Neural coding of time-varying interaural time differences and its relation to perception

by

Nathaniel J. Zuk

B.S. Biomedical Engineering
University of Rochester, 2011

Submitted to the Harvard-MIT Program in Health Sciences and Technology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2016

Signature of author ……
Harvard-MIT Program in Health Sciences and Technology
September 2nd, 2016

Certified by ……...……...
Bertrand Delgutte, PhD
Professor of Otolaryngology, Harvard Medical School
Thesis Supervisor

Accepted by ……...……..
Emery N. Brown, MD, PhD
Director, Harvard-MIT Program in Health Sciences and Technology
Professor of Computation Neuroscience and Health Sciences and Technology

Neural coding of time-varying interaural time differences and its relation to perception

by

Nathaniel J. Zuk

Submitted to the Harvard-MIT Program in Health Sciences and Technology
on September 2, 2016 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy
in Speech and Hearing Bioscience and Technology

ABSTRACT

In natural environments, sounds are often not static. Usually, moving objects require the most attention, e.g. for identifying the presence and direction of a moving vehicle, or detecting and tracking the trajectory of a predator or prey. Faster time-varying location cues can occur in acoustic environments containing many spatially distributed sound sources, like at a cocktail party. In this case, we can identify the locations of the sources by "glimpsing" at short-duration localization cues when the sound energy from one source dominates the mixture. Even faster time-varying spatial cues result from reverberation in an echoic environment and we perceive them as spatially diffuse. We qualitatively perceive motion, a cocktail party, and reverberation differently, and these three percepts are determined by how quickly the spatial cues are moving. How these percepts come about in the auditory system is unknown.

Here, we studied how neurons encode time-varying location cues and how the neural code relates to perception. Our focus was on time-varying interaural time differences (ITD), one of the main cues for localizing sounds in the horizontal plane. We recorded from single neurons in the inferior colliculus (IC) in the auditory midbrain of unanesthetized rabbits. The IC is the site of an obligatory synapse in the auditory pathway and one of the first stages of processing following the initial extraction of spatial cues in the brainstem. We hypothesized that the IC exhibits limitations in its ability to encode time-varying ITD that give rise to these different percepts.

First, we show that IC neurons are more "sluggish" on average at synchronizing to the time-varying ITD than to amplitude modulations presented at a static ITD. Binaural sluggishness has been proposed based on human psychophysics but never validated neurophysiologically in the IC. Second, we show that most neurons are unable to synchronize to the time-varying ITD at speeds where humans no longer perceive fluctuations. Instead, neurons exhibit a change in average firing rate that corresponds to binaural decorrelation of the noise for very fast time-varying ITD, and this may explain the percept of a spatially diffuse sound at these speeds.

We further recorded neural responses to slow-moving ITDs in opposite directions within the range of perceived motion. Using a generalized linear model to parse the neuron's response into ITD-following and direction selectivity components, we show that the

responses of IC neurons are dominated by their ability to follow the ITD more than direction selectivity. In parallel experiments, we asked human participants to either identify the motion direction or detect the slow-moving ITD in the same stimuli and determined the threshold durations for direction identification and for detection for each participant. Direction identification threshold durations were larger than detection threshold durations. We then implemented neural classifiers that either identified the motion direction or detected the slow-moving ITD based on single-neuron responses to the stimuli, and we found that the classifier exhibited duration thresholds that matched human thresholds on both tasks.

Together, these results suggest that temporal limitations of neural responses in the IC may give rise to the limiting speeds of time-varying localization cues where we perceive motion, "glimpse" the position of a source amidst a mixture, and perceive a spatially diffuse background in a reverberant environment.

Thesis Supervisor: Bertrand Delgutte, PhD
Title: Professor of Otolaryngology, Harvard Medical School

Acknowledgements

Thanks to Bertrand Delgutte for his mentorship, constant availability, thoughtful criticisms, and his love of puns that I can relate to.

Thanks to my committee members Josh McDermott, John Assad, Chris Brown, and Steve Colburn for their comments on this work and also for many fun discussions about auditory neurophysiology, psychophysics, sound localization, visual motion perception, future career plans, and gardening. Thanks especially to Josh McDermott, the committee chair, for also letting me use his sound booths for the human psychophysics work.

Thanks to Mitch Day and Yoojin Chung for teaching me the protocols for the neurophysiology work and for giving me advice on statistical analysis and surgical techniques throughout the past five years. Thanks also to the rest of the B lab, Michaël Slama, Brian Buechel, Yaqing Su, Kameron Clayton, Camille Shaw, and Alice Gelman for their many comments that have shaped this work to where it is today.

Thanks to Melissa McKinnon for her help with surgeries, Mike Ravicz for his help with the acoustic setup for neurophysiology, and Ken Hancock for writing the code we use for sound generation and neural recording during the neurophysiology experiments. Thanks also to Mike Kaplan, Stephanie Ventura, and the rest of the ACF for their constant attention to the well being of the rabbits.

Thanks to Uri Eden and Ross Williamson for their thoughts and help on the generalized linear modeling work, and to Ross for the discussions on modeling and spike sorting.

Thanks to James Traer and Kevin Woods for their help with setting up and testing for the psychophysics work. Thanks also to the other members of Josh McDermott's lab for many fun discussions about auditory psychophysics: Dana Boebinger, Sam Norman-Haignere, Alex Kell, Max Seigel, Malinda McPherson, Wiktor Młynarski, Nori Jacoby, and Maddie Cusimano.

Thanks to my SHBT class Sonam Dilwali, Jordan Whitlock, Rachelle Horwitz-Martin, and Koeun Lim for their support during the first year of SHBT and beyond.

And thanks to my parents Susan and William Zuk, my brother Josh Zuk, and to Jasmine Florentine for cheering me on throughout graduate school.

Please forgive me for any omissions.

Table of Contents

Abstract 3
Acknowledgements 5
Chapter 1: General Introduction 7
Chapter 2: Neural coding of time-varying interaural time differences and time-varying amplitude in the inferior colliculus 17
Abstract 17
Introduction 18
Methods 20
Results 35
Discussion 55
Chapter 3: Neural coding of time-varying interaural time differences in the inferior colliculus explains human performance in motion direction identification and binaural gap detection 61
Abstract 61
Introduction 62
Methods 65
Results 86
Discussion 107
Appendix 113
Chapter 4: General Discussion and Future Directions 118
References 124

Chapter 1

General Introduction

Time-varying localization cues in the natural environment:

In natural environments, the most ethologically important objects are usually not static. Detecting and identifying the trajectory of prey or a predator is important for an animal's survival. Similarly, detecting an oncoming car or identifying which way an ambulance is going affects our decision to cross the street or drive through an intersection. How the auditory system makes these inferences, and how important it is ecologically, is still not entirely clear. Nevertheless, we can track the motion with our head, albeit less accurately than vision (Leung et al. 2015). Motion improves our ability to focus on a source when stationary distractors are present (Saberi et al. 2002; Davis et al. 2016), and self motion improves our ability to localize sounds in the periphery (Wightman & Kistler 1999). So the presence of motion does improve our ability to parse an auditory scene.

Time-varying location cues do not solely come from spatially moving objects. When there are many spatially distributed and time-varying sources present, like at a cocktail party, what we hear are localization cues quickly varying in time (Faller & Merimaa 2004). Even in these environments, we can still identify where different sources are coming from (Santala & Pulkki 2011; Yost & Brown 2013). We may do this by catching the short "glimpses" of a sound when its energy is particularly strong (Yost & Brown 2013). When a "target" noise alternates with another "masker" noise and the target and masker are at different locations, the locations discriminable when the target is

presented for only 5 ms, but the discriminability rapidly drops below 20 ms and larger separations are needed (Bernstein et al. 2001; Reed et al. 2016)

Even faster variations in localization cues (below 5 ms) can come from reflections of the sound bouncing off objects or walls in the environment, also known as reverberation. When two clicks are presented quickly in succession (< 5 ms) from two separate locations, listeners hear a fused image, suggesting that our auditory system is adept at interpreting these very fast variations in localization cues as a single source rather than two separate sources (Litovsky et al. 1999). The inability to distinguish between quickly alternating localization cues may also result from spatially diffuse sources of sound like some sound textures (McDermott & Simoncelli 2011), including running water in a nearby river or the noise of a nearby flock of seagulls. Indeed, natural binaural sounds contain variations in interaural phase differences that are beyond those considered ethologically useful (Młynarski & Jost 2014) because the interaction of sound coming from many locations can binaurally decorrelate the incoming sounds. Binaural decorrelation also occurs in reverberant environments. The amount of binaural decorrelation between the two ears can change the spatial width of the image that we perceive (Culling et al. 2001; Whitmer et al. 2012). Thus, the perception of a spatially diffuse sound may be related to binaural decorrelation resulting from localization cues that are varying quickly in time.

Intuitively, because we perceive motion, spatially separated sound sources, and reverberation or spatially diffuse sounds differently, time-varying localization cues should result in three different ethologically relevant percepts that depend upon the speed of the variation. Qualitatively, this is what others have observed for sounds with a time-

varying interaural time difference (ITD) (Licklider 1950; Perrott & Musicant 1977; Grantham & Wightman 1978; Siveke et al. 2008) or a time-varying interaural level difference (ILD) (Grantham 1984). When listeners are presented with a noise or pure tone that has an oscillating localization cue and the speed of that variation is changed, three different percepts are perceived: motion at slow speeds (<8-10 Hz), modulation or "flutter" without motion at fast speeds (10 Hz - 50 Hz), and a spatially diffuse sound at very fast speeds (> 50 Hz). If both a time-varying ITD and a time-varying interaural correlation are varying, fluctuations may be perceived up to 1 kHz (Siveke et al. 2008).

Our ability to hear time-varying binaural cues does not extend to the same high speeds where we can hear time-varying monaural cues (Grantham & Wightman 1978; Grantham & Wightman 1979; Grantham 1984). Monaural amplitude modulation of noise can be detected up to 1 kHz (Viemeister 1979). Additionally, masking of a diotic probe tone persists for longer durations following an instantaneous transition from correlated to anticorrelated noise than for a noise with an instantaneous decrease in sound level (Kollmeier & Gilkey 1990). These results suggest that binaural processing in the auditory system is "sluggish" relative to monaural processing. How the different percepts for time-varying locations cues arise in the auditory system, and how they may differ from the coding of amplitude modulation, is still unclear.

Variations in interaural level differences (ILD) and variations in monaural spectral cues are also important for detecting, discriminating, and identifying the direction of motion (Harris & Sergeant 1971; Blauert 1972). Additionally, people are better at detecting fast modulations in ILD than fast modulations in ITD (Grantham 1984). Thus, the presence of time-varying ILD and time-varying monaural cues

potentially improves motion perception and glimpsing in the natural environment. Nevertheless, the perceptions of motion, fluctuations, and a spatially diffuse sound occur for sounds with where only the ITD is changing without any monaural fluctuations in level. Because our interest is in the temporal limitations of binaural processing, from here on our focus will solely be on time-varying ITD and time-varying interaural correlation.

Temporal limitations in binaural processing:

Where might these limitations arise? ITDs are initially encoded in the superior olivary complex (SOC) (Grothe et al. 2010). At low frequencies most neurons in the SOC encode ITDs of the sound's fine structure, and at high frequencies most neurons in the SOC encode ITDs of on the sound's envelope. The nuclei in the SOC project to the inferior colliculus (IC) in the midbrain, where neurons are sensitive to both fine-structure and envelope ITD (Kuwada & Yin 1983; Yin et al. 1984; Joris et al. 2006; Devore & Delgutte 2010). We can perceive motion and fluctuations resulting from time-varying ITD in both the fine structure (Perrott & Musicant 1977; Licklider 1950) and the envelope (McFadden & Pasanen 1975), so both types of ITD cues are potentially relevant for broadband sounds in the natural environment. Limitations in coding that reduce the detectability of time-varying ITD at high speeds may be present at any of these stages as well as more central stages. The inputs to the SOC are bandpass filtered by the cochlea and lowpass filtered in the auditory nerve, which limits the speed of the both phase and level fluctuations. Beyond the SOC, lowpass filtering via synaptic inputs can further limit the speeds of possible variations in firing rate. If the perceptual limitations of time-

varying ITD arise in the SOC or the IC, evidence of these limitations should be present in the coding abilities of IC neurons.

Interestingly, no evidence of binaural sluggishness has been found in the IC. Joris et al. (2006) measured neurons in the IC of anesthetized cats while presenting broadband noise with a sinusoidally time-varying interaural correlation. Several of the neurons could keep up with fluctuations in interaural correlation up to 1 kHz, well above the modulation frequencies that we can detect (Grantham 1982). Additionally, Shackleton & Palmer (2010) recorded from the IC of anesthetized guinea pigs while playing stimuli where a "target" tone is presented soon after an instantaneous inversion in interaural correlation of a background masking noise. They expected to see a masking effect of the response to the probe tone following this transition, since this masking effect has been observed psychophysically in humans (e.g. Kollmeier & Gilkey 1990). They found instead that neurons could respond to the probe tone even at the shortest time scales measured (minimum of 30 ms). Based on these results, both Joris et al. and Shackleton and Palmer proposed that binaural sluggishness might arise at higher levels of the auditory system above the IC.

In the dorsal nucleus in the lateral lemniscus of gerbils, however, Siveke et al. (2008) found that more neurons significantly synchronized to amplitude modulated noise at frequencies above 64 Hz than to the noise stimuli with oscillating binaural cues. This was true even for their "phasewarp" stimuli with both a time-varying ITD and time-varying interaural correlation, which was detectable at high modulation frequencies similar to amplitude modulation detection. They found that detection based on firing rate could account for the difference in detectability of the "phasewarp" stimulus relative to

other stimuli with oscillating correlation, but performance using firing rates did not match human performance at the highest modulation frequencies (500-1000 Hz).

Comparing neurophysiological limits in time-varying ITD and perceptual limits is also complicated by the fact that so many different thresholds have been observed, and the thresholds particularly depend upon the experimental paradigm. For tasks where participants detect motion relative to a static source or identify motion direction, performance decreases rapidly for stimuli with durations below 100 ms (see Carlile & Leung 2016 for review). When the task is to detect a brief change in the interaural correlation of a noise or "binaural gap", thresholds are even shorter and vary depending upon the correlation of the noise played during the "gap". When a correlated or anticorrelated noise is the "gap" and is flanked by binaurally uncorrelated noise, thresholds are between 20-40 ms (Boehnke et al. 2002). When the uncorrelated noise is used as the "gap" and is flanked by correlated or anticorrelated noise, thresholds can be as low as 3-5 ms (Akeroyd & Summerfield 1999; Boehnke et al. 2002). Other studies have found threshold durations for detectable time-varying binaural ITD and time-varying interaural correlation spanning the range of 15-400 ms (for review see Shackleton & Palmer 2010; Lüddemann et al. 2016). A proper comparison of the neural data to these psychophysical thresholds will clearly depend on the types of stimuli used as well as the questions asked.

Direction selectivity in the auditory system:

There is some evidence that IC neurons are selective to the direction of a time-varying ITD. The first observation of direction selectivity to sound motion in the inferior

colliculus was found in anesthetized cats. Altman (1968) measured responses to a click train with a time-varying ITD. He found numerous neurons whose average firing rates were larger when the click train moved towards 0 μs ITD than when it moved away from 0. Also in cats, Yin and Kuwada (1983) presented binaural beats with carrier frequencies equal to the best frequency of the low-frequency neurons (< 3 kHz). Binaural beats are pure tone stimuli where the interaural phase difference changes linearly from 0 to 360° in once cycle. 14% of the neurons they recorded from fired more for one direction of motion than for the other. Similar results were observed in unanestheized rabbits (Fitzpatrick et al. 2009), even for binaural beat speeds well above those where humans can no longer hear modulation of the binaural beat (Licklider 1950). Spitzer and Semple (1991, 1993) extended these findings in both anesthetized gerbils and cats. Using interaural phase modulated tones, similar to binaural beats but with a narrower range of interaural phases, they found differences in firing rate for different directions of motion. These results were not observed for phase-locking neurons in the superior olive that likely project to the IC, but were observed for non-phase locking neurons in the superior olivary complex and in the IC (Spitzer & Semple 1998). However, the firing rate of neurons may also be suppressed following a transient increase in firing irrespective of what is going on in the stimulus. The suppression in firing is also known as spike-rate adaptation. Others have argued that "direction selective" responses can largely be explained by adaptation rather than direction selectivity per se (Cai et al. 1998; McAlpine et al. 2000; Borisyuk et al. 2002).

While spike-rate adaptation may explain the dynamic responses observed in the IC for time-varying ITD and stimuli with time-varying interaural phases, it does not

discount the presence of direction selectivity. But more importantly, it is still unclear how the patterns of firing may be used to ultimately identify the trajectory of motion. The pattern of firing available in the IC must be used somehow in order to identify the motion direction of a sound source. Humans can do this, and there is some evidence that true direction selectivity for auditory motion may be present in other cortical areas of the brain of monkeys (Ahissar et al. 1992; but see Scott et al. 2009) and humans (Jiang et al. 2014). If direction selectivity in the IC is relevant for perceiving motion direction, we would expect IC neurons to show distinct firing rates for different motion directions, like in the retina (Barlow & Levick 1965; Grzywacz & Amthor 2007). Additionally, a classifier should be able to properly identify the motion direction of a time-varying ITD based on the firing rates of this population of neurons.

Relevance of neural noise:

Very few of the studies listed above have taken into account inherent neural noise as a possible factor for limiting the detectability of a time-varying ITD or the ability to identify its direction. For example, when averaged over a long period of time, a neuron may be appear to significantly synchronize to a noise with a modulating ITD or interaural correlation at speeds that can no longer be perceived, but the strength and phase of the synchronization may vary on a trial-by-trial basis, as has been observed for neural responses to amplitude modulation (Yin et al. 2011). In vision, classification of motion direction based on single neuron responses in the medial temporal lobe has been used to successfully predict direction identification performance in monkeys (Britten et al. 1992). We are only aware of a small of studies that have attempted to classify the motion

direction of sounds based on cortical responses measured with fMRI in humans (Jiang et al. 2014; Jiang et al. 2016). Whether or not it is possible to classify the motion direction of a sound based on single neuron responses seems to be completely unexplored.

Summary of thesis:

In Chapter 2, we compare the firing rates and synchronization strengths of IC neurons in unanesthetized rabbits to time-varying ITD and amplitude modulations at the same set of modulation frequencies. Contrary to previous literature (and our initial hypothesis), we find that neurons in the IC significantly synchronize to higher modulation frequencies for amplitude-modulated noise than for time-varying ITD. This could be a neural correlate for binaural sluggishness. Furthermore, while nearly all neurons are unable to synchronize to the time-varying ITD at high modulation frequencies, they exhibit a change in firing rate that tends towards the rate for binaurally uncorrelated noise. An optimal process using single-neuron responses to detect the dynamic ITD reached its worst performance around 64 Hz, which is close the modulation frequency where human listeners had the most difficulty detecting a time-varying ITD in the classic study by Grantham and Wightman (1978).

In Chapter 3, we examine the ability for IC neurons to encode linearly time-varying ITD ("ITD sweep") within binaurally uncorrelated noise. We then use a generalized linear model to separate the contributions of ITD following, direction selectivity, and spike rate adaptation to the responses of these neurons. We find that the temporal response patterns of neurons in the IC can be described better by their ability to follow the ITD of the sweep than by direction selectivity, suggesting that direction

selectivity is of little to no importance in encoding the ITD sweeps. Furthermore, we assess the performance of an optimal classifier that identifies the motion direction or detects the ITD sweep based on single neuron responses, and we compare the performance of this classifier to human performance by identifying the ITD sweep durations for both tasks where performance is at threshold (75% correct responses). Both the neural classifier and human participants exhibit similar threshold durations for direction identification and detection. Additionally, both neural and perceptual threshold durations for detection are significantly lower than the thresholds for direction identification.

This thesis provides evidence for temporal limitations for binaural processing in the IC that match those proposed in the psychophysics literature. We also provide evidence for distinct thresholds for direction identification and for detection that can partly explain the wide range of threshold durations observed across psychophysics studies (see Shackleton & Palmer 2010; Carlile & Leung 2016; Lüddemann et al. 2016). The results from this work identify possible neural correlates for the three percepts qualitatively observed for time-varying ITD: motion, fluctuation, and spatially diffuse sound.

Chapter 2

Neural coding of time-varying interaural time differences and time-varying amplitude in the inferior colliculus

---

Sound sources in natural environments are usually not static, but our understanding of how the auditory system processes time-varying sound position is still lacking. Here, we directly compared neural responses in the inferior colliculus (IC) of unanesthetized rabbits to broadband noise with time-varying interaural time differences (ITD) with responses to noise with sinusoidal amplitude modulation (SAM) over a wide range of modulation frequencies. Based on prior research, we hypothesized that the IC, one of the first stages to exhibit tuning of firing rate to modulation frequency, might use a common mechanism to encode time-varying information in general. Instead, we found weaker temporal coding for dynamic ITD compared to amplitude modulation and stronger effects of adaptation for amplitude modulation. The differences in temporal coding of dynamic ITD compared to SAM at the single-neuron level could be a neural correlate of "binaural sluggishness", the reduced detectability of time-varying binaural cues at high modulation frequencies, for which a physiological explanation has so far remained elusive. At ITD-variation frequencies of 64 Hz and above where the effectiveness of a temporal code was reduced, noise with a dynamic ITD could still be distinguished from noise with a constant ITD through differences in average firing rate in many neurons, suggesting a frequency-dependent tradeoff between rate and temporal coding of time-varying binaural information.

Introduction:

In natural environments, interaural time differences (ITD), a dominant cue for localizing sounds in the horizontal plane (Wightman & Kistler 1992), usually change over time. Slow changes in ITD produced by a moving sound source or head motion are usually perceived as motion (Blauert 1972; Perrott & Musicant 1977). Fast changes in ITD occur in the presence of multiple spatially distributed sound sources (Faller and Merimaa, 2004; Yost and Brown, 2013), and very fast changes in ITD come from the spatial dispersion of acoustic reflections resulting in reverberation (Litovsky et al. 1999). Clearly our perceptions of motion, spatially distributed sound sources, and reverberation are different, but the time scale at which this change in perception occurs and how it is represented in the brain are still unclear.

The detection of modulations in binaural cues has been studied psychophysically (for review, see: Bernstein 1997; Shackleton & Palmer 2010; Carlile & Leung 2016). In a classic study, Grantham and Wightman (1978) measured the subjects' ability to discriminate broadband noise with a sinusoidally-varying ITD from noise with a static ITD. Performance worsened with increasing modulation frequency up to around 20-50 Hz but then improved at higher modulation frequencies up to 500 Hz. These high-frequency improvements could not be explained by monaural cues alone. Subjects reported detecting the time-varying ITD at high frequencies through a difference in spatial extent of the sound image rather than from a percept of motion. In contrast, monaural amplitude modulations can be detected up to 1-2 kHz with a gradual, monotonic degradation in performance (Viemeister 1979). Grantham and Wightman concluded that the performance at low modulation frequencies may represent the processing of auditory motion, which is "sluggish" relative to monaural processing. At high modulation frequencies a secondary binaural mechanism may be at play.

A neural correlate and explanation for binaural sluggishness, however, remains elusive. Joris et al. (2006) examined the sensitivity of neurons in the inferior colliculus (IC) of anesthetized cats to time-varying interaural correlations of broadband noise. They found that the upper frequency limit of phase locking to the sinusoidal correlation was much higher than the detection limits observed psychophysically and close to the neural limits for phase locking to monaural amplitude modulation. Siveke et al. (2008) showed high phase-locking limits for neurons in the dorsal nucleus of the lateral lemniscus for "phasewarp" broadband stimuli having both a time-varying ITD and a time-varying interaural correlation. Shackleton and Palmer (2010) also found shorter time constants in the IC of anesthetized guinea pigs in response to tone-in-noise stimuli with abrupt transitions in interaural phase than expected from psychophysical studies of binaural detection. These studies have led to the suggestion that the sluggishness of binaural processing may arise at higher stages of the auditory system. They also raise the possibility that time-varying spatial cues and amplitude modulation may be processed similarly in the IC. Other studies have demonstrated that amplitude modulation can affect the range of firing rates and tuning to binaural beats (d'Angelo et al 2003; Sterbing et al 2003), source azimuth, and distance (Kuwada et al 2012). Dietz et al. (2014) examined the effects of amplitude modulation on the localization of binaural beats at high modulation frequencies, but only when the amplitude modulation and the binaural beats were at the same modulation frequency. The sensitivity of IC neurons to time-varying ITD has not been directly compared with sensitivity to amplitude modulation in the same neurons, which would be a more direct test of binaural sluggishness (see Fitzpatrick et al (2009) for this comparison in the auditory cortex).

Here, we recorded from neurons in the IC of unanesthetized rabbits in response to broadband noise with a time-varying ITD, and compared with responses to sinusoidally

amplitude modulated (SAM) broadband noise for the same modulation frequencies. Our results suggest that the coding mechanisms for time-varying ITD and SAM are largely distinct, which could be a correlate of binaural sluggishness. We also suggest that a rate code might explain the detectability of rapidly varying ITDs within the range of reverberation.

Methods:

Dynamic ITD stimuli

Stimuli with time-varying ITD (henceforth "dynamic ITD" stimuli) were derived from a broadband noise (0-50 kHz) carrier synthesized with a 100 kHz sampling rate. The noise was first segmented into 10.9 ms windows with 50% overlap between successive windows. In each window, the signal was upsampled to 1 MHz and then resampled with the desired time-varying delay at a resolution of 1 μs (**Figure 1a**). Only a half segment starting 0.9 ms after the window onset was used in order to avoid edge effects associated with upsampling. Concatenating these half segments created broadband noise with a time-varying delay relative to the original sound. A dynamic ITD was produced by presenting the original broadband noise signal in the ear contralateral to the recording site and the processed signal with time-varying delay to the ipsilateral ear.

With this method, the dynamic ITD could follow any arbitrary ITD trajectory. We used triangular modulations, which produce constant rates of change in ITD in each direction. The accuracy of the ITD trajectory was verified by computing the running cross-correlation function between the signals at the two ears using half-overlapping 2 ms Hanning windows (**Figure 1b**). We chose triangular modulations to facilitate comparison with previous neurophysiological

studies that used pure tone stimuli with linearly-varying ITD (Spitzer & Semple 1993; McAlpine et al. 2000).

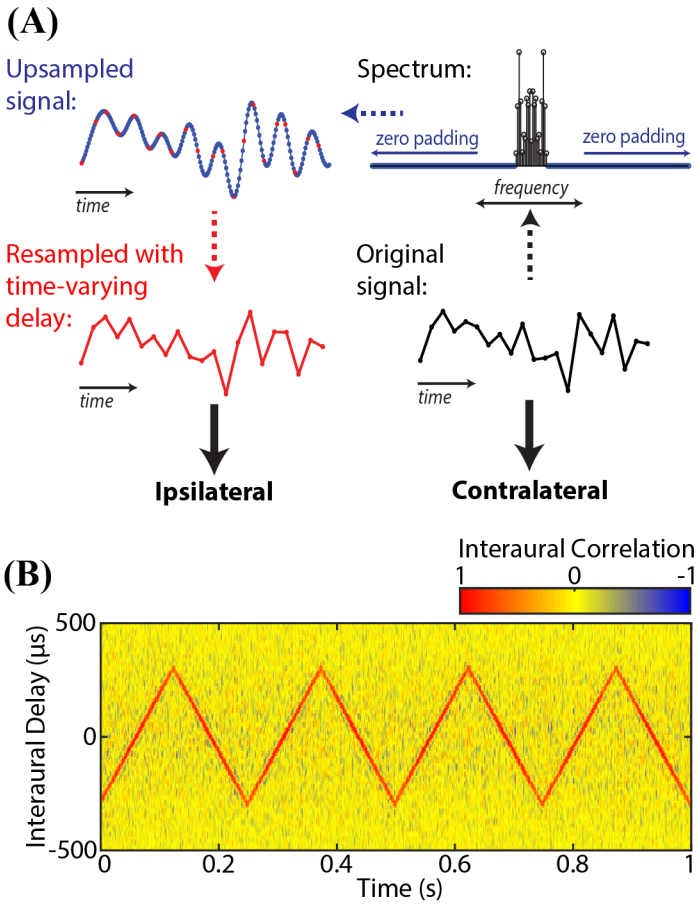


Figure 1: (A) Creation of a time-varying interaural delay in a window of broadband noise. The original signal was zero-padded in the frequency domain to ten times its original length resulting in a 10x upsampled signal. The upsampled signal was then resampled with a time-varying delay. The original noise was presented through the contralateral channel, and the delayed noise was presented through the ipsilateral channel, resulting in a dynamic ITD. (B) Cross-correlation of the two channels for 1 second of a 4 Hz dynamic ITD stimulus as a function of time. For the stimulus shown in this figure the magnitude of the spectrum below 50 Hz and above 17 kHz was set to zero. Each cross-correlation was performed within 2 ms cosine windows. Neighboring windows were half-overlapping.

Subjectively, listening to dynamic ITD stimuli over headphones produces different percepts depending on the rate of modulation (Grantham & Wightman 1978; Siveke et al. 2008). Up to ~2 Hz the interaural location of the broadband noise can be explicitly tracked as it moves between the two ears. Between 2-8 Hz, the dynamic ITD stimulus is perceived to move back and forth between the ends of its trajectory. Above 8 Hz, the noise no longer sounds like it is

moving, but a time-varying quality of the sound (a faint beating or flutter) can still be perceived. Between 32-500 Hz, the perception of flutter fades, and the noise is perceived to widen spatially. Around 500-1000 Hz the noise sounds binaurally uncorrelated, as if independent samples of noise were presented at each ear. Because these high modulation frequencies produce spectral fluctuations that are audible monaurally, we did not examine frequencies above 256 Hz in this study.

In an early version of the dynamic ITD stimulus, the time-varying delay was created without upsampling by shifting the broadband noise in one ear by one sample point for every 10 μs ITD step in the triangular ITD trajectory. These early dynamic ITD stimuli were used for 12 neurons. For 4 neurons, the shifting was done only in one ear. For these stimuli, a clear "buzzing" could be perceived monaurally for frequencies above 100 Hz, so we restricted our dynamic ITD and SAM recordings to 64 Hz and below. For the 8 remaining neurons, the shift was performed in one ear or the other depending on whether the ITD was increasing or decreasing, so that the time-varying delay in each ear was either decreasing or constant. No monaural effects were heard for these stimuli at 128 and 256 Hz so these frequencies were used for recordings.

A major goal of this study was to compare responses to dynamic ITD stimuli with responses to sinusoidal amplitude modulation (SAM) in the same neurons. Compared to sinusoidal modulations, triangular modulations contain additional frequency components at integer multiples of the fundamental modulation frequency. These components arise from discontinuities in the rate of change of ITD every half period. In principle, a neuron could respond to the higher frequency components of the modulation, making comparison to responses to SAM stimuli hard to interpret. Indeed, Spitzer and Semple (1993) found slightly sharper

peaks in firing rate for sinusoidal motion than for triangular motion in gerbil IC neurons, which they attributed to the accelerations of ITD that are present in the sinusoidal modulation. To address the difference, we separately recorded the responses of 9 neurons to both triangularly-modulated and sinusoidally-modulated dynamic ITD and compared their responses and looked primarily at temporal and adaptation metrics.

Animal preparation

All surgical and experimental procedures using Dutch-belted rabbits were approved by the Animal Care Committee of Massachusetts Eye and Ear. Three female rabbits were used in these experiments. Surgeries for implantation of a head bar and cylinder and the subsequent craniotomy were as described previously (Day et al. 2012; Slama & Delgutte 2015). Briefly, under ketamine/xylazine anesthesia the rabbit's skull was exposed, and a stainless steel bar and cylinder were fixed to the skull using dental cement. The cylinder was positioned along bregma in order to provide access to either IC. After a week of post-operative care, the rabbit was habituated to the experimental setup over a 7-8 day period. A craniotomy was then made under isoflurane anesthesia, and recording sessions began 2-3 days afterwards. The rabbit was monitored by a video camera during the recording sessions which lasted 2.5 hours and took place over a period of 8-12 months.

Experimental Preparation

All neural recordings were done in a soundproof chamber. The rabbit was restrained in a spandex sleeve and its head clamped through the head bar during the recording sessions. Sound was delivered via custom made ear inserts, either composed of a vinyl polysiloxane impression

material (DENTSPLY International Reprocil) or constructed from polyethylene foam inserts (Comply P-Version Foam Tips). Stimuli were generated in MATLAB, converted into an analog signal using a 24-bit digital-to-analog converter (National Instruments PXI-4461), and delivered to speakers (Beyerdynamic DT-48) attached to the ear inserts via tubes. In-ear calibrations were performed using an Etymotics ER-7C probe tube microphone. The probe tube was positioned in each ear insert such that the end of the probe tube was 1.5-2 cm from the tympanic membrane (Kuwada et al. 1987). The in-ear calibrations were used to design an inverse digital filter that was applied to all broadband stimuli in order to produce a flat spectrum over the range 50 Hz to 17 kHz at the tip of the probe tube.

Electrophysiology

Most neurons were recorded using a flexible polyimide linear microelectrode array with 4 platinum/iridium-based contacts (Microprobe). The impedance of each contact was between 0.2-1 MΩ and the contacts were spaced 150 μm apart. The electrode was advanced through the dura into the IC using a micropositioner (Kopf Model 650). Signals from the microelectrode were sent through a Plexon head stage and PBX preamplifier before being digitized at 50 or 100 kHz for computer processing. In early experiments, 4 neurons were isolated with epoxy-insulated tungsten microelectrodes (A-M Systems). The procedure for processing signals from these electrodes has been described previously (Day et al. 2012).

Single neurons in the IC were detected using a search stimulus consisting of two 200 ms bursts of broadband noise at 50-65 dB SPL with ITDs of 0 and +300 μs (contralateral leading), respectively. Once a neuron with stable spike shape and amplitude was isolated, spike times were measured by a manually set voltage threshold crossing and stored for later analysis. During

post-hoc analysis, neurons were excluded if > 1% of the interspike intervals were less than 1 ms in order to ensure that they were single neurons.

Stimulation protocols: Static stimuli:

Once a neuron was isolated, a static ITD tuning curve was measured: 300 ms broadband noise bursts were presented every 600 ms with ITDs ranging from –900 to +900 μs in increments of 100 μs. Each ITD was presented 10 times and the ITDs were randomly interleaved. The sound level was usually 40-60 dB SPL, which was typically 20-35 dB above rate-level threshold for broadband noise. A neuron was included only if its static ITD tuning was statistically significant by a one-way ANOVA of the spike counts with ITD as factor ($p < 0.001$) (Hancock et al. 2010; Day et al. 2012; Chung et al. 2016). We also measured rate-level curves using 300 ms broadband noise bursts over 5-70 dB SPL in 5 dB steps. Each sound level was presented 10 times in random order. The noise for rate-level characterization was usually set at the ITD exhibiting the maximal firing rate within the range of ITDs chosen for the dynamic ITD stimulus (see below). Frequency tuning was characterized either by an automatic threshold tracking procedure with 100 ms tone pips (Kiang & Moxon 1974), or, when this method failed (e.g. due to suppressive responses), by measuring firing rate as a function of frequency for a 100 ms tone at 10-20 dB above rate threshold. For simplicity, we use the term "best frequency (BF)" to refer either to the frequency exhibiting the lowest threshold with the tracking procedure or to the frequency eliciting the maximum firing rate in the iso-level paradigm. Based on their sharp frequency selectivity, and an orderly increase in best frequency with electrode depth on most penetrations, it is likely that most of the neurons we recorded from were located in the central nucleus of the IC (Aitkin et al. 1972; 1975).

Stimulation protocols: Dynamic stimuli:

Both SAM noise and dynamic ITD stimuli were presented at 40-60 dB SPL, at the same level as for measuring static ITD tuning. For both stimuli, modulation frequency was varied over 2-256 Hz with octave spacing. Each modulation frequency was presented 5 times for 5 seconds with 4 ms sine-square onset and offset ramps and 1 second of silence in between presentations. For dynamic ITD stimuli, the range of ITD variation was always ±300 μs around a center ITD so as to roughly encompass the physiological ITD range for Dutch-belted rabbits (Day et al. 2012). The center ITD was chosen separately for each neuron so that the ITD range would be located within a portion of the static ITD tuning curve where firing rate varied nearly monotonically with ITD. The blue arrow in **Figure 2a** illustrates the ITD range of the dynamic ITD stimulus relative to the static ITD tuning curve in an example neuron. This choice was intended to elicit a temporal pattern of response that would resemble the expected response pattern to SAM noise, namely a unimodal distribution over the course of the modulation period. Across our neuronal sample, the center of the ITD range of the dynamic ITD stimulus was chosen between –600 to +600 μs, but most often it was either 0 μs (22/66 neurons) or +300 μs (contralateral leading, 19 neurons). With these choices, the firing rate increased nearly monotonically as the ITD increased in 51/66 neurons, and the firing rate decreased with increasing ITD in the remaining 15 neurons.

In addition to the dynamic ITD stimuli with different modulation frequencies, two "unmodulated" stimuli of equal duration and repetition rate were also presented: one with a static ITD at the center of the dynamic ITD trajectory, and one consisting of binaurally uncorrelated

noise. The presentations of all dynamic ITD and unmodulated stimuli were randomly interleaved.

SAM broadband noises were presented with an ITD located at the point within the dynamic ITD's trajectory that produced the maximum firing rate to static ITD stimuli (red arrow in **Figure 2a**). SAM stimuli with different modulation frequencies were randomly interleaved with an unmodulated (constant amplitude) broadband noise at the same ITD.

When time permitted, static ITD tuning was measured a second time after testing all dynamic stimuli to verify stability of recording. Data were excluded if the average difference between the new static ITD tuning curve and the original one was >20% of the original ITD curve's firing rate maximum.

Data analysis

To compute temporal modulation transfer functions (tMTFs) for dynamic ITD and SAM stimuli, the strength of phase locking to each modulation frequency was quantified using the vector strength (VS) (Goldberg & Brown 1969) computed over the entire 5-s stimulus duration excluding the onset response. The duration of the onset response, which ranged from 10-200 ms, was determined individually for each neuron by visually inspecting the raster plot. The analysis interval always consisted of an integer number of modulation periods. Statistical significance of the VS was determined using the Rayleigh test (Mardia & Jupp 2000) with $p < 0.001$ criterion. The "synchronization limit" was determined for each neuron and each stimulus type (SAM or dynamic ITD) by linearly interpolating the Rayleigh statistic between the highest modulation frequency where VS was significant and the next highest frequency to find the frequency intercept corresponding to p=0.001 (**Figures 2c, 3c**). In addition, the temporal best modulation

frequency (tBMF) was determined by quadratic interpolation around the maximum of the tMTF and the two adjacent points. Lastly, the 6 dB cutoff frequency was determined by linear interpolation at the frequency where VS was equal to half the maximum. If the synchronization limit or 6 dB cutoff frequencies were beyond the range of measured values, they were excluded from population analysis.

In order to obtain rate modulation transfer functions (rMTF) for SAM and dynamic ITD stimuli, the average firing rate was computed for each modulation frequency over the stimulus duration excluding the first 500 ms of each 5-s trial. Standard error bars (**Figure 2e** for example) were computed by dividing the 5 stimulus trials × 4.5 s/trial into 45 segments of length 500 ms. We did this because a 500-ms stimulus duration is often used in psychophysical experiments and we wanted to define average firing rate in a way that would be appropriate for ROC analysis (see below). The average firing rate during the 1 second in between stimuli was also computed and is labeled "Stimulus Off" in the plots.

We classified the shapes of the rMTFs for SAM relative to their firing rate to the static noise based on classification schemes used in previous studies (Carney et al. 2015; Kim et al. 2015). A neuron was "band-enhanced" if its maximum firing rate was 1 SD above the distribution of firing rates for the static noise, a neuron was "band-reject" if its minimum firing rate was 1 SD below the static noise rates, and a neuron was "complex" if it passed both criteria. Neurons that did not pass either of these criteria were classified as "flat". We also computed the minimum and the maximum modulation frequencies of the rMTFs using quadratic interpolation.

We quantified the strength of rate coding of modulation frequency by the mutual information (MI) in bits:

$$MI = \sum_{s \in S} \sum_{r \in R} p(s,r) log_2 \frac{p(s,r)}{p(s)p(r)}$$

where *p(s)* is the probability distribution of modulation frequencies (assumed to be uniform), *p(r)* is the probability distribution of spike counts across all modulation frequencies, and *p(s,r)* is the joint distribution of spike count and modulation frequency. Bias in MI estimates due to low sampling sizes was removed using a bootstrap resampling method (Chase & Young 2005; Devore & Delgutte 2010). MI is always bounded upward by the entropy of the stimulus set, which was 3 bits since there were 8 equally probable modulation frequencies. Reaching this bound would mean that each modulation frequency is uniquely coded by a particular spike count that is identical on every trial.

Analysis of adaptive effects

Previous studies of responses of IC neurons to tonal stimuli with time-varying ITD (Spitzer & Semple 1998; McAlpine et al. 2000; Ingham & McAlpine 2004) have suggested that firing rate adaptation may underlie some of the observed selectivity to the direction of ITD motion. If an IC neuron shows firing rate adaptation, one would expect the response to a dynamic ITD stimulus to occur earlier and show a sharper rising edge when the stimulus moves from an ITD evoking a low firing rate in the static condition to an ITD evoking a large firing rate than when the stimulus moves in the opposite direction. To test this prediction, two metrics were derived from temporal response patterns (period histograms) to dynamic ITD stimuli: the mean response phase and the skewness of the spike distribution.

The mean response phase to SAM and dynamic ITD stimuli was computed from the period histogram for each modulation frequency as described by Goldberg and Brown (1969). A straight line was fit to the mean phase as a function of modulation frequency to estimate the mean group delay and phase intercept (**Figure 2d**). Because of issues with phase unwrapping

resulting from the wide spacing between tested modulation frequencies, the fitting was done in three steps: 1) the line was first fit to the lower-frequency phases that increased monotonically with modulation frequency, 2) integer numbers of cycles were added or subtracted to the phases at the higher modulation frequencies to minimize their distance to the fitted line, 3) the line was refit to the phases at all frequencies for which VS was statistically significant. Typically this method produced good fits: the median R^2 was 0.996 for dynamic ITD stimuli and 0.992 for SAM. Occasionally, however, R^2 could be as low as 0.7. The latter usually occurred when the mean phases were not strictly increasing at low modulation frequencies. If such cases occurred (16/116 responses to either SAM or dynamic ITD stimuli), the set of frequencies used in step (1) was selected based on visual identification of the most appropriate phases for initially fitting the straight line, usually between 2 and 32 Hz.

The slope of the fitted line represents the mean group delay of response, reflecting neural transmission delays as well as the cochlear traveling wave delay. The 0-Hz intercept of the fitted line represents the neuron's preferred phase after removing the group delay. Our reference for phase intercepts was always the phase at which the excitatory drive is maximum, i.e. the phase of maximum amplitude for SAM stimuli, and the ITD within the trajectory evoking the maximum firing rate in the static condition for dynamic ITD stimuli. With this convention, the phase intercept is expected to be zero in the absence of dynamic processes such as adaptation. Adaptation would cause the response phase to lead relative to this reference, resulting in a negative phase intercept.

A second metric derived from the period histogram that may serve as an indicator of adaptation is circular skewness (Pewsey 2004):

$$skewness = \frac{1}{N}\sum_{n=1}^{N} sin\,2(\theta_n - \bar{\theta})$$

where θ_n is the phase of each spike and $\bar{\theta}$ is the mean phase (**Figure 2f**). Skewness can range from –1 to +1. In the absence of dynamic processes such as adaptation, the temporal firing patterns to dynamic ITD stimuli would be symmetric around the mean phase so the skewness would be 0.

Receiver operating characteristic (ROC) analysis:

We used receiver operating characteristic (ROC) analysis (Green & Swets 1988) to quantify the performance achievable by an ideal observer of single-neuron responses in two discrimination tasks: (1) discrimination of a dynamic ITD stimulus from a static ITD stimulus located at the center of the dynamic ITD's trajectory; (2) discrimination of a SAM stimulus from unmodulated noise (see also Nelson & Carney 2007; Carney et al. 2014). For each task, performance based on rate and temporal codes was compared. We describe the ROC analysis for the dynamic ITD detection task; methods for the SAM detection task are identical except for the stimuli being discriminated.

For ROC analysis of rate coding, spike counts from each neuron were computed in 500 ms response bins, excluding the first 500-ms of each 5-s trial, resulting in a distribution of 45 spike counts for each stimulus. This was done for dynamic ITD stimuli at each modulation frequency and also for the static ITD stimulus. In many cases, the mean firing rate to dynamic ITD stimuli is higher than the rate for the static ITD stimulus. Assuming that the dynamic ITD is detected when the spike count on a given trial exceeds a certain spike count criterion, the probability that a dynamic ITD stimulus is correctly detected is the fraction of trials in which the spike counts exceeds the criterion. Similarly, the fraction of trials in which the spike count for the static ITD stimulus exceed the criterion represents the probability of "false alarms" (incorrect

detection of the static ITD stimulus). The ROC curve is obtained by plotting the probability of correct detection against the probability of false alarm as the spike count criterion is systematically varied from the lowest to the highest spike count. In psychophysics, the area under the ROC curve is equal to the percent correct in a 2-interval, 2-alternative-forced-choice detection task (Green & Swets 1988). Thus, we refer to the area under the neural ROC curve as the neural percent correct score. Because an adaptive threshold procedure widely used in psychophysics (Levitt 1971) converges to 71% correct probability, we also use this value as a threshold for neural detectability.

The above reasoning assumes that the dynamic ITD stimulus evokes a higher firing rate than the static ITD stimulus, which is indeed what is observed in a majority of cases. However, in some cases, the dynamic ITD stimulus evoked a lower firing rate than the static ITD stimulus, resulting in a neural percent correct score lower than 50%. Since a decrease in firing rate as well as an increase provides a cue for detection that would be available to an ideal observer, we consider that a neuron detects the dynamic ITD if its percent correct score is either greater than 71% or less than 29%.

For ROC analysis based on a temporal code, the VS and mean response phase were computed for each of the same 45 response bins as in the rate analysis. ROC curves were constructed from the distribution of "phase-projected vector strength" (VS_{pp}), defined by the following equation (Yin et al. 2011):

$$VS_{pp} = VS_i \cos(\theta_i - \bar{\theta})$$

where VS_i is the vector strength for the i^{th} response trial, θ_i is the mean response phase for that trial, and $\bar{\theta}$ is the mean response phase averaged across all trials. By projecting the complex vector strength (magnitude and phase) for each trial along the direction defined by the mean

Figure 2:

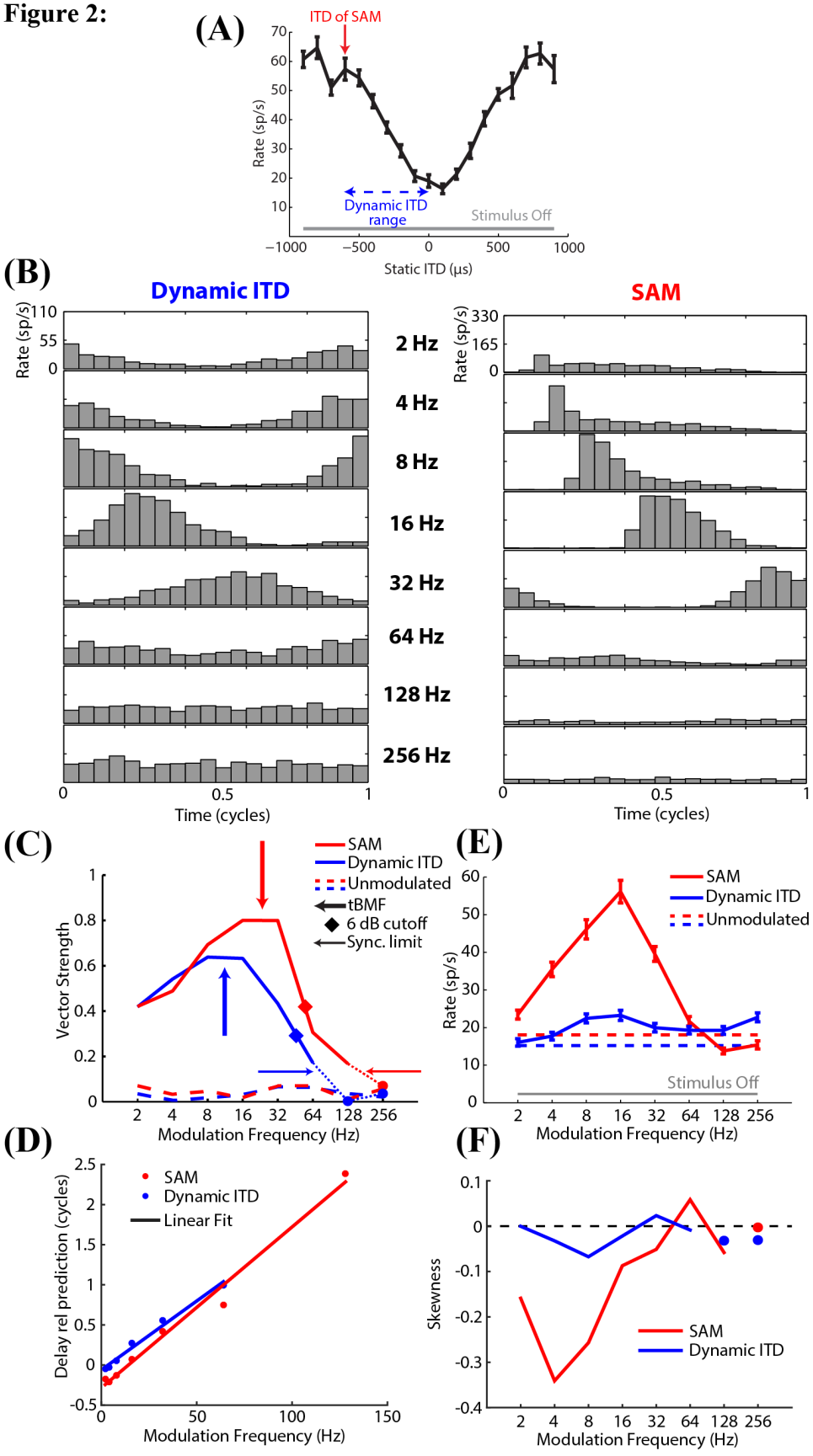

(A)
ITD of SAM
Rate (sp/s)
Dynamic ITD range
Stimulus Off
Static ITD (μs)
(B)
Dynamic ITD
SAM
Rate (sp/s)
2 Hz
4 Hz
8 Hz
16 Hz
32 Hz
64 Hz
128 Hz
256 Hz
Time (cycles)
(C)
SAM
Dynamic ITD
Unmodulated
tBMF
6 dB cutoff
Sync. limit
Vector Strength
Modulation Frequency (Hz)
(D)
SAM
Dynamic ITD
Linear Fit
Delay rel prediction (cycles)
Modulation Frequency (Hz)
(E)
SAM
Dynamic ITD
Unmodulated
Rate (sp/s)
Stimulus Off
Modulation Frequency (Hz)
(F)
Skewness
SAM
Dynamic ITD
Modulation Frequency (Hz)

Figure 3:

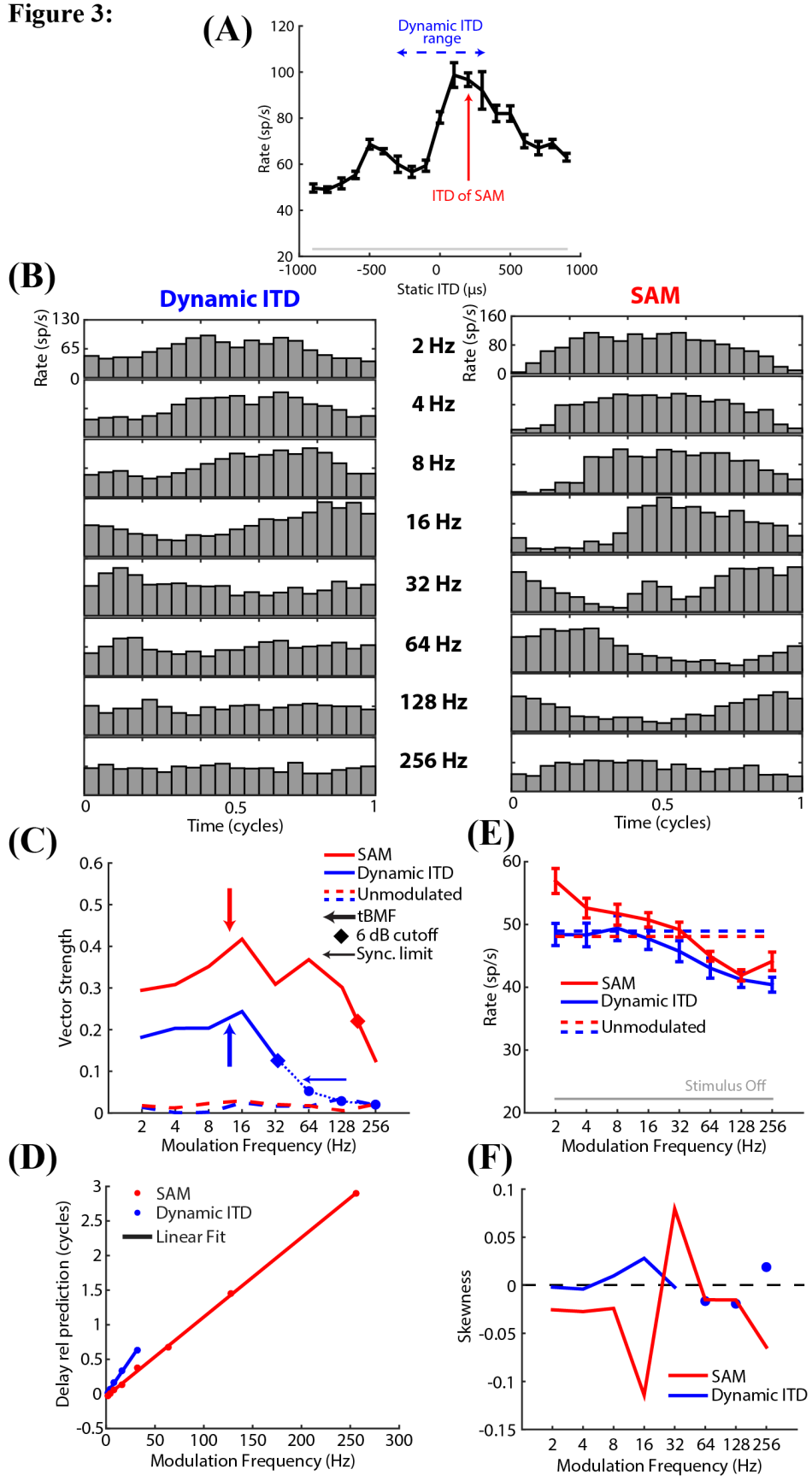

(A)
Dynamic ITD range
ITD of SAM
Rate (sp/s)
Static ITD (μs)
(B)
Dynamic ITD
SAM
2 Hz
4 Hz
8 Hz
16 Hz
32 Hz
64 Hz
128 Hz
256 Hz
Rate (sp/s)
Time (cycles)
(C)
SAM
Dynamic ITD
Unmodulated
tBMF
6 dB cutoff
Sync. limit
Vector Strength
Moulation Frequency (Hz)
(D)
SAM
Dynamic ITD
Linear Fit
Delay rel prediction (cycles)
Modulation Frequency (Hz)
(E)
SAM
Dynamic ITD
Unmodulated
Stimulus Off
Rate (sp/s)
Modulation Frequency (Hz)
(F)
SAM
Dynamic ITD
Skewness
Modulation Frequency (Hz)

Figure 2: (A) Static ITD tuning curve measured in an example neuron (BF = 3512 Hz). The dashed blue arrow designates the ITD range of the dynamic ITD stimulus presented for this neuron. The solid red arrow designates the ITD of the SAM stimulus. (B) Period histograms for the dynamic ITD and SAM for this neuron. (C) tMTFs for SAM (red) and dynamic ITD (blue). The dotted lines connect to vector strengths that are not significant, marked by circles ($p > 0.001$ by the Rayleigh test with Bonferroni correction). Thick arrows mark the tBMFs, diamonds mark the 6 dB cutoff, and thin arrows mark the synchronization limits. For this neuron, both the tBMF and the synchronization limit are higher for SAM than for dynamic ITD (Sync. Limit: SAM = 169 Hz, Dynamic ITD = 82 Hz; Peak VS: SAM = 0.81, Dynamic ITD = 0.65; 6 dB cutoff: SAM = 56 Hz, Dynamic ITD = 43 Hz; tBMF: SAM = 23 Hz, Dynamic ITD = 11 Hz). (D) rMTFs for SAM (red) and dynamic ITD (blue) with STE. The average rates to the unmodulated stimuli for both SAM and dynamic ITD are shown as dashed lines. For this neuron, the MI was higher for SAM than dynamic ITD (SAM = 0.95, Dynamic ITD = 0.10). (E) Mean phases for SAM (red) and dynamic ITD (blue) along with the linear fit (SAM: group delay = 20.1 ms, phase-intercept = -0.28 cycles, $R^2 = 0.98$; Dynamic ITD: group delay = 17.2 ms, shifted phase-intercept = -0.06 cycles, $R^2 = 0.98$). Only phases where phase-locking was significant are shown. (F) Skewness as a function of modulation frequency for SAM (red) and dynamic ITD (blue). Skewness values at modulation frequencies with non-significant phase-locking are labeled with dots.

Figure 3: Example a low BF neuron (BF = 842 Hz). Figures are plotted in the same way as Figure 2. Sync. Limit: SAM was beyond measured values, Dynamic ITD = 52 Hz; 6 dB cutoff: SAM = 184 Hz, Dynamic ITD = 35 Hz; tBMF: SAM = 15 Hz, Dynamic ITD = 14 Hz; Peak VS: SAM = 0.42, Dynamic ITD = 0.25. SAM MI = 0.42; Dynamic ITD MI = 0.24. SAM: group delay = 11.5 ms, shifted phase-intercept = -0.04 cycles, $R^2 = 0.997$; Dynamic ITD: group delay = 20.5 ms, shifted phase-intercept = -0.09 cycles, $R^2 = 0.9996$.

---

response phase, VS_{pp} alleviates the ill-conditioned nature of the unprojected VS for low spike counts (for example, a trial with just one spike always yields a vector strength of 1 regardless of the spike phase). VS_{pp} was computed for each dynamic ITD stimulus and for the static ITD stimulus to construct an ROC curve and determine a neural percent correct score for each modulation frequency in the same way as for the rate analysis. The threshold for neural detectability based on a temporal code was set to 71% since VS_{pp} is expected to always be at least as large for dynamic ITD stimuli as for static ITD stimuli.

Results:

We recorded from 297 ITD-sensitive single neurons or multi-neuron clusters in the IC of 3 unanesthetized rabbits. These neurons showed statistically significant tuning of their firing rates to static ITD stimuli (ANOVA, $p < 0.001$). Responses to both SAM and dynamic ITD

stimuli were measured in 113 of these recordings; 66 of these were identified as single neurons. The 66 neurons spanned a wide range of BFs (95 Hz to 19 kHz); 24 had BFs below 2 kHz, where ITD sensitivity of rabbit IC neurons with broadband noise stimuli is primarily dependent on the temporal fine structure (Devore & Delgutte 2010). Of those low frequency neurons, 17 had BFs below 1 kHz.

Temporal coding of SAM and dynamic ITD stimuli

Figure 2b shows temporal response patterns (period histograms) from one neuron (BF = 3512 Hz) for SAM and dynamic ITD stimuli as a function of modulation frequency. The period histograms for SAM had sharper peaks and appeared more skewed than those for dynamic ITD. The firing rate was nearly constant with modulation frequency for dynamic ITD stimuli but showed much greater variations for SAM stimuli (**Figure 2e**). For both stimuli, the neuron's ability to phase-lock to the modulation period was quantified by the VS, which was plotted against modulation frequency to obtain a temporal modulation transfer function (tMTF) (**Figure 2c**). In this neuron the tMTFs had bandpass shapes for both stimuli, but both the peak VS and the best modulation frequency were larger for SAM than for dynamic ITD. In addition, both the 6 dB cutoff frequency (where VS falls to 50% of maximum) and the synchronization limit (the highest frequency where VS is significantly greater than zero (Rayleigh test, $p > 0.001$)) were higher for SAM than for dynamic ITD (see **Figure 2c**).

The BF of the neuron in **Figure 2** was 3512 Hz, in a range where static ITD sensitivity for broadband noise in rabbit IC is dependent on the cochlear-induced envelope (Devore & Delgutte 2010). Similar trends were observed for lower frequency neurons, where ITD sensitivity is predominantly dependent on the temporal fine structure. **Figure 3** shows results

from a lower-frequency neuron (BF=842 Hz) in a range where ITD tuning is dependent on the temporal fine structure. As in **Figure 2**, the tMTFs have bandpass shapes for both SAM and dynamic ITD (**Figure 3c**) and the peak VS, 6 dB cutoff and synchronization limit are all higher for SAM than for dynamic ITD, indicating stronger temporal coding of modulations for SAM. However, unlike in **Figure 2**, the range of firing rates across frequency observed for SAM was similar to the range for dynamic ITD (**Figure 3e**), which was uncommon across our sample of neurons.

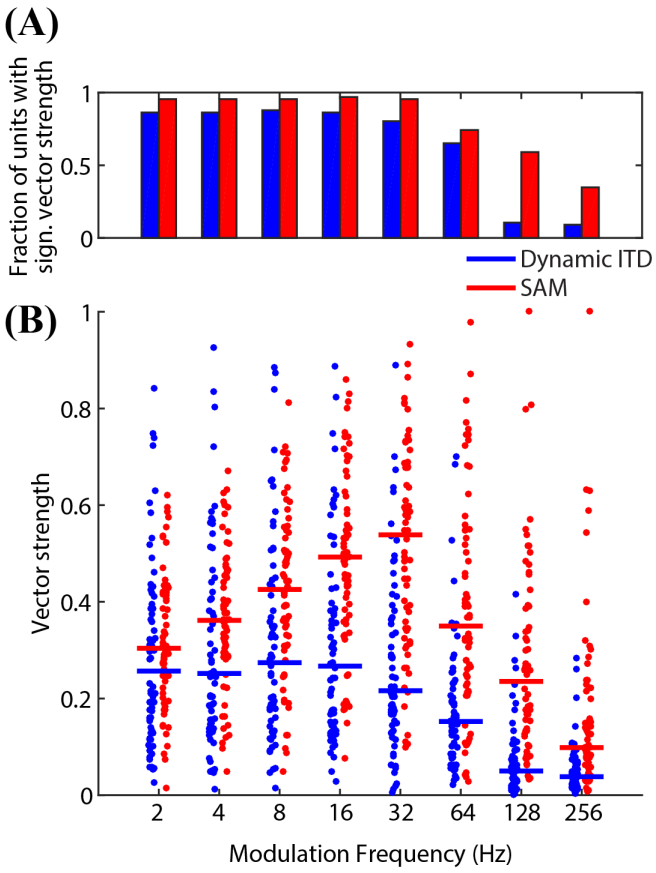


Figure 4: (A) The fraction of neurons out of 66 with significant vector strengths at each modulation frequency for SAM (red) and dynamic ITD (blue). (B) Vector strength of all of the neurons measured (n=66) for SAM and dynamic ITD at each modulation frequency. Lines mark median values.

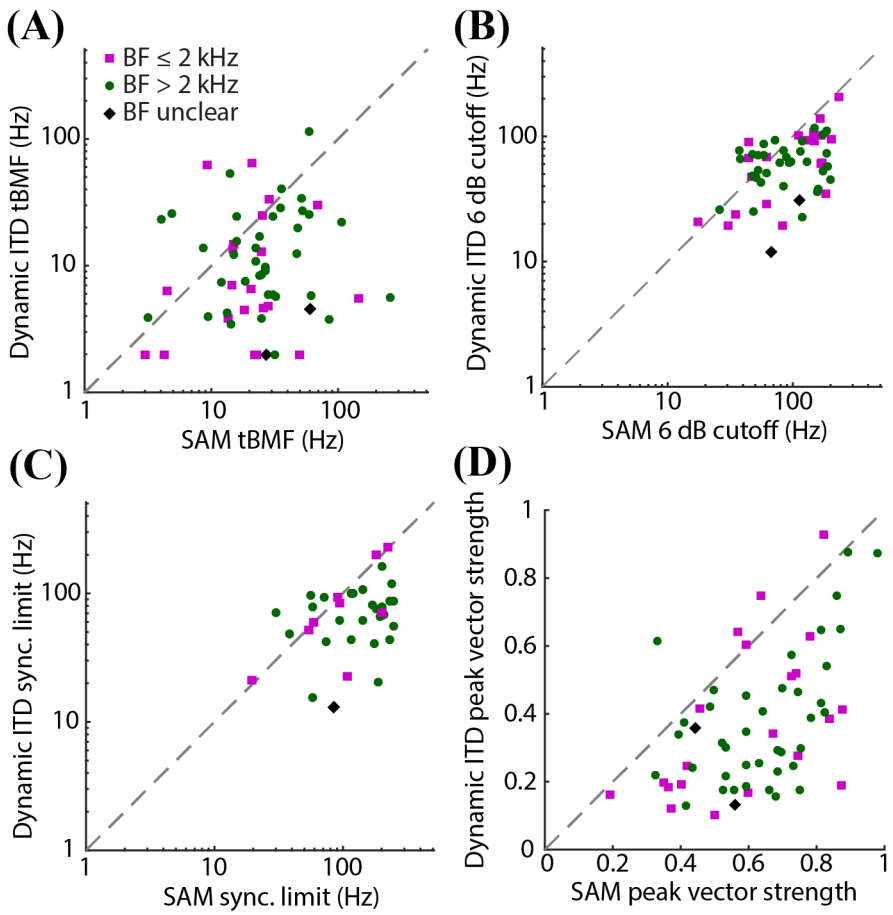


Figure 5: Comparison of SAM and dynamic ITD tBMFs (n = 62) (A), 6 dB cutoffs (n = 55) (B), synchronization limits (n = 36) (C), and peak vector strengths (n = 62) (D). Only neurons with tBMFs, 6 dB cutoffs, and synchronization limits below 256 Hz are included. All of these metrics were significantly higher for SAM than dynamic ITD across the population (Wilcoxon signed-rank test: $p < 0.001$). In all of these plots, neurons with BF below 2 kHz are shown in purple squares, neurons with BF above 2 kHz are shown in green dots, and neurons where the BF was unclear (2 neurons) are shown in black diamonds.

On average, neurons showed stronger phase locking to SAM than to dynamic ITD (**Figure 4**), and more neurons significantly phase locked to SAM than dynamic ITD at 128 and 256 Hz. Among 62 neurons that significantly phase locked to at least one modulation frequency for both SAM and dynamic ITD stimuli, the median tBMFs (**Figure 5a**), 6 dB cutoffs (**Figure 5b**), and synchronization limits (**Figure 5c**) were significantly higher for SAM than for dynamic ITD (Wilcoxon signed-rank, two-tailed: tBMF, n = 62, W = 1649; 6 dB cutoff, n = 55, W = 1892; sync. limit, n = 36, W = 572; $p < 0.001$ for all measures). Additionally, the median peak vector

strength (**Figure 5d**) was significantly higher for SAM than for dynamic ITD across the neuronal sample (n = 62, W = 1892, $p < 0.001$), showing that temporal coding was generally more salient for SAM. Peak vector strengths for SAM and dynamic ITD were significantly correlated across the sample (Kendall's $\tau = 0.41$, $p < 0.001$). There was also a weak correlation for 6 dB cutoffs ($\tau = 0.27$, $p = 0.004$). However, there was no significant correlation between tBMFs ($\tau = 0.09$, $p = 0.30$) nor synchronization limits ($\tau = 0.22$, $p = 0.06$). These weak correlations, along with the clear differences in median coding metrics between the two stimuli suggest that, while the temporal coding of SAM and dynamic ITD may be influenced by a common factor, the two are likely to be largely mediated by distinct mechanisms. When comparing low BF (< 2 kHz) to high BF (> 2 kHz) neurons, we found no significant difference between tBMFs, 6 dB cutoffs, synchronization limits, and peak vector strengths for both SAM and for dynamic ITD (Mann-Whitney U test: $p > 0.05$), suggesting that temporal coding of dynamic ITD is similar whether ITD sensitivity is dependent on the envelope or the temporal fine structure.

Effect of Adaptation

Previous studies of the coding of pure tone stimuli with time-varying ITD have suggested a role for firing rate adaptation in creating apparent sensitivity to the direction of ITD motion in the responses of IC neurons (Spitzer & Semple 1993; Cai et al. 1998; McAlpine et al. 2000; Borisyuk et al. 2002; Ingham & McAlpine 2004; Wang & Peña 2013). Here we used two metrics derived from temporal response patterns (period histograms) to assess a possible influence of adaptation on responses to SAM and dynamic ITD stimuli.

First, a straight line was fit to the mean phase of the period histogram as a function of modulation frequency to compute a group delay and a phase-intercept for both dynamic ITD and

SAM (**Figures 2d, 3d)**. This was done for 58 neurons. Although these linear fits were generally very good, on average, phase data for dynamic ITD were fit better than the SAM response phases: The median R^2 were 0.996 and 0.992 for dynamic ITD and SAM respectively, and the difference was significant (Mann-Whitney U-test, U = 3997, $p < 0.001$). In 8 neurons, a linear fit could not be obtained because either there were too few data points exhibiting significant phase locking or a straight line poorly fit the phase data.

Across the neuronal sample (**Figure 6a**), there was no significant difference between the median group delays for SAM and dynamic ITD (SAM = 17.0 ms, dynamic ITD = 18.2 ms, Wilcoxon signed-rank, two-tailed: W = 774, p = 0.53), and the two delays were significantly correlated (Kendall's tau: $\tau = 0.32$, $p < 0.001$). Even so, SAM group delays could be very different from dynamic ITD group delays in some neurons. There was no significant difference between group delays for low BF and high BF neurons, both for SAM and for dynamic ITD (Mann-Whitney U test: SAM, U = 513; dynamic ITD, U = 505, $p > 0.05$).

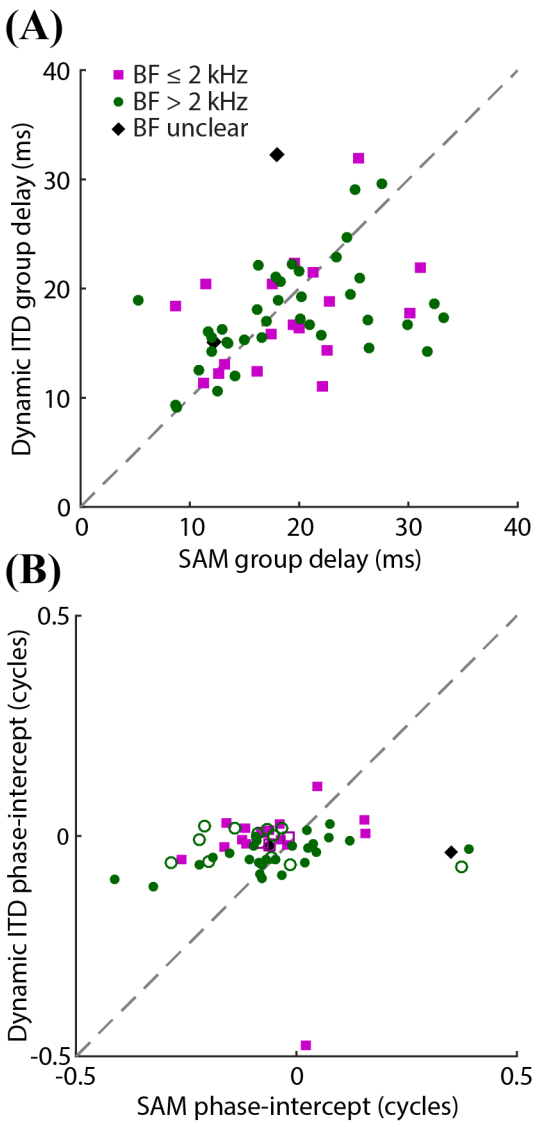


Figure 6: (A) Scatter plot of group delays computed for the SAM and dynamic ITD responses in each neuron (58/66 neurons). (B) Scatter plot of phase intercepts. Neurons that had an excitatory maximum was at 0.5 cycles for the dynamic ITD are shown as filled symbols, open symbols correspond to neurons where the excitatory maximum was at 0 cycles (see Methods). Colors and shapes are the same as in Figure 5.

In the absence of adaptation, phase intercepts (**Figure 6b**) are expected to be near zero when a neuron fires maximally at the peak amplitude of a low-frequency SAM stimulus or when the instantaneous ITD in a dynamic stimulus approaches the ITD producing the maximum firing rate with static presentation. Across the neuronal sample, the median phase intercept for dynamic ITD stimuli (–0.02 cycles) was significantly less than 0 (Wilcoxon signed-rank, two-tailed: W = 336, $p < 0.001$). This result is consistent with the idea that adaptive effects cause phase leads for time-varying binaural temporal cues (McAlpine et al. 2000; Borisyuk et al. 2002), but the effect was weak. The median phase intercept for SAM (–0.07 cycles) was smaller than 0 by a greater amount (W = 408, $p < 0.001$), suggesting stronger adaptation effects for SAM. Indeed, across the neuronal sample phase intercepts were significantly greater

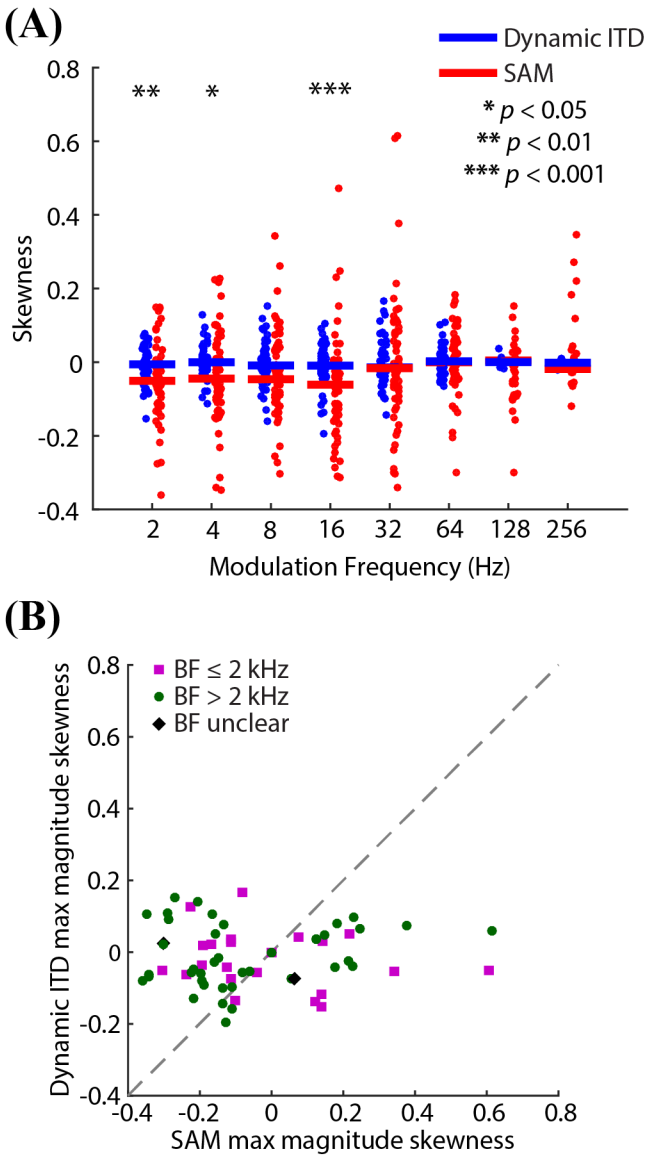


Figure 7: (A) Skewness across the population (n = 66) at each modulation frequency (red for SAM, blue for dynamic ITD). Skewness is only shown at modulation frequencies with significant phase locking. Solid lines designate the median values of each distribution. The medians of the skewness for SAM were significantly different than dynamic ITD for 2, 4, and 16 Hz (p values are based on the two-tailed Mann-Whitney U test with Bonferroni correction). None of the dynamic ITD medians were significantly different than zero. (B) The maximum magnitude skewness values for SAM and dynamic ITD across the population (n = 62). Colors are the same as in Figures 5 and 6.

(less negative) for dynamic ITD than for SAM (Wilcoxon signed-rank, two-tailed: W = 1152, p = 0.022). Whereas phase intercepts for dynamic ITD stimuli were relatively tightly clustered near 0, intercepts for SAM showed much greater scatter, so that there was no significant correlation between intercepts for the two stimuli (Kendall's tau: τ = 0.12; p = 0.19). In 5 neurons, the phase intercept for SAM was close to 0.5 cycles (**Figure 6b**), suggesting the SAM stimulus suppressed neural responses near the amplitude maxima. There was a slightly significant difference between phase intercepts of low and high BF neurons for dynamic ITD (Mann-Whitney U test: U = 669, p = 0.006), but not for SAM (U = 531, p = 0.76).

For both SAM and dynamic ITD stimuli, the period histograms could be very asymmetric for low modulation frequencies but the asymmetry tended to disappear at higher modulation frequencies (e.g. **Figure 2b**). This asymmetry was quantified by the circular skewness of the period histogram at modulation frequencies where a neuron exhibited significant phase-locking (Rayleigh test with Bonferroni correction, $p < 0.001$). For SAM stimuli, skewness was often negative for low modulation frequencies (**Figure 7a**), meaning that the rise in firing rate towards the peak of the distribution was steeper than the fall in firing rate beyond the peak, consistent with an adaptive process. At higher modulation frequencies, a few neurons showed large positive skewness values, but these were unusual. When pooled across the population, the median skewness for SAM stimuli was significantly less than the median skewness for dynamic ITD stimuli for 2, 4, and 32 Hz (Mann-Whitney U test with Bonferroni correction for each modulation frequency, see **Figure 7a**).

We compared the skewness with the largest magnitude across modulation frequencies for SAM relative to dynamic ITD in the 62 neurons (**Figure 7b**). The median maximum skewness across the neuron sample was weakly but significantly less than 0 for SAM (median = –0.12,

Wilcoxon signed-rank, two-tailed: W = 671, $p = 0.03$) but not for dynamic ITD (median = –0.04, W = 783, $p = 0.17$). There was no significant difference in median skewness between SAM and dynamic ITD (Wilcoxon signed-rank, two-tailed: W = 1205, $p = 0.11$), nor a significant correlation between the two (Kendall's tau: $\tau = 0.03$, $p = 0.71$).

Since adaptation is expected to result in a suppression of firing following an initial rise in activity (McAlpine et al. 2000), these results suggest that adaptation is stronger for SAM than dynamic ITD, particularly below 32 Hz. However, because vector strengths were also larger for SAM than dynamic ITD (**Figure 4**), the greater skewness for SAM stimuli may have been a by-product of the stronger phase locking. Indeed, there was a weak negative correlation between skewness and VS when data for the two stimuli were combined (Pearson's correlation: $\rho = -0.11$, $p = 0.003$). However, after partialing out the effect of VS on skewness, the median skewness for SAM and dynamic ITD were still significantly different at 2 and 16 Hz (Mann-Whitney U test with Bonferroni correction: 2 Hz, U = 3605, $p < 0.01$; 4 Hz, U = 3293, $p = 0.078$; 8 Hz, U = 3628, $p = 0.294$; 16 Hz, U = 3690, $p < 0.001$; 32-256 Hz, $p > 0.05$).

Comparison of responses to triangular and sinusoidal modulations of dynamic ITD:

In 9 neurons (BF = 277 – 6553 Hz), separate from the 66 neurons studied with SAM, we recorded responses to both sinusoidal and triangular ITD modulation with the same ITD range and center ITD. While we found differences in temporal coding for SAM and dynamic ITD, we wanted to ensure that these differences were not simply due to our choice of dynamic ITD modulation.

Figure 8:

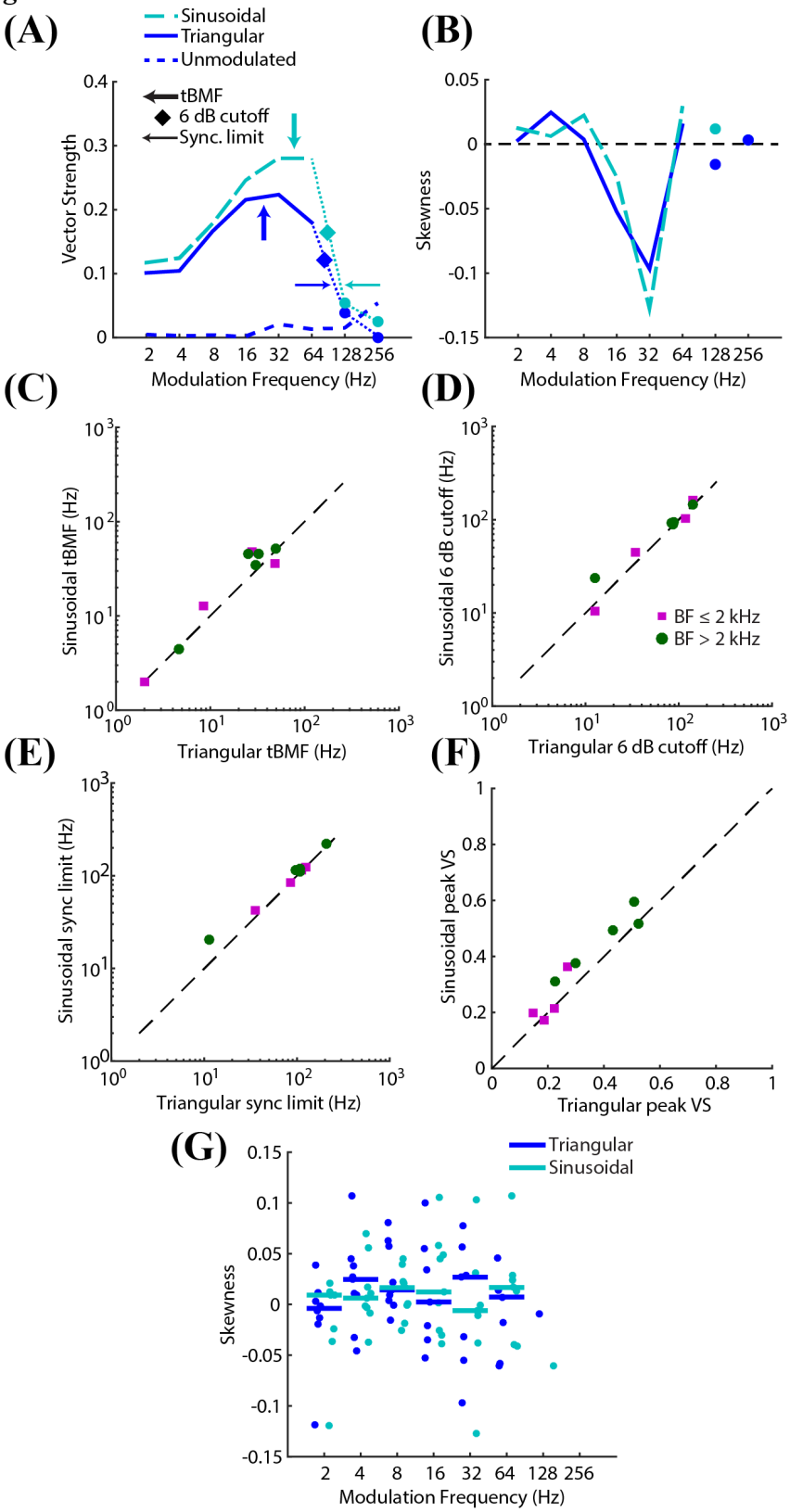

(A)
Sinusoidal
Triangular
Unmodulated
tBMF
6 dB cutoff
Sync. limit
Vector Strength
Modulation Frequency (Hz)
(B)
Skewness
Modulation Frequency (Hz)
(C)
Sinusoidal tBMF (Hz)
Triangular tBMF (Hz)
(D)
Sinusoidal 6 dB cutoff (Hz)
BF ≤ 2 kHz
BF > 2 kHz
Triangular 6 dB cutoff (Hz)
(E)
Sinusoidal sync limit (Hz)
Triangular sync limit (Hz)
(F)
Sinusoidal peak VS
Triangular peak VS
(G)
Triangular
Sinusoidal
Skewness
Modulation Frequency (Hz)

Figure 8: (A) and (B) show the vector strengths and skewness for dynamic ITD with a sinusoidal trajectory (light blue dashed) and a triangular trajectory (blue dashed) for a single neuron (BF = 4296 Hz). The labels of tBMF, 6 dB cutoff, and synchronization limit are the same as in Figure 2 and 3. (tBMF: Triangular = 25 Hz, Sinusoidal = 45 Hz; 6 dB cutoff: Triangular = 88 Hz, Sinusoidal = 94 Hz; Sync. Limit: Triangular = 108 Hz, Sinusoidal = 119 Hz; Peak VS: Triangular = 0.23, Sinusoidal = 0.31). (C) tBMF, (D) 6 dB cutoff, (E) synchronization limit, and (F) peak vector strengths for triangular and sinusoidal dynamic ITD modulations in 9 neurons. Colors and shapes are labed by BF identically to Figures 6-8. Only synchronization limits were significantly different (Wilcoxon signed-rank test, p = 0.016) but by a small amount. (G) Skewness for sinusoidal (light blue) and triangular (blue) modulations for the same 9 neurons. We have plotted only skewness at modulation frequencies where the vector strength was significant (Rayleigh test with Bonferroni correction, $p < 0.001$). The lines are plotted at median values. Only one neuron significantly synchronized to 128 Hz and no neurons significantly synchronized to 256 Hz. From 2-64 Hz, there were no significant differences between the median skewness (Mann-Whitney U test, $p > 0.05$).

Figures 8a and b show the vector strength and skewness respectively sinusoidal and triangular modulations of dynamic ITD in a single neuron (BF = 4296 Hz). This neuron showed some of the largest differences in temporal coding of the population with a larger peak vector strength and higher tBMF for sinusoidal than triangular modulation. For the sampling of neurons, however, there was no significant difference between tBMFs, peak VS, or 6 dB cutoffs between the two types of ITD modulation (Wilcoxon signed-rank test, $p > 0.05$) (**Figures 8c-f**). Synchronization limits for the sinusoidal modulation were significantly larger than for triangular modulation, but the difference was very small (0.19 octaves, $W = 0$, $p = 0.016$). Likewise, there were no significant differences between the median skewness at all modulation frequencies measured (Mann-Whitney U test, $p > 0.05$) (**Figure 8g**). We conclude from this that there are no significant differences in temporal coding for sinusoidal and triangular dynamic ITD modulations. Furthermore, the differences observed between dynamic ITD and SAM temporal coding cannot be explained by differences in triangular and sinusoidal modulation.

Rate coding of SAM and dynamic ITD stimuli:

Figures 2e and 3e show the average firing rate as a function of modulation frequency for SAM and dynamic ITD for two example neurons. In **Figure 2e**, the firing rate for SAM stimuli is clearly enhanced over the unmodulated response over a wide range of modulation frequencies (2-64 Hz). Such neurons with "band enhanced" SAM tuning are frequently observed in IC (Langner & Schreiner 1988; Nelson & Carney 2007). In contract, the variations in firing rate with modulation frequency are much less pronounced for dynamic ITD and the firing rate remains close to that for a static-ITD stimulus. While such clear difference between SAM and dynamic ITD is not observed in **Figure 3e**, this neuron was unusual in this respect. We classified SAM rate tuning curves relative to their firing rates for the unmodulated noise, based on the classification scheme in previous studies (Carney et al. 2015; Kim et al. 2015). Across the population, 13 SAM rMTFs neurons were classified as band-enhanced, 25 as band-reject, 14 as complex, 10 neurons had flat rMTFs. The median frequency exhibiting the maximum firing rate was 18 Hz (interquartile range of 10-56 Hz), and the median frequency with a firing rate minimum was 103 Hz (interquartile range of 64-129 Hz).

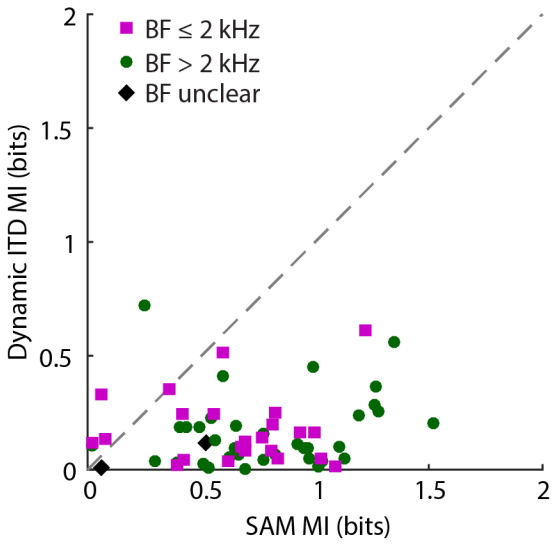


Figure 9: MIs for SAM and dynamic ITD across the population (n = 66). MIs were on average significantly larger for SAM than dynamic ITD (Wilcoxon signed-rank test: $p < 0.001$). Colors and shapes are identical to Figures 5-8.

We used mutual information (MI) to quantify the rate coding of modulation frequency in each neuron (see Methods). The median MI across the neuronal sample was significantly larger for SAM than for dynamic ITD (n = 66, SAM: MI =

0.68, ITD: MI = 0.11; Wilcoxon signed-rank, two tailed: W = 2151, $p < 0.001$) (**Figure 9**). There was also no significant difference in MI between low BF and high BF neurons for either stimulus (Mann-Whitney U test: SAM, U = 683; dynamic ITD, U = 822; $p > 0.05$ for both).

Despite the weaker variation in firing rates for dynamic ITD compared to SAM, many neurons showed small but systematic changes in firing rate with modulation frequency for dynamic ITD stimuli, particularly at high modulation frequencies. For the two neurons in **Figure 10**, the mean firing rate to dynamic ITD stimuli was close to the rate for a static ITD stimulus located at the center of the dynamic ITD trajectory for modulation frequencies below 16 Hz. At higher modulation frequencies, the firing rate to dynamic ITD stimuli approached the firing rate to interaurally uncorrelated noise, whether the latter was higher (**Figure 10a**) or lower (**Figure 10b**) than the rate to the static ITD stimulus.

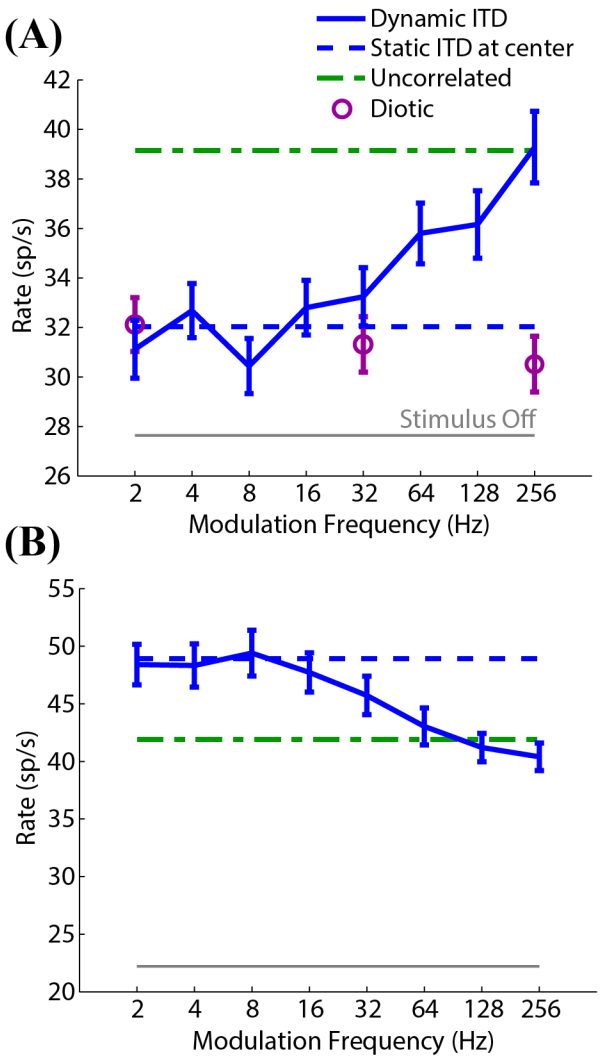


Figure 10: Dynamic ITD firing rates measured for two different neurons: (A) CF = 3250 Hz, (B) CF = 842 Hz. Error bars are SE. The neuron in (B) is the same as the neuron in Figure 3. For many neurons, average firing rates at high modulation frequencies tended towards their average firing rate for uncorrelated noise (green dot-dashed line). For (A), the noise with the time-varying delay, typically presented to the ipsilateral channel, was presented diotically. A diotic presentation of the time-varying delayed signal could not explain the increase in firing rate (purple circles).

To test whether the changes in firing

rate at high modulation frequencies might be caused by spectrotemporal distortion in the dynamic ITD stimulus, the channel of the dynamic ITD stimulus processed to have a time varying delay (which is normally presented to the ipsilateral ear) was presented diotically in 6 neurons. If the change in rate was due to monaural features of the stimulus with time-varying delay, we would expect to see similar changes in firing rate at high modulation frequencies with diotic presentation. However, such changes were not observed for any of these neurons (see open circles in **Figure 10a** for an example).

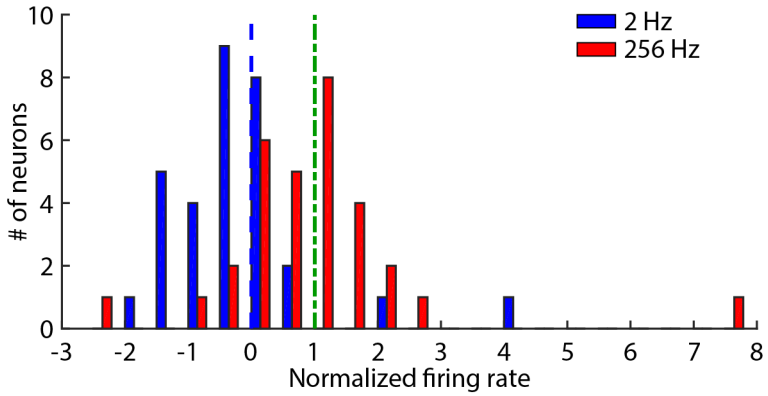


Figure 11: For each neuron, its firing rate for the 2 and 256 Hz dynamic ITD stimuli was normalized by the its firing rate for the static ITD and its firing rate for the binaurally uncorrelated noise (see Results). If the normalized firing rate was 0, the firing rate was equal to the rate for the static ITD (marked by the blue dashed line in the histogram). If the normalized firing rate was 1, the firing rate was equal to the rate for binaurally uncorrelated noise (marked by the dot-dashed green line). One neuron had a normalized firing rate of 35 for 256 Hz; this rate is not shown.

We compared the firing rates for the dynamic ITD stimuli at 2 and 256 Hz with the firing rates for the static ITD stimulus and for binaurally uncorrelated noise. To make the firing rates comparable across the population, we normalized the firing rates to dynamic ITD stimuli in each neuron so that a normalized rate of 0 means the raw firing rate equals the rate for the static ITD stimulus, while a normalized rate of 1 corresponds to the rate for uncorrelated noise (**Figure 11**). The median normalized firing rates at 2 Hz were not significantly different from 0 (Wilcoxon signed-rank test: W = 240, $p = 0.88$) but they significantly smaller than 1 (W = 60, $p < 0.001$).

Additionally, the median normalized firing rates at 256 Hz were not significantly different from 1 (W = 177, $p = 0.16$) but significantly larger than 0 (W = 424, $p < 0.001$). From this we conclude that, at low modulation frequencies, the average firing rate for the dynamic ITD tends to approach the firing rate for a static ITD stimulus, which produces a spatially focused percept. In contrast, at high modulation frequencies, the firing rate to dynamic ITD stimuli approaches the rate binaurally uncorrelated noise, which produces a spatially diffuse percept.

For most neurons, the change in firing rate to dynamic ITD stimuli away from the rate evoked by a static ITD stimulus and towards the rate produced by uncorrelated noise occurred at around 64 Hz. This is roughly the frequency range where subjects' performance in the Grantham and Wightman (1978) study started to improve again after decreasing at lower frequencies. This observation suggests that a rate code may be available for detecting dynamic ITD stimuli at high modulation frequencies where temporal coding is less effective.

ROC analysis:

ROC analysis was used to quantify the ability of an ideal observer to discriminate a dynamic ITD stimulus from a static ITD stimulus positioned at the center of the dynamic ITD trajectory based on the responses of a single neuron. This task is similar to that performed by the subjects of Grantham and Wightman (1978), except we used a fixed ±300 μs ITD modulation depth while these authors varied modulation depth to find threshold. We examined ideal observer performance separately for the rate code based on spike counts and a temporal code based on the phase-projected vector strength (VS_{pp}, see Methods). For each modulation frequency, the percent correct score of the ideal observer in a two-interval, two-alternative forced choice task was determined based on the distributions of spike counts and VS_{pp} across stimulus

trials. Note that the dynamic ITD trajectory was usually chosen for each neuron so as to lie along a portion of the static ITD tuning curve showing a steep change in firing rate with ITD. Because this choice leads to large changes in instantaneous firing rate depending on instantaneous ITD, we are approximately assessing each neuron's optimal performance using a temporal code in this task.

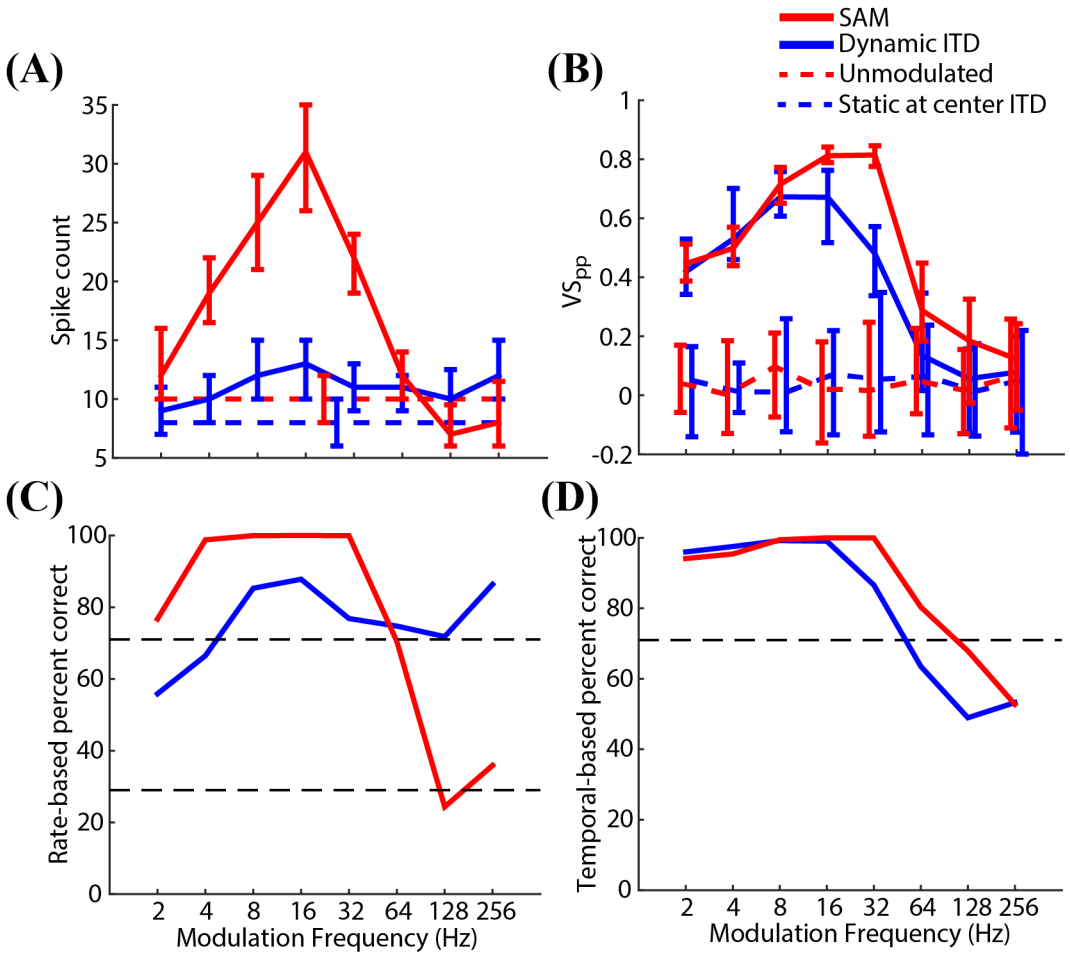


Figure 12: Median spike counts (A) and phase-projected vector strengths (B) from the 45 500 ms bins for the neuron from Figure 2. Solid lines are the responses for each modulation frequency (blue for dynamic ITD, red for SAM), and dashed lines are the responses for the noise positioned at the center of the ITD trajectory (blue) and the unmodulated noise at the same ITD as the SAM stimulus (red). Error bars mark the 75th and 25th percentiles of the distributions. Percents correct computed based on this distributions are shown below (C and D respectively). Dashed lines mark 71% and 29%.

Blue lines in **Figure 12** show the spike counts (**Figure 12a**), VS_{pp} (**Figure 12b**), and percent correct scores based on rate (**Figure 12c**) and temporal (**Figure 12d**) codes as a function

of modulation frequency for the example neuron from **Figure 2**. The temporal code (**Figure 12b**) provides good discrimination (large percent correct) at low modulation frequencies, but performance falls to near chance (50%) above 64 Hz. In contrast, the rate code (**Figure 12a**) performs poorly at 2 and 4 Hz, but provides sufficient information to exceed 71% correct at higher frequencies. This value corresponds to asymptotic performance in the 2-up, 1-down threshold tracking procedure widely used in psychophysics (Levitt 1971).

The results of the ROC analysis for discrimination of dynamic ITD from static ITD stimuli are summarized for our neuronal sample in **Figure 13a**. The percent correct scores using the rate code (purple dots) and temporal code (green dots) are shown for every neuron in the lower panel, along with the fraction of neurons that met the 71% correct threshold criterion as a function of modulation frequency in the upper panel. Since stimuli can be detected through either an increase or a decrease in firing rate, percent correct scores below 29% were also considered to meet the threshold criterion for the rate code (see Methods). Over half the neurons had percent correct scores above 71% using the temporal code for frequencies of 32 Hz and below, but this fraction rapidly fell to near 0 at higher modulation frequencies. In contrast, with the rate code, performance increased with increasing modulation frequency with most of the increase occurring above 64 Hz. However, only 41 neurons (66%) had rate-based percent correct scores outside of the 29-71% threshold boundary for at least one modulation frequency. In contrast, 55 neurons (89%) had percent correct scores above 71% using the temporal code, suggesting the temporal code is more effective overall.

Figure 13:

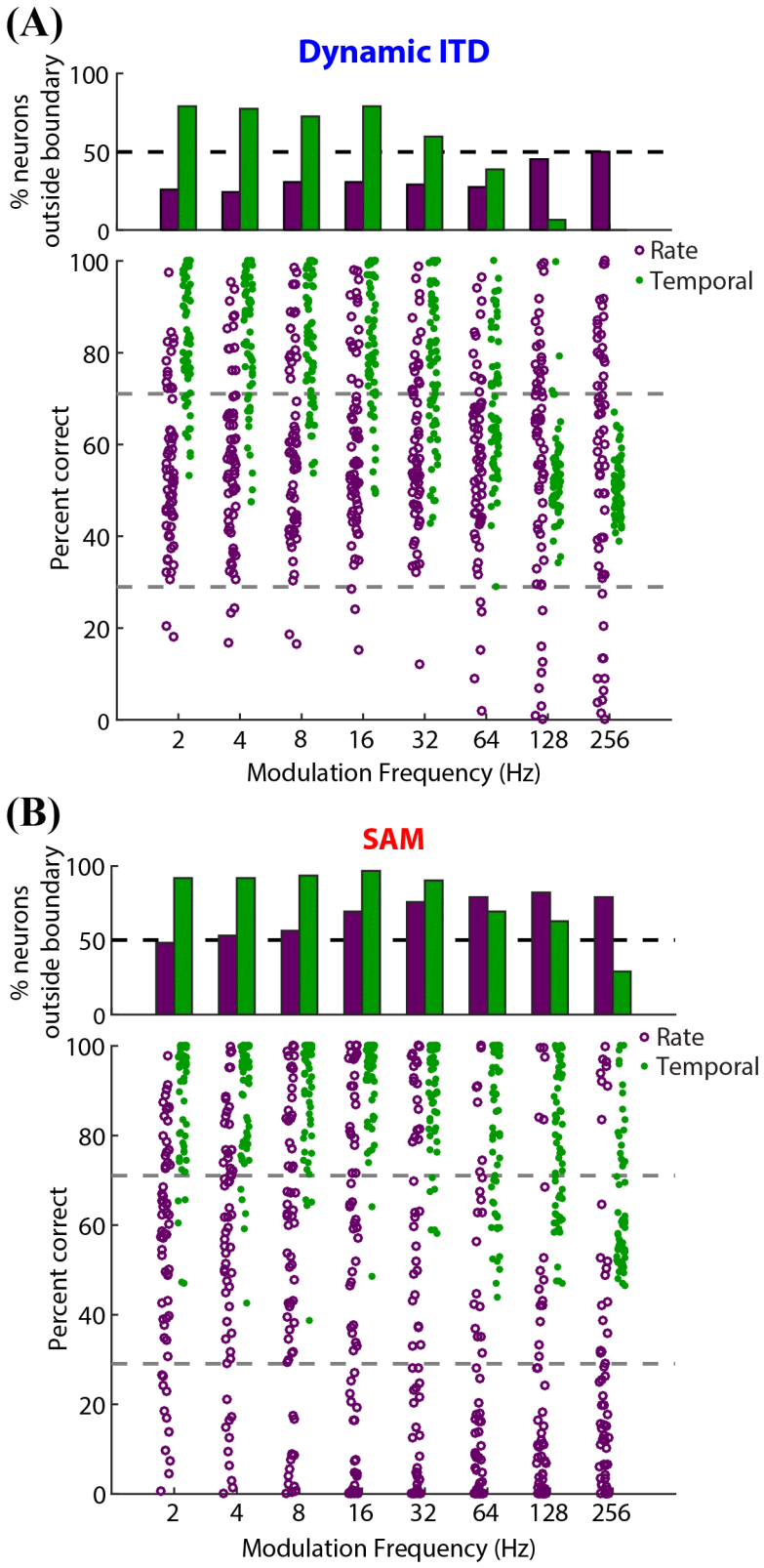

(A)
Dynamic ITD
% neurons outside boundary
100
50
0
Rate
Temporal
Percent correct
100
80
60
40
20
0
2
4
8
16
32
64
128
256
Modulation Frequency (Hz)
(B)
SAM
% neurons outside boundary
100
50
0
Rate
Temporal
Percent correct
100
80
60
40
20
0
2
4
8
16
32
64
128
256
Modulation Frequency (Hz)

Figure 13: (A) Population ROC analysis for detecting the dynamic ITD stimulus in a two-alternative forced choice task with a static ITD noise positioned at the center of the ITD trajectory. Bottom plots show the percent correct for all neurons using a rate (purple circles) and temporal (green dots) code for discrimination (n = 62). Upper plots show the percentage of neurons whose percentages for correct detection for each code were outside of the boundary labeled with a grey dashed line in the bottom plot (>71% or <29% for rate, >71% for temporal). (B) Population ROC analysis for detecting SAM for a two-alternative forced choice task with an unmodulated noise at the same ITD (n = 62).

Discrimination of the dynamic ITD from the static ITD appears to be possible based on the responses of single neurons, but the performance of single neurons is noisy (for example, see **Figure 12c**). More realistically, human listeners are likely to pool information across multiple neurons to perform the task. The performance of a classifier that optimally combines information across neurons will generally be dominated by the neurons that provide the most reliable information for the task (Green & Swets 1988), either by rate or by temporal coding. Taking the maxima of the fractions of neurons that met the 29-71% threshold using the rate and temporal codes to estimate performance based on the better of the two codes yields a nonmonotonic curve with a shallow minimum at 64 Hz. This minimum roughly matches the modulation frequency where the performance of Grantham and Wightman's (1978) subjects in detecting sinusoidally-varying ITD was poorest, although the nonmonotonicity in performance was more pronounced in the perceptual data than in the neural data.

For comparison, we also applied ROC analysis to characterize the performance of an ideal observer using single-neuron responses to discriminate SAM noise from unmodulated noise in a two-interval, two-alternative forced choice task. This is the same analysis as performed by Nelson and Carney (2007) and Carney et al. (2014) based on responses of rabbit IC neurons, except we are keeping the modulation depth at 100% rather than adjusting it to find the detection threshold. Red lines in **Figures 12c** and **12d** show the percent correct scores of the example neuron from **Figure 2** in AM detection using the rate and temporal codes, respectively.

Performance using the temporal code is very high for modulation frequencies up to 32 Hz, but degrades rapidly at higher frequencies to approach chance at 256 Hz. A slight decrease in performance reflecting the bandpass shape of the tMTF is also discernible at low frequencies. Performance using the rate code exceeds 71% correct for frequencies up to 32 Hz, and then falls abruptly to around 29% correct at 128 Hz, reflecting the suppression of firing rate at high frequencies relative to the response to static ITD stimuli in this neuron (**Figure 2e**).

The results for AM detection are summarized for the neuronal sample in **Figure 13b**. Using the temporal code, percent correct scores were high for modulation frequencies up to 32 Hz, above which the fraction of neurons with percent correct scores over 71% fell gradually with increasing frequency. In contrast, performance using the rate code gradually improved with increasing modulation frequency up to 32 Hz, above which 70-80% of neurons had percent correct scores outside the 29-71% boundary. At these high modulation frequencies (64 Hz and above), the performance was driven by rate suppression in a majority of neurons, as shown by the large number of percent correct scores below 29%. Suppression of firing rate to AM stimuli is consistent with previous reports on responses of IC neurons to SAM (Krishna & Semple 2000; Nelson & Carney 2007; Kim et al., 2015). Unlike the ROC analysis for dynamic ITD stimuli, where the population performance using the best of the rate or temporal codes was nonmonotonic, population performance for SAM noise using the better of the two codes was nearly constant over the entire 2-256 Hz range. This stability contrasts with the behavioral modulation detection thresholds measured in rabbits by Carney et al. (2014), which have a nonmonotonic dependence on modulation frequency with optimum performance at 64 Hz (see Discussion).

Discussion:

We directly compared the responses of IC neurons in unanesthetized rabbits to SAM and dynamic ITD stimuli for the same modulation frequencies. We observed higher tBMFs, 6 dB cutoffs, synchronization limits, and peak vector strengths for SAM than dynamic ITD in most neurons. We also observed, on average, both larger skewness in the period histograms and larger deviations from 0 in phase intercepts for SAM than dynamic ITD, suggesting a greater influence of adaptive processes for SAM stimuli. Although peak vector strengths and 6 dB cutoffs for SAM and dynamic ITD were weakly correlated, no correlation was observed for any of the other response metrics, contrary to the hypothesis that sensitivity to SAM and dynamic ITD are shaped by a common mechanism. Lastly, we found stronger rate coding for SAM than for dynamic ITD. Together, these results suggest that the mechanisms underlying sensitivity of IC neurons to SAM and dynamic ITD are largely distinct.

We also measured the responses to triangular and sinusoidal dynamic ITD and found that temporal coding was very similar for both trajectories. This agrees with Spitzer & Semple's (1993) finding that the responses to sinusoidal and triangular time-varying ITDs for pure tones are similar. Based on this, we think that our use of triangular modulations instead of sinusoidal modulations for dynamic ITD does not affect the comparison of dynamic ITD coding with SAM. In any case, compression in cochlear mechanics and synaptic transmission severely distort modulation waveforms, so that the modulations in the firing rates of auditory-nerve fibers in response to SAM noise are clearly not sinusoidal, making the goal of precisely matching monaural and binaural modulation waveforms hard to achieve.

Based on previous IC studies using tone stimuli with time-varying ITD in anesthetized animals (Spitzer & Semple 1993; McAlpine et al. 2000) we expected both the skewness and the

phase-intercepts to deviate markedly towards the leading side in the presence of a dynamic ITD. Instead, we saw a small but significant shift in the median phase intercept across the population without any significant effects on skewness. It is possible that the modulation frequencies we tested were too high to observe the strongest adaptive effects. Both IC data (Spitzer & Semple 1998) and a model of directional selectivity in IC (Borisyuk et al. 2002) show the largest phase leads at 1 Hz, while the lowest frequency we tested was 2 Hz. On the other hand, Fitzpatrick et al. (2009) only found weak direction selectivity with binaural beat stimuli in IC of Dutch-belted rabbits. There are too many differences between paradigms to identify the reason for the differences in results across studies.

Although changes in firing rate with modulation frequency were small for dynamic ITD stimuli, they occurred fairly consistently at high modulation frequencies. If the initial cross-correlation stage for ITD detection is unable to keep up with fast changes in ITD, it may respond as if the inputs are binaurally uncorrelated. The frequency at which this transition occurs would depend upon the width of the coincidence window and the time course of envelope fluctuations in cochlear output (Joris 2003). In contrast, changes in rate with frequency for SAM may be caused by inhibitory mechanisms (Nelson & Carney 2004).

Our results suggest a temporal code could be used at low modulation frequencies and a rate code at high frequencies for detecting both SAM and dynamic ITD stimuli. While there is considerable evidence for dual codes for AM stimuli in both IC (Langner & Schreiner 1988; Nelson & Carney 2007) and auditory cortex (Liang et al. 2002; Yin et al. 2011), to our knowledge such a scheme has not been explicitly proposed for binaural processing.

Because rate coding of dynamic ITD stimuli improved at high modulation frequencies where temporal coding was degraded, the fraction of neurons with a percent correct score

exceeding the 29-71% threshold criterion based on either code showed a weakly nonmonotonic dependence on modulation frequency, with a minimum at 64 Hz (**Figure 13a**, top). While this is qualitatively similar to the trend in human detection thresholds measured by Grantham and Wightman (1978), a rigorous comparison is not possible because we measured neural detection performance for a fixed range of ITD variation (600 μs) of adjusting the range of dynamic ITDs to find a detection threshold. This was done so that detection performance could be characterized over a wide range of modulation frequencies while also measuring responses to SAM noise in every neuron within the recording time typically achievable in our unanesthetized preparation. Nevertheless, the frequency dependence of performance measured with a fixed ITD range could in principle parallel the trend in thresholds measured by varying the ITD range if the percent correct scores varied similarly with ITD range across frequencies.

For modulation frequencies of 64 Hz and above, fewer than half the neurons reached the 29-71% threshold criterion using a 600 μs range of ITD variation. In this frequency range, the subjects of Grantham and Wightman (1978) had dynamic ITD detection thresholds of 40-100 μs. Thus, only a very small fraction of our neurons, if any, would be expected to reach 29-71% correct performance for dynamic stimuli with ITD ranges near human thresholds, making it necessary to invoke pooling of information across neurons with subthreshold performance to account for human performance (Parker & Newsome 1998). A caveat is that we are comparing performance of rabbit IC neurons with human performance. Although, to our knowledge, behavioral detection thresholds for time-varying ITDs have not been measured in rabbit or any other nonhuman mammal, both static ITD discrimination thresholds (Ebert et al. 2008) and AM detection thresholds (Carney et al. 2014) are higher in rabbits than humans, making it likely that detection thresholds for dynamic ITDs would also be higher in rabbits.

For SAM detection, neural percent correct scores based on the temporal code remained relatively high up to 128 Hz, while rate-based scores were already high by 16 Hz, so that performance based on the best of the two codes remained nearly constant over the entire frequency range (**Figure 13b**). Again, a direct comparison with behavioral results is difficult because we measured neural performance for a fixed 100% modulation depth, while behavioral modulation detection thresholds (as in Viemeister (1979) for humans and Carney et al. (2014) for rabbits) were obtained by adjusting the stimulus modulation depth. Carney et al. (2014) measured modulation detection thresholds of rabbit IC neurons, but only for one modulation frequency at each recording site. Because neural modulation detection thresholds roughly matched rate-based thresholds but fell below synchrony-based thresholds, they argued that rabbits do not make use of a temporal code in modulation detection. However, for low modulation frequencies (2-32 Hz), their rate based thresholds for SAM noise were always higher than the behavioral thresholds, while some of their synchrony thresholds were within the range of behavioral thresholds. Moreover, even if neural synchrony thresholds are lower than behavioral thresholds, this does not rule out that a temporal code may be used for detection, albeit inefficiently. Without more suitable data to assess the frequency dependence of neural thresholds, a role for temporal coding at the midbrain level in AM detection cannot be ruled out, especially at low modulation frequencies where the temporal code is largely maintained in the auditory cortex (Liang et al. 2002).

The lower synchronization limits and weaker phase locking observed in IC neurons for dynamic ITD stimuli compared to SAM noise could be interpreted as a neural correlate of the "binaural sluggishness" observed psychophysically with dynamic ITD stimuli (Grantham & Wightman 1978). The performance of Grantham and Wightman's subjects improved for

frequencies above 50 Hz, and we have suggested that this transition in performance may correspond to a switch from a temporal code to a rate code. Rate coding might also explain the detectability of high-frequency time-varying binaural cues observed in other studies. For example, using a "phasewarp" broadband noise stimulus having both a time-varying ITD and time-varying interaural correlation, Siveke et al. (2008) found that humans could detect modulations up to 1 kHz. In contrast, studies using stimuli with time-varying interaural correlation alone (Grantham 1982; Siveke et al. 2008) have observed lower detection limits (50-128 Hz).

As suggested by Siveke et al. (2008), it is possible that time variations in ITD are specifically necessary to produce sufficiently strong rate cues to distinguish stimuli with time-varying binaural cues from a static source at high frequencies. Using ROC analysis of rate coding, Siveke et al. (2008) found that neurons in the dorsal nucleus of the lateral lemniscus (DNLL) of anesthetized gerbils could detect phasewarp stimuli against binaurally uncorrelated noise at higher modulation frequencies than they detected noise stimuli with sinusoidally-varying interaural correlation. In contrast, Joris et al. (2006) found that neurons in the IC of anesthetized cat rarely showed changes in firing rate for stimuli with time-varying correlation without a time-varying ITD relative to a static stimulus.

In summary, at low modulation frequencies, temporal coding of time-varying ITDs by IC neurons was "sluggish" relative to the temporal coding of SAM. The temporal coding at these frequencies may be important for encoding motion at slow speeds. For the mid-range of modulation frequencies, this may produce the perception of modulation and allow spatially separated sources to remain distinguishable perceptually. At higher modulation frequencies, a secondary mechanism, namely a rate code, allowed time-varying ITDs to remain detectable

against a static sound. The secondary mechanism may be important for encoding reverberant or other spatially diffuse sounds in everyday environments where localization cues change rapidly.

Chapter 3

Neural coding of time-varying interaural time differences in the inferior colliculus explains human performance in motion direction identification and binaural gap detection

---

While motion is important for parsing a complex auditory scene, how it is encoded in the auditory system is still unclear. Perceptual studies suggest that our ability to identify the direction of motion is limited by the duration of the moving sound, yet we can detect the presence of brief changes in interaural correlation with much higher temporal precision. Here, we recorded from single neurons in the inferior colliculus of unanesthetized rabbits when presented with noise stimuli with brief, linearly time-varying interaural time difference ("ITD sweep"). We also presented these sounds to human listeners and measured their ability to either detect the ITD sweeps or identify their motion direction. We found that the neurons respond primarily by following the instantaneous ITD rather than by exhibiting true direction selectivity. Furthermore, using an optimal classifier to decode the single-neuron responses, we find that single-neuron threshold durations of ITD sweeps for both direction identification and detection in IC neurons match human threshold durations in these tasks.

Introduction:

In the previous chapter we observed that neurons in the IC exhibit a decrease in temporal coding and an increase in rate coding of dynamic ITD between 32-128 Hz. This transition occurred roughly where participants in Grantham & Wightman's (1978) identified the noise with flutter to detect the dynamic ITD at lower modulation frequencies and identified the spatially diffuse noise at high modulation frequencies. But there is a second transition in our perception of time-varying ITD at even lower modulation frequencies: we perceive the sound as moving at slow speeds and we perceive fluctuation without motion at faster speeds. As we have shown, IC neurons can significantly synchronize beyond the perception of "motion". So how does the transition from "motion" to "fluctuation" occur?

The limit of human perception of motion has been studied for pure tones with a time-varying interaural phase (Perrott & Nelson 1969; Perrott & Musicant 1977), broadband noise with a time-varying ITD (Grantham & Wightman 1978; Siveke et al. 2008), click trains (Blauert 1972), and harmonic complex tones (Feron et al. 2010) (for review see Carlile & Leung, 2016; Shackleton & Palmer, 2010). In these studies, a sound is no longer perceived as moving above about 8 Hz or for switching periods of around 200 ms (Feron et al, 2011, measured a limit of 3 Hz, although participants continued to hear the sounds moving back and forth between the ears at higher speeds). Similarly, Grantham (1986) and Chandler & Grantham (1992) showed that the minimum angle at which listeners identify a sound as "moving" asymptotes for sound durations above 100-600 ms, suggesting that the perception of motion may be limited by a minimum integration time in the auditory system. Interestingly, this temporal limit is also similar to

the limit for identifying the direction of motion (Carlile & Leung, 2016, Fig. 2). Thus, the limiting duration of around 100 ms may be the duration below which the trajectory of a sound can no longer be identified. If there are direction-selective neurons in the auditory system, they should exhibit a similar limitation.

The earliest stage of the auditory system to exhibit selectivity to the direction motion is the inferior colliculus. Direction sensitivity has been observed in this nucleus in cats (Altman 1968; Spitzer & Semple 1991; Spitzer & Semple 1993), gerbils (Spitzer & Semple 1991; Spitzer & Semple 1993), guinea pigs (McAlpine et al. 2000; Ingham et al. 2001), rabbits (Douglas C Fitzpatrick et al. 2009), and owls (Wagner & Takahashi 1992; Wang & Peña 2013). Specifically, these studies have observed differences in firing rates at a given position of a moving sound depending upon the direction of motion. However, many modeling studies have shown that this apparent direction selectivity can be explained by rate-based adaptation (Cai et al. 1998; Borisyuk et al. 2002; Wang & Peña 2013), since high firing rates result in a firing rate suppression at later times, irrespective of the starting position of the stimulus (McAlpine et al. 2000). Thus, the "direction selectivity" observed here is different from true direction selectivity observed in the visual system, where selectivity to a particular direction of motion is the same irrespective of the spatial extent of the motion or stimulus type (Barlow & Levick 1965; Albright 1984). That said, the auditory system can adapt to the direction of motion over a long period of time, resulting in motion aftereffects (Grantham 1989; Dong et al. 2000), which suggests that there may still be weak direction selectivity present in the responses of these neurons or others at higher levels of the auditory system. In order to properly identify direction selectivity, a neuron's "direction selective" response needs

to be dissociated from its response to the instantaneous position of the sound and its rate-dependent adaptation.

Additionally, while we may not be able to perceive motion in short duration sounds, we should be able to identify that they are present even amidst a noisy background. This should be true if we can "glimpse" the spatial position of multiple sources for time-varying ITD faster than speeds where we perceive motion. More specifically, if we present a moving "probe" embedded in a binaurally uncorrelated noise "masker", our ability to detect the probe should extend to speeds where we can no longer identify its direction. The ability to detect a short probe noise embedded in binaurally uncorrelated noise is known as a "binaural gap detection" task (Boehnke et al, 2002; Luddemann et al 2016). For the specific condition where the masker is binaurally uncorrelated noise, threshold durations are between 20-40 ms (Boehnke et al, 2002; Luddemann et al 2016), which is shorter than the threshold durations for motion detection and direction identification measured in previous studies. How the difference between these two threshold durations comes about is still unclear, but the difference suggests that a neuron may fire at a distinct rate for the moving sound relative to the uncorrelated noise "masker" even when the time-varying ITD is too fast for the neuron to encode motion direction.

Here, we recorded from neurons in the IC of unanesthetized rabbits in response to linearly time-varying ITDs ("ITD sweep") embedded in binaurally uncorrelated noise. To capture both fast and gradual dynamics in firing rate that cannot be captured with a typical histogram approach, we modeled each neuron's response using a point process model whose spiking probability was determined by ITD-following, direction-selective,

and spike history components. We show that the responses of neurons in the IC are largely dominated by ITD-following. Furthermore, we presented these same ITD sweeps to human listeners and asked them in two separate tasks to either identify the motion direction of the sweeps or detect the presence of a sweep. We show that a classifier that uses single neuron responses matches human thresholds for both identifying the sweep direction and detecting the sweep.

Methods:

ITD Sweep stimuli:

All ITD sweep stimuli used broadband noise (50 Hz – 15 kHz). The "sweeps" were a linearly time-varying ITD that changed from -300 to +300 μs ITD (the "positive-going" sweep, or "+ sweep") or from +300 to -300 μs (the "negative-going" sweep, or "- sweep") (see **Figure 1a** and **Figure 2** for an examples). For the psychophysical task, a positive ITD refers to an ITD that is leading in the right ear. For the neurophysiology, a positive ITD is leading in the ear contralateral to the recording site. Sweep durations varied from 1000 ms to 31.25 ms in octave steps. Because the range of the ITD sweep was always the same (-300 to +300 μs ITD), shorter sweep durations corresponded to faster rates-of-change of ITD. Each of these sweeps was preceded and followed by binaurally uncorrelated broadband noise such that the total duration of each stimulus was 2 s. Qualitatively, the binaurally uncorrelated noise was perceived as spatially diffuse or split into two different noises, one at each ear. During the sweep for long sweep durations the noise sounded like a moving, spatially focused image.

We generated a time-varying ITD in these noise stimuli in a similar manner to the dynamic ITD stimuli in Chapter 2. To create the sweeps, an unfiltered noise was initialized at a 100 kHz sampling rate (40 kHz for the psychophysics). In one ear (the left ear for psychophysics, the ipsilateral ear for neurophysiology), the noise was upsampled by 10x to a sampling rate of 1 MHz (400 kHz for psychophysics) using an FFT. From the upsampled noise, the noise was then resampled with delays based on the linear ITD trajectory of the sweep with a resolution of 1 μs (2.5 μs for psychophysics). The noise with the time-varying delay was then preceded and followed by an independent noise, resulting in a binaurally uncorrelated noise in each ear before and after the sweep. The overall stimulus was then filtered between 50 Hz and 15 kHz in both ears. Qualitatively we were unable to perceive any spectral distortions in the ear containing the time-varying delay, and monaurally the ITD sweep could not be detected.

Psychophysics:

All psychophysics experiments were done in a soundproof booth and sounds were presented using Sennheiser HD 280 pro over-ear headphones. All ITD sweeps started and ended with binaurally uncorrelated noise such that the overall duration of each stimulus was 2 s. The sweeps had a span of 600 μs with a center ITD of 0, -300 or +300 μs. The different center ITDs were included to prevent participants from simply identifying the motion direction based on the lateral position of the sweep at a particular moment in time (e.g. at the end of the trajectory). Participants were trained to identify the motion direction of 2-second long ITD sweeps that started and ended with 0.5 seconds of binaurally uncorrelated noise (total stimulus duration of 3 seconds). Each

block contained 12 trials, six of which were centered on 0 µs, three centered on -300 µs and the last three centered on +300 µs. Center ITDs were randomly varied within a block. The order of the center ITDs within a block was random. Participants completed 2 – 6 blocks of training. Initially, 6/33 participants were trained instead on 1-second long sweeps that started and ended with 0.5 seconds of uncorrelated noise, identical to the longest-duration sweeps tested. For these stimuli, each block consisted of 20 trials: ten centered on 0 µs, five centered on -300 µs and five centered on +300 µs. We switched to the other training regimen after the first 6 participants because the 1-second sweep duration was not hypothesis neutral.

After training, participants did the ITD sweep direction identification task using the 2-second long ITD sweep stimuli (**Figure 1a**). Within a block, all ITD

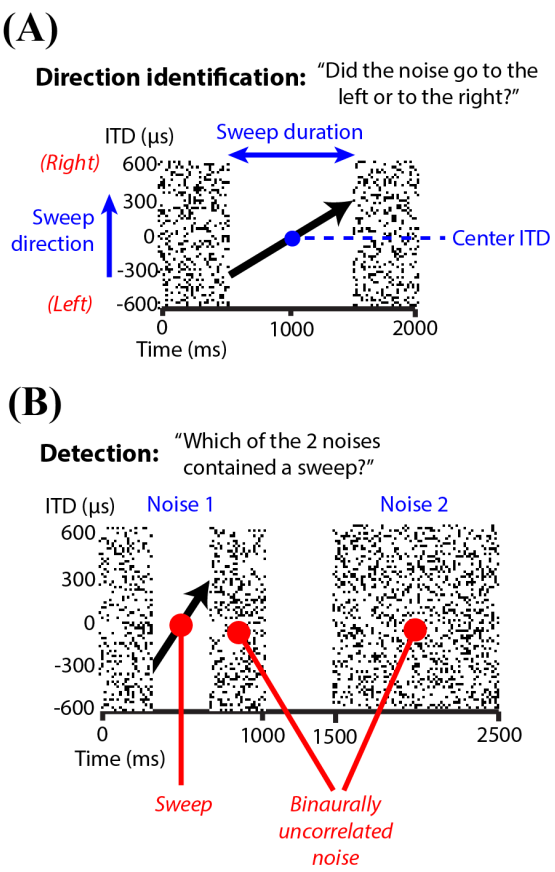


Figure 1: (A) Stimulus for the direction identification task. Participants were presented with a 2 second long noise that contained a sweep (black arrow) flanked by binaurally uncorrelated noise (speckled region). Participants were asked to identify the direction of motion of the sweep, either to the left or to the right; a sweep moving to the right is shown here. The corresponding ITDs are shown. Center ITDs were either -300, 0, or 300 µs and randomly varied between trials in a block; a center ITD of 0 µs is shown here. Sweep durations randomly varied across blocks. The center ITD of the sweep always occurred at 1000 ms. (B) After the direction identification tasks, participants performed a sweep detection task. They were presented with 2 1-second long noises separated by 500 ms of silence. One of the two noises had a sweep, which always had a center ITD of 0 µs and the sweep direction could be to the left or to the right. Participants were asked to identify which of the two noises contained a sweep. Sweep durations varied across blocks of trials.

sweeps had the same duration. Each block consisted of 20 trials: for ten trials, the ITD sweeps were centered on 0 μs, for five trials sweeps were centered on -300 μs, and for five trials sweep were centered on 300 μs. Initially, listeners were given blocks where ITD sweep durations successively decreased, from 1000 ms to 31.25 ms durations in octave steps. After these first 6 blocks the ITD sweep durations were randomized. Participants completed 24 blocks of this task, four for each of the six sweep durations, giving a total of 80 trials for each duration.

The shortest sweep durations were close to binaural gap detection thresholds for correlated noise embedded in binaurally uncorrelated noise (20-40 ms, see Boehnke et al. 2002; Lüddemann et al. 2016). Because of this, direction identification could theoretically be limited by the participant's ability to detect the sweep rather than an inability to identify motion direction per se. Alternatively, the thresholds for sweep detection and direction identification may be distinct. To examine these possibilities, participants performed a sweep detection task following the direction identification task (**Figure 1b**). Participants were presented with two 1-second long noises on each trial separated by 0.5 seconds of silence. We used shorter stimulus duration for the detection task in order to reduce testing time. One of the two noises contained a short segment containing a sweep that was 125 ms, 62.5 ms or 31.25 ms in duration. The segment was padded with binaurally uncorrelated noise in order to maintain an overall stimulus duration of 1 second. The sweeps were centered on 0 μs and the motion direction was randomized across trials. The other interval on each trial contained binaurally uncorrelated noise with no sweep. Participants were asked to identify which of the two noises contained a "sweep", a term that they were familiar with from the previous

direction identification task. Typically participants mentioned that they performed this task by identifying a gap in the binaurally uncorrelated noise. Sweep durations were constant within each block and randomized across blocks. Listeners completed 12 blocks of this task, four for each sweep duration.

Psychophysics analysis:

We computed the percent correct responses for direction identification for each sweep duration and for each center ITD. Nearly all participants showed a sigmoidal variation in performance going from around 100% correct at the longest sweep durations to around 50% correct at the shortest durations. However, there was a wide variability in the slopes and midpoints of these curves. We used nonlinear least squares to fit a sigmoid model to the percentage of correct responses for sweeps centered on 0 μs. The model took the form:

$$PC(d) = \frac{50}{1 + \exp\left[\frac{-(\log_2 d - \log_2 d_{75})}{\tau}\right]} + 50$$

We looked at two parameters: d_{75}, the threshold duration at 75% correct, and the slope through d_{75} equal to $50/4\tau$.

Similarly, the percentage of correct responses for the sweep detection task tended to be near 100% around the longest sweep durations tested and dropped below 75% at 31.25 ms. We fit the same sigmoid model to the sweep detection percentage of correct responses in order to identify d_{75} and the slope through d_{75} for detection.

Neurophysiology:

We recorded from single neurons in the IC of two unanesthetized Dutch-belted rabbits. The procedure for attaching a headpost, craniotomy, and neural recording is identical to that used in the previous chapter and has been described in other studies (Devore & Delgutte 2010; Day et al. 2012). All recordings here were done using a flexible polyimide linear multielectrode array (Microprobe).

IC neurons were identified based on their response to a search stimulus consisting of two 200 ms bursts of broadband noise, one at 0 μs and one at +300 μs, separated by 200 ms and repeating every 1000 ms. Spike times were measured by threshold crossing.

Stimuli, neurophysiology:

Once a neuron was isolated, we measured its frequency response map to 100 ms pure tone bursts separated by 100 ms gaps of silence. The tones were randomly varied between frequencies of 20 Hz and 18 kHz in ¼-octave steps and sound levels of 5 dB SPL to 70 dB SPL in 5 dB steps. Each frequency-level pairing was repeated 1-3 times. We think many of our recordings were made in the central nucleus of the IC because of an observed increase in best frequency with increasing depth and a lack of habituation to successive repetitions of the tonal stimuli, which are characteristics of the central nucleus (Aitkin et al. 1972; Aitkin et al. 1975).

ITD tuning was measured using 300 ms bursts of broadband noise separated by 300 ms of silence. The ITD of the noise was randomly varied between -1500 and +1500 μs in 150 μs steps, and each ITD was presented ten times. We looked at the average firing rate during the stimulus presentation as a function of static ITD. The significance

of the ITD tuning was assessed based on an ANOVA of the firing rates (Hancock et al. 2010; Day et al. 2012; Chung et al. 2016). Only neurons with significant ITD tuning were tested with sweep stimuli ($p < 0.001$).

The presentations of each sweep duration and direction were randomly interleaved. In addition to the positive-going sweep and negative-going sweep stimuli, 2-second-long "0 μs ITD" stimuli were also presented, randomly interleaved with the ITD sweep stimuli (see **Figure 2** for all stimuli). The 0 μs ITD stimuli started and ended with binaurally uncorrelated noise and a noise at a constant 0 μs ITD was presented for the sweep duration instead of a sweep. In 46/62 neurons, we also presented a 2-second long binaurally uncorrelated noise stimulus (**Figure 2**). Each of the ITD sweeps and the uncorrelated noise were presented 10-15 times.

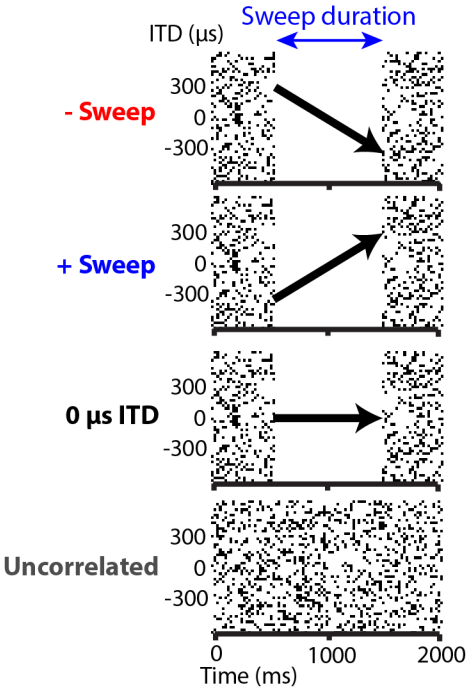


Figure 2: We presented each neuron with a positive-going sweep, a negative-going sweep, a "0 μs ITD" stimulus, and 2 seconds of binaurally uncorrelated noise. Each of these stimuli was presented 15 times for 6 different stimulus durations: 1000, 500, 250, 125, 62.5, and 31.25 ms. All of the sweeps had a range of 600 μs and were centered on 0 μs ITD. Positive ITDs are for ITDs contralateral to the IC recording site.

Single unit spike isolation and stability:

Each unit required at least 25 minutes of recording time during which the spike height and shape could vary. The isolation of the neuron and the stability of the

recording were tested post hoc based on two criteria. Firstly, at each spike time the spike waveforms were extracted in a 1 ms window (50 sample points) with the peak centered at 0.5 ms. The peak-to-peak voltage (V_{pp}) for each spike was then calculated. Then, a number of "noise" segments equal to the number of spikes was retrieved by randomly sampling segments of the signal that did not overlap with spikes or with each other. The V_{pp} of each of these noise segments was also calculated. Within each 10-second period of the recording, the signal-to-noise ratio (SNR) of the mean V_{pp} of the spikes and the mean V_{pp} of the noise was calculated. A neuron was excluded if its SNR dropped below 2.5 during any 10-second period of the overall recording time.

Secondly, we assessed the stability in the spike shape. The signal and noise segments were normalized so that each had a minimum of 0 and a maximum of 1. Principal components analysis was then used to transform these normalized waveforms into a set of orthogonal dimensions that maximized their overall variance. As a result, each spike has a set of coordinates ("scores") for each dimension of the orthogonal space, transformed by a set of coefficients. The distance between the noise and the spike distribution of scores was calculated using:

$$d' = \sqrt{\sum_i \frac{(\mu_{i,spike} - \mu_{i,noise})^2}{\sigma_{i,spike}\sigma_{i,noise}}}$$

where each μ_i and σ_i are the mean and standard deviation of the distribution of scores for the signal and noise along one of the orthogonal dimensions. Similarly, we calculated the *d'* to assess the consistency of the spikes between recordings for different stimulus presentations (pure tones, static ITD noise, ITD sweeps). The normalized spikes were transformed using the same coefficients as those identified for spikes in an earlier

recording, and the *d'* was calculated between the reference spikes and the later spikes. A recording was rejected if either the *d'* between the spikes and the noise was below 4 or if the *d'* between successive recordings of the same neuron was greater than 3. Typically the *d'* between different units had values of 5 and above.

A neural recording was identified as a single unit based on visual inspection of the spike shape variance, spike sorting following PCA on the original non-normalized and the normalized spikes, and based on the percentage of inter-spike intervals less than 1 ms which was typically <1% of the total number of ISIs (Day & Delgutte 2016). While the latter criterion was passed for all recordings for the static ITD and ITD sweep stimuli, it occasionally failed for recordings of the frequency response areas using pure tones that sometimes had 1-2% of ISIs below 1 ms. When this occurred, it was often because there were fewer spikes, so we did not exclude these neurons.

Characteristic frequency:

The characteristic frequency (CF) of the neuron was identified from frequency response areas based on the method described in Palmer et al. (2013). First, the mean firing rate during the stimulus presentation was calculated for each frequency and level. A distribution of "spontaneous" firing rates was calculated by computing the mean and standard deviation of the average firing rates of the neuron at the lowest sound level (5 dB) for all frequencies. The frequency response area (FRA) was then filtered with a 2-D kernel that was triangular in level with a 2.5 dB half-width and Gaussian in frequency with a standard deviation of 1/8 octave. For each frequency we identified the lowest "threshold" level for which the firing rate was either greater than the mean plus 4 SD of

the spontaneous firing rate or greater than the mean spontaneous rate plus 0.15 times the range of the original, unfiltered firing rates (see Palmer et al, 2013). The CF was then calculated by quadratic interpolation of the original firing rate at the frequency corresponding to the lowest threshold and the neighboring two frequencies at the same level. The CF was then validated by visual inspection. For 2 inhibitory neurons where tone bursts only reduced the firing rate, the CF was identified by visual inspection.

Generalized linear model of ITD sweep response:

We expected that the instantaneous spiking probability of each unit would be a function of the instantaneous ITD of the ITD sweep ("ITD-following"), the direction of motion of the ITD sweep ("direction selectivity"), and the recent history of spiking. Here, we assume that both the contribution of ITD-following and direction selectivity are dependent on the ITD and the sweep duration. The ITD-following component is a one-to-one mapping of the instantaneous ITD to the neuron's instantaneous spike probability that should hypothetically match the neuron's response to the static ITDs. If a neuron just followed the ITD of the sweep and was not direction selective, the response to the negative-going sweep would be a time-reversed version of its response to the positive-going sweep (**Figure 3a**). The direction selectivity component captures differences in spike probability associated with the direction of ITD motion at the instantaneous ITD of the sweep. A perfectly direction selective neuron that does not follow the ITD of the sweep would show enhanced firing in one direction of motion and suppressed firing in the other direction at the same instantaneous ITD (**Figure 3b**).

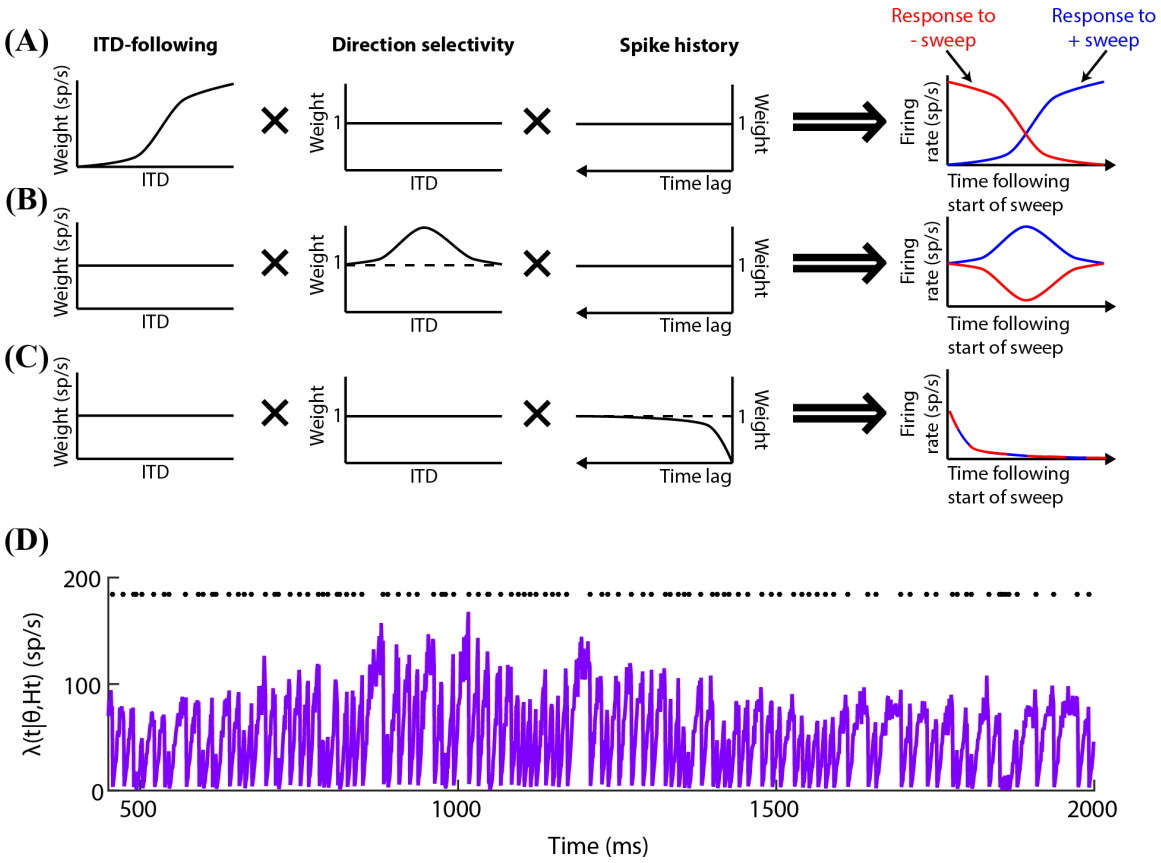


Figure 3: (A) If a neuron shows a variation in ITD following weights and the direction selectivity and spike history weights are constant at 1, then the neuron's response to the "- sweep" would be time-reversed version of the response to the "+ sweep". (B) If a neuron did not follow the ITD (the ITD-following weights were constant), and its direction selectivity was greater than 1, then the neuron's firing rate would be enhanced for the "+ sweep" and suppressed for the "- sweep". (C) If the neuron did not show ITD-following or direction selectivity and its spike history components were below 1 for short pre-spike times, then we would expect the average response to have an immediate suppression in firing following the transient between the binaurally uncorrelated noise and the sweep. (D) Example conditional intensity function for a neuron responding to a 1000-ms positive-going sweep. In this case the spike history weights are suppressive. Spike times are shown by the black dots at the top. Notice that the spike history weights reduce the probability of firing immediately after each spike.

In previous studies, neurons in the IC have shown apparently direction-selective responses to variations in the motion direction of a time-varying ITD, but these responses could qualitatively be explained by firing rate adaptation (Cai et al. 1998; Borisyuk et al. 2002; Wang & Peña 2013). We therefore included a spike history dependent component to separate the contribution of rate adaptation from a potential contribution of direction selectivity. Spike history has been used in the past to model the effects of firing rate

adaptation (Truccolo et al. 2005; Pillow et al. 2005). Even if a neuron showed dependence of firing rate on ITD and no direction selectivity, it could still respond dynamically following the change from binaurally uncorrelated noise to the sweep, because a step increase in firing will cause a suppression of firing at later times (**Figure 3c**). The result is on average a monotonically decaying firing rate, but locally the spike history acts by suppressing (or enhancing for some neurons) the spike probability following each spike (**Figure 3d**).

These three components (ITD-following, direction selectivity, and spike history) determine the neuron's instantaneous probability of spiking. We used a generalized linear model to estimate the time-varying conditional intensity function of this inhomogeneous Poisson process (Brown et al. 2002). The conditional intensity function was modeled as:

$$\lambda(t|\theta, H_t) = \exp\left[w_{ITD}(ITD(t), d) + sgn(DIR)\times w_{DIR}(ITD(t), d) + \sum_{u=i}^{U} h_u n_{t-u}\right]$$

The rightmost term is the spike history dependence, which weights each of the spike counts n_{t-u} at time lags up to U = 300 ms preceding the current time t, by the weights h_u. The weights in w_{ITD} and w_{DIR} are both functions of the instantaneous ITD ($ITD(t)$) and the sweep duration (d). These weights were defined for a discrete set of ITDs, specifically -300, -150, 0, 150, and 300 μs. At any time t when the instantaneous ITD was not equal to one of these discrete ITDs, w_{ITD} and w_{DIR} were determined using linear interpolation:

$$w(ITD(t), d) = w_i(d)\frac{ITD(t) - ITD_{i-1}}{ITD_i - ITD_{i-1}} + w_{i-1}(d)\frac{ITD_i - ITD(t)}{ITD_i - ITD_{i-1}}$$

where w_i and w_{i-1} correspond to the weights for ITD_i and ITD_{i-1} respectively for $ITD_i > ITD_{i-1}$. The direction selectivity component w_{DIR} is multiplied by the sign of the motion direction (*sgn(DIR)*), which is 1 for a positive-going sweep, -1 for a negative-going sweep, and 0 for a stationary stimulus. During the binaurally uncorrelated noise preceding and following each sweep (or 0 μs ITD stimulus), the conditional intensity function was fit by the sum of a constant weight for the binaurally uncorrelated noise (w_{Uncor}), a direction selectivity weight of 0, and the same spike history weights as in Equation (1). Overall, the parameters in θ include the ITD-dependent weights (w_{ITD}), the weight for the uncorrelated noise (w_{Uncor}), the direction selectivity weights (w_{DIR}), and the spike history weights (h).

A full description of model fitting can be found in the Appendix. Briefly, model fitting was done in two steps. First, the model was fit to all 0 μs ITD stimuli simultaneously (six durations, 10-15 trials), which fit w_{Uncor} and h in the absence of any ITD motion. Next, the remaining components, w_{ITD} and w_{DIR}, were estimated by fitting the model separately to the sweeps for each sweep duration while holding w_{Uncor} and h at their previously fitted values. In doing so, we assume that the intensity function for the uncorrelated noise and the history component are invariant to the sweep duration or ITD motion direction. All model fitting was done on the spike times binned at 1 ms. For one neuron we did not include the history components in the fitted model because the fitted history components were either equal to or greater than 1 which resulted in a monotonically increasing average model response that could not reliably be used for classifying direction identification and detection (see *Direction identification of ITD sweeps from neural responses* later in Methods).

The conditional intensity function equation above can also be expressed as:

$$\lambda(t|\theta, H_t) = \exp(w_{ITD}[ITD(t), d]) \times \exp\,(sgn(DIR) w_{DIR}[ITD(t), d]) \times \prod_{u=1}^{U} \exp\,(h_u n_{t-u})$$

We use this formulation of the model when representing each of the components (as in **Figure 3**). The exponential of the ITD-following weight describes the neuron's firing rate at each instantaneous ITD and sweep duration. We express this component in spikes per second. This quantity is multiplied by the exponential of the direction selectivity weight times the sign of the ITD sweep direction, which acts to facilitate or suppress the firing rate. Lastly, the exponential of the history components multiplicatively suppresses (or enhances for some neurons) the firing rate based on the recent spiking history.

After the model was fit for each neuron, we identified the minimum time lag for each neuron that had a spike history weight less than 0.5 (the "halfpoint"). We used this to compare the range of time lags during which spike history components were suppressive across the population. We chose to use a nonparametric approach because there was a wide variation in the shapes of the spike history components.

Goodness-of-fit:

We assessed the goodness-of-fit of the model using the time-rescaling test (Brown et al. 2002; Truccolo et al. 2005). The interspike intervals (ISI) were rescaled by the monotonic transform:

$$z_i = 1 - \exp\,[-\int_{t=s_{i-1}}^{s_i} \lambda(t|\theta, H_t) dt]$$

where the conditional intensity function is integrated from the previous spike time s_{i-1} to the current spike time s_i. If the conditional intensity function properly captures the variation in firing rate, the distribution of z_i should be uniform over [0,1] (Brown et al. 2002). From this we assessed the goodness of fit using a Kolmogorov-Smirnov test comparing the cumulative distribution function of z to a uniform distribution of an equal number of ISIs. We computed the goodness-of-fit separately for each ITD sweep stimulus separately. We also used this test to assess the goodness-of-fit to the binaurally uncorrelated noise which was not included in the initial model fit.

Effects of stimulus duration on ITD-following:

We quantified the changes in the ITD-following weights with stimulus duration in two ways. Let R_{ITD} be the ITD-following weights expressed as a firing rate, namely:

$$R_{ITD}(ITD_i, d) = \exp[w_{ITD}(ITD_i, d)]$$

where ITD_i is -300, -150, 0, 150, or 300 μs ITD. First we computed the duration-dependent "dynamic gain" (also called "gain") equal to the ratio of the standard deviation of the ITD-following rates to the standard deviation of the static ITD firing rates. The gain quantifies the increase in the range of ITD-following weights with changes in stimulus duration. A gain > 1 means that the range is larger than the rates for static ITD tuning, and a gain < 1 means that there is less variability.

We also computed the Pearson correlation coefficient between the ITD-following weights and the firing rates for the same set of ITDs in the static condition, which quantifies the similarity between the shapes of the firing rates for the ITD-following weights and the static ITD tuning. A maximum correlation of 1 implies that the firing

rate for the ITD-following weights and the static ITD tuning covary with respect to ITD, irrespective of the magnitude of the variability.

We then examined how many neurons had correlations that were significantly > 0 for each sweep duration. The "threshold" correlation value (0.805) was computed for a significance value of $p = 0.05$ using a one-tailed Student's t test. To ensure that the percentage of neurons passing this threshold correlation was not due to chance, we bootstrap resampled the ITD-following weights and static ITD firing rates with replacement to get a sampling of 62 correlations, equal to the number of recorded neurons, and then computed the percentage of correlations that passed the threshold for significance. This was repeated 500 times to get a distribution of percentages corresponding to chance.

Quantifying the relative contributions of ITD-following and direction selectivity:

Once the model was fit to the sweep responses and zero ITD responses at each sweep duration, we created "ITD-following only" and "direction selectivity only" models with conditional intensity functions λ_{ITD} and λ_{DIR} respectively. To create the "ITD-following only" model we set the direction selectivity components to zero ($w_{DIR}(ITD,d) = 0$). For the "direction selectivity only" model we set all ITD-following components to ITD-following weight at 0 μs ITD for that sweep duration ($w_{ITD}(ITD,d) = w_{ITD}(0,d)$). The history components (h) were included in both models. We then computed the likelihood of each model to the spike trains in all trials for each ITD sweep stimulus:

$$l[n|\lambda] = \prod_t \frac{\lambda_t^{n_t} e^{-\lambda_t}}{n_t!}$$

where λ_t is the model conditional intensity function at time t, and n_t is the spike count at time t. Lastly, we calculated the likelihood ratio (LR) between the ITD-following only model and the direction selectivity only model:

$$LR = \frac{l[n|\lambda_{ITD}]}{l[n|\lambda_{DIR}]}$$

A likelihood ratio greater than 1 implies that the ITD-following only model captures the timing of spikes better than the direction selectivity only model. To compare the model fits across neurons for each sweep duration, we summed the log-likelihood ratios for the positive-going sweep and the negative-going sweep.

Direction identification of ITD sweeps from neural responses:

Neurons in the IC respond to ITD sweeps with a time-varying firing rate that provides information about the direction of motion of the sweep. While the spiking of these neurons is highly variable, on average there is a trend in firing that should be distinct for different sweep directions if the neuron is properly capturing the variation in ITD. Theoretically, a part of the brain that is responsible for judging the direction of the sweep has implicit knowledge about the average response of the neuron and uses this information to assess how likely the positive-going sweep or the negative-going sweep explains the neuron's response. Thus, in order to assess how well the ITD sweep direction could be decoded from a neuron's response on each trial, we compared the spike train to an average response template generated by the full model, including ITD-following, direction selectivity, and spike history components, and quantified performance assuming an optimal classifier is using the single-neuron responses to do the task.

For each sweep duration, we fit the full model to the positive-going, negative-going, and 0 μs ITD stimuli with one trial of each stimulus left out. Using the parameters from the generalized linear model, a new conditional intensity function was simulated by stepping through each time bin in the trial, from 450 ms following stimulus onset to 2000 ms, and computing the conditional intensity function. The average intensity function, or $\Lambda(t)$, was created by iteratively regenerating the conditional intensity function and spike counts for each trial and averaging the functions:

$$\Lambda(t) = \frac{1}{M}\sum_{m=1}^{M} \lambda(t|\theta, \widetilde{H}^{m}{}_{t})$$

where $\widetilde{H_t^m}$ is the history of simulated spikes generated on iteration m, and M=500. We computed the average conditional intensity function for each sweep direction ($\Lambda_+(t)$ for the positive-going sweep and $\Lambda_-(t)$ for the negative-going sweep).

We then scored the response in each left-out trial based on its similarity to the average response of the model to each sweep direction by computing the log of the likelihood ratio between the array of spike counts for the trial and each of the average responses for each possible sweep direction, which is the optimal classification for a 2-alternative forced choice task (Green & Swets 1988). The log of the likelihood ratio was computed by:

$$Score_{DIR} = \log\left(p[n|\Lambda_+]\right) - \log\left(p[n|\Lambda_-]\right)$$

where n is the array of spike counts for the left-out trial, and the likelihood of the average intensity function given n is:

$$p[n|\Lambda(t)] = \prod_t \frac{\Lambda(t)^{n_t}\exp\left[-\Lambda(t)\right]}{n_t!}$$

where n_t is the spike count at time t. Each likelihood was calculated for the response over the duration of the sweep. Then, each trial was classified as a positive-going sweep if the score was greater than or equal to 0 and a negative sweep otherwise. The percentage of correct classifications was then computed by dividing by the total number of trials (2x the number of trials for each sweep direction, 30 trials for most neurons). This percentage was computed for each stimulus duration.

ITD sweep detection from neural responses:

Similarly, we assessed the neuron's ability to detect the sweep relative to binaurally uncorrelated noise by assuming that a classifier compares the response during the trial to the expected response (average intensity function) to either sweep and the expected response to binaurally uncorrelated noise. The average intensity function to either sweep ($\Lambda_{+}(t)$ and $\Lambda_{-}(t)$) and the average intensity function to binaurally uncorrelated noise ($\Lambda_{unc}(t)$) were generated iteratively as described above. We then simulated a 2-interval 2-alterative forced choice task similar to the psychophysical detection task in the following manner.

First, we picked one trial containing a sweep in either direction (n_1) and one containing binaurally uncorrelated noise (n_2). Next, we computed the average response of the model for both sweep directions ($\Lambda_{+}(t)$ and $\Lambda_{-}(t)$) as well as the average response to binaurally uncorrelated noise ($\Lambda_{uncor}(t)$). The average intensity functions for each sweep direction were generated based on a model fit to all trials except the selected trial. Then for each of the trials we computed the likelihood of the average intensity function given the array spike counts n as described above. The likelihood was computed between 500

and 1500 ms of the response on each trial, equivalent to the stimulus durations used in the psychophysical task. The score for sweep detection was given by:

$$Score_{DET} = \log\left[\frac{(p[n_1|\Lambda_+(t)] + p[n_1|\Lambda_-(t)]) \times p[n_2|\Lambda_{unc}(t)]}{(p[n_2|\Lambda_+(t)] + p[n_2|\Lambda_-(t)]) \times p[n_1|\Lambda_{unc}(t)]}\right]$$

The detection score is equivalent to the log-likelihood of a 2-interval 2-alternative forced choice task (Green & Swets 1988) where each interval is twice as likely to contain binaurally uncorrelated noise than either sweep and detection can correspond to either a positive-going sweep or a negative-going sweep. We calculated this score for each pairing of trials with binaurally uncorrelated noise and trials for either sweep (2x the number of trials squared, 450 pairings for most neurons). The percentage of correct responses was calculated by the number of scores greater than or equal to 0 divided by the total number of pairings. The percentage of correct responses was calculated separately for each sweep duration.

Neurometric functions for direction identification and detection:

After computing the percentage of correct responses for direction identification and detection at each sweep duration, the same sigmoid model used to analyze the psychophysical data was fit to the percent correct curves for each neuron. We then identified the threshold duration and slope of the neurometric function for both direction identification and for detection.

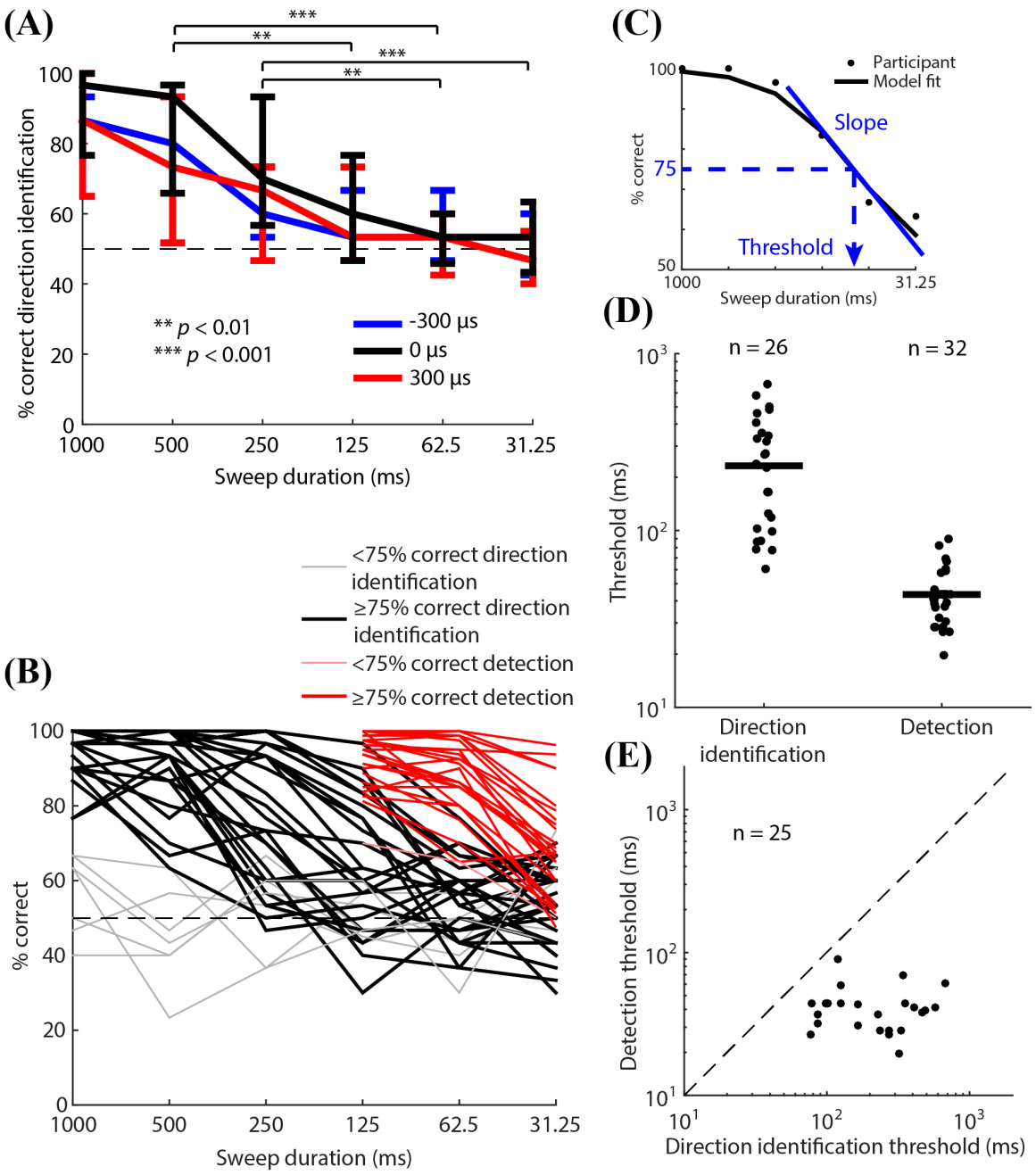


Figure 4: (A) Average performance of the participants for direction identification for center ITDs of 300 µs (red), 0 µs (black), and -300 µs (blue). Significance values were calculated by a Tukey's HSD test. (B) Overall average performance for direction identification (black) and sweep detection (red). Thick lines show participants the performed at or above 75% correct for at least one sweep duration, and thin pale likes show participants with percents correct below this criterion. The participants that performed at or above 75% correct were included in later analyses. (C) The sigmoid model was fit to the percents correct in order to compute the threshold, the duration resulting in 75% correct, and the slope through the threshold duration. (D) Thresholds for direction identification and detection for the participants that passed the 75% criterion. (E) Comparison of individual direction identification and detection thresholds in participants that passed the 75% criterion for both tasks.

Results:

Psychophysics:

Participants were 33 self-described normal hearing listeners (21 male, 12 female), ages from 21 - 64 years old (median = 31 years). The direction identification and sweep detection tasks were performed in blocks of 20 trials where each block contained stimuli of the same duration. Listeners were presented with 4 blocks for each sweep duration. Across all participants, there was a weak but significant increase in performance across blocks in direction identification at the 1000 ms sweep duration (Kruskal-Wallis test with Bonferroni correction for 6 sweep durations, 1000 ms: $\chi^2 = 14.0$, $p = 0.026$) but not at the other durations ($p > 0.05$). However, a difference was only found between the first block and the last 2 blocks (Mann-Whitney U test with Bonferroni correction: $p < 0.05$). For this reason, we chose to look at percent correct scores for direction identification averaged over the last 3 blocks of each sweep duration after the participant's performance seemed to have reached an asymptote. This left a total of 30 trials per duration at a center ITD of 0 μs ITD, 15 trials for -300 μs ITD, and 15 trials for 300 μs ITD. There was no effect of the block number on performance for detection (Kruskal-Wallis test with Bonferroni correction, $p > 0.05$). For detection, 4 blocks of 80 trials were included for each sweep duration.

On average, there was a clear decrease in direction identification performance with a decrease in sweep durations for all center ITDs (Kruskal-Wallis test, -300 μs: $\chi^2 = 43.5$, $p < 0.001$; 0 μs: $\chi^2 = 65.7$, $p < 0.001$; 300 μs: $\chi^2 = 52.6$, $p < 0.001$) (**Figure 4a**). The largest changes in performance occurred between 125-250 ms. We also found a significant difference in performance with center ITD (Friedman test: $\chi^2 = 7.32$, $p =$

0.026) in that the performance for sweeps centered on 0 μs ITD was significantly better than for sweeps centered on -300 and +300 μs ITD (Wilcoxon signed-rank with Bonferroni for 3 comparisons, -300 to 0 μs: $W = 3273$, $p < 0.001$; 0 to +300 μs: $W = 8710$, $p < 0.001$; -300 to +300 μs: $W = 5366$, $p = 0.26$). Even though we did two different training regimens (6 participants were trained with 1000 ms sweeps, 27 were trained with 2000 ms sweeps), we found no significant difference in performance between the two groups for the 1000 ms sweeps during testing (Mann-Whitney U: $U = 107.5$, $p = 0.63$).

In each block we presented sweeps with center ITDs of either -300 or 300 μs in addition to 0 μs ITD in order to prevent listeners from identifying the direction of motion based on the lateralization of the sounds at the beginning or end of the trajectory. However, this did not completely rule out the possibility of judging the direction based on lateralization. For instance, for a sweep moving from -600 μs to 0 μs ITD, the listener could theoretically identify the motion as rightward (a positive-going ITD) by recognizing that -600 was the left-most (most negative) possible ITD of the set of ITDs in the experiment. The same logic can be applied if the sweep ended at -600 μs. To test whether participants used such a strategy, the percentages of correct responses for direction identification were recalculated based on the lateral position at the start of the sweep. If listeners identified the direction based on the lateral position at the start of the sweep, they should do better for sweeps starting from a position on the left or a position on the right than for sweeps with a starting position in the center. If the sweep started in the center, this strategy would result in chance performance. Alternatively, if they identified the direction based on the end of the sweep, they should show the opposite trend in performance. We found no effect of starting position on performance (Friedman

test: $\chi^2 = 4.08$, $p = 0.13$). Thus, we conclude participants were indeed identifying the direction of the ITD sweep based on the change in ITD rather than identification of specific ITDs in the trajectory.

While all participants showed a decrease in performance with decreasing sweep duration, the durations at which this decrease was steepest varied across individuals. As shown in **Figure 4b**, most individuals who could perform the task well (> 75% correct detection for at least one sweep duration) showed a sigmoid trend in performance. The threshold duration at which the sigmoid crossed 75% correct varied across individuals, resulting in a large variability in performance around 250 and 125 ms sweep durations. Furthermore, the median performance for detection was significantly better than direction identification at 125 ms and below (**Figure 4b**) (Mann-Whitney U test with Bonferroni correction: 125 ms, $U = 642.5$; 62.5 ms, $U = 581$; 31.25 ms, $U = 785$; $p < 0.001$ for all sweep durations). Thus, the variability in performance for direction identification could not be explained by an inability to hear the sweep amidst the flanking binaurally uncorrelated noise. Indeed, the threshold durations for direction identification and detection appeared markedly distinct.

We quantified the dependence of sweep duration on performance for direction identification and detection for each individual by fitting a sigmoid model to the percentages of correct scores ("psychometric function") (**Figure 4c**). The model assumed that all participants exhibited a maximum performance of 100% correct for long sweep durations and a minimum performance of 50% correct for short sweep durations. Model fitting yielded two parameters for each individual: the sweep duration resulting in a performance of 75% correct ("threshold duration") and the slope (%correct/octave)

through that midpoint. We only fit this model for individuals that performed at 75% or above for at least one sweep duration (n=26/33 for direction identification, n=32/33 for detection). On average, threshold durations were 5.3 times higher for direction identification than for detection (**Figure 4d**), and the difference between them was significant (median direction identification threshold = 232.3 ms, median detection threshold = 43.4 ms; Mann-Whitney U test: $U = 1172$, $p < 0.001$). Slopes of the neurometric functions were similar for both tasks (median direction identification slope = 24.9 %correct/octave, median detection slope = 20.0 %correct/octave; $U = 809$, $p = 0.52$). The better performance in detection relative to direction identification can be clearly seen by comparing direction identification and detection threshold durations for participant who passed the 75% criterion for both tasks (n=26) (**Figure 4e**). For these participants, there was no correlation between the direction identification and detection threshold durations (Kendall's tau: $\tau = 0.14$, $p = 0.33$) or between the slopes of each ($\tau = 0.16$, $p = 0.43$). This suggests that the factors limiting thresholds are distinct for the two tasks.

Modeling the neural responses to ITD sweeps:

We recorded from 101 ITD-sensitive neurons, determined based on an ANOVA test ($p < 0.001$) (Hancock et al. 2010; Day et al. 2012; Chung et al. 2016). For 62 of these neurons, responses to ITD sweep stimuli were measured for at least 10 trials, and the

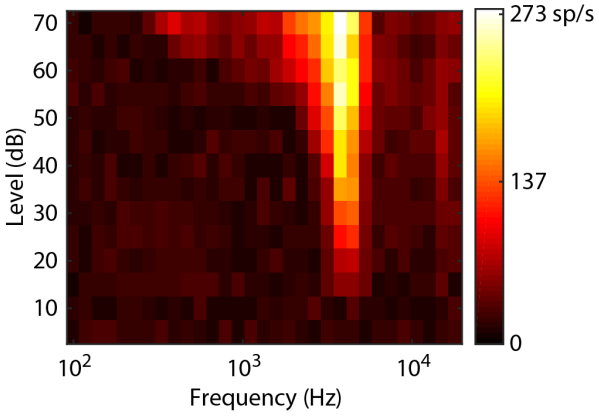


Figure 5: Example FRA for one neuron. The FRA is color coded by the average firing rate during the stimulus, shown in the color bar on the right of each plot. CF = 4150 Hz.

recordings were deemed stable enough and sufficiently well isolated to be included in later analysis. Characteristic frequencies ranged from 293 Hz to 18.1 kHz (median 3.81 kHz, 19 neurons with CF < 2 kHz, 43 with CF > 2 kHz) (see **Figure 5** for example FRAs).

Figure 6a shows the static ITD tuning for an example neuron (CF = 4150 Hz) and **Figure 6b** shows its response to the positive-going (red), negative-going (blue) sweeps, and 0 µs ITD stimulus (black), as well as the response to binaurally uncorrelated noise (grey) at the bottom of the figure. We've plotted the average firing rate across all 15 trials when the average firing rate is computed in a 50 ms bin. The trend in firing rate is clear with the larger bin size. As expected based on the neuron's static ITD tuning, the firing rate starts high and

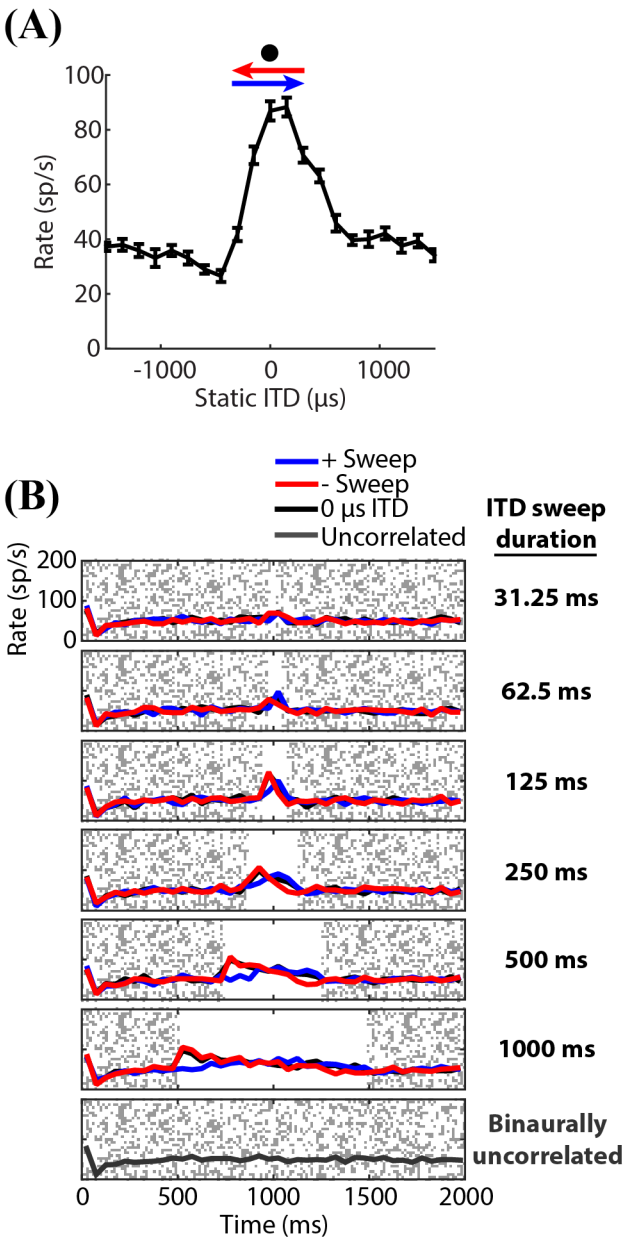


Figure 6: (A) Static ITD tuning curve of the example neuron shown in Figure 5 (CF = 4150 Hz). The trajectories of the ITD sweep stimuli are shown on above the tuning curve (blue arrow for positive-going sweep, red arrow for negative-going sweep, black dot for the 0 µs ITD stimulus). (B) Response of the example neuron in (A) to the positive-going (blue) and negative-going (red) ITD sweeps, the 0 µs ITD stimulus (black), and the binaurally uncorrelated noise in the bottom plot (grey). In each plot the grey speckled regions mark when the noise is binaurally uncorrelated, and the white regions are during the ITD sweeps. These post-stimulus time histograms were calculated using 50 ms bins.

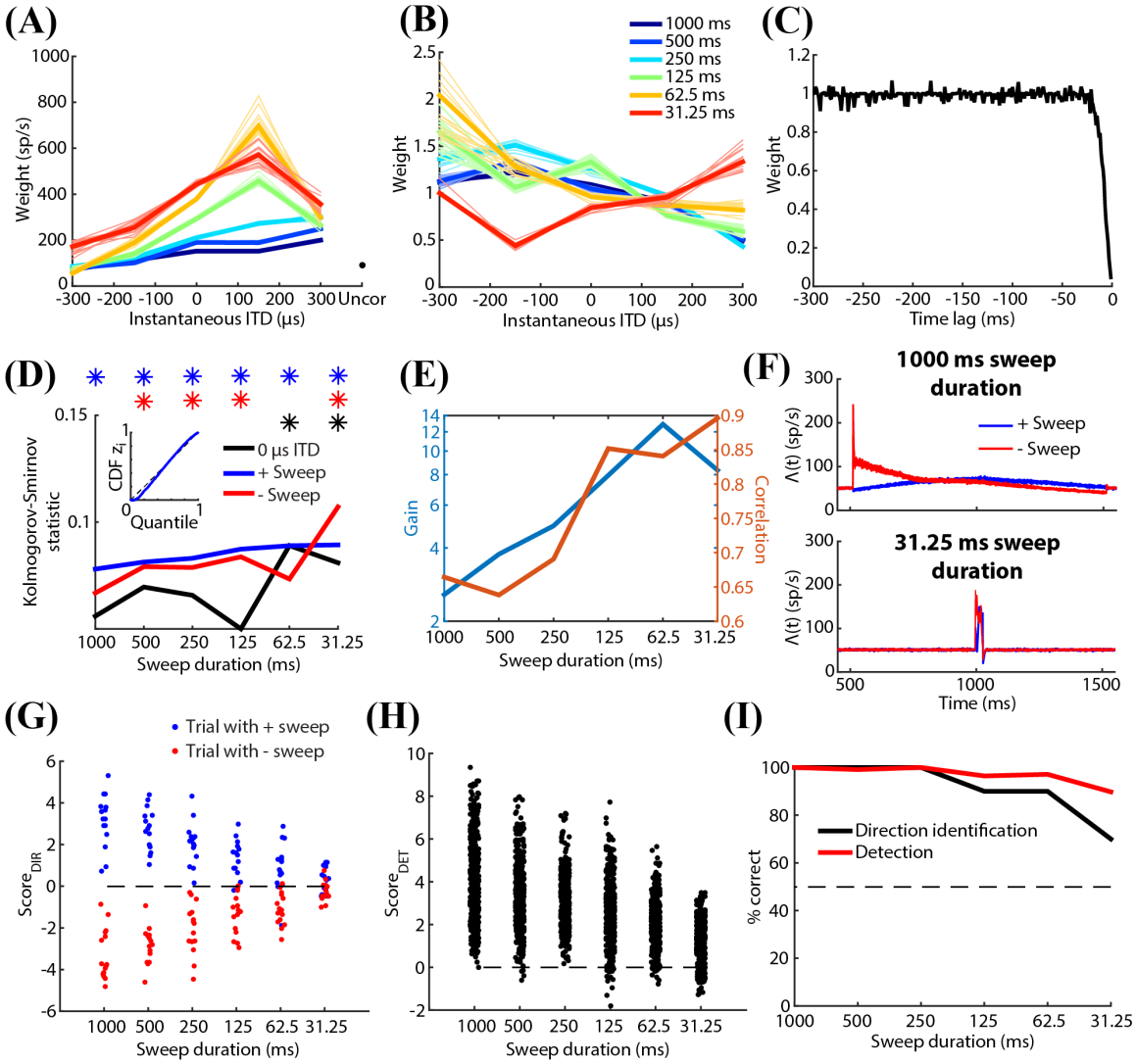
decreases over time as the negative-going sweep goes from positive ITDs to negative ITDs. Similarly, the firing rate for the positive-going sweep increases over time as expected based on the static ITD tuning. Some of the dynamics at the beginning of the sweep are also captured by the response to the 0 μs ITD stimulus, suggesting that some of the response to the negative-going sweep may be explained by adaptation.

However, for the larger bin size, the trends are difficult to see for the shortest sweep durations. If we reduce the bin size, it becomes clear that the neuron does respond very briefly at the shortest sweep durations (not shown). On the other hand, the shorter bin size captures more neural noise and makes the subtle trends in firing at the longest sweep durations hard to see. Our intention in this study was to examine the neural responses to these ITD sweep stimuli and ultimately classify the sweep directions based on these responses. Clearly, examining the firing rate using a histogram approach as shown in **Figure 6b** highly depends upon the bin size. For this reason, we chose to model the firing rates as a point process using a generalized linear model. The model reduced the overall dynamics in the response to a smaller number of parameters in order to avoid overfitting while capturing fast dynamics in the response that could be present at the shortest sweep durations.

We modeled the responses of each of the 62 neurons to the ITD sweep stimuli under the assumption that the log probability of firing at each time point was a linear combination of a set of weights dependent on the instantaneous ITD ("ITD-following"), components that quantified the magnitude of enhancement or suppression in firing depending upon the direction of the ITD sweep ("direction selectivity"), and suppression or enhancement based on spike history. Data from example neuron are shown in **Figure**

7 (CF = 4150 Hz). This is the same neuron shown in **Figure 6**. The ITD-following components (**Figure 7a**) follow the same trend in firing rate that is expected based on the neuron's static ITD tuning (**Figure 7a**). Additionally, the range of the ITD-following weights increases with decreasing sweep duration.

Figure 7:



We modeled the direction selectivity component as a function of both the instantaneous ITD and sweep duration. If a neuron "preferred" one of the two motion directions, all of its direction selectivity components would be either greater than 1 or less than 1. In the rabbit retina, for example, direction selective ganglion cells have a

Figure 7: GLM modeling and analysis for the same example neurons shown in Figure 5 and 6. (A) ITD-following, (B) direction selectivity, and (C) spike history weights for this neuron. The thick lines in A and B show the weights for the GLM fit to all trials. Each of the thin lines are the weights for the models fit to all trials with one trial left out. The halfpoint for this neuron was -7 ms. (D) The Kolmogorov-Smirnov (KS) statistic from the time-rescaling test is shown for each sweep duration and each sweep type (black for 0 μs ITD, blue for positive-going sweep, red for negative-going sweep). This statistic is equal to the maximum distance between the cumulative distribution function for the time-rescaled ISIs and the unity line – an example is shown in the inset. The stars designate KS statistics that fail the time-rescaling test ($p < 0.05$ with Bonferroni correction for 18 conditions). The time rescaling test failed for most sweep durations and sweep types. (E) Gain (blue) and correlation (orange) as a function of sweep duration. Interestingly, the correlation increased with decreasing sweep duration, which was uncommon in the population of neurons studied. (F) Average responses of the model for the positive- and negative-going sweeps with a1000 ms sweep duration (top) and 31.25 ms sweep duration (bottom). The GLM shows the general trends in firing rate for the longest sweep durations while capturing the brief responses for the shortest sweep duration. (G) Direction selectivity scores computed for responses to trials when a positive-going sweep (blue) or negative-going sweep (red) was presented. If the score is equal to or above 0 then the trial is classified as a positive-going sweep, and otherwise it's classified as a negative-going sweep. (H) Detection scores for trial pairings between a sweep and binaurally uncorrelated noise. Scores greater than 0 are classified correctly. (I) Percentage of correct responses for direction identification (black) and detection (red) for this neuron based on the direction identification scores and detection scores shown in (G) and (H). This neuron was produced one of the best performances for ITD sweep detection.

---

preferred direction irrespective of the initial position of the stimulus (Barlow & Levick 1965). Whether or not neurons in the IC have a preferred direction is unclear. There is some evidence for preferred motion directions for time-varying ITDs in IC neurons using rotating stimuli such as binaural beats (Yin & Kuwada 1983; Fitzpatrick et al. 2009; Dietz et al. 2014, Fig. 3d & f at 32 and 64 Hz). Additionally, we expected the shape of the direction selectivity components to be consistent for all sweep durations. In the rabbit retina, for example, a neuron's direction selectivity is consistent for a wide range of motion speeds and displacements (Grzywacz & Amthor 2007).

In the neurons we recorded, we did not find evidence of "preferred" directions for IC neurons, and the shape of the direction selectivity components was often varied across sweep durations. For the example neuron shown in **Figure 7**, there is a shallow but consistent slope that is above 1 for negative ITDs and below 1 for positive ITDs (**Figure**

7b). The slope is present for all sweep durations except the shortest one, where the slope is inverted. We observed this sloping shape in the direction selectivity components in several other neurons.

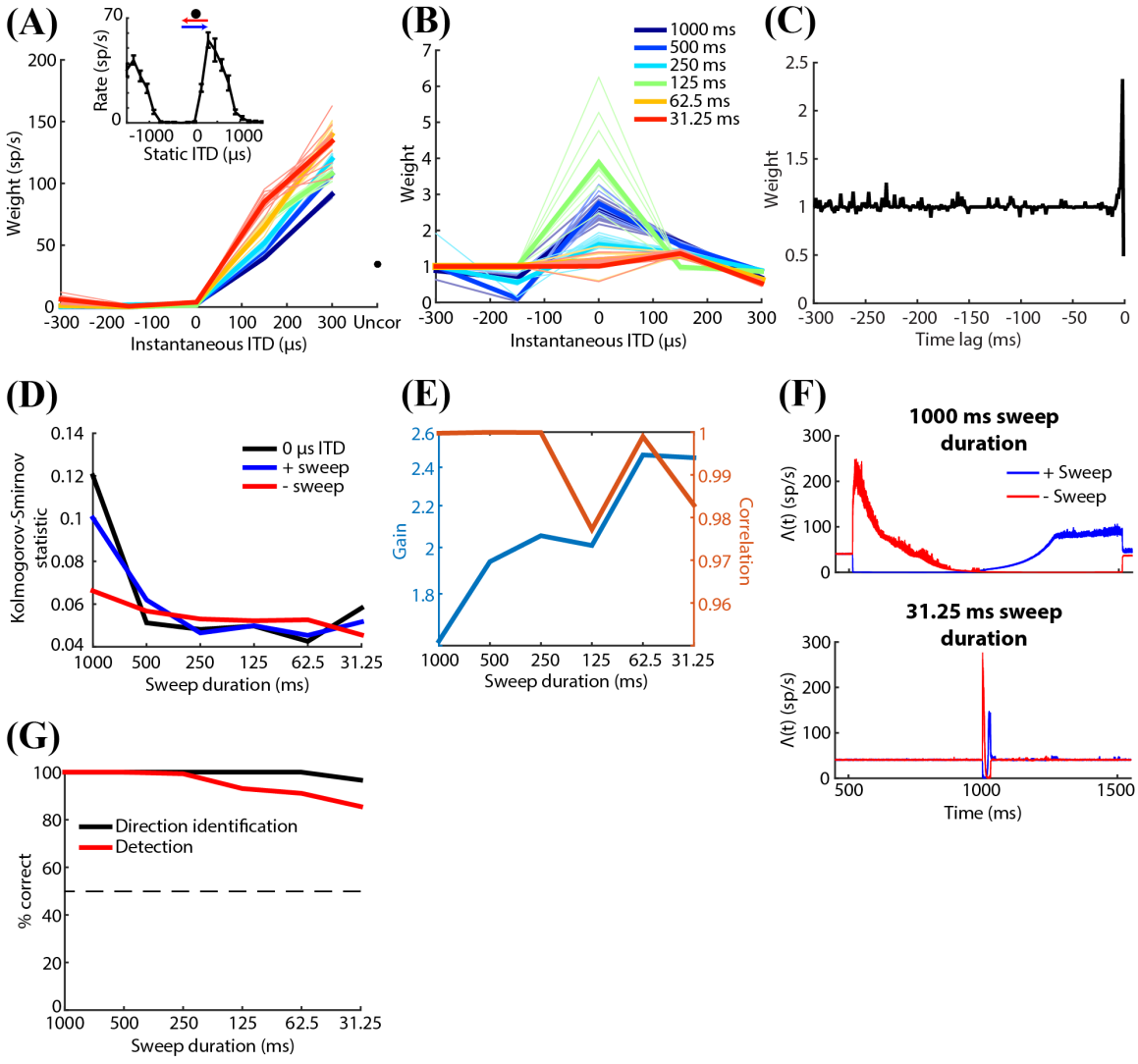


Figure 8: GLM model and analysis for another example neuron with a low CF (CF = 475 Hz), plotted identically to Figure 7. The static ITD tuning curve is shown as an inset to (A). This neuron was one of 8 that had spike history weights > 1.4. The halfpoint was -1 ms. For this neuron, the model significantly fit the data for all sweep durations and sweep types (D) even though the KS statistics are very similar to the neuron in Figure 7. This neuron also had a lower firing rate than the neuron in Figure 7, as can be seen from the model's average response (F). This neuron produced the best performance for direction identification of all of the neurons studied (G).

Data from an example neuron with a low CF (CF = 475 Hz) are shown in **Figure 8**. Again, the neuron's ITD-following component qualitatively matched what was expected based on the static ITD tuning curve (**Figure 8a**, static ITD tuning shown in the inset): for negative ITDs the firing rate was low, and the firing rate increased as the ITD increased. There were no consistent trends in the direction selectivity components for this neuron (**Figure 8b**).

Often the spike history components had an effect only at short time lags. The spike history components rapidly asymptote to 1 for both example neurons shown here (**Figure 7c and 8c**). The neuron in **Figure 8c** showed an enhancement in spiking probability at time lags of -2 to -3 ms rather than the usual suppression, as for the neuron in **Figure 7c**. We occasionally but infrequently observed enhancement at short time lags in our population of neurons: 8 neurons had spike history components that exceeded a magnitude of 1.4. To compare the range of time lags over which the spike history components in our population of neurons were suppressive, we identified the minimum time lag for each neuron that had a spike history weight less than 0.5 (the "halfpoint"). Of the 50/62 neurons that had spike history components below 0.5, the median of the halfpoints was -7 ms (interquartile range of -3 to -9 ms). This is much shorter than the adaptation time constants measured for instantaneous changes in interaural phase difference measured in the IC of anesthetized guinea pigs (52.9 ± 26.4 ms, Ingham & McAlpine 2004).

Taken together with the trends in the direction selectivity components we observed, we think that the spike history components captured very short-term adaptation associated with refractoriness whereas the direction selectivity components may have

captured some longer-term adaptation. Specifically, we observed sloping direction selectivity components, as in **Figure 7b**, in several of the neurons we recorded. If we consider their effect on a neuron where the ITD-following weights are constant, the resulting response would be a steady decrease in firing rate over time for both the positive-going and negative-going sweeps. This trend in firing rate could also be explained by an adaptive component associated with subthreshold events that was not properly captured by spike history.

Assessing goodness-of-fit:

We assessed the goodness-of-fit of each model using the time-rescaling test (Brown et al. 2002; Truccolo et al. 2005). The interspike intervals (ISI) of the response were rescaled based on the predicted firing rate (conditional intensity function) of the model, and the rescaled ISIs were compared to a uniform distribution by a Kolmogorov-Smirnov (KS) test (see inset of **Figure 7d**). If the model properly fits the spiking variability, the rescaled ISIs will be uniformly distributed, and there will be no difference between cumulative distribution for the rescaled ISIs and a uniform distribution with the same number of ISI (the KS plot will be the identity line). The KS statistic is equal to the largest distance between the identity line and the KS plot. The model produced a good fit for many of the neurons in the population ($p > 0.05$ with Bonferroni correction for all 18 ITD sweep stimuli). For example, the fit for the neuron in **Figure 8d** was good for all stimuli. However, there were several neurons for which the model did not produce a good fit for the responses to many of the sweep stimuli, as for the neuron in **Figure 7d**. Comparing the KS statistics to the mean firing rates for each ITD stimulus showed that

the test failed most often when the firing rates were high (**Figure 9**). This is also true for the neuron in **Figure 7**, which had a high firing rate. It is likely that the test failed primarily because there were simply more ISIs for these neurons, since the significance of the KS statistic is inversely proportional to the square root of the number of samples (Feller 1948). We also found that the model fit the responses to binaurally uncorrelated noise well, even though those responses were not included when fitting the model (**Figure 9**, see Appendix for more details).

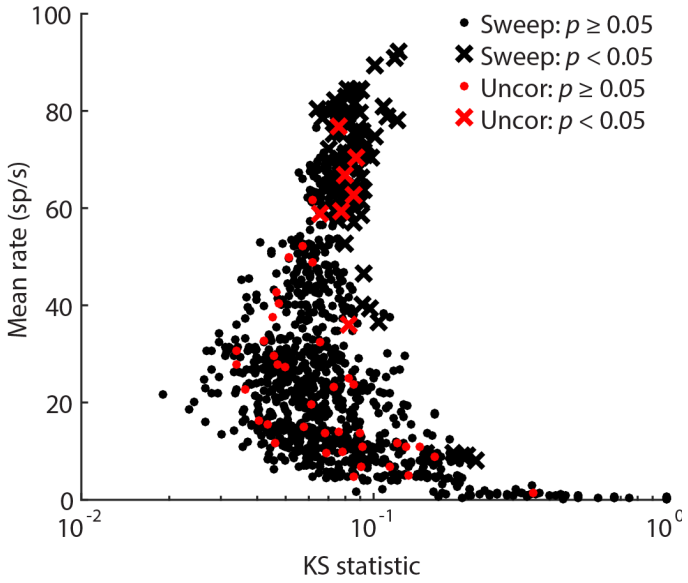


Figure 9: Mean firing rates are plotted against KS statistics for KS statistics that passed (dot) or failed (x's) the time rescaling goodness-of-fit test. Black dots are the plots for the fits to each sweep of each neuron. Failed tests here had KS statistics with $p < 0.05$ with Bonferroni correction for all 1116 comparisons. Red dots and x's are the passed and failed time rescaling tests for the binaurally uncorrelated noise ($p < 0.05$ with Bonferroni for 46 comparisons).

Based on this analysis, our model properly captures the variance in firing for most of the data, but fails primarily for high-firing neurons. Our model only captured spike-dependent adaptation but did not account for other sources of adaptation such as adaptation associated with subthreshold events. Including a stimulus-dependent time lag component (e.g. Weber & Pillow 2016) may further improve the model fit for the high-firing neurons.

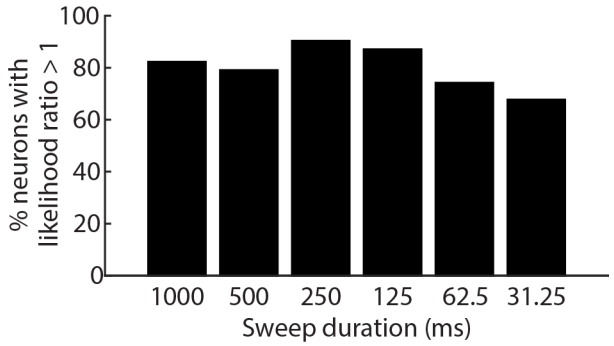


Figure 10: For each neuron, we computed the ratio of the likelihood of ITD-following only model given the neural response to the likelihood of the direction selectivity only model given the neural response. A likelihood ratio greater than 1 means that the ITD-following only model captures the time-varying spike probability better than the direction selectivity only model. The percentage of neurons with likelihood ratios greater than 1 is shown for each sweep duration (n = 62). Many of the neural responses had higher likelihoods using the ITD-following only model than the direction selectivity only model.

Contribution of ITD-following and direction selectivity:

We assessed the contribution of the ITD-following components and the direction selectivity components on each neuron's overall response by computing the likelihood ratio between the model where the direction selectivity weights were removed ("ITD-following only") given the data and a model where the ITD-following weights were removed ("direction selectivity only") given the data. If the neuron's response was dominated by ITD following, then the ITD-following only model will better capture the variability in the neuron's response than the direction selectivity only model, and the likelihood ratio will be greater than 1. If the neuron's response was dominated by direction selectivity, then the likelihood ratio will be less than 1. Many of the neurons had likelihood ratios greater than 1 for all sweep durations (**Figure 10**). There were a number of neurons with likelihood ratios below 1 for at least one sweep duration, but they typically did not go far below zero. Thus, nearly all of the neurons in the population were dominated by ITD following.

Effect of sweep duration on ITD following component:

Since the ITD-following only model captured the firing variability in most neurons, we next analyzed how the ITD-following components were affected by sweep duration. If the inability to identify the motion direction of the ITD sweeps was due to an inability for the auditory system to follow the ITD, then we would expect the range of firing rates in the ITD-following components to decrease with decreasing sweep duration. Instead, we found that the range of the weights of the ITD-following components increased in magnitude with decreasing stimulus duration for many of the neurons we recorded, including the example neurons in **Figures 7 and 8**.

First, we quantified the increase in variability of the ITD-following components by the "dynamic gain", relating the standard deviation in the ITD-following components to the standard deviation in firing rates for ITD tuning. For both example neurons, the gain increased with decreasing sweep duration (blue in **Figure 7e and 8e**). This trend was also apparent across the population of neurons (**Figure 11a**): there was a significant increase in the median gain with decreasing sweep duration (Kruskal-Wallis test: $\chi^2 = 31.8$, $p < 0.001$). Only 8/62 neurons showed a gain for the 31.25 ms sweep duration that was smaller than the gain for the 1000 ms sweep duration. Nevertheless, the median gain at 1000 ms was larger than 1 in many neurons, implying that even at the slowest sweep duration some neurons had a stronger dependence on ITD than predicted from the static ITD tuning.

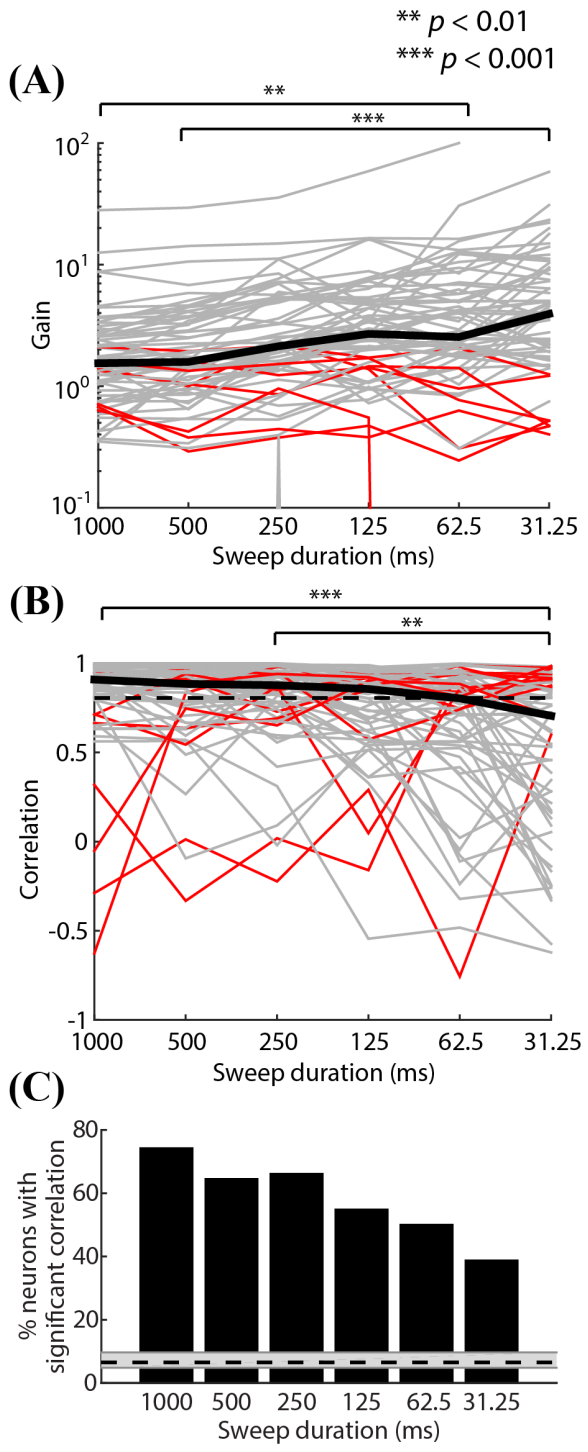


Figure 11: (A) Gain for the population of 62 neurons. Red lines designate neurons with the gain decreased from 1000 ms to 31.25 ms. On average though, the gain increased with decreasing sweep duration. (B) Correlations for the population. Red lines designate neurons where the correlation increased with sweep duration, contrary to the average. In both plots, the thick black line marks the median values. Significance values are based on the Tukey HSD test. A dashed black line shows the threshold for significant correlations with $p <$ 0.05 based on a one-tailed Student's t test. (C) Percentage of neurons whose correlations were significant based on the criterion shown in B. The black dashed line shows the percentage of neurons that would occur by chance, and the grey area designates the upper and lower quartiles of the distribution of percentages (see Methods). For all sweep durations, it is clear that many neurons have ITD-following components hat are significantly correlated with their static ITD tuning.

Second, we used the correlation between the ITD-following weights and the static ITD tuning to quantify changes in the shape of tuning with sweep duration (**Figure 11b**, see **Figure 7e and 8e** in orange for examples). For the majority of neurons correlation decreased with decreasing sweep duration, although 12/62 neurons showed an increase in correlation (as in **Figure 7e**). The median correlation across the population significantly decreased with decreasing sweep duration (Kruskal-Wallis test: $\chi^2 = 31.5$, $p < 0.001$). However, a high percentage of the neurons had ITD-following components that were significantly correlated with their static ITD tuning (Student's t-test, $p < 0.05$, see **Figure 11b** for correlation threshold

and **Figure 11c** for percentages of neurons). Even at 31.25 ms, 40% of neurons had correlations that were significant, and this proportion was well above chance. Taken together, neurons remain similarly tuned to ITD at short sweep durations as they are for static ITDs and in fact show larger variations in firing rate for shorter sweeps. Both of these are contrary to the hypothesis that direction identification is limited by the IC's ability to follow the ITD.

Direction identification and detection performance based on the neural response:

In general, the IC neurons we recorded appear to follow the ITD and maintain their static ITD tuning for all sweep durations. These attributes may be sufficient to allow a classifier both to identify the direction of the sweep and to detect the presence of a sweep. A neuron's average response should be characteristic for a particular sweep direction, and the likelihood that the response on any trial corresponds to a positive-going sweep, a negative-going sweep, or binaurally uncorrelated noise (in the detection task) depends on the likelihood that the spike train on that trial matches some average expected response for either of the sweeps. Thus, we modeled the direction identification task based on the neural spike train in a single trail by computing log of the likelihood ratio of the spike train in the trial given the average responses for the positive-going sweep and the negative-going sweep. We used the log of the likelihood ratio to quantify the performance of an optimal classifier that is using single-neuron responses to do the direction identification and detection tasks, both of which are 2-alternative forced choice tasks (Green & Swets 1988).

We first calculated the average response of the neuron to each sweep direction and each sweep duration by iteratively generating a response to each sweep and averaging the responses for 500 iterations using the model fit to all trials except the left-out trial (see **Figure 7f and 8f**). We then scored the response of each trial by computing the log of the likelihood ratio of the response on the left-out trial given average response to the positive-going sweep relative to the likelihood that the response matches the average response to the negative-going sweep, called $Score_{DIR}$ (**Figure 7g**). A positive $Score_{DIR}$ means that it is more likely that the spike train on a given trial matches the average response to the positive-going sweep, and a negative $Score_{DIR}$ means that it is a better match to the negative-going sweep. We then classified the trial based on whether its score was positive (for "+ sweep") or negative (for "- sweep"). We calculated the percentage of correct responses over all trials (20-30 trials) (**Figure 7i and 8g**).

Similarly, we calculated the average response of the neuron to a constant binaurally uncorrelated noise and then scored the response of a pair of left-out trials. The first one contained the response for a sweep in either direction and the second contained a response for binaurally uncorrelated noise. The score was based on the likelihood ratio corresponding to a 2-interval 2-alternative forced choice task (Green & Swets 1988), called $Score_{DET}$ (**Figure 7h**). This is identical to the paradigm performed by the human participants in our study. When calculating $Score_{DET}$, we paired the trials such that the sweep always came first. A positive $Score_{DET}$ means that it was more likely that the sweep came first (the correct order), and a negative $Score_{DET}$ means the sweep came second (the incorrect order). The percent correct was equal to the number of positive $Score_{DET}$ out of the total number of pairings (200-450 total pairings) (**Figure 7i and 8g**).

Figure 12:

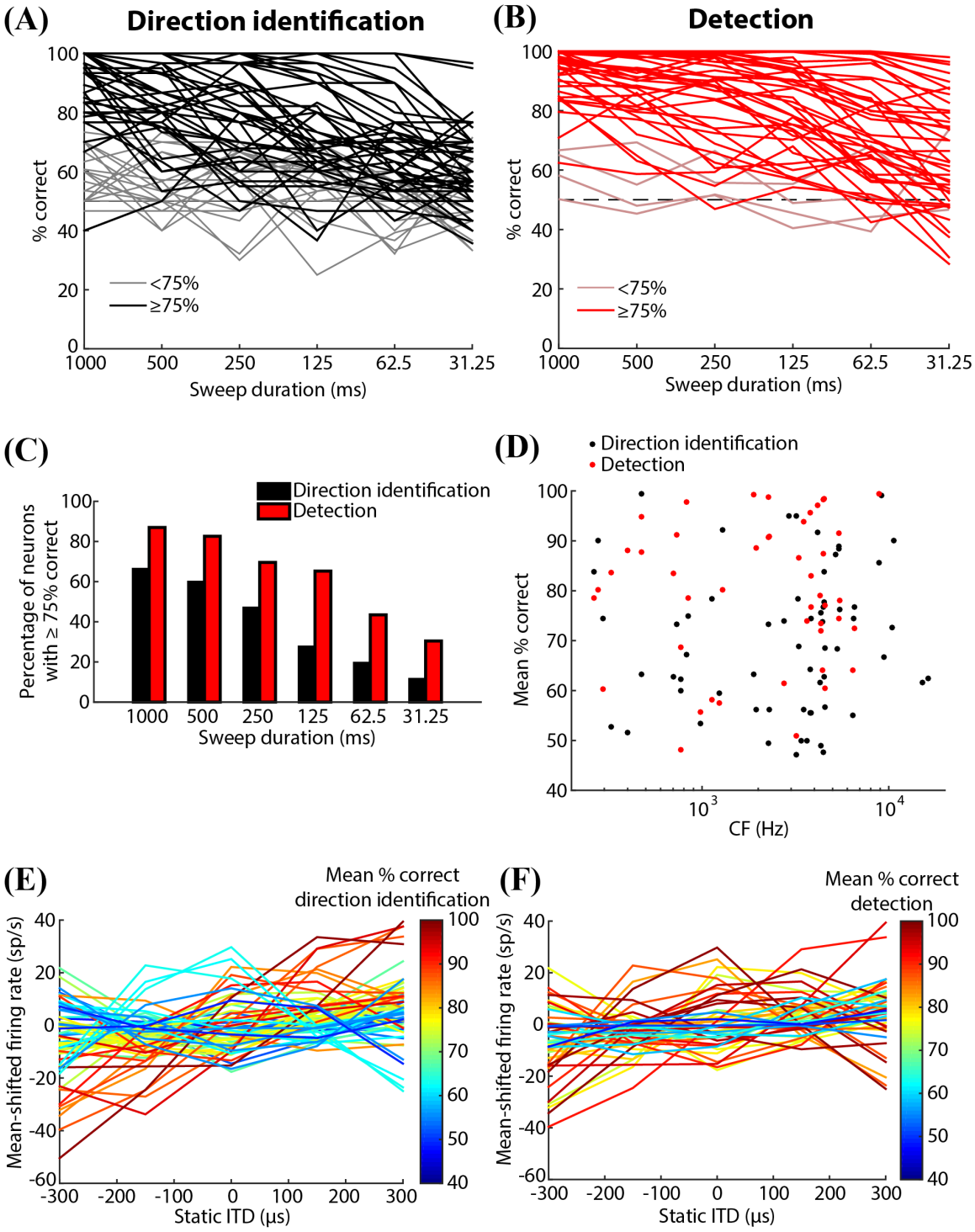

(A)
Direction identification
% correct
<75%
≥75%
Sweep duration (ms)
(B)
Detection
% correct
<75%
≥75%
Sweep duration (ms)
(C)
Direction identification
Detection
Percentage of neurons with ≥ 75% correct
Sweep duration (ms)
(D)
Direction identification
Detection
Mean % correct
CF (Hz)
(E)
Mean % correct direction identification
Mean-shifted firing rate (sp/s)
Static ITD (µs)
(F)
Mean % correct detection
Mean-shifted firing rate (sp/s)
Static ITD (µs)

Figure 12: (A) Percent correct direction identification for 62 neurons and (B) percent correct detection for 46 neurons. Thick lines mark neurons where the performance was at or above 75% correct for at least one sweep duration. (C) For all sweep durations there was a higher percentage of neurons where classifiers performed above 75% for detection than for direction identification. These numbers decreased with decreasing sweep duration. (D) Average percent correct for direction identification (black) and detection (red) as a function of the neuron's CF. (E & F) Static ITD tuning curves color-coded by average performance for the direction identification and detection tasks. The highest performance is shown in red and the lowest performance in blue – color bars are shown on the right of each plot. The best performing neurons for direction identification tended to have monotonic tuning curves between -300 and 300 μs. The best performing neurons for detection had large variations in firing rate in their static ITD tuning, irrespective of shape.

We performed the direction classification on 62 neurons and the detection classification on 46 neurons. **Figure 12a and 12b** show the overall percentage of correct responses for direction identification and detection respectively. Performance varied widely across neurons. In particular, several neurons performed direction identification very accurately; the neuron in **Figure 8g** produced the one of the highest average performance for direction identification. Likewise, the performance of many of the neurons for sweep detection was extremely high; the neuron in **Figure 7i** is an example. **Figure 12c** shows the percentage of neurons for which the percentage of correct responses was above 75%. Overall, a higher percentage of neurons were better for detecting a sweep than for identifying its direction. There was no significant correlation between neural CF and the average percent correct for direction identification (Kendall's tau: $\tau = 0.09$, $p = 0.32$) and there was also no significant correlation with detection ($\tau = 0.03$, $p = 0.80$) (**Figure 12d**). There was no significant difference between low CF (< 2 kHz) and high CF (> 2 kHz) neurons in their average percent correct for direction identification (Mann-Whitney U: $U = 585.5$, $p = 0.85$) and no significant difference

between these two populations in their average percent correct for detection ($U = 423$, $p = 0.61$).

We next examined the relationship between a neuron's static ITD tuning and the average performance of a classifier using the neuron's responses for direction identification or detection. We did this to identify trends in ITD tuning that may relate to performance. The neurons with the best direction identification performance tended to have large, monotonic tuning within the range of the ITD sweep (**Figure 12e**). In contrast, neurons with symmetric tuning curves typically did not provide good classification for direction identification. If a neuron's firing rate followed the ITD of the sweep and its ITD tuning curve was monotonic, it would produce a discernable response for the positive-going sweep relative to the negative going sweep (see **Figure 3a** for example). If instead the neuron's ITD tuning curve was symmetrical, then the response would be identical for either sweep direction. The trend in static ITD tuning curves suggests that classification of the ITD sweep direction is primarily based on the neuron's ability to follow the ITD. This is consistent with our finding that ITD following is stronger the direction selectivity. For detection, neurons with a high range of firing rates in their static ITD tuning produced the best performance (**Figure 12f**).

Comparison of neural and psychophysical thresholds:

Similar to our analysis of the psychophysics, we measured threshold durations for direction identification by fitting a sigmoid model to the percentage of correct responses for each neuron ("neurometric function"). We did this only for neurons that performed at or above 75% correct for at least one sweep duration. Despite the large spread in

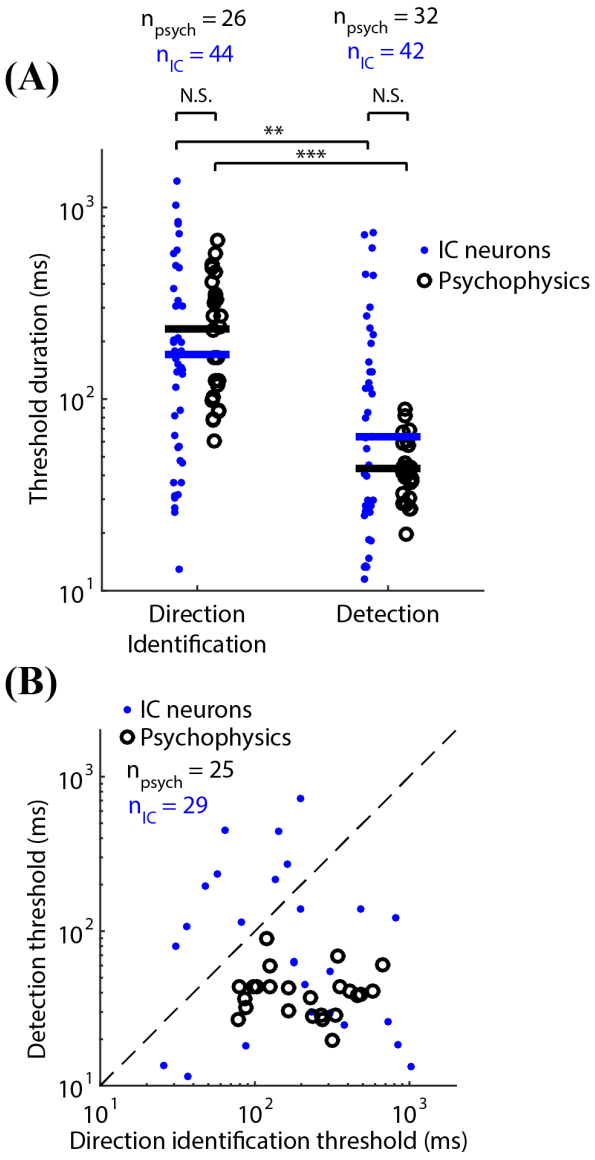


Figure 13: (A) Thresholds for direction identification and detection computed based on the classifier performance using the IC neuron responses (blue dots) and threshold for psychophysics (black circles). Neural thresholds and psychophysical thresholds are only included if the percent correct was at or above 75% for at least one sweep duration. Two direction identification thresholds were > 2000 ms, two detection thresholds were < 10 ms and one detection threshold was > 2000 ms – these threshold durations are not shown. The black circles are identical to the data plotted in Figure 4d. Significance values are based on Mann-Whitney U test. (B) Thresholds for direction identification and detection plotted for individual neurons (blue dots) and individual participants (black circles). The black circles are identical to the data plotted in Figure 4e.

threshold durations for direction identification and for detection, the medians were significantly different (**Figure 13a**) (median direction identification threshold = 171 ms, median detection threshold = 64 ms; Mann-Whitney U: $U = 2254$, $p = 0.0034$). There were 29 neurons that produced performances at or above 75% correct for both direction identification and detection (**Figure 13b**). The difference between threshold durations for direction identification and for detection was borderline significant (Wilcoxon signed-rank: $W = 308$, $p = 0.050$) and there was no correlation between their thresholds (Kendall's tau: $\tau = -0.16$, $p = 0.23$). We also found no significant difference between the neural slopes (median direction identification slope = 11.5 %correct/octave, median detection slope = 12.4 %correct/octave; Mann-Whitney U: $U = 2332$, $p = 0.23$).

We then compared the neurometric threshold durations and slopes with those calculated from the psychometric functions. There was no significant difference between the neural and psychophysical thresholds for direction identification ($U = 1496$, $p = 0.43$) or for detection ($U = 1706$, $p = 0.15$) (**Figure 13a**). This was surprising given that neurons were able to follow the ITD of the sweep at even the shortest durations for many neurons. Neurometric slopes were significantly shallower than psychometric slopes for both direction identification ($U = 1270$, $p < 0.001$) and detection ($U = 1267$, $p < 0.001$).

Together, these results suggest that the temporal limits in the IC may explain the limits of motion direction identification and ITD sweep detection in humans. However, we found that neurons in the IC can still follow the ITD quite well at the shortest sweep durations. Thus, we think that the limits in direction identification appear to be due to limitations in the ability to identify the motion direction from single-trial neural responses resulting from neural noise. For shorter sweep durations, neurons fire fewer spikes, and the inherent variability in firing makes the responses less reliable for direction identification.

Discussion:

We recorded from neurons in the IC of unanesthetized rabbits in response to broadband noises containing a linearly time-varying ITD ("ITD sweep"). We also presented these stimuli to human participants and asked them to identify the motion direction of the ITD sweeps or detect the sweeps in two separate tasks. We found that neurons in the IC responded dynamically to the ITD sweeps and that their responses were dominated by their ability to follow the ITD rather than by true direction selectivity.

Classifiers that use the responses of single neurons to either identify the motion direction of the ITD sweeps or detect the sweeps exhibit on average the same threshold durations as the human participants on both tasks.

Our results suggest that direction identification may be limited by neural noise: many IC neurons could follow the ITD of the sweeps at the shortest sweep durations we measured, but they could not be used reliably for direction identification. However, it is unlikely that the participants in our study were using one IC neuron at a time in order to perform the task. More likely multiple neurons in the IC are observed simultaneously or "pooled" when a participant performs either of the tasks. When neurons are pooled together, performance generally improves (Delgutte 1996; Parker & Newsome 1998). Performance may be limited by correlations in neural noise between pairs of neurons, also called noise correlations (Averbeck et al. 2006). Only a couple of studies using anesthetized gerbils have looked at noise correlations in the IC, and noise correlations appear to be very small and do not limit the information gained by pooling (Garcia-Lazaro et al. 2013; Belliveau et al. 2014). Additionally, neurons in the IC may be pooled, but information may be further limited at regions of the auditory system above the IC, such as the thalamus or cortex. Understanding how single-neuron information can be used to perform a task is essential to understand how information is encoded in the brain (see Parker & Newsome 1998), but how this information my be pooled across neurons requires further study.

The equivalence of neurometric and psychometric threshold durations has interesting implications for the coding of time-varying ITD. Several previous studies have argued that the temporal precision of firing in the IC allows neurons to respond to

time-varying binaural cues, such as changes in interaural correlation, faster than humans can detect them psychophysically (Shackleton & Palmer 2010). In our population, a number of neurons had threshold durations for direction identification and detection that were lower than human threshold durations. Because there are some neurons that can be used to identify the motion direction well at the shortest sweep durations, it does appear that direction identification may be suboptimally coded. Our human participants received very little training, and training greatly improved direction identification of moving sounds (Perrott & Marlborough 1989; for comparison to other studies see Carlile & Leung 2016). With additional training, it is possible that participant performance might improve to levels close to the best performing neurons.

We found that a point process model with ITD-following and no direction selectivity matched the response for nearly all neurons better than a model with direction selectivity and no ITD-following, suggesting that neurons primarily respond by following the ITD. Furthermore, the neurons that produced the best performance for direction identification were usually neurons with sloped static ITD tuning curves within the range of the ITD sweeps, again implying that ITD-following is dominating performance for direction identification. Other studies had found evidence for direction selectivity using rotating tones (Yin & Kuwada 1983; Douglas C Fitzpatrick et al. 2009; Dietz et al. 2014), but only a small number of neurons were direction selective and the speeds at which direction selectivity was present were inconsistent with human perception of motion (e.g. Perrott & Musicant 1977). Furthermore, direction selectivity observed in other studies using other oscillating stimuli could be explained by stimulus-independent parameters such as rate adaptation (Cai et al. 1998; McAlpine et al. 2000; Borisyuk et al. 2002;

Wang & Peña 2013). Overall, it is unlikely that direction selective responses in the IC are used for inferring the trajectory of the ITD sweeps.

Our focus here was on manipulations of ITD to examine motion direction identification without monaural changes in loudness, but motion direction identification is possible monaurally when other localization cues are present (Harris & Sergeant 1971). Auditory motion aftereffect is stronger when the adapting stimulus is a time-varying ILD instead of an ITD sweep (Carlile & Leung 2016), which suggests that neural direction selectivity may be stronger when binaural and monaural level cues are present. This may also be the case for IC neurons.

Grantham (1986) argued that the auditory system detects motion by taking "snapshots" of the moving sound at two different times and identifying a difference between the two locations in each snapshot (see also Middlebrooks & Green 1991). Our results are in agreement with this hypothesis: At the level of the IC neural responses are dominated by ITD following rather than direction selectivity, so it is likely that a higher area of the brain is determining the direction of motion by comparing the firing rates of the neurons at two different points in time. There is some evidence that cortical areas in humans such as the medial temporal area are direction selective because auditory motion direction can be decoded from fMRI activity in these areas (Jiang et al. 2014; Jiang et al. 2016). It is possible that this comparison occurs somewhere between the IC and cortex, but even in the primary auditory cortex of macaques direction selectivity in single neurons appears to be fairly weak (Ahissar et al. 1992; Scott et al. 2009). Thus, exactly where the decoding of motion direction occurs is still unclear.

In the visual system, detecting and identifying the direction of spatial changes in luminance, also called "first-order motion", appears to utilize velocity selective and direction selective neurons in the visual cortex (Cavanagh & Mather 1989). Interestingly, the perception of auditory motion is more similar to the more complex "second-order motion" based on motion of textures or color patterns when the average luminance is constant (Cavanagh & Mather 1989). While the detection of first-order motion is limited by velocity (Seiffert & Cavanagh 1998), detection of second-order motion is limited by displacement (Seiffert & Cavanagh 1998) and direction identification of second-order motion is limited by stimulus duration (Derrington et al. 1993). Motion detection and direction identification in audition also seem to be limited by displacement and duration rather than by velocity (Grantham 1986). Additionally, the perception of second-order motion in vision is attention-dependent (Cavanagh 1992). In our study, 26/33 the participants were unable to do the direction identification task, and several participants mentioned that the direction identification task required a great deal of attention (in contrast, participants mentioned that the detection task was easier and were usually listening for a "gap" in the noise and not necessarily the sweep). It is possible that attention may be important for motion perception in audition, similar to the perception of second-order motion in vision.

Many psychophysical studies have used various methods to estimate the binaural integration time of the auditory system within which interaural time differences and interaural correlations are calculated (for review see Shackleton & Palmer 2010 and Carlile & Leung 2016). The integration time can be quantified by the equivalent rectangular duration (ERD) that contains the same amount of energy as the temporal

window that the participants use. In a review of the literature on binaural integration times, Lüddemann et al. (2016) pointed out that the ERD estimates vary widely across studies, with individual values between 30-400 ms and mean values of 50-210 ms (Shackleton & Palmer (2010) cite studies with resulting ERD of 15-50 ms as well). Lüddemann et al. argue that the variation in ERD may be explained by both differences in paradigms across studies and differences in the way the ERDs were calculated. The median psychophysical threshold durations and the neural threshold durations that we measured for direction identification and detection fall at either end of this range (psychophysical and neural threshold durations respectively: 43 and 64 ms for detection, 232 and 171 ms for direction identification). This supports the view that the variance in ERD is due in part to differences in paradigms.

We found that neurometric threshold durations for direction identification and for detection on average matched human performance. A caveat to this result is that we are comparing neural data in rabbits with psychometric measurements in humans. It is possible that psychophysical threshold durations for direction identification and for detection are larger in rabbits than in humans. It has been suggested that rabbits make suboptimal use of temporal codes for other tasks, such as amplitude modulation detection (Carney et al. 2014). Furthermore, inputs from cortical areas could affect the responses of IC neurons during a behavioral task – indeed, corticocollicular projections appear to be important for plasticity during learning for a sound localization task (Bajo et al. 2010). Further experiments where neural recordings and behavior are done simultaneously would more closely elucidate the importance of ITD-following in the IC for direction identification.

Our results demonstrate that the differences between threshold durations for both ITD sweep direction identification and detection are consistent with differences in classification performance using single-neuron responses in the IC. The coding schemes described here may relate closely to our ability perceive time-varying ITD as well as detect glimpses of objects of interest in a complex auditory scene.

Appendix:

Determining response latency:

All model fitting and data analysis was performed in Matlab (Mathworks). Spikes were binned at 1 ms such that there were either 0, 1 spikes in nearly all bins. Some neurons did show 2 spikes in a single bin, but this occurred very rarely (the percentage of bins with 2 spikes ranged from 0% to 0.019% across the sample of neurons, with a median of 0.0006%). First we estimated the latency of the response by fitting the static ITD responses with a model weighted by shifted versions of the static ITD stimulus, namely:

$$\lambda_{st}(t|w_d) = \exp\left[\sum_{d=0}^{D} w_d U(t-d)\right]$$

U(t) equals 1 when the stimulus is on and 0 otherwise, equivalent to the difference in unit steps (*U(t) = u(t-T)-u(t)*) for a static ITD duration of *T*. w_d is the weighted contribution of stimulus delay on the response at time *t*, resulting in a time-varying conditional intensity function λ_{st}. The model was fit simultaneously to all static ITD durations (1000, 500, 250, 125, 62.5 and 31.25 ms) for delays between 0 and 150 ms without regularization (using "glmfit" in Matlab). Because the resulting weight vector was often noisy, the weights w_d were smoothed using a Gaussian kernel with a standard deviation

of 2 ms. We identified the latency by the maximum w_d for d > 5 ms if the response to the static ITD was excitatory relative to the binaurally uncorrelated noise or the minimum w_d if the response was inhibitory. In general latencies were between 10-30 ms. We used this latency in subsequent model fitting to adjust the stimulus vectors so that the timing of the stimulus variables matched the timing of the expected response.

Model fitting with regularization:

Because of the large number of model parameters (5 ITD-following weights + 1 uncorrelated noise weight + 5 direction selectivity weights + 300 spike history weights = 311 parameters) and the relatively few spikes exhibited by some neurons, we used elastic net regularization when fitting the model (Zou & Hastie 2005; Friedman et al. 2010). The model was fit by minimizing the deviance $D(\theta)$ between the model and the measured spike train (McCullagh & Nelder 1989; Dobson 1990):

$$D(\theta) = 2\sum_t n_t \log n_t - n_t \log\left[\lambda(t|\theta, H_t)\right] - \left[n_t - \lambda(t|\theta, H_t)\right]$$

where n_t is the spike count at time t, and $\lambda(t|\theta,H_t)$ is the conditional intensity function of the model and θ is the set of parameters in the model (ITD-following, direction selectivity, weight for uncorrelated noise, and spike history). Minimizing the deviance is equivalent to maximizing the log-likelihood of the model given the data (Dobson 1990). We minimized the deviance using elastic-net regularization (Zou & Hastie 2005; Friedman et al. 2010), namely:

$$\hat{\theta} = \arg\min_{\theta} \sum_t D(\theta) - \gamma P_\alpha(\theta)$$

where γ is the regularization parameter and $P_\alpha(\theta)$ is defined by:

$$P_\alpha(\theta) = \frac{1}{2}(1-\alpha)\|\theta\|_{l_2}^2 + \alpha\|\theta\|_{l_1}$$

For $\alpha \rightarrow 1$, this form of regularization enforces a sparsity penalty. However, it does not exhibit wild behavior when coefficients are correlated, which is a problem with other sparsity-enforcing regularization techniques (LASSO for example) (Friedman et al. 2010). While we expected many of the linear coefficients to be close to zero, preliminary analysis showed that there were weak correlations between ITD-following and direction-selectivity coefficients. Thus, we chose $\alpha = 0.9$ in order to capture both the sparsity and the correlations between the coefficients (Friedman et al. 2010).

The fitting was done using "lassoglm" in Matlab which uses interatively reweighted least squares (Friedman et al. 2010) to find the weights that minimize the deviance for an array of γ values. This program also includes an additional constant term, λ_0, to the conditional intensity function. When the regularization parameter is very large, all of the model weights will be set to zero and $\lambda(t|\theta,H_t) = \lambda_0$. For each fitting we chose the largest γ that minimized the deviance. Because the sum of the linear interpolant used to compute the ITD-following weight is 1 at each time point, the constant term included for regularization can be completely accounted for by the ITD-following weights in our model. After model fitting, we added λ_0 to each of the ITD-following weights for the fitted sweep duration, $w_{ITD}(ITD_i,d)$ in order to get the final ITD-following weights for the model.

Determining weights for spike history and binaurally uncorrelated noise:

Model fitting was done in a two-step process: first we fit the weights for the uncorrelated noise and the history components to the static ITD responses, then we fit the

ITD-following and direction selectivity weights separately for each ITD sweep duration. This two-step process reduced the number of parameters being fit to the ITD sweep types for each sweep duration, greatly reducing computation time.

After computing the latency, we first fit weights for the ITD-following component at 0 μs ITD, the binaurally uncorrelated noise (w_{Uncor}), and 300 spike history components (up to a time lag of -300 ms) to the neuron's response to all stationary stimuli, with stationary ITD durations of 1000, 500, 250, 125, 62.5, and 31.25 ms and 10-15 trials of each. The stationary stimulus can be described by a step function from the binaurally uncorrelated noise to a 0 μs ITD and back to binaurally uncorrelated noise, needing only 2 weights to map the stimulus to the response. Any dynamics in the response to the stationary stimuli is attributed to adaptation, which is in principle captured by the spike history components.

In the second step, we fit the full model to the responses to the ITD sweeps while holding w_{Uncor} and h fixed to the values found previously, under the assumption that the response to the binaurally uncorrelated noise and the spike history weights are the same irrespective of the sweep duration or ITD sweep duration. This assumption was validated later by testing the goodness-of-fit of the model to the response to the binaurally uncorrelated noise stimulus, which was not included in the model fitting. Specifically, we fit the ITD-following weights (w_{ITD}) and the direction selectivity weights (w_{DIR}), a total of 10 parameters, to the 0 μs ITD, positive-going sweep, and negative-going sweep stimuli for each sweep duration separately.

Goodness-of-fit for responses to binaurally uncorrelated noise:

We also used the time-rescaling test to assess the goodness-of-fit of the model to the response for binaurally uncorrelated noise, which was not included in the model fitting. Responses to binaurally uncorrelated noise were recorded in 46 neurons. The KS statistics for the binaurally uncorrelated noise showed a similar dependence on firing rate as for the ITD sweep stimuli (**Figure 9**). In 7/46 neurons the test failed for uncorrelated noise ($p < 0.05$ with Bonferroni correction for 46 neurons). For these neurons, the test also failed for at least one of the ITD sweep stimuli. The KS statistic for the binaurally uncorrelated noise was consistently smaller than the maximum KS statistic for the ITD sweeps. This suggests that the model properly fit the responses to the binaurally uncorrelated noise for most neurons, and when the time-rescaling test failed the fits were no worse than the fits to the responses for the ITD sweeps.

Chapter 4

General Discussion and Future Directions

In this thesis, we recorded from neurons in the inferior colliculus of unanesthetized rabbits while presenting noise with a time-varying ITD. We assessed how well these neurons could be used to identify the motion direction of the time-varying ITD and encode fluctuations in these stimuli.

In Chapter 2, we found that IC neurons were unable to keep up with the fluctuations in a periodically time-varying ITD "dynamic ITD" to the same modulation frequencies that they could for diotic amplitude modulated noises. This finding may be a neural correlate for binaural sluggishness. We also found that neurons responded to dynamic ITD at high modulation frequencies with a firing rate similar to that produced by binaurally uncorrelated noise, and often this firing rate differed from the rate for a noise with a static ITD. As a result, firing rate could be used to detect the dynamic ITD relative to a static ITD at high modulation frequencies where temporal coding was insufficient. The worst performance using either code occurred around 64 Hz, roughly where humans had the most trouble detecting the dynamic ITD in Grantham & Wightman's (1978) study.

In Chapter 3, we presented broadband noises with a linearly time-varying ITD ("ITD sweep") and examined how the motion direction of these noises was encoded. In two separate tasks, we also measured the threshold durations at which human participants could identify the motion direction of the sweeps or detect the sweeps. We used a generalized linear model to parse the contributions of ITD-following and direction

selectivity to the neuron's overall firing rate. We found that the temporal firing patterns for nearly all neurons could be better explained by the neuron's ability to follow the ITD more than by true direction selectivity. However, neurons could follow the ITD to the shortest sweep durations measured, so our inability to identify the direction of motion of the ITD sweep at short durations does not appear to be limited by the IC's ability to track the ITD. We then examined how well an optimal classifier could use the single-neuron responses to identify the motion direction or detect the ITD sweeps, similar to the tasks performed by the human participants. On average, the threshold duration for detecting the ITD sweeps flanked by binaurally uncorrelated noise was lower than the threshold duration for identifying its direction in both perceptual and neural data. Moreover, there was no significant difference between the threshold durations using the neural classifier and human psychophysical performance for either of these tasks. Thus, it is more likely that motion direction identification of a time-varying ITD is limited by neural noise than by limitations in ITD coding.

In our stimuli, the ITD "sweep" was flanked at the start and end by binaurally uncorrelated noise. In order to detect the sweep at short durations, participants mentioned that they listened for a "gap" in the flanking noise because the ITD sweep sounded more spatially focused than the binaurally uncorrelated noise. This task is similar to a binaural detection task where listeners are required to detect a binaural "gap" of correlated or anticorrelated noise that is flanked at the beginning and the end by uncorrelated noise (Boehnke et al. 2002; Lüddemann et al. 2016). The median threshold duration for the detecting the ITD sweeps was 43 ms, which is close to the threshold durations of 20-40 ms measured in the earlier studies. Interestingly, if the binaural gap

detection task is performed for a stimulus where the "gap" contains binaurally uncorrelated noise and the flankers are either correlated or anticorrelated, binaural gaps are detectable down to 3-7 ms durations (Akeroyd & Summerfield 1999; Boehnke et al. 2002; Lüddemann et al. 2016). Some studies have proposed a neural correlate for the differences in detectability for these two binaural gap detection tasks. For example, Lüddemann et al. (2016) show that late-evoked auditory potentials in humans show this expected trend: the threshold duration at which the potential is significantly larger than noise is shorter for gaps with uncorrelated noise than for gaps with correlated or anticorrelated noise. Similarly, Chait et al. (2005) found that peak latencies in MEG qualitatively matched the speed at which humans could detect a change in interaural correlation, which was shorter for transitions from correlated to uncorrelated noise than for transitions from uncorrelated to correlated noise.

Are there neural correlates in the IC for the differences in thresholds for binaural gap detection? This seems possible since the average neurometric threshold durations for ITD sweep detection matched the psychophysical threshold durations in humans, and the durations were longer than the threshold durations previously measured for detecting binaural gaps containing uncorrelated noise (3-7 ms). Shackleton & Palmer (2010) showed that neurons can respond to probe tones presented 30 ms after an instantaneous inversion in the interaural correlation of a masking noise, but they did not examine responses for transitions to or from binaurally uncorrelated noise. Wang & Li (2015) showed that IC neurons might respond to brief gaps of binaurally uncorrelated noise when surrounded by diotic noise, but they did not examine these responses for various

gap durations. Thus, a more thorough examination of the neural coding of binaural gaps in the IC is needed.

Several studies have suggested that we may be able to distinguish between multiple sources in a complex environment (a cocktail party, for example) by catching "glimpses" of the location of each source when its amplitude dominates the mixture (Faller & Merimaa 2004; Yost & Brown 2013). But these environments usually also contain background noise that is not restricted to a particular location (e.g. reverberation or multi-talker babble). Listening for a single talker amidst a spatially diffuse background is also an important task for a "cocktail party"-like listening situation. The binaural gap detection task with binaurally uncorrelated noise as the "masker" is a simplified version of this type of situation – the listener is required to detect a spatially focused sound amidst a diffuse background. How the auditory system encodes a spatially focused target in this task, as far as we are aware, is still largely unstudied.

Here we show that neurons in the IC can respond quickly to the change in spatial extent, allowing the ITD sweep with flanking binaurally uncorrelated noise to remain detectable at short durations. The ability for neurons to temporally lock in this way may degrade after sensorineural hearing loss, particularly with background noise present (Henry & Heinz 2012). On the other hand, hearing loss can also increase the strength of temporal coding to the sound envelope in the auditory periphery (Henry et al. 2014). How sensorineural hearing loss affects the ability for IC neurons to respond to a spatially focused sound amidst a diffuse background is also worth studying in the future.

Several studies have shown that if listeners are presented with a moving source in one direction for a long period of time, participants identify a static sound played

afterwards as moving in the opposite direction (Grantham & Wightman 1979; Grantham 1989; Dong et al. 2000). These studies have identified the effect as a motion aftereffect that has been interpreted as evidence for motion direction sensors in the auditory system. We found that the responses of neurons in the IC are better explained by ITD following than direction selectivity, so if direction selectivity exists in the auditory system it may occur at higher levels. Motion aftereffect is also present in vision and it is well known that there are neurons throughout the visual system, even in the retina, that respond more strongly to one direction of motion than the other (Hubel & Wiesel 1962; Barlow & Levick 1965; Albright 1984). However, motion aftereffect in vision is characterized by the presence of motion without a change in location (Wohlgemuth 1911; Nakayama 1985). It is unclear from the studies on auditory motion aftereffect whether or not this is the case, since these studies asked participants to judge the motion direction and but not identify the position. This could be examined further by having participants identify the direction of motion of a target sound on some trials or the position on others to validate if the motion aftereffect is similar to that observed in vision.

Overall, our results identify duration thresholds and modulation thresholds that may relate to the boundaries of hearing motion, fluctuations, and spatially diffuse noise in environments with time-varying ITDs of various speeds. We also think that these percepts for artificial time-varying ITD stimuli have ecological validity for identifying the trajectories of moving sounds, identifying the positions of multiple sounds sources in a complex acoustic environment, and suppressing the masking effects of diffuse background sounds such as reverberation. At this point, however, the extent to which these time-varying ITDs are present in a real environment is speculation. We can better

understand the distribution of time-varying ITDs in natural scenes by quantifying the statistics of time-varying ITDs in binaural recordings of various environments and identifying the sources of these statistics (for example, see Młynarski & Jost 2014 for the statistics of static binaural cues). We predict that the distributions of time-varying statistics within the range of our perception of "fluctuating" time-varying ITD come from spatially discriminable sound sources, and time-varying ITDs within the range of our perception of motion come from moving objects or self-motion.

Additionally, how the statistics of time-varying ITDs are encoded in auditory neurons would be very interesting to study in the future. It may be that our ability to focus on a single talker at a cocktail party, or identify the trajectory of an ambulance going by, relates directly the auditory system's ability to encode the most important time-varying ITDs in the environment while suppressing very fast time-varying ITDs, like reverberation, that are not ethologically relevant (for example, see Slama & Delgutte 2015). Neurons in the IC may be sensitive to the statistics of localization cues, since they can adapt their tuning based on the distribution of interaural level differences presented (Dahmen et al. 2010). While our results suggest ways for how these time-varying ITDs are encoded, examining how these mechanisms deal with real time-varying ITDs in the natural environment can have a large impact on understanding how we parse everyday complex scenes.

References

Ahissar, M., Ahissar, E., Bergman, H. & Vaadia, E., 1992. Encoding of sound-source location and movement: Activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. *J Neurophysiol*, 67(1), pp.203–15.

Aitkin, L.M., Fryman, S., Blake, D.W., & Webster, W.R., 1972. Responses of neurones in the rabbit inferior colliculus. I. Frequency-specificity and topographic arrangement. *Brain Res*, 47(1), pp.77–90.

Aitkin, L.M., Webster, W.R., Veale, J.L. & Crosby, D.C., 1975. Inferior colliculus. I. Comparison of response properties of neurons in central, pericentral, and external nuclei of adult cat. *J Neurophysiol*, 38(5), pp.1196–1207.

Akeroyd, M.A. & Summerfield, A.Q., 1999. A binaural analog of gap detection. *J Acoust Soc Am*, 105(5), pp.2807-2820.

Albright, T.D., 1984. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol*, 52(6), pp.1106–1130.

Altman, J.A., 1968. Are there neurons detecting direction of sound source motion? *Exp Neurol*, 22(1), pp.13–25.

Averbeck, B.B., Latham, P.E. & Pouget, A., 2006. Neural correlations, population coding and computation. *Nature Rev Neurosci*, 7(5), pp.358–366.

Bajo, V.M., Nodal, F.R., Moore, D.R. & King, A.J., 2010. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nature Neurosci*, 13(2), pp.253–260.

Barlow, H.B. & Levick, W.R., 1965. The mechanism of directionally selective units in rabbit's retina. *J Physiol*, 178(3), pp.477–504.

Belliveau, L.A.C., Lyamzin, D.R. & Lesica, N.A., 2014. The neural representation of interaural time differences in gerbils is transformed from midbrain to cortex. *J Neurosci*, 34(50), pp.16796–16808.

Bernstein, L.R., 1997. Detection and discrimination of interaural disparities: Modern earphone-based studies. In: *Binaural and Spatial Hearing in Real and Virtual Environments* (Gilkey, R. & Anderson, T.R., eds), pp 117–138. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Bernstein, L.R., Trahiotis, C., Akeroyd, M.A. & Hartung, K., 2001. Sensitivity to brief changes of interaural time and interaural intensity. *J Acoust Soc Am*, 109(4), pp.1604–1615.

Blauert, J., 1972. On the lag of lateralization caused by interaural time and intensity differences. *Audiology*, 11(5-6), pp.265–270.

Boehnke, S.E., Hall, S.E. & Marquardt, T., 2002. Detection of static and dynamic changes in interaural correlation. *J Acoust Soc Am,* 112(4), pp.1617-1626

Borisyuk, A., Semple, M.N. & Rinzel, J., 2002. Adaptation and inhibition underlie responses to time-varying interaural phase cues in a model of inferior colliculus neurons. *J Neurophysiol*, 88(4), pp.2134–2146.

Britten, K.H., Shadlen, M.N., Newsome, W.T. & Movshon, J.A., 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci,* 12(12), pp.4745–4765.

Brown, E.N., Barbieri, R., Ventura, V., Kass, R. & Frank L., 2002. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput*,

14(2), pp.325–346.
Cai, H., Carney, L.H.. & Colburn, H.S., 1998. A model for binaural response properties of inferior colliculus neurons. II. A model with interaural time difference-sensitive excitatory and inhibitory inputs and an adaptation mechanism. *J Acoust Soc Am*, 103(1), pp.494-506.
Carlile, S. & Leung, J., 2016. The perception of auditory motion. *Trends Hear*, 20, pp.1-19.
Carney, L.H., Li, T. & McDonough, J.M., 2015. Speech coding in the brain: Representation of vowel formants by midbrain neurons tuned to sound fluctuations. *eNeuro*, 2(4). doi: 10.1523/eneuro.0004-15.2015.
Carney, L.H., Zilany, M.S.A., Huang, N.J., Abrams, K.S. & Idrobo, F., 2014. Suboptimal use of neural information in a mammalian auditory system. *J Neurosci,* 34(4), pp.1306–1313.
Cavanagh, P., 1992. Attention-based motion perception. *Science*, 257(5076), pp.1563–1565.
Cavanagh, P. & Mather, G., 1989. Motion: The long and short of it. *Spat Vis*, 4(2-3), pp.103–129.
Chait, M., Poeppel, D., Cheveigné, A 2005. Human auditory cortical processing of changes in interaural correlation. *J Neurosci*, 25(37), pp.8518–8527.
Chandler, D.W. & Grantham, D.W., 1992. Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. *J Acoust Soc Am*, 91(3), p.1624.
Chase, S.M., Young, E.D., 2005. Limited segregation of different types of sound localization information among classes of units in the inferior colliculus. *J Neurosci*, 25(33), pp.7575–7585.
Chung, Y., Hancock, K.E. & Delgutte, B., 2016. Neural coding of interaural time differences with bilateral cochlear implants in unanesthetized rabbits. *J Neurosci*, 36(20), pp.5520–5531.
Culling, J.F., Colburn, H.S. & Spurchise, M., 2001. Interaural correlation sensitivity. *J Acoust Soc Am*, 110(2), pp.1020-1029.
Dahmen, J.C., Keating, P., Nodal, F.R., Schulz, A.L. & King, A.J., 2010. Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron*, 66(6), pp.937–948.
Davis, T.J., Grantham, D.W. & Gifford, R.H., 2016. Effect of motion on speech recognition. *Hear Res*, 337, pp.80–88.
Day, M.L., Koka, K., Delgutte, B., 2012. Neural encoding of sound source location in the presence of a concurrent, spatially separated source. *J Neurophysiol*, 108(9), pp.2612–2628.
Day, M.L. & Delgutte, B., 2016. Neural population encoding and decoding of sound source location across sound level in the rabbit inferior colliculus. *J Neurophys*, 115(1), pp.193–207.
Delgutte, B., 1996. Physiological Models for Basic Auditory Percepts. In *Auditory computation* (Hawkins H.L., McMullen, T.A. & Fay, R.R., eds), pp. 157–220. New York, New York : Springer.
Derrington, A.M., Badcock, D.R. & Henning, G.B., 1993. Discriminating the direction of second-order motion at short stimulus durations. *Vision Res*, 33(13), pp.1785–1794.

Devore, S. & Delgutte, B., 2010. Effects of reverberation on the directional sensitivity of auditory neurons across the tonotopic axis: influences of interaural time and level differences. *J Neurosci*, 30(23), pp.7826 –7837.
Dietz, M., Marquardt, T., Stange, A., Pecka, M., Grothe, B. & McAlpine, D., 2014. Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds II: Single-neuron recordings. *J Neurophysiol*, 111(10), pp.1973–1985.
Dobson, A.J., 1990. *An introduction to generalized linear models.*, London, United Kingdom : Chapman and Hall.
Dong, C.J., Swindale, N.V, Zakarauskas, P., Hayward, V. & Cynader, M.S., 2000. The auditory motion aftereffect: Its tuning and specificity in the spatial and frequency domains. *Percept Psychophys*, 62(5), pp.1099–1111.
Ebert, C.S., Blanks, D.A., Patel, M.R., Coffey, C.S., Marshall, A.F. & Fitzpatrick, D.C., 2008. Behavioral sensitivity to interaural time differences in the rabbit. *Hear Res*, 235, pp.134–142.
Faller, C. & Merimaa, J., 2004. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *J Acoust Soc Am*, 116(5), pp.3075-3089.
Feller, W., 1948. On the kolmogorov-smirnov limit theorems for empirical distributions. *Ann Math Stat*, 19(2), pp.177–189.
Feron, F.-X., Frissen, I., Boissinot, J. & Guastavino, C., 2010. Upper limits of auditory rotational motion perception. *J Acoust Soc Am*, 128(6), pp.3703–3714.
Fitzpatrick, D.C., Roberts, J.M., Kuwada, S., Kim, D.O. & Filipovic, B., 2009. Processing temporal modulations in binaural and monaural auditory stimuli by neurons in the inferior colliculus and auditory cortex. *J Assoc Res Otolaryngol*, 10(4), pp.579–593.
Friedman, J., Hastie, T. & Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*, 33(1), pp.1–22.
Garcia-Lazaro, J.A., Belliveau, L.A.C. & Lesica, N.A., 2013. Independent population coding of speech with sub-millisecond precision. *J Neurosci*, 33(49), pp.19362–19372.
Goldberg, J.M. & Brown, P.B., 1969. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization. *J Neurophysiol*, 32(4), pp.613–636.
Grantham, D.W., 1982. Detectability of time-varying interaural correlation in narrow-band noise stimuli. *J Acoust Soc Am*, 72(4), pp.1178-1184.
Grantham, D.W., 1984. Discrimination of dynamic interaural intensity differences. *J Acoust Soc Am*, 76(1), pp.71-76.
Grantham, D.W., 1986. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J Acoust Soc Am*, 79(6), pp.1939-1949.
Grantham, D.W., 1989. Motion aftereffects with horizontally moving sound sources in the free field. *Percept Psychophys*, 45(2), pp.129–136.
Grantham, D.W. & Wightman, F.L., 1979. Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. *J Acoust Soc Am*, 65(6), pp.1509-1517.
Grantham, D.W. & Wightman, F.L., 1979. Auditory motion aftereffects. *Percept*

Psychophys, 26(5), pp.403–408.
Grantham, D.W. & Wightman, F.L., 1978. Detectability of varying interaural temporal differences. *J Acoust Soc Am*, 63(2), pp.511-523.
Green, D.M. & Swets, J.A., 1988. *Signal detection theory and psychophysics*. Los Altos, California : Peninsula Pub.
Grothe, B., Pecka, M. & McAlpine, D., 2010. Mechanisms of sound localization in mammals. *Physiol Rev*, 90(3), pp.983–1012.
Grzywacz, N.M. & Amthor, F.R., 2007. Robust directional computation in on-off directionally selective ganglion cells of rabbit retina. *Visual Neurosci*, 24(4), pp.647–661.
Hancock, K.E., Noel, V., Ryuog, D.K. & Delgutte, B., 2010. Neural coding of interaural time differences with bilateral cochlear implants: Effects of congenital deafness. *J Neurosci*, 30(42), pp.14068–14079.
Harris, J.D. & Sergeant, R.L., 1971. Monaural/binaural minimum audible angles for a moving sound source. *J Speech Lang Hear Res*, 14(3), pp.618-629.
Henry, K.S. & Heinz, M.G., 2012. Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nat Neurosci*, 15(10), pp.1362–1364.
Henry, K.S., Kale, S. & Heinz, M.G., 2014. Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers. *Front Sys Neurosci*, 8(20). doi:10.3389/fnsys.2014.00020.
Hubel, D.H. & Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160(1), pp.106–154.
Ingham, N.J., Hart, H.C. & McAlpine, D., 2001. Spatial receptive fields of inferior colliculus neurons to auditory apparent motion in free field. *J Neurophysiol*, 85(1), pp.23–33.
Ingham, N.J. & McAlpine, D., 2004. Spike-frequency adaptation in the inferior colliculus. *J Neurophysiol*, 91(2), pp.632–645.
Jiang, F., Stecker, G.C., Boynton, G.M. & Fine, I., 2016. Early blindness results in developmental plasticity for auditory motion processing within auditory and occipital cortex. *Front Hum Neurosci*, 10:324. doi:10.3389/fnhum.2016.00324
Jiang, F., Stecker, G.C. & Fine, I., 2014. Auditory motion processing after early blindness. *J Vis*, 14(13):4, pp.1-18.
Joris, P.X., 2003. Interaural time sensitivity dominated by cochlea-induced envelope patterns. *J Neurosci*, 23(15) pp.6345–6350.
Joris, P.X., de Sande, B.V., Recio-Spinoso, A. & van der Heijden, M., 2006. Auditory midbrain and nerve responses to sinusoidal variations in interaural correlation. *J Neurosci*, 26(1), pp.279–289.
Joris, P.X., de Sande, B.V., Louage, D.H. & van der Heijden, M., 2006. Binaural and cochlear disparities. *Proc Natl Acad Sci USA*, 103(34), pp.12917–12922.
Kiang, N.Y.S. & Moxon, E.C., 1974. Tails of tuning curves of auditory-nerve fibers. *J Acoust Soc Am*, 55(3), pp.620–630.
Kim, D.O., Zahorik, P., Carney, L.H., Bishop, B.B. & Kuwada, S., 2015. Auditory distance coding in rabbit midbrain neurons and human perception: Monaural amplitude modulation depth as a cue. *J Neurosci*, 35(13), pp.5360–5372.
Kollmeier, B. & Gilkey, R.H., 1990. Binaural forward and backward masking: evidence for sluggishness in binaural detection. *J Acoust Soc Am*, 87(4), pp.1709-1719.

Krishna, B.S. & Semple, M.N., 2000. Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol*, 84(1), pp.255–273.
Kuwada, S., Stanford, T.R. & Batra, R., 1987. Interaural phase-sensitive units in the inferior colliculus of the unanesthetized rabbit: Effects of changing frequency. *J Neurophysiol*, 57(5), pp.1338–1360.
Kuwada, S. & Yin, T.C.T., 1983. Binaural interaction in low-frequency neurons in inferior colliculus of the cat. I. Effects of long interaural delays, intensity, and repetition rate on interaural delay function. *J Neurophysiol*, 50(4), pp.981–999.
Langner, G. & Schreiner, C.E., 1988. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol*, 60(6), pp.1799–1822.
Leung, J., Wei, V., Burgess, M. & Carlile, S., 2015. Head tracking of auditory, visual, and audio-visual targets. *Front Neurosci,* 9(493). doi:10.3389/fnins.2015.00493
Levitt, H., 1971. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, 49(2), pp.467-477.
Liang, L., Lu, T. & Wang, X., 2002. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *J Neurophysiol*, 87(5), pp.2237–2261.
Licklider, J.C.R., 1950. On the frequency limits of binaural beats. *J Acoust Soc Am*, 22(4), pp.468–473.
Litovsky, R.Y., Colburn, H.S., Yost, W.A. & Guzman, S.J., 1999. The precedence effect. *J Acoust Soc Am*, 106(4), pp.1633-1654
Lüddemann, H., Kollmeier, B. & Riedel, H., 2016. Electrophysiological and psychophysical asymmetries in sensitivity to interaural correlation gaps and implications for binaural integration time. *Hear Res*, 332, pp.170–187.
Mardia, K.V. & Jupp, P.E., 2000. *Directional statistics*. Chichester, New York : J. Wiley.
McAlpine, D., Jiang, D., Shackleton, T.M. & Palmer, A.R., 2000. Responses of neurons in the inferior colliculus to dynamic interaural phase cues: evidence for a mechanism of binaural adaptation. *J Neurophysiol*, 83(3), pp.1356–1365.
McCullagh, P. & Nelder, J.A., 1989. *Generalized linear models.*, London, United Kingdom : Chapman and Hall.
McDermott, J.H. & Simoncelli, E.P., 2011. Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron*, 71(5), pp.926–940.
McFadden, D. & Pasanen, E.G., 1975. Binaural beats at high frequencies. *Science*, 190(4212), pp.394–396.
Middlebrooks, J.C. & Green, D.M., 1991. Sound localization by human listeners. *Annu Rev Psychol*, 42, pp.135–159.
Młynarski, W. & Jost, J., 2014. Statistics of natural binaural sounds. *PLoS ONE*, 9(10), e108968. doi:10.1371/journal.pone.0108968
Nakayama, K., 1985. Biological image motion processing: A review. *Vision Res*, 25(5), pp.625–660.
Nelson, P.C. & Carney, L.H., 2004. A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *J Acoust Soc Am*, 116(4), pp.2173–2186.
Nelson, P.C. & Carney, L.H., 2007. Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus.

J Neurophysiol, 97(1), pp.522–539.
Palmer, A.R., Shackleton, T.M., Sumner, C.J., Zobay, O. & Rees, A., 2013. Classification of frequency response areas in the inferior colliculus reveals continua not discrete classes. *J Physiol*, 591(16), pp.4003–4025.
Parker, A.J. & Newsome, W.T., 1998. Sense and the single neuron: Probing the physiology of perception. *Annu Rev Neurosci*, 21, pp.227–277.
Perrott, D.R. & Marlborough, K., 1989. Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source. *J Acoust Soc Am*, 85(4), p.1773.
Perrott, D.R. & Musicant, A.D., 1977. Rotating tones and binaural beats. *J Acoust Soc Am*, 61(5), pp.1288-1292
Perrott, D.R. & Nelson, M.A., 1969. Limits for the detection of binaural beats. *J Acoust Soc Am*, 46(6), pp.1477-1481.
Pewsey, A., 2004. The large-sample joint distribution of key circular statistics. *Metrika*, 60(1), pp.25–32.
Pillow, J.W., Paninski, L., Uzzel, V.J., Simoncelli, E.P. & Chichilnisky, E.J., 2005. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J Neurosci*, 25(47), pp.11003–11013.
Reed, D.K., Dietz, M., Josupeit, A. & van de Par, S., 2016. Lateralization of stimuli with alternating interaural time differences: The role of monaural envelope cues. *J Acoust Soc Am*, 139(1), p.30-40.
Saberi, K., Tirtabudi, P., Petrosyan, A., Perrott, D.R. & Strybel, T.Z., 2002. Concurrent motion detection based on dynamic changes in interaural delay. *Hear Res*, 174(1), pp.149–157.
Santala, O. & Pulkki, V., 2011. Directional perception of distributed sound sources. *J Acoust Soc Am*, 129(3), pp.1522–1530.
Scott, B.H., Malone, B.J. & Semple, M.N., 2009. Representation of dynamic interaural phase difference in auditory cortex of awake rhesus macaques. *J Neurophysiol*, 101(4), pp.1781–1799.
Shackleton, T.M. & Palmer, A.R., 2010. The time course of binaural masking in the inferior colliculus of guinea pig does not account for binaural sluggishness. *J Neurophysiol*, 104(1), pp.189–199.
Seiffert, A.E. & Cavanagh, P., 1998. Position displacement, not velocity, is the cue to motion detection of second-order stimuli. *Vision Res*, 38(22), pp.3569–3582.
Siveke, I., Ewert, S.D., Grothe, B. & Wiegrebe, L., 2008. Psychophysical and physiological evidence for fast binaural processing. *J Neurosci*, 28(9), pp.2043–2052.
Slama, M.C.C. & Delgutte, B., 2015. Neural coding of sound envelope in reverberant environments. *J Neurosci*, 35(10), pp.4452–4468.
Spitzer, M.W. & Semple, M.N., 1991. Interaural phase coding in auditory midbrain: influence of dynamic stimulus features. *Science*, 254(5032), pp.721–724.
Spitzer, M.W. & Semple, M.N., 1993. Responses of inferior colliculus neurons to time-varying interaural phase disparity: effects of shifting the locus of virtual motion. *J Neurophysiol*, 69(4), pp.1245–1263.
Spitzer, M.W. & Semple, M.N., 1998. Transformation of binaural response properties in the ascending auditory pathway: Influence of time-varying interaural phase

disparity. *J Neurophysiol*, 80(6), pp.3062–3076.
Truccolo, W., Eden, U.T., Fellows, M.R., Donoghue, J.P. & Brown, E.N., 2005. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol*, 93(2), pp.1074–1089.
Viemeister, N.F., 1979. Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am*, 66(5), pp.1364-1380.
Wagner, H. & Takahashi, T., 1992. Influence of temporal cues on acoustic motion-direction sensitivity of auditory neurons in the owl. *J Neurophysiol*, 68(6), pp.2063 – 2076.
Wang, Q. & Li, L., 2015. Auditory midbrain representation of a break in interaural correlation. *J Neurophysiol*, 114(4), pp.2258–2264..
Wang, Y. & Peña, J.L., 2013. Direction selectivity mediated by adaptation in the owl's inferior colliculus. *J Neurosci*, 33(49), pp.19167–19175.
Weber, A.I. & Pillow, J.W., 2016. Capturing the dynamical repertoire of single neurons with generalized linear models. Available at: http://arxiv.org/abs/1602.07389.
Whitmer, W.M., Seeber, B.U. & Akeroyd, M.A., 2012. Apparent auditory source width insensitivity in older hearing-impaired individuals. *J Acoust Soc Am*, 132(1), pp.369–379.
Wightman, F.L. & Kistler, D.J., 1992. The dominant role of low-frequency interaural time differences in sound localization. *J Acoust Soc Am*, 91(3), pp.1648-1661.
Wightman, F.L. & Kistler, D.J., 1999. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *J Acoust Soc Am*, 105(5), pp.2841–2853.
Wohlgemuth, A., 1911. *On the after-effect of seen movement*, London, United Kingdom : Cambridge University Press.
Yin, P., Johnson, J.S., O'Connor, K.N & Sutter, M.L., 2011. Coding of amplitude modulation in primary auditory cortex. *J Neurophysiol*, 105(2), pp.582–600.
Yin, T.C.T. & Kuwada, S., 1983. Binaural interaction in low-frequency neurons in inferior colliculus of the cat. II. Effects of changing rate and direction of interaural phase. *J Neurophysiol*, 50(4), pp.1000–1019.
Yin, T.C.T., Kuwada, S. & Sujaku, Y., 1984. Interaural time sensitivity of high-frequency neurons in the inferior colliculus. *J Acoust Soc Am*, 76(5), pp.1401-1410.
Yost, W.A. & Brown, C.A., 2013. Localizing the sources of two independent noises: Role of time varying amplitude differences. *J Acoust Soc Am*, 133(4), pp.2301–2313.
Zou, H. & Hastie, T., 2005. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*, 67(2), pp.301–320.