparison of every predicted and actual AB response (Fig. 13G) the model struggles to predict the response for some AB combinations. For example, panel 13B shows the model predicts a firing rate ~ 50 Hz lower than the actual response to the 500+2000 condition and the model largely fails to capture the variability in the AB responses of the neuron shown in panel 13E. The failure of the linear combination model to capture the response of every cell is indicated by the large proportion of neurons with low adjusted R2 values (Fig. 13H) and, the failure of the model to capture every condition is indicated by the significant structure in the residuals (one-way ANOVA: $df = 15$; $F = 4.84$; $p < 0.05$), where the model underestimates the responses to AB conditions with large frequency separation (Fig. 13I). These results suggest that many IC neurons are well described by a linear combination; however, after close inspection of the fits it is clear that to fully capture the variance of every neuron's dual sound response will require more complex models.

By evaluating the time and trial averaged responses of IC neurons while monkeys localized concurrent sounds, we have shown that many, but not all, IC neurons respond somewhere between the two component sound responses. This type of response normalization is a suboptimal representation of multiple stimuli (Orhan & Ma, 2015), limiting the information in the macaque IC concerning each sound (chapter 3).

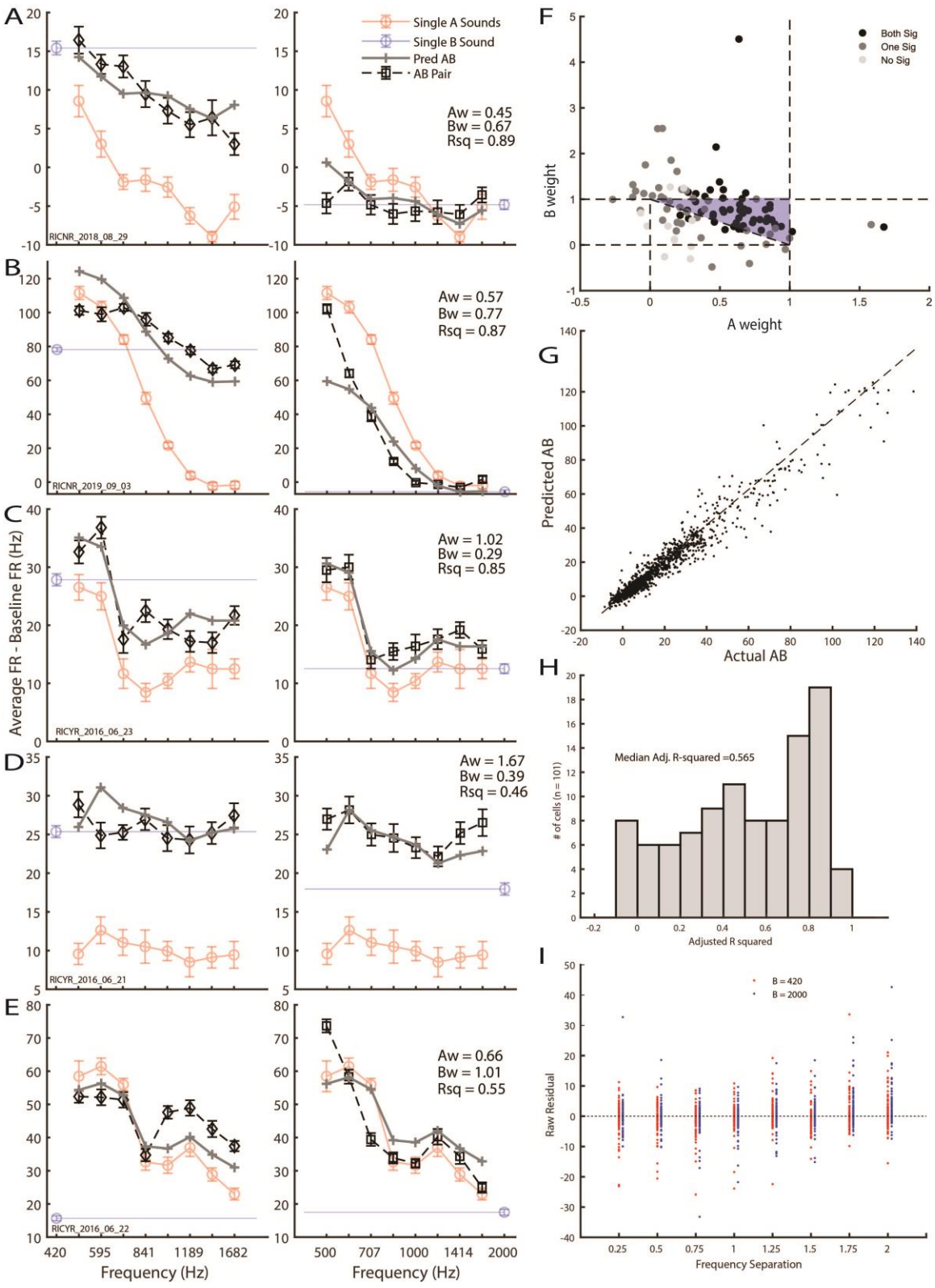

A
B
C
D
E
F
G
H
I
Single A Sounds
Single B Sound
Pred AB
AB Pair
Aw = 0.45
Bw = 0.67
Rsq = 0.89
RICNR_2018_08_29
Aw = 0.57
Bw = 0.77
Rsq = 0.87
RICNR_2019_09_03
Aw = 1.02
Bw = 0.29
Rsq = 0.85
RICYR_2016_06_23
Aw = 1.67
Bw = 0.39
Rsq = 0.46
RICYR_2016_06_21
Aw = 0.66
Bw = 1.01
Rsq = 0.55
RICYR_2016_06_22
Average FR - Baseline FR (Hz)
Frequency (Hz)
Frequency (Hz)
Both Sig
One Sig
No Sig
B weight
A weight
Predicted AB
Actual AB
Median Adj. R-squared =0.565
# of cells (n = 101)
Adjusted R squared
B = 420
B = 2000
Raw Residual
Frequency Separation

Figure 13: The time and trial averaged dual sound reposes are mostly well fit by a linear combination of the single sound responses. (A-E) The same example neurons seen in figure 11A-E with the addition of the predicted AB response (grey curve) based on the weights fit through the linear combination model. It can be seen the predicted AB responses generally match the actual AB responses (black curve), though some conditions are poorly fit (e.g. panel B, right column 2000 +500). Aw = A weight; Bw = B weight; Rsq = Adjusted R-squared value. (F) The A and B weight distribution for all responsive neurons (N = 101). The region demarcated by the shaded triangle corresponds to a number of different combination strategies. Weights of [0,1] or [1, 0], the points of the triangle, correspond to winner-take-all responses, while points along the hypotenuse correspond to a weighted average of the two component sounds (with the sum of the weights = 1). The point [1, 1] corresponds to a perfect sum. As expected, most of the winner-take-all responses have a single significant weight (grey points). The bulk of the neurons are fit by weights that fall within the triangle, mostly correspond to a response normalization, though some of these could be response enhancement depending on the particular weight and A and B responses. (G) The actual AB response versus the predicted AB response for every condition for each responsive cell (101 neurons x 16 AB conditions = 1616 points). If the predicted AB response fit the data well the data would fall upon the unity line (black dashed line). Although some data points are quite distant from unity, the means of each distribution are not significantly different (actual AB mean = 14.92 ± 0.5186 S.E.M vs. predicted AB mean = 14.72 ± 0.5078 S.E.M; t-test, p-value = 0.094). (H) The distribution of adjusted R-squared values for all 101 neurons. (I) The residuals plotted as a function of frequency separation and paired B sound. There is significant structure in the residual (one-way ANOVA: df = 15; F = 4.84; $p < 0.05$) which can be seen as higher values for larger frequency separations. Note, all data was modeled using the baseline subtracted, time and trial averaged response from 0 to 500 ms after sound onset.

Recent work that accounts for trial-by-trial structure in the response of macaque IC neurons, finds that, in response to concurrent sounds, neurons alternate their firing rate between the rates corresponding to each component sound (Caruso et al., 2018). These response alternations are, essentially, a winner-take-all that varies across time and could underlie the ability to encode simultaneous sounds. However, this work (Caruso et al., 2018) presented sounds of limited frequency separation which could bias the IC towards

response alternations. Therefore, we asked if the, apparent, normalization seen in our data was the result of neurons alternating their firing rate across or within trials and if these response alternations occur for paired sounds with large degrees of frequency separation.

We first addressed if AB responses were well described by across trial fluctuations. Across trial fluctuations would appear as AB responses that respond as if only A was present on some trials and respond as if only B was present on other trials, in other words a time-varying winner-take-all. We used a previously described Bayesian model comparison method (Caruso et al., 2018; Mohl et al., 2020) to classify each AB response distribution as 1 of 4 different models:

1. Mixture: This model corresponds to across trial fluctuations (time-varying winner-take-all). The neuron's response distribution for a given dual sound condition is described as draws from the two component sound response distributions (Fig. 14A – 'Mixture'). Note, the number of draws from each component sound distribution are not necessarily equivalent so the mixture could be biased towards a particular single sound response.
2. Intermediate: This model corresponds to a weighted average of the two component sound responses. The neuron's response distribution for a given dual sound condition is somewhere between the two component sound response

distributions (Fig. 14A – 'Intermediate'). AB conditions captured by this model could potentially fluctuate within trials.

3. Outside: This model corresponds to enhancement or inhibition. The neuron's response distribution for a given dual sound condition is larger (enhancement) or lower (inhibition) than either of the two component sound response distributions (Fig. 14A – 'Outside').
4. Single: This model corresponds to winner-take-all. The neuron's response distribution for a given dual sound condition matches one of the two component sound response distributions (Fig. 14A – 'Single').

Importantly, we make two assumptions about the single sound response distributions. One is that they are well described by a Poisson distribution. The other is that the A and B response distributions are separable; if the A and B responses were equivalent than the single, intermediate, and mixture models converge. 378 AB conditions from 64 neurons pass these two assumptions. To summarize, we compute the posterior probability of the 4 models for each AB condition, labeling the AB condition as the model with the highest posterior probability (which also serves as confidence measure).

The largest number of AB conditions was classified as single (183 out of 378) while the smallest number of AB conditions was classified as outside (16 out of 378). Importantly, a substantial proportion (100 out of 378) of the AB responses were well

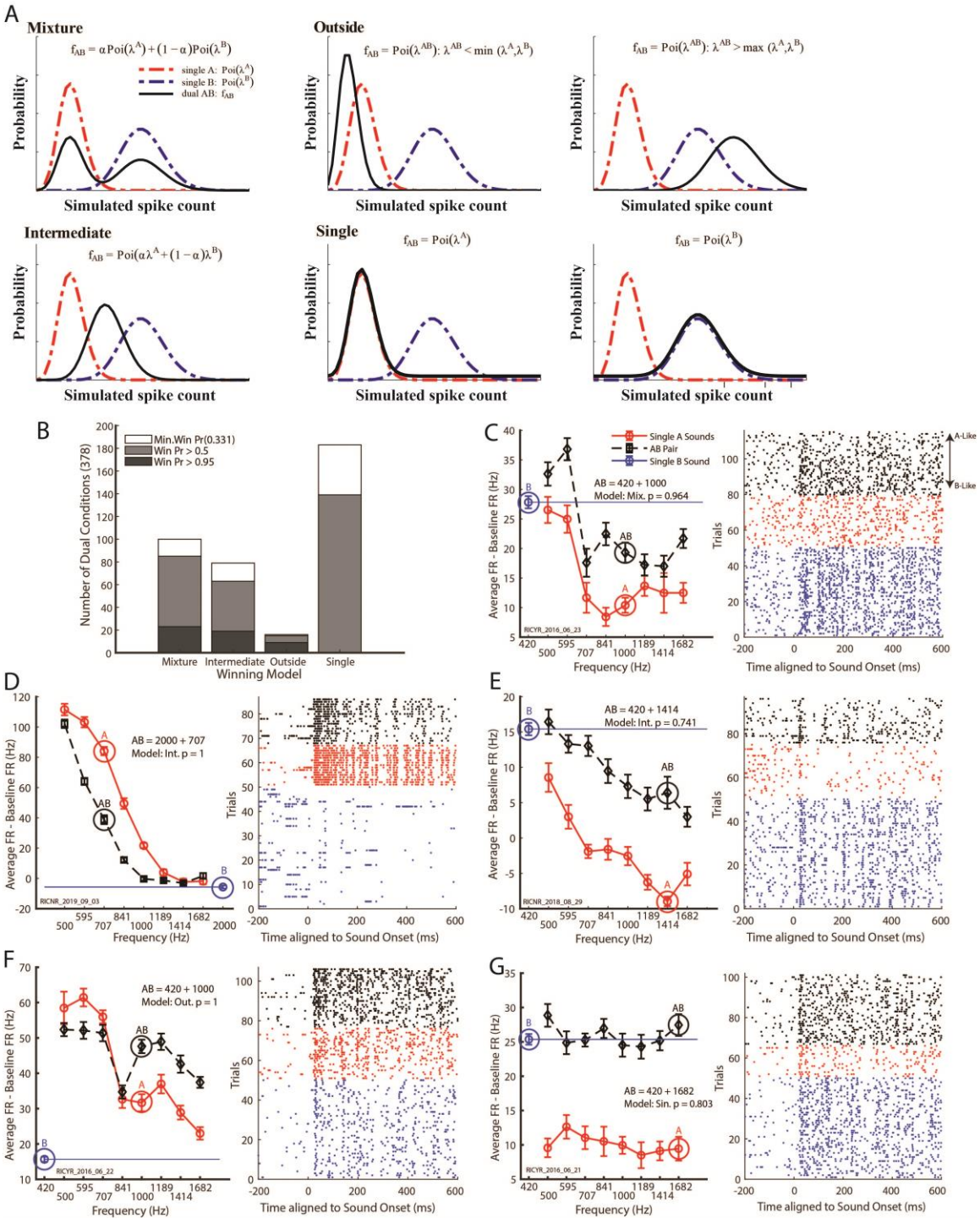


Figure 14: Normalization seen in time and trial average responses can be described by across trial response alternations. (A) A schematic of the four competing hypotheses modeling the AB spike count distributions. Mixture corresponds to an AB

distribution that samples from the A distribution on some trials and the B distribution on other trials. This corresponds to across trial fluctuations, or a time-varying winner-take-all response. Outside corresponds to an AB distribution that is either lower (an inhibition) or higher (an enhancement) than either single sound response distribution. Intermediate corresponds to an AB distribution that is between the two single sound response distributions. Single corresponds to an AB distribution that exactly matches one of the single sound response distributions. (B) The distribution of model classifications for each AB condition that passed the Poisson and separation filters (378 from 64 neurons). Win Pr. corresponds to the posterior probability, or confidence, with which the AB condition was classified. Note, the single model is the rate limiting model for the 3 other models, therefore it cannot achieve a posterior probability > 0.95. (C) An example unit with a select condition classified as mixture. The left column shows the time and trial averaged response function. The right column is a raster with A (red), B (blue) and AB (black) trials. The AB trials are sorted by their probability to come from either component sound distribution. The first AB trial is the most B-like trial while the last AB trial is the most A-like trial. There appears to be a discrete shift from B-like to A-like trials in this example neuron. (D & E) Formatted the same as (C) but for two example neurons with a select condition classified as Intermediate. (F) Formatted the same as (C) but for an example neuron with a select condition classified as Outside. (G) Formatted the same as (C) but for an example neuron with a select condition classified as Single. Note, that only the neuron in panel C appears to have a discrete shift from B-like to A-like trials in its AB response. These example neurons were also presented in figures 11 & 13.

described as a mixture of the component sound responses (Fig. 14B). Out of the 51 conditions classified with high confidence (posterior probability > 0.95), the most were classified as mixtures (23/51), while the next most supported model was intermediate (19/51) and the last remaining conditions were classified as outside (9/51). The single category received no high confidence classifications, due to the strict definition of this category. Nevertheless, the different models capture the variety of neural response patterns observed in the example neurons quite well (Fig. 14C-G). Interestingly, it seems that the across trial fluctuations may even be observable in the raw data. The AB trials,

sorted by their likelihood of being drawn from either component sound response distribution, seem to discreetly shift from higher to lower firing rates around trial 100, or a shift from B-like to A-like responses (Fig. 14C raster plot, black points). These results confirm that across trial fluctuations are present in this population of IC neurons. However, this whole trial analysis can only capture response alternations across trials; reliable alternations that occur on a timescale shorter than a trial would be labeled as intermediate. Therefore, we now investigate if AB responses were well described by within trial fluctuations.

To determine if intermediate labeled AB conditions exhibited within trial fluctuations, we first removed AB conditions that the whole trial analysis classified with low confidence (posterior probability < 0.67), leaving 153 AB conditions from 49 neurons. Approximately, 55% of mixture and intermediate conditions passed this filter (Fig. 15A). We, then, used a previously described statistical approach, the Dynamic Admixture of Poisson Process (DAPP) model (Caruso et al., 2018; Glynn et al., 2019) to detect if AB responses, particularly those classified as mixture or intermediate, remained stable or fluctuated within trials. The DAPP algorithm treats each spike train, for a given single sound condition, as independent realizations of a nonhomogeneous Poisson process with unknown rate $\lambda A(t)$ and $\lambda B(t)$ for the two component sounds. Dual sound trials were modeled as a weighted combination of the component sound rates: $\lambda AB(t) = (1-\alpha(t))\,\lambda A(t) + \alpha(t)\,\lambda B(t)$ (Fig. 15B). The weight function $\alpha(t)$ was fit to each AB trial and

it quantified the relative contribution of either component sound. If $\alpha(t)$ was 0 then the response at time t was dominated by the A response, while if $\alpha(t)$ was 1 then the response at time t was dominated by the B response. Consequently, the dynamics of $\alpha(t)$ capture the dynamics of the AB response on each trial.

An advantage to fitting the $\alpha(t)$ function for every trial of an AB condition is that it allowed for a Bayesian estimation of the underlying distribution of $\alpha(t)$ parameters (see Methods). This let us simulate $\alpha(t)$ curves the corresponding cell was likely to produce on future trials of that AB condition, which were summarized by two features: the mean and range of $\alpha(t)$ values for a given future trial. The AB condition was then tagged based on the distribution of these two features over the predicted $\alpha(t)$ curves (Fig 15B). The AB condition was tagged as wavy or flat if the distribution of $\alpha(t)$ ranges was large (maximum of 1) or small (minimum of 1), and as central or extreme if the distribution of mean $\alpha(t)$ values peaked close to 0.5 or had one or two peaks at the extremes (values near 0 or 1). Therefore, the DAPP analysis should tag mixtures as flat-extremes (Fig. 15C); but, intermediates could be labeled as a flat-central if the neuron averaged the two component responses (Fig. 15D) or as a wavy-central if the neuron alternated between the two component responses within the trial (Fig. 15E).

The DAPP analysis confirmed the results of the whole trial classifications. Roughly 70% of the AB conditions labeled as mixtures were classified as flat-extreme (Fig. 15F). However, the intermediate classifications contained more heterogeneity. The

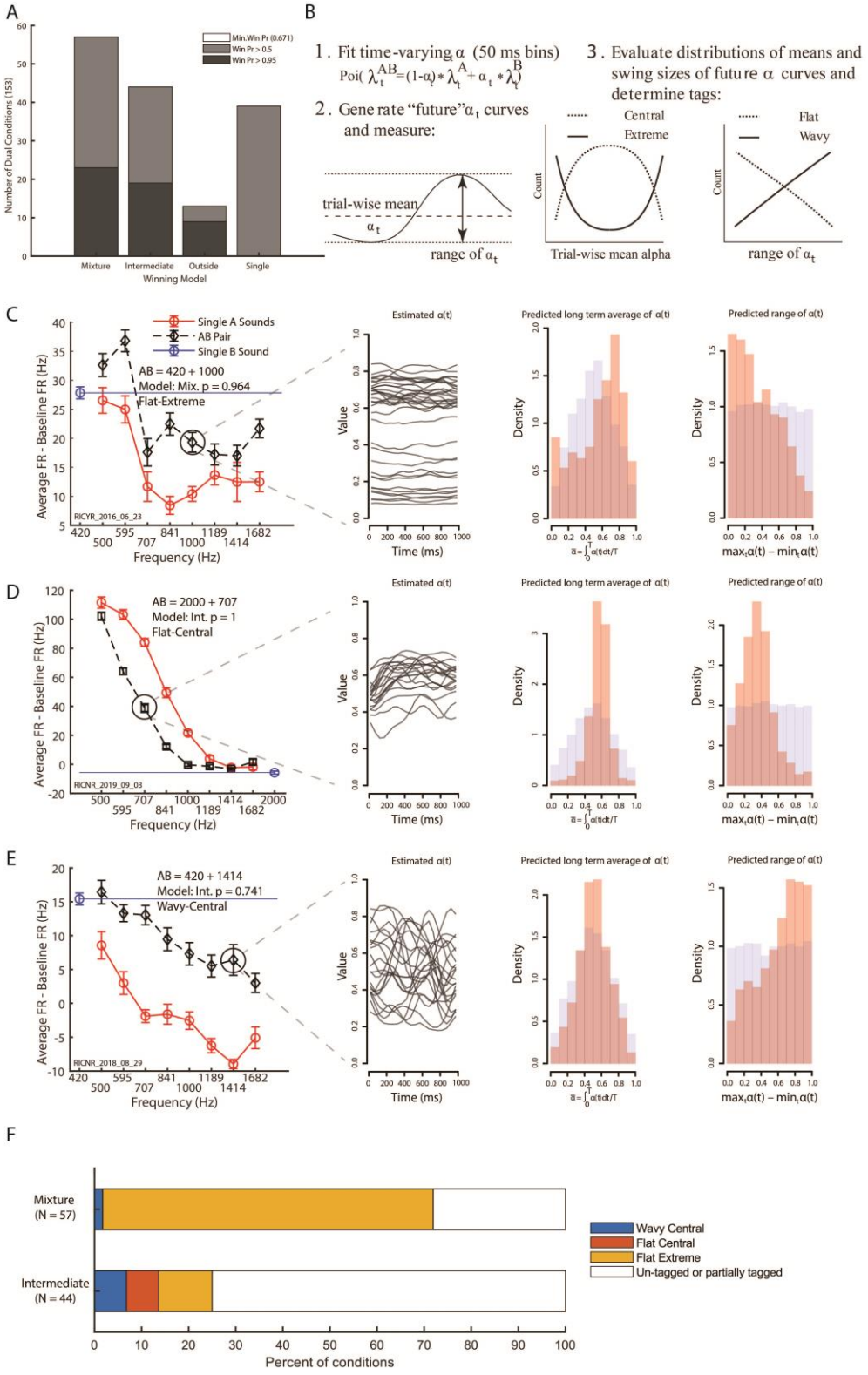

A
Number of Dual Conditions (153)
Min.Win Pr (0.671)
Win Pr > 0.5
Win Pr > 0.95
Mixture
Intermediate
Outside
Single
Winning Model
B
1. Fit time-varying α (50 ms bins)
Poi(λ_t^AB=(1-α_t)*λ_t^A+α_t*λ_t^B)
2. Gene rate "future"α_t curves and measure:
trial-wise mean
α_t
range of α_t
3. Evaluate distributions of means and swing sizes of future α curves and determine tags:
Central
Extreme
Count
Trial-wise mean alpha
Flat
Wavy
Count
range of α_t
C
Single A Sounds
AB Pair
Single B Sound
AB = 420 + 1000
Model: Mix. p = 0.964
Flat-Extreme
Average FR - Baseline FR (Hz)
RICYR_2016_06_23
Frequency (Hz)
Estimated α(t)
Value
Time (ms)
Predicted long term average of α(t)
Density
Predicted range of α(t)
Density
max_t α(t) – min_t α(t)
D
AB = 2000 + 707
Model: Int. p = 1
Flat-Central
Average FR - Baseline FR (Hz)
RICNR_2019_09_03
Frequency (Hz)
Estimated α(t)
Value
Time (ms)
Predicted long term average of α(t)
Density
Predicted range of α(t)
Density
max_t α(t) – min_t α(t)
E
AB = 420 + 1414
Model: Int. p = 0.741
Wavy-Central
Average FR - Baseline FR (Hz)
RICNR_2018_08_29
Frequency (Hz)
Estimated α(t)
Value
Time (ms)
Predicted long term average of α(t)
Density
Predicted range of α(t)
Density
max_t α(t) – min_t α(t)
F
Mixture
(N = 57)
Intermediate
(N = 44)
Percent of conditions
Wavy Central
Flat Central
Flat Extreme
Un-tagged or partially tagged

Figure 15: The DAPP model confirms the across trial fluctuations and uncovers a small group of neurons that fluctuate within trials. (A) The whole trial model classifications for each AB condition classified with a posterior probability > 0.67 (153 conditions from 49 neurons). A posterior probability of 0.67 means the winning model is, at minimum twice as likely as the next highest model. (B). A schematic outlining the steps of the DAPP model. (C). An example neuron with a select condition classified as mixture (left column). The Estimated $\alpha(t)$ panel shows the $\alpha(t)$ functions fit to the AB trials. The predicted long-term average of $\alpha(t)$ panel shows the mean trial-averaged α value for simulated AB trials, while the range of $\alpha(t)$ panel corresponds to how flat the predicted $\alpha(t)$ functions are. The AB response for this neuron showed a bimodal distribution of α values and a range distribution skewed towards 0, leading this condition to be tagged as Flat-Extreme – which is expected given its mixture classification. Note, the red histograms correspond to the density of the predicted $\alpha(t)$ values while the blue histogram shows the prior distribution. (D) Formatted the same as (C) but for the neuron except it classified as an intermediate but tagged as a Flat-Central, which corresponds to an AB response that averages the two component sound responses. (E) Formatted the same as (C) but for the neuron except classified as an intermediate but tagged as a Wavy-Central, which corresponds to an AB response that alternates within trial. (F) The percent of AB conditions labelled as mixture or intermediate that received a specific tag. Note, the example units shown here were presented in the previous figures.

bulk of the intermediates (75%) were not completely classified; of those that were tagged, ~11% were classified as flat-extreme while each of the wavy-central and the flat-central categories accounted for ~7% of the tagged conditions (Fig. 15F). Crucially, the mixture flat-extreme conditions had two peaks in the distribution of mean α(t) values indicating these responses fluctuated across trials, while the intermediate flat-extreme conditions had a single peak that was biased towards one of the two component sound responses. Therefore, the intermediate conditions tagged as flat-extreme or flat-central correspond to weighted averages of the two component sounds with biased or equal weights, respectively. To summarize, the DAPP results confirmed the existence of across

trial fluctuations and true-average responses, while also uncovering a small group of neurons that fluctuation their AB responses within trials.

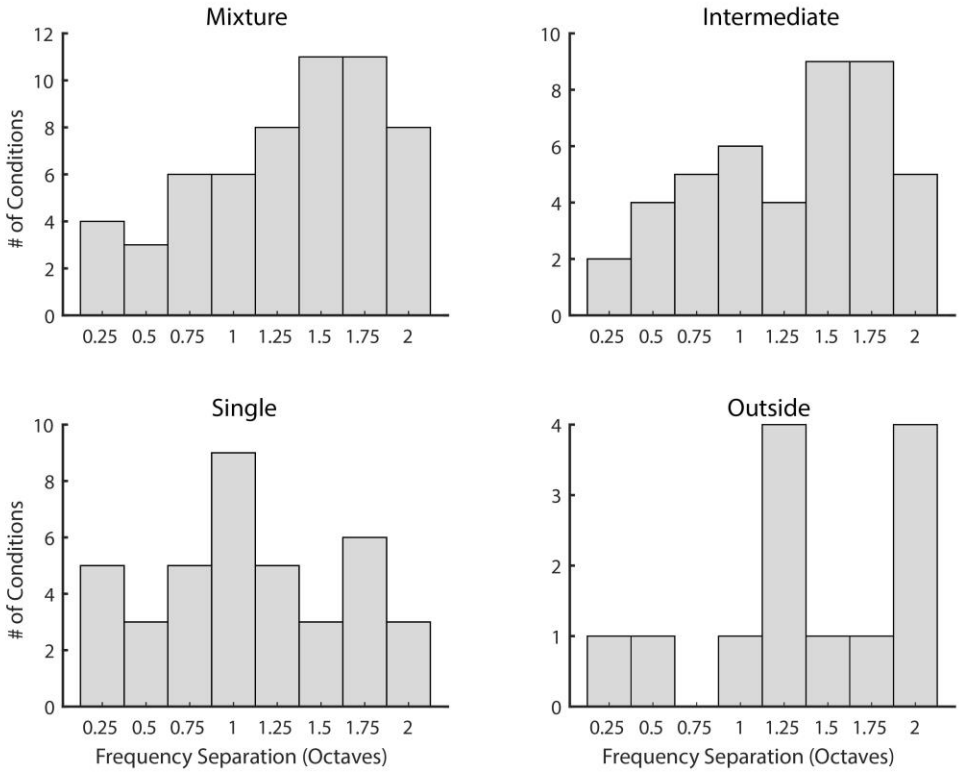


Figure 16: Each model classifies conditions across the range of frequency separations presented. No model is specialized to encode sound pairs of a particular frequency separation.

These results largely reproduce the findings of Caruso et al. (2018) and confirm that the normalization observed with canonical analyses (Fig. 12 & Fig. 13) can be partially explained by neurons that alternate their firing rate. However, are these response fluctuations specialized for encoding sound pairs proximal in frequency or are

they a more general computation for encoding all concurrent sounds? To answer this question, we plotted the number of AB conditions (N = 153 with posterior $p > 0.67$) by frequency separation for each model class (Fig. 16). It is clear response alternations, conditions classified as mixture, occur for the entire 0.25-2 octave range tested (Fig. 16 – 'Mixture'). In fact, AB conditions were observed for every model across the range of presented frequency separations. Although many of the models seem biased towards large frequency separations, this is likely explained by intrinsic properties of IC neurons and the nature of the whole trial analysis. Many macaque IC neurons are tuned to sound frequency (e.g. Fig. 11A-C & E). Therefore, their firing rates become more separable as frequency separation increases, leading to more AB conditions that pass the separability assumption and classifying more AB conditions with higher confidence (Mohl et al., 2020). Though these results suggest that response fluctuations are a more general computation occurring for sound pairs as far as 2 octaves apart, it remains to be seen whether IC neurons consistently contribute to a particular model category or implement a number of different models across AB conditions.

An advantage to presenting a large number of conditions during each recording session is that it allows for an interrogation of the consistency of the dual sound response patterns. To determine how stable dual sound responses were across conditions and the IC population, we first plotted model counts for each of the 49 neurons with AB conditions classified with a posterior probability greater than 0.67 (Fig.

17A). Though there is heterogeneity in the AB response patterns across the IC population, roughly 63% of neurons (31 out of 49) have at least one AB condition classified as a mixture. Thus, across trial response alternations are a general mechanism many IC neurons use to encode concurrent sounds.

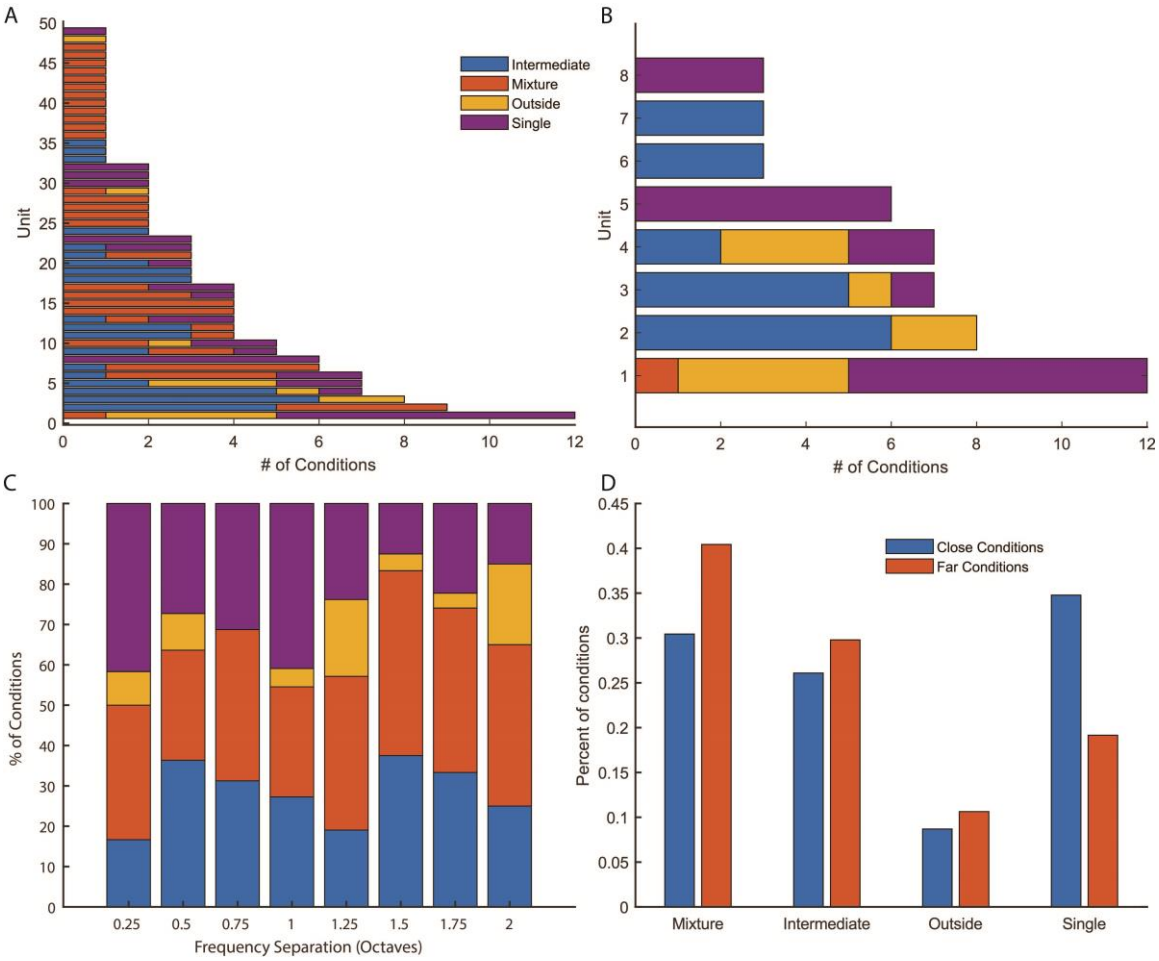


Figure 17: Mixtures are a predominate model across the population and their contribution to the population response pattern is stable over sound pairs of different frequency separations. . (A) The counts for each model for all 49 cells with an AB condition classified with a posterior probability > 0.67. It is apparent that many neurons have at least 1 condition labeled as mixture. (B) The neurons who were significantly

different based on a chi-square goodness of fit test between the neurons model counts and the counts expected from chance given the population model distribution (Fig. 15A). Note, this test was done on all neurons that had at least 3 AB conditions (N = 23). Note, some of these neurons are only fit by a single model. (C) The percent each model contributes to the population distribution across sound pairs of different frequency separations. The single conditions seem to contribute less at higher separations while the mixture conditions seem to contribute slightly more. (D) We combined the two furthest frequency separations (Far - 1.75 & 2 octaves) and the two closest frequency separations (Close – 0.25 & 0.5 octaves) and performed a chi-square test of independence and found these two distributions are not significantly different (chi-squared test, degrees of freedom = 3, chi-statistic = 2.102, p-value = 0.552).

We next asked if the AB responses of any IC neurons showed a bias towards a particular model. Of the 32 neurons with at least two AB conditions classified with a posterior probability greater than 0.67, 14 neurons had AB conditions well described by only a single model. To statistically test if a given neuron implemented a different distribution of models than predicted from chance, we performed a chi-square goodness of fit test between a given neuron's model distribution and the population's model distribution (Fig. 15A). Though we first subtracted the conditions from the tested neuron so as to not bias the population's distribution. Due to small sample size we, also, excluded neurons that had less than 3 AB conditions, leaving 23 neurons (Fig. 17A – first 23 units). Roughly 35% of neurons (8 out of 23; Fig. 15B) implemented a distribution of AB response patterns significantly different than the IC population (chi-squared test, degrees of freedom = 3, chi-statistic > 7.81 , p-value < 0.05). Many of these neurons are biased towards intermediate or single AB responses. Due to the small sample size and the large number of mixtures in the population's model distribution, the two neurons

(unit 14 & 15) with only mixture responses only trended toward significance (chi-squared test, degrees of freedom = 3, chi-statistic =7.245, p-value = 0.0645). Nonetheless, there seems to be substantial minority of neurons that are biased towards a particular model, but the majority of the neuron's implement a combination of the different model classifications.

We then asked if the population's model distribution depended on the frequency separation of the sound pair (Fig. 17C). It appears that as frequency separation increases the single model contributes less and the mixture model contributes more to the population response. To statistically test this, we performed a chi-squared test of independence by comparing the model distribution for the two most distant frequency pairs to the model distribution for the two most proximal frequency pairs. Although these two distributions seem to differ (Fig. 17D), they are not significantly different (chi-squared test, degrees of freedom = 3, chi-statistic = 2.102, p-value = 0.552). Therefore, the frequency separation of the paired sounds does not seem to significantly affect the way the IC encodes concurrent sounds. To summarize, the IC response to concurrent sounds is heterogenous. Many neurons encode simultaneous sounds with a number of combination rules, the most common of which is across trial response alternations (mixtures), and the population response pattern remains largely stable across all dual sound pairs.

4.4 Discussion

How the brain encodes multiple objects simultaneously is an oft ignored question in neuroscience. Perhaps the most relevant, but slightly tangential, field concerns how populations of neurons encode a single object in the presence of other objects. In this case, it is abundantly clear, across species and brain regions, that neurons respond less strongly to a stimulus in the presence of other stimuli, a behavior known as normalization (Carandini & Heeger, 2012). Indeed, we observed similar behavior in the average response of IC neurons when monkeys successfully reported the presence of two concurrent sounds, with many IC neurons responding between the responses of either component sound (e.g. Fig. 11, 12 & 13). Although stimuli can be accurately decoded with normalized responses, it likely requires a preservation of individual neuron's selectivity (Li, Cox, Zoccolan, & DiCarlo, 2009) or a homogenously tuned population (Bao & Tsao, 2018), neither of which are observed in the monkey IC (Bulkin & Groh, 2011; Versnel et al., 2009; chapter 3). More likely, the IC is exploiting dual sound responses that fluctuate between the response to each component sound. These response alternations were observed (Fig. 14 & 15) in roughly 63% of IC neurons and their contribution to the population response remained stable across the full, 2 octave, range of frequency separations tested (Fig. 16 & 17).Thus it appears that response alternations are a general mechanism used by the IC to facilitate the encoding of multiple sounds. More broadly these results support response alternations a general mechanism seen

across many brain areas including V1 (Jun, Ruff, Tokdar, Cohen, & Groh, 2019), IT cortex (Caruso et al., 2018), hippocampus (Kay et al., 2020) and may potentially be observed in areas involved in decision making (Ashwood et al., 2020; Kang et al., 2020).

A limitation of the current study arises due to the two assumptions required for the whole trial classification of AB conditions; we assume the component responses are Poisson and separable. These assumptions result in much of the data being lost (roughly 76%), making it difficult to interrogate the consistency of a neurons combination rule. The Poisson assumption can be relaxed by implementing modeling responses with other distributions (e.g. negative binomial). However, the separation assumption cannot be relaxed and is fundamental to the interpretation of the whole trial results. It is unknown what roles the neurons that do not differentially respond to each component sound play in the perception of sounds. Nevertheless, future analyses will need to be developed to prevent loss of data and statistically probe the role of non-separable responses.

Broadly, these response fluctuations are possibly connected with cognitive processes, potentially carried through the large corticofugal projection into the IC (Huffman & Henson, 1990). For example, the alternations bear some resemblance to the results of selective attention (Moran & Desimone, 1985) where neurons selectivity responds to a particular stimulus or the rhythmic theory of attention where attention fluctuates between salient stimuli 3-8 times a second (Fiebelkorn & Kastner, 2019). It is important to note that we do not control for attention in this study. However, to

correctly perform on dual sound trials the monkey must report the location of both sounds, so the representation of each sound should be maintained for the length of the trial. That being said, it's quite unclear what is meant by attention most of the time (Hommel et al., 2019).

The growing support for response alternations calls for revisiting data where a trial averaged combined response is modeled as a function of the component responses. For example, we found that even though some neuron's trial averaged AB responses were well captured by a linear combination model (e.g. Fig. 13C), trial-by-trial structure of the AB responses clearly indicate that a given neuron may alternate their responses between the two component sounds (e.g. Fig. 14C & 15C). However, it also true that some neurons do average the component sound responses under some conditions (Fig. 13B, 14D & 15D). Indeed, there are a number of computational benefits to the normalization of responses (Carandini & Heeger, 2012). Additionally, the degradation of information concerning component stimuli can be alleviated through heterogeneity in combination weights or tuning properties and specific correlational structure (Ecker, Berens, Tolias, & Bethge, 2011; Orhan & Ma, 2015). It is possible, and likely, that alternating populations and averaging populations could differ in connectivity or physiology and may underlie different behaviors. Therefore, it is vital to our understanding of how brains function in cluttered environments that studies account for trial-by-trial structure in neural responses.

5. Conclusions

5.1 Summary

The fact that animals can localize simultaneously occurring sounds challenges prevailing assumptions that sound location is encoded via a rate code. Neurons cannot fire at two rates simultaneously; so how could they encode concurrent sound sources? In the previous chapters I presented work investigating how a coding bottleneck, the monkey inferior colliculus (IC), could overcome this constraint of rate coding during the localization of a single, or two concurrent, sound source(s).

I first explored whether IC neural sensitivity to sound frequency could facilitate the encoding of concurrent sounds. I hypothesized that on dual sound trials IC neurons would increase their selectivity to sound frequency, reducing population overlap and allowing for segregated populations to encode each sound. I first confirmed that monkeys could localize paired sounds that differed in center frequency from 0.25-2 octaves. Monkeys performed, on average, 86% of dual sound trials correctly, and performance improved as the frequency separation of the sound pair increased. I then verified that the presented sounds recruited a largely overlapping population of broadly tuned neurons. Although there was heterogeneity in the neural responses to dual sounds, it was evident that most IC neurons did not sharpen their frequency receptive fields but instead broadened them. These modulations of the frequency response functions reduced the information available to a maximum-likelihood decoder, which

performed worse on dual sound trials. I concluded that IC selectivity to sound frequency is unlikely to underlie the ability to localize concurrent sounds and proposed that recently reported (Caruso et al., 2018) alternations of IC responses may provide a more promising avenue of research.

I next investigated if these response alterations were a general mechanism the IC implemented to encode concurrent sounds or if they were a specialized computation used only to encode sound pairs proximal in frequency (< 1 octave), like the frequencies used in Caruso et al. (2018). Response alternations would appear as trial averaged responses that fall between the two component sound responses. I, first, confirmed this by modelling the dual sound responses as a linear combination of the two component sound responses and found most neurons were well fit by a response normalization. I then verified that across and within trial fluctuations were present in the data. Indeed, these response alterations were the predominate computation in the IC, being observed in roughly 63% of IC neurons and across the full, 2 octave, range of frequency separations. Importantly, their contribution to the population response was stable across sound pairs of all frequency separations; suggesting that response alternations are a general computation exploited by IC neurons to facilitate the encoding of concurrent sounds

5.2 Implications, limitations, and future directions

The predominance of across trial response fluctuations in the monkey IC has interesting implications for the representation of auditory space reference frames. A peculiar finding throughout many auditory areas, including the IC, is that they do not strictly represent auditory space in head-centered coordinates, instead they represent auditory space somewhere between head- and eye-centered reference frames (for review see:Willett, Groh, & Maddox, 2019). Interestingly, recent work has shown that neurons in the ventral intraparietal area are capable of dynamically switching between world- and self-coordinate systems depending on the demands of the task (Sasaki, Anzai, Angelaki, & DeAngelis, 2020). So therefore, it is possible that the origin of the hybrid reference frame seen in the IC (Groh et al., 2001) and other auditory areas could emerge due to latent across trial fluctuations between the head- or eye-centered coordinate systems. However, reference frames can only, currently, be measured using across trial averaged activity. Therefore, new analyses will need to be developed before the contribution of response alternations in this field can be addressed.

The majority of work investigating the responses of neurons to the presence of multiple stimuli within their receptive field finds a reduction in firing rate. This characteristic response puzzled the field for some time because animals seem to trivially recognize objects in cluttered environments, yet the representation of the object is not invariant to the presence of other objects. In fact, the particular way the multi-object

responses changed reduces the information concerning each object within the scene (Orhan & Ma, 2015, chapter 3). To account for this reduction in firing rate (Li et al., 2009) proposed a readout that depended not on firing rate per se but instead on the rank order of object preference. Although the stability of object preference was important for the successful decoding of stimuli it required a different decoding threshold, which implies the existence of two neural codes, one for objects in isolation and one for objects in clutter. How would the decoder know which threshold to use at a given time if it does not know how many objects are in a given scene beforehand? Importantly, the current work offers an alternative explanation, that the response normalization could be explained by across time response alternations, effectively a time-varying winner-take-all code. A winner-take-all code is known to outperform weighted averaging and divisive normalization in the representation of multi-object responses (Orhan & Ma, 2015) and would allow for the instantiation of a single downstream decoder robust to multiple objects in a scene.

A strong assumption of the current work is that the neural code fundamentally depends on a stable relationship between a neuron's firing rate and the stimulus space. However, this is clearly not always the case. For example IC neurons in the current work can enhance their responses under some conditions, and it is known throughout the brain that firing rates can be modulated by cognitive processes (Humphreys et al., 1998), behavioral states (e.g. arousal or locomotionVinck, Batista-Brito, Knoblich, & Cardin,

2015), or stimulus statistics (Malmierca, Cristaudo, Pérez-González, & Covey, 2009). Alternatives to this interpretation of firing rate exist. For instance, probabilistic population codes assume firing rate corresponds to confidence a particular stimulus is present (Ma, Beck, Latham, & Pouget, 2006). On the other hand there is evidence of neurons encoding stimulus information through spike timing (Gollisch & Meister, 2008) and it is known that spike timing contributes to the encoding of sound features of the inferior colliculus (Chase & Young, 2006). Nevertheless, future work must investigate the contribution of firing rate and spike timing to neural representations (Brette, 2015).

Another limitation of the current study is that we largely ignore the known diversity of cell types within the IC (Peruzzi, Sivaramakrishnan, & Oliver, 2000; Allen Ryan & Miller, 1977). In the auditory system, many cell types are known to functionally contribute to the processing of specific aspects of sound (Cant & Benson, 2003). Do response alternations occur across the diverse number of cells in the inferior colliculus or is it specialized to only a particular class? Future work should address not only if particular classes of cells implement response alternations but also determine how their connectivity allows for the implementation of time-varying winner-take-all signals.

References

Aitkin, L. M., & Phillips, S. C. (1984). Is the inferior colliculus and obligatory relay in the cat auditory system? *Neuroscience Letters, 44*(3), 259-264. doi:https://doi.org/10.1016/0304-3940(84)90032-6

Ashwood, Z. C., Roy, N. A., Stone, I. R., Churchland, A. K., Pouget, A., & Pillow, J. W. (2020). Mice alternate between discrete strategies during perceptual decision-making. *bioRxiv*, 2020.2010.2019.346353. doi:10.1101/2020.10.19.346353

Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature Communications, 9*(1), 1774. doi:10.1038/s41467-018-04126-7

Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian analysis, 1*(3), 385-402.

Best, V., Schaik, A. v., & Carlile, S. (2004). Separation of concurrent broadband sound sources by human listeners. *The Journal of the Acoustical Society of America, 115*(1), 324-336. doi:10.1121/1.1632484

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*: MIT press.

Brette, R. (2015). Philosophy of the Spike: Rate-Based vs. Spike-Based Theories of the Brain. *Frontiers in Systems Neuroscience, 9*(151). doi:10.3389/fnsys.2015.00151

Bulkin, D. A., & Groh, J. M. (2011). Systematic mapping of the monkey inferior colliculus reveals enhanced low frequency sound representation. *Journal of Neurophysiology, 105*(4), 1785-1797. doi:10.1152/jn.00857.2010

Bulkin, D. A., & Groh, J. M. (2012a). Distribution of eye position information in the monkey inferior colliculus. *Journal of Neurophysiology, 107*(3), 785-795. doi:10.1152/jn.00662.2011

Bulkin, D. A., & Groh, J. M. (2012b). Distribution of visual and saccade related information in the monkey inferior colliculus. *Frontiers in Neural Circuits, 6*(61). doi:10.3389/fncir.2012.00061

Busse, L., Wade, A. R., & Carandini, M. (2009). Representation of Concurrent Stimuli by Population Activity in Visual Cortex. *Neuron, 64*(6), 931-942. doi:https://doi.org/10.1016/j.neuron.2009.11.004

Cant, N. B., & Benson, C. G. (2003). Parallel auditory pathways: projection patterns of the different neuronal populations in the dorsal and ventral cochlear nuclei.

Brain Research Bulletin, 60(5), 457-474. doi:https://doi.org/10.1016/S0361-9230(03)00050-9

Capuano, U., & McIlwain, J. T. (1981). Reciprocity of receptive field images and point images in the superior colliculus of the cat. *Journal of Comparative Neurology, 196*(1), 13-23. doi:10.1002/cne.901960103

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience, 13*(1), 51-62. doi:10.1038/nrn3136

Caruso, V. C., Mohl, J. T., Glynn, C., Lee, J., Willett, S. M., Zaman, A., . . . Groh, J. M. (2018). Single neurons may encode simultaneous stimuli by switching between activity patterns. *Nature Communications, 9*(1), 2715. doi:10.1038/s41467-018-05121-8

Chase, S. M., & Young, E. D. (2006). Spike-Timing Codes Enhance the Representation of Multiple Simultaneous Sound-Localization Cues in the Inferior Colliculus. *The Journal of Neuroscience, 26*(15), 3889. doi:10.1523/JNEUROSCI.4986-05.2006

Chettih, S. N., & Harvey, C. D. (2019). Single-neuron perturbations reveal feature-specific competition in V1. *Nature, 567*(7748), 334-340. doi:10.1038/s41586-019-0997-6

Connell, M. N., Barczak, A., Schroeder, C. E., & Lakatos, P. (2014). Layer Specific Sharpening of Frequency Tuning by Selective Attention in Primary Auditory Cortex. *The Journal of Neuroscience, 34*(49), 16496. doi:10.1523/JNEUROSCI.2055-14.2014

Day, M. L., & Delgutte, B. (2013). Decoding Sound Source Location and Separation Using Neural Population Activity Patterns. *The Journal of Neuroscience, 33*(40), 15837. doi:10.1523/JNEUROSCI.2034-13.2013

Day, M. L., Koka, K., & Delgutte, B. (2012). Neural encoding of sound source location in the presence of a concurrent, spatially separated source. *Journal of Neurophysiology, 108*(9), 2612-2628. doi:10.1152/jn.00303.2012

Ecker, A., Berens, P., Tolias, A., & Bethge, M. (2011). The effect of noise correlations in populations of diversely tuned neurons. *Nature Precedings*. doi:10.1038/npre.2011.6170.1

Fiebelkorn, I. C., & Kastner, S. (2019). A Rhythmic Theory of Attention. *Trends in Cognitive Sciences, 23*(2), 87-101. doi:https://doi.org/10.1016/j.tics.2018.11.009

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2005). Experience-Dependent Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex. *Cerebral Cortex, 16*(11), 1631-1644. doi:10.1093/cercor/bhj100

Glynn, C., Tokdar, S. T., Zaman, A., Caruso, V. C., Mohl, J. T., Willett, S. M., & Groh, J. M. (2019). Analyzing second order stochasticity of neural spiking under stimuli-bundle exposure. arXiv:1911.04387. Retrieved from https://ui.adsabs.harvard.edu/abs/2019arXiv191104387G

Gollisch, T., & Meister, M. (2008). Rapid Neural Coding in the Retina with Relative Spike Latencies. *Science, 319*(5866), 1108. doi:10.1126/science.1149639

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience, 5*(11), 887-892. doi:10.1038/nrn1538

Groh, J. M., Kelly, K. A., & Underhill, A. M. (2003). A Monotonic Code for Sound Azimuth in Primate Inferior Colliculus. *Journal of Cognitive Neuroscience, 15*(8), 1217-1231. doi:10.1162/089892903322598166

Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R., & Inati, S. (2001). Eye Position Influences Auditory Responses in Primate Inferior Colliculus. *Neuron, 29*(2), 509-518. doi:https://doi.org/10.1016/S0896-6273(01)00222-7

Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of Sound Localization in Mammals. *Physiological Reviews, 90*(3), 983-1012. doi:10.1152/physrev.00026.2009

Hartline, H. K., & Ratliff, F. (1957). Inhibitory interaction of receptor units in the eye of Limulus. *The Journal of general physiology, 40*(3), 357-376. doi:10.1085/jgp.40.3.357

Henry, C. A., & Kohn, A. (2020). Spatial contextual effects in primary visual cortex limit feature representation under crowding. *Nature Communications, 11*(1), 1687. doi:10.1038/s41467-020-15386-7

Hofman, P. M., Van Riswick, J. G. A., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience, 1*(5), 417-421. doi:10.1038/1633

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics, 81*(7), 2288-2303. doi:10.3758/s13414-019-01846-w

Hudspeth, A. J. (2014). Integrating the active process of hair cells with cochlear function. *Nature Reviews Neuroscience, 15*(9), 600-614. doi:10.1038/nrn3786

Huffman, R. F., & Henson, O. W. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. *Brain Research Reviews, 15*(3), 295-323. doi:https://doi.org/10.1016/0165-0173(90)90005-9

Humphreys, G. W., Duncan, J., Treisman, A., Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353*(1373), 1257-1270. doi:doi:10.1098/rstb.1998.0281

Imig, T. J., Irons, W. A., & Samson, F. R. (1990). Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. *Journal of Neurophysiology, 63*(6), 1448-1466. doi:10.1152/jn.1990.63.6.1448

The Inferior Colliculus. (2005). (J. A. Winer & C. E. Schreiner Eds. 1 ed.): Springer-Verlag New York.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience, 9*(5), 690-696. doi:10.1038/nn1691

Jenkins, W. M., & Masterton, R. B. (1982). Sound localization: effects of unilateral lesions in central auditory system. *Journal of Neurophysiology, 47*(6), 987-1016. doi:10.1152/jn.1982.47.6.987

Judge, S. J., Richmond, B. J., & Chu, F. C. (1980). Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Research, 20*(6), 535-538. doi:https://doi.org/10.1016/0042-6989(80)90128-5

Jun, N. Y., Ruff, D. A., Tokdar, S. T., Cohen, M. R., & Groh, J. M. (2019). Patterns of neural correlations in V1 vary with the number of objects. *bioRxiv*, 777912. doi:10.1101/777912

Kang, Y. H. R., Löffler, A., Jeurissen, D., Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2020). Multiple decisions about one object involve parallel sensory acquisition but time-multiplexed evidence incorporation. *bioRxiv*, 2020.2010.2015.341008. doi:10.1101/2020.10.15.341008

Kay, K., Chung, J. E., Sosa, M., Schor, J. S., Karlsson, M. P., Larkin, M. C., . . . Frank, L. M. (2020). Constant Sub-second Cycling between Representations of Possible Futures in the Hippocampus. *Cell, 180*(3), 552-567.e525. doi:https://doi.org/10.1016/j.cell.2020.01.014

Kelly, J. B., & Kavanagh, G. L. (1994). Sound localization after unilateral lesions of inferior colliculus in the ferret (Mustela putorius). *Journal of Neurophysiology, 71*(3), 1078-1087. doi:10.1152/jn.1994.71.3.1078

Knudsen, E. I., & Konishi, M. (1978). A neural map of auditory space in the owl. *Science, 200*(4343), 795. doi:10.1126/science.644324

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron, 93*(3), 480-490. doi:https://doi.org/10.1016/j.neuron.2016.12.041

Lee, C.-C., & Middlebrooks, J. C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. *Nature Neuroscience, 14*(1), 108-114. doi:10.1038/nn.2713

Li, N., Cox, D. D., Zoccolan, D., & DiCarlo, J. J. (2009). What Response Properties Do Individual Neurons Need to Underlie Position and Clutter "Invariant" Object Recognition? *Journal of Neurophysiology, 102*(1), 360-376. doi:10.1152/jn.90745.2008

Lim, H. H., Lenarz, M., & Lenarz, T. (2009). Auditory Midbrain Implant: A Review. *Trends in Amplification, 13*(3), 149-180. doi:10.1177/1084713809348372

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*(11), 1432-1438. doi:10.1038/nn1790

Malmierca, M. S., Cristaudo, S., Pérez-González, D., & Covey, E. (2009). Stimulus-Specific Adaptation in the Inferior Colliculus of the Anesthetized Rat. *The Journal of Neuroscience, 29*(17), 5483. doi:10.1523/JNEUROSCI.4153-08.2009

Mason, A. C., Oshinsky, M. L., & Hoy, R. R. (2001). Hyperacute directional hearing in a microscale auditory system. *Nature, 410*(6829), 686-690. doi:10.1038/35070564

Maunsell, J. H. R. (2015). Neuronal Mechanisms of Visual Attention. *Annual Review of Vision Science, 1*(1), 373-391. doi:10.1146/annurev-vision-082114-035431

McAlpine, D., Jiang, D., & Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nature Neuroscience, 4*(4), 396-401. doi:10.1038/86049

Mickey, B. J., & Middlebrooks, J. C. (2003). Representation of Auditory Space by Cortical Neurons in Awake Cats. *The Journal of Neuroscience, 23*(25), 8649. doi:10.1523/JNEUROSCI.23-25-08649.2003

Middlebrooks, J. C., Clock, A. E., Xu, L., & Green, D. M. (1994). A panoramic code for sound location by cortical neurons. *Science, 264*(5160), 842. doi:10.1126/science.8171339

Middlebrooks, J. C., & Pettigrew, J. D. (1981). Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. *The Journal of Neuroscience, 1*(1), 107. doi:10.1523/JNEUROSCI.01-01-00107.1981

Mohl, J. T., Caruso, V. C., Tokdar, S. T., & Groh, J. M. (2020). Sensitivity and specificity of a Bayesian single trial analysis for time varying neural signals. arXiv:2001.11582. Retrieved from https://ui.adsabs.harvard.edu/abs/2020arXiv200111582M

Moore, B. C. (2012). *An introduction to the psychology of hearing*: Brill.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782. doi:10.1126/science.4023713

Orhan, A. E., & Ma, W. J. (2015). Neural Population Coding of Multiple Stimuli. *The Journal of Neuroscience, 35*(9), 3825. doi:10.1523/JNEUROSCI.4097-14.2015

Pages, D. S., Ross, D. A., Puñal, V. M., Agashe, S., Dweck, I., Mueller, J., . . . Groh, J. M. (2016). Effects of Electrical Stimulation in the Inferior Colliculus on Frequency Discrimination by Rhesus Monkeys and Implications for the Auditory Midbrain Implant. *The Journal of Neuroscience, 36*(18), 5071. doi:10.1523/JNEUROSCI.3540-15.2016

Perrott, D. R. (1984a). Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity. *The Journal of the Acoustical Society of America, 75*(4), 1201-1206. doi:10.1121/1.390771

Perrott, D. R. (1984b). Discrimination of the spatial distribution of concurrently active sound sources: Some experiments with stereophonic arrays. *The Journal of the Acoustical Society of America, 76*(6), 1704-1712. doi:10.1121/1.391617

Peruzzi, D., Sivaramakrishnan, S., & Oliver, D. L. (2000). Identification of cell types in brain slices of the inferior colliculus. *Neuroscience, 101*(2), 403-416. doi:https://doi.org/10.1016/S0306-4522(00)00382-1

Pfingst, B. E., Laycock, J., Flammino, F., Lonsbury-Martin, B., & Martin, G. (1978). Pure tone thresholds for the rhesus monkey. *Hearing Research, 1*(1), 43-47. doi:https://doi.org/10.1016/0378-5955(78)90008-4

Porter, K. K., Metzger, R. R., & Groh, J. M. (2006). Representation of Eye Position in Primate Inferior Colliculus. *Journal of Neurophysiology, 95*(3), 1826-1842. doi:10.1152/jn.00857.2005

Porter, K. K., Metzger, R. R., & Groh, J. M. (2007). Visual- and saccade-related signals in the primate inferior colliculus. *Proceedings of the National Academy of Sciences, 104*(45), 17855. doi:10.1073/pnas.0706249104

Rayleigh, L. (1907). XII. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 13*(74), 214-232.

Ress, D., & Chandrasekaran, B. (2013). Tonotopic Organization in the Depth of Human Inferior Colliculus. *Frontiers in Human Neuroscience, 7*(586). doi:10.3389/fnhum.2013.00586

Robinson, D. A. (1963). A Method of Measuring Eye Movemnent Using a Scieral Search Coil in a Magnetic Field. *IEEE Transactions on Bio-medical Electronics, 10*(4), 137-145. doi:10.1109/TBMEL.1963.4322822

Ryan, A., & Miller, J. (1977). Effects of behavioral performance on single-unit firing patterns in inferior colliculus of the rhesus monkey. *Journal of Neurophysiology, 40*(4), 943-956.

Ryan, A., & Miller, J. (1978). Single unit responses in the inferior colliculus of the awake and performing rhesus monkey. *Experimental Brain Research, 32*(3), 389-407. doi:10.1007/BF00238710

Sasaki, R., Anzai, A., Angelaki, D. E., & DeAngelis, G. C. (2020). Flexible coding of object motion in multiple reference frames by parietal cortex neurons. *Nature Neuroscience, 23*(8), 1004-1015. doi:10.1038/s41593-020-0656-0

Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing Familiar Objects in Real World Scenes: The Role of Temporal Cortical Neurons in Natural Vision. *The Journal of Neuroscience, 21*(4), 1340. doi:10.1523/JNEUROSCI.21-04-01340.2001

Spezio, M. L., Keller, C. H., Marrocco, R. T., & Takahashi, T. T. (2000). Head-related transfer functions of the Rhesus monkey. *Hearing Research, 144*(1), 73-88. doi:https://doi.org/10.1016/S0378-5955(00)00050-2

Stebbins, W. C., Green, S., & Miller, F. L. (1966). Auditory Sensitivity of the Monkey. *Science, 153*(3744), 1646. doi:10.1126/science.153.3744.1646-a

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*(4), 255-266. doi:10.1038/nrn2331

Thompson, S. P. (1882). LI. On the function of the two ears in the perception of space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 13*(83), 406-416.

Versnel, H., Zwiers, M. P., & van Opstal, A. J. (2009). Spectrotemporal Response Properties of Inferior Colliculus Neurons in Alert Monkey. *The Journal of Neuroscience, 29*(31), 9725. doi:10.1523/JNEUROSCI.5459-08.2009

Vinck, M., Batista-Brito, R., Knoblich, U., & Cardin, Jessica A. (2015). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron, 86*(3), 740-754. doi:https://doi.org/10.1016/j.neuron.2015.03.028

Wallisch, P. (2014). Chapter 21 - Neural Decoding I: Discrete Variables. In P. Wallisch, M. E. Lusignan, M. D. Benayoun, T. I. Baker, A. S. Dickey, & N. G. Hatsopoulos (Eds.), *MATLAB for Neuroscientists (Second Edition)* (pp. 329-336). San Diego: Academic Press.

Weinberger, N. M. (1995). Dynamic Regulation of Receptive Fields and Maps in the Adult Sensory Cortex. *Annual Review of Neuroscience, 18*(1), 129-158. doi:10.1146/annurev.ne.18.030195.001021

Willett, S. M., & Groh, J. M. (2020). Multiple sounds degrade the frequency representation in monkey inferior colliculus. *bioRxiv*, 2020.2007.2003.187021. doi:10.1101/2020.07.03.187021

Willett, S. M., Groh, J. M., & Maddox, R. K. (2019). Hearing in a "Moving" Visual World: Coordinate Transformations Along the Auditory Pathway. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes: The Auditory Perspective* (pp. 85-104). Cham: Springer International Publishing.

Zhong, X., & Yost, W. A. (2017). How many images are in an auditory scene? *The Journal of the Acoustical Society of America, 141*(4), 2882-2892. doi:10.1121/1.4981118

Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple Object Response Normalization *in Monkey Inferotemporal Cortex. The Journal of Neuroscience, 25(36), 8150. doi:10.1523/JNEUROSCI.2058-05.2005*

Zwiers, M. P., Versnel, H., & Van Opstal, A. J. (2004). Involvement of Monkey Inferior Colliculus in Spatial Hearing. The Journal of Neuroscience, 24(17), 4145. doi:10.1523/JNEUROSCI.0199-04.2004

Biography

Shawn Willett first became interested in the study of brains during a neurophysiology lecture in his junior year psychology course at Plano Senior High School. In 2009 he went on to pursue the study of both neuroscience and psychology at the University of Texas at Dallas. During his tenure at UTD he completed an honors thesis investigating potential therapies for anxiety disorders in the lab of Christa McIntyre, under the tutelage of David Peña. In 2013 he received two bachelor of science degrees from the University of Texas at Dallas, one in neuroscience and the other in psychology and took a lab managerial position in the lab of Christa McIntyre. He entered the cognitive neuroscience admitting program at Duke University in 2014, joined the lab of Jennifer Groh in 2015, affiliated with the Department of Neurobiology in 2016 and completed his PhD work in 2020. He was awarded the Duke Chancellor's Award in 2014 and his PhD work was supported by NIDCD grants awarded to Jennifer Groh. He accepted a post-doctoral position at the University of Pittsburgh in the lab of Patrick Mayo.