Physiological and Computational-Modeling Studies of Timbre Encoding in the Inferior Colliculus

by

Johanna B. Fritzinger

Submitted in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Professor Laurel H. Carney

Department of Neuroscience

School of Medicine and Dentistry

University of Rochester

Rochester, New York

2025

Table of Contents

Biographical Sketch

Johanna Fritzinger graduated from Case Western Reserve University in 2018 with a Bachelor of Science degree in Electrical Engineering and a minor in music. They began their doctoral studies in the Department of Neuroscience at the University of Rochester in 2019 and received a Master of Science degree in Neuroscience in 2021. They were awarded an NIH F31 Predoctoral Fellowship in 2022. Beginning in 2020, Johanna has pursued a thesis focused on understanding how timbre is encoded in the auditory midbrain under the supervision of Dr. Laurel H. Carney.

The following publications were a result of work conducted during doctoral study:

Journal Publications:

Fritzinger, J. B., & Carney, L. H. (2025). Mechanisms of Tone-in-Noise Encoding in the Inferior Colliculus. Journal of Neuroscience.

Acknowledgements

I want to thank my advisor, Dr. Laurel Carney, for her support and guidance throughout this journey. Her unwavering excitement for science and dedication to teaching and students has made her a role model and an inspiration to me, and I have learned an incredible amount about science and about life under her guidance.

Thanks also to Dr. Ken Henry, who has been an integral part of my graduate experience and has provided fantastic feedback at our weekly graduate meetings. I would like to thank Dr. Ross Maddox and Dr. Samuel Norman-Haignere for their feedback and perspectives on my projects and advice.

I am deeply grateful for all of the lab members I have had the pleasure to work with, Dr. Afagh Farhadi, Dr. Paul Mitchell, Dr. Braden Maxwell, Dr. Daniel Guest, Dr. Swapna Agarwalla, Kris Abrams, and Doug Schwarz. Their constant support, insights, and comradery throughout this journey have meant the world to me. I have appreciated all of the help in the lab, feedback on my projects, bike rides, and happy hours that we have had over the years. I would also like to thank the graduate students in Dr. Ken Henry's lab, Yingxuan Wang and Leslie Gonzales, for their support, feedback, and friendship. I am also thankful for all of the undergraduate and high school students who have been in our lab over the years. Their joy and enthusiasm have been a light in this journey. I have learned an incredible amount from everyone around me, and I will be forever grateful for everyone I have been able to work with over these six years.

I would like to thank other teachers that have made an impact on my career as a scientist. Mary Handley, my first piano teacher, introduced me to teaching through

teaching young piano students, and that experience was unforgettable. Thanks to Dr. Hillel Chiel, without whom I would not have applied to graduate school. He was my first introduction to neuroscience, in class and in lab, and his careful consideration every scientific question is a quality I wish to emulate.

Next, I would not be here without my family. I feel very lucky to have chosen a university near immediate and extended family, and I am incredibly thankful for their support and encouragement. My parents, Amy and Bill Fritzinger, and my siblings, Will and Katrina Fritzinger have been a breath of fresh air whenever anxiety and stress were overwhelming. I want to thank our family dogs, Sundae, Pepper, and Greta, for forcing me back to the present moment. And to my extended family, there has been so much to celebrate over the years, and I am grateful for every party, gathering, and adventure. Aunt Ginny and Aunt Helene, your love for science makes me want to be the best scientist I can be. Thanks to my friends in the neuroscience program and beyond. I am thankful for chats, ski trips, camping excursions, and more that brought me joy during stressful times.

Thanks to my therapists for teaching me how to handle stress and social anxiety throughout graduate school. For some people, fear of public speaking fades over time. For me, I needed support from others, a great lab environment, and a helpful advisor. I am very grateful to everyone who helped me find my voice.

Lastly, thanks to Michael, my wonderful partner. The amount of steadiness and calm he brought throughout both of our graduate school experiences has changed my life. Our adventures have brought me so much joy and your kindness during hardships has helped me overcome many challenges.

Abstract

Timbre is the quality of sound that allows listeners to distinguish instruments playing the same note and is a critical aspect of music enjoyment and speech communication. Timbre perception is impacted by spectral properties of a sound; for example, the center of mass of a spectrum changes the brightness percept. Despite its importance, how timbre is represented in the auditory system is unclear. This thesis examines whether two mechanisms, neural fluctuation (NF) sensitivity and amplitude-modulation (AM) tuned broad inhibition, contribute to timbre encoding in the inferior colliculus (IC). The IC is a 'hub' in the subcortical auditory system where signals converge and is an important area of study.

NF sensitivity and AM-tuned broad inhibition mechanisms were investigated by recording IC extracellular responses to tones in narrowband or wideband noise. Results revealed that tones in narrowband noise were encoded through NF sensitivity, but responses to tones in wideband noise were consistent with AM-tuned broad inhibition. To test these mechanisms further, an IC computational model that replicates NF sensitivity was modified to include AM-tuned broad inhibition. The modified model predicted IC responses to AM and tones in narrowband and wideband noise.

Next, responses were recorded to a harmonic tone complex with a triangular spectral envelope, representing a controlled musical stimulus, to test if these mechanisms encoded spectral peaks related to timbre. Spectral peaks were encoded in rate responses and results were consistent with AM-tuned broad-inhibition predictions. The modified computational model more accurately predicted these responses than the previous model.

Additionally, many neurons had stable rate responses across suprathreshold levels, and neural discrimination thresholds were similar to psychophysical thresholds in humans.

Lastly, IC responses to natural musical sounds were recorded to investigate how IC neurons encode timbre and pitch (fundamental frequency, F0) in natural sounds. Decoding models were used to investigate information present in IC responses. Timbre and F0 were robustly represented in a population of rate responses. Temporal information in a population could also be used to decode instrument and F0, but decoding accuracy decreased at higher F0s. Additionally, both timbre and F0 information could be decoded at the same time, again using the rate population.

Results from these projects have advanced our understanding of mechanisms in the IC that encode complex sounds. Spectral peaks in musical sounds, related to the percept of brightness, were encoded in the IC. The modified computational IC model predicted responses to synthetic stimuli, but improvements will be necessary to predict responses to natural sounds. Hearing aids and cochlear implants are not currently designed for timbre perception, and these results could lead to novel strategies for restoring music enjoyment in listeners with hearing loss.

Contributors and Funding Sources

This work was supported by a dissertation committee consisting of Professor Laurel H. Carney (advisor) of the Departments of Neuroscience and Biomedical Engineering, Professor Kenneth S. Henry of the Departments of Otolaryngology, Neuroscience, and Biomedical Engineering, Professor Samuel Norman-Haignere of the Departments of Biostatistics and Computational Biology, Biomedical Engineering, and Neuroscience, Professor Mark Bocko of the Department of Electrical and Computer Engineering, and Professor Ross Maddox in the Department of Otolaryngology at the University of Michigan. Professor Elize Piazza of the Department of Brain and Cognitive Sciences and Neuroscience contributed during the qualifying exam proposal.

All work conducted for the dissertation was completed by the student independently, under the supervision of Professor Laurel H. Carney.

Work was supported by the National Institute of Health and National Institute on Deafness and Other Communication Disorders under grant numbers NIH R01-DC010813 and NIH F31-DC020630.

List of Tables

List of Figures

Chapter 1: Introduction

Timbre is the quality of sound that allows listeners to distinguish two instrument sounds that are identical in note, duration, and loudness (ANSI, 1994) and it is critical for speech perception, music enjoyment, and sound-source recognition. However, how timbre is represented in the auditory system is not yet understood. The percept of timbre has been shown to degrade with hearing loss and is not compensated for in hearing aids and cochlear implants, motivating this work to investigate timbre encoding (Gfeller et al., 2002; Emiroglu and Kollmeier, 2008).

Timbre encoding was investigated specifically in the inferior colliculus (IC), a midbrain nucleus. The IC is a nearly obligatory pathway in the auditory system (Aitkin and Phillips, 1983). IC neurons are sensitive to many stimulus features, including amplitude modulation (AM), binaural features, frequency, and frequency-sweep direction and velocity (Krishna and Semple, 2000; Joris et al., 2004; Review: Winer and Schreiner, 2005; Andoni et al., 2007; Kim et al., 2020; Mitchell et al., 2023). These features are contained in instrument sounds with timbre and in many other complex sounds such as speech. The IC is also hypothesized to transform temporal codes from earlier auditory nuclei into rate representations used in the thalamus and cortex (Joris et al., 2004). This thesis work investigated both rate and temporal representations of timbre along with mechanisms that could drive the temporal-to-rate transformation. Projects included analyzing extracellular IC responses to synthetic and natural stimuli with timbre and improving upon IC computational models to better represent encoding mechanisms.

1.1 Background

Timbre is a perceptual attribute, and initial psychophysics work focused on decomposing timbre into sound characteristics that contribute to timbre perception (Review: McAdams, 2019). First, listeners would rate different qualities of instrument sounds to separate perceptual dimensions of timbre. Then, those dimensions were correlated with attributes of sound using multi-dimensional scaling techniques (Grey, 1977; Grey and Gordon, 1978; McAdams et al., 1995). Timbre dimensions can be categorized by three types of sound attributes: temporal, spectral, and time-varying spectral (Review: McAdams, 2019). Temporal attributes include attack time and decay time (Fig. 1.1a); spectral attributes include the spectral centroid ('brightness' percept), spectral slope, jaggedness of the spectral envelope, and more (Fig 1.1b). Time-varying spectral attributes include spectral flux, defined as the variability of the spectrum over time (Fig 1.1c).

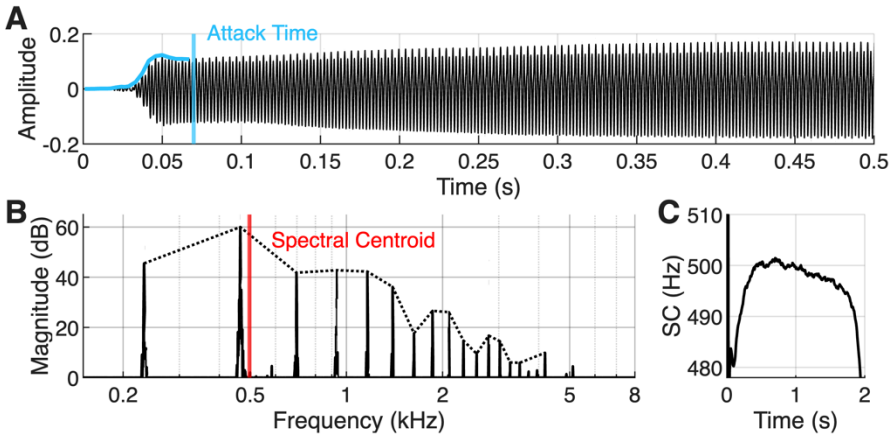

Figure 1.1 Example analysis of timbre attributes from a bassoon note using Timbre Toolbox (Peeters et al., 2011). (A) Temporal waveform with the attack/onset (blue). (B) Magnitude spectrum and spectral envelope highlighted (black dotted line) and spectral

centroid estimate (red). (C) Spectral-temporal attribute, spectral centroid changes over the duration of the stimulus.

This thesis focuses specifically on spectral attributes that contribute to timbre perception, due to the foundation of psychophysical work on this aspect of timbre. The spectral centroid, or the center of mass of the sound spectrum, is highly correlated with the 'brightness' percept: an instrument with a higher centroid, such as a trumpet, will sound brighter than an instrument with a lower centroid, such as a clarinet (McAdams, 2019). Spectral centroids are also correlated with peaks in the sound spectrum, the centroid is influenced by harmonics with the highest amplitudes. We hypothesized that spectral peaks are encoded in the IC through two specific mechanisms based on previous physiological and computational modeling work.

The IC is the 'hub' of the subcortical auditory system. Much focus of IC research has included binaural investigations (Palmer and Kuwada, 2005), but neurons in the IC exhibit sensitivity to many other features, such as frequency, AM, and direction of fast frequency sweeps, called chirps (Krishna and Semple, 2000; Joris et al., 2004; Andoni et al., 2007; Kim et al., 2020; Mitchell et al., 2023). AM sensitivity in IC neurons is quantified by average rate responses to modulated noise, called modulation transfer functions (MTFs). These MTFs fall into four general classifications: neurons with band-enhanced (BE) MTFs are excited, or enhanced, over a band of modulation frequencies in comparison to an unmodulated stimulus; band-suppressed (BS) neurons are suppressed over a band of modulation frequencies; hybrid neurons contain bands of both

enhancement and suppression, and flat neurons are not tuned to modulation (Kim et al., 2020; Fan et al., 2021; Fritzinger and Carney, 2025).

More generally, inhibition is a key factor in shaping IC responses to complex sounds: blocking inhibition in the IC broadens tuning curves, removes frequency-sweep direction selectivity, and alters binaural responses (Yang et al., 1992; Palombi and Caspary, 1996; LeBeau et al., 2001; Xie et al., 2005; Pollak, 2011). Two-tone paradigms reveal inhibitory sidebands in IC neurons as well (Egorova et al., 2001; Portfors and Felix, 2005; Alkhatib et al., 2006; Palmer et al., 2013). Broad inhibition is also theorized to help encode harmonic complex tones with a flat spectral envelope (Su and Delgutte, 2020).

Investigating timbre encoding in the subcortical auditory system is critical for future work understanding how the percept of timbre, specifically the spectrum of a sound, is altered with hearing loss. Hearing-impaired listeners have perceptual deficits in instrument and vowel identification tasks (Emiroglu and Kollmeier, 2008; Chintanpalli et al., 2016). In one study, increased just-noticeable-differences were found for participants with hearing loss discriminating a continuum of two instruments (such as horn-trombone or cello-sax), compared to normal hearing participants (Emiroglu and Kollmeier, 2008). In another study, participants with hearing loss had deficits identifying two vowels played at the same time (Chintanpalli et al., 2016). Possible mechanisms for perceptual changes in hearing loss include the loss of cochlear sensitivity due to outer-hair-cell damage (Chen et al., 2008), and loss of inhibition due to aging (Caspary et al., 2008). Potential strategies to compensate for deficits in timbre perception differ depending on which

mechanisms encode timbre in the auditory midbrain, thus motivating research into timbre encoding in the IC.

Another motivation for timbre encoding research is to advance cochlear implants (CI). Music is generally less enjoyable for cochlear implant users compared to normal hearing listeners, but there is large variability (Looi and She, 2010; Review: Marozeau and Lamping, 2019). Some CI users report never listening to music, whereas other users report listening to music often and getting great enjoyment out of music (Looi and She, 2010). Instrument identification is highly variable among CI users, with one study finding an average of 47% accuracy when identifying 8 instruments, with a range of 11-100% accuracy across individual participants (Gfeller et al., 2002). Multidimensional scaling techniques have been used to further understand where deficits in timbre perception exist in CI users (Kong et al., 2011, 2012). The first dimension, related to attack time, and the second dimension, related to brightness and spectral centroid, were present in these studies (Kong et al., 2011, 2012). However, correlation between spectral centroid and brightness perception were highly variable between subjects. The third dimension, spectral flux, related to differences in spectrum over time, was not present in CI users. These results indicate deficits in timbre perception related to spectral dimensions of timbre in CI users, further motivating this study.

The remaining sections of Chapter 1 introduce background relevant to Chapter 2, 3 and 4. First, Chapter 2 focuses on possible mechanisms that influence complex sound encoding by investing IC responses to tones in noise (Fig 1.2a). Then, Chapter 3 investigates how a synthetic, controlled stimulus with a brightness timbre percept is

encoded in the IC (Fig. 1.2b). Lastly, Chapter 4 broadens the scope of this thesis to include IC responses to natural instrument sounds (Fig. 1.2c).

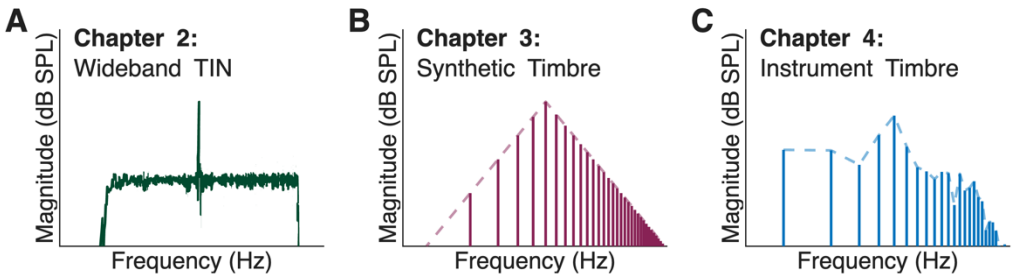

Figure 1.2. Overview of stimuli used in Chapters 2, 3, and 4. (A) Tones in wideband noise were used to investigate IC encoding mechanisms. (B) Harmonic tone complexes with a triangular spectral envelope were used to investigate spectral timbre encoding. (C) Bassoon and oboe sounds were used to investigate how IC neurons represent timbre and pitch.

1.2 What encoding mechanisms could impact IC timbre encoding?

In the first project, I investigated mechanisms influencing complex sound encoding in the IC. The process of encoding complex sounds starts at the auditory nerve (AN), the first auditory area where action potentials encode sound. Two general encoding strategies are used in the AN: rate-place and phase-locking to temporal fine structure or to stimulus envelopes. One hypothesis was that spectral peaks could be represented in AN fiber rate: AN fibers tuned near spectral peaks receive the most energy and therefore have the largest firing rate (Sachs and Young, 1979). This pattern is seen in the AN at low sound levels, but at normal speaking levels the high-spontaneous-rate and medium-spontaneous-rate fibers are saturated and this code is degraded. Phase-locking to temporal fine structure, specifically the harmonics near spectral peaks, is robust over level (Young

and Sachs, 1979), also creating a representation of spectral peaks in the AN. However, the frequency limit for phase-locking decreases as signals ascend through the auditory pathway, and some transformation is required to take advantage of AN temporal responses.

The AM sensitivity in the IC could take advantage of neural fluctuation (NF) profiles, another temporal aspect AN fibers, to transition from a timing to rate encoding. NFs are low-amplitude fluctuations in AN fibers that often lock to the envelope (Fig. 1.3, Carney, 2018, 2024). These NFs can be reduced, or 'captured', by a tone tuned to the characteristic frequency (CF) of the AN fiber (Fig. 1.3, Carney, 2018, 2024). A tone can reduce the NF response to a AN fiber tuned near the tone due to nonlinearities in the inner ear, including inner-hair-cell saturation (Fig. 1.3, (Zilany and Bruce, 2007; Carney, 2018). The AN fibers tuned away from the tone will produce NFs, creating a pattern across AN fibers of fluctuations and captured responses that carry information about spectral peaks. Critically, this NF profile is present at suprathreshold sounds and in high-spontaneous-rate fibers, which are the largest proportion of fiber type in the AN (Liberman, 1978).

The IC can take advantage of the NF profile due to strong AM sensitivity and support the transformation from a temporal to a rate code. In one study, tones in narrowband 1/3 octave noise were presented at the CF of an IC neuron (Fan et al., 2021). As SNR increased, rates of IC neurons increased or decreased dependent on MTF type. BE neurons decreased in rate as SNR increased, consistent with NFs at the tone frequency flattening due to IHC saturation and other nonlinearities (Fig. 1.2). BS neurons

increased in rate as SNR increased, again consistent NFs being reduced due to capture (Fan et al., 2021). The spectral peak of a harmonic stimulus may also lead to the same type of encoding as the tone, with a rate profile that increases or decreases at the spectral peak dependent on IC MTF type.

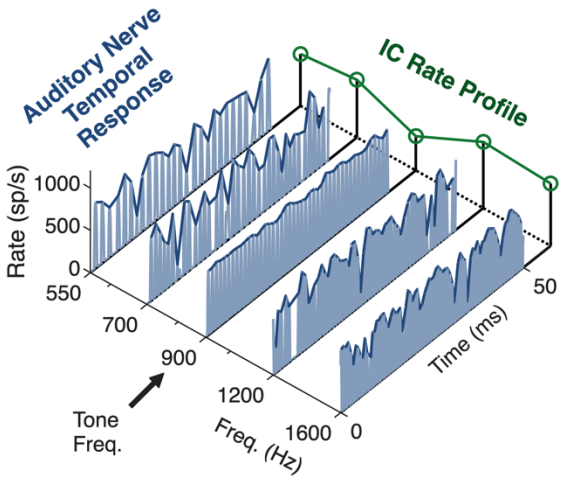

Figure 1.3. Model AN and IC BE cell responses to a tone in wideband noise. AN fibers tuned away from the tone will fluctuate, but the ANF tuned at the tone frequency has decreased fluctuations due to saturation of IHCs at 900 Hz. The IC BE cell is sensitive to AM and decreases in rate when no modulation is present, at the 900 Hz tone. This results in a rate profile across frequency in the IC with information that carries information about the tone frequency.

Inhibition is a second, more general, mechanism that influences complex sound encoding in the IC. IC neurons receive inhibition from ascending ipsilateral and contralateral pathways and local circuits. Neurons in the superior olivary complex and in the ventral nucleus of the lateral lemniscus inhibit ipsilateral IC neurons (Saint-Marie and Baker, 1990; Winer et al., 1995; Oliver, 2000), and projections from the dorsal nucleus of the lateral lemniscus inhibit both contralateral and ipsilateral IC neurons (Adams and

Mugnaini, 1984). Additionally, intrinsic connections and commissural connections contain inhibitory GABAergic projections to IC neurons (Adams, 1980). Blocking inhibition broadens tuning curves, removes frequency-sweep direction sensitivity, and alters binaural responses (Andoni et al., 2007; Pollak, 2011; Williams and Fuzessery, 2011; Su and Delgutte, 2020). Modeling work has hypothesized that inhibition is also critical for AM sensitivity in the IC (Nelson and Carney, 2004; Dicke et al., 2007) and for encoding complex sounds such as flat harmonic tone complexes (Su and Delgutte, 2020).

To make initial hypotheses about how NF profiles and inhibition impact sound encoding in the IC, we can use computational models of the IC. One IC model used throughout this thesis is a phenomenological model created to replicate IC responses to AM (Nelson and Carney, 2004; Carney and McDonough, 2019). This model, called the same-frequency inhibition-excitation (SFIE) model, is made up of a ventral cochlear nucleus (VCN) stage and IC BE and BS cell stages. The input to the SFIE model is an AN model fiber output from the Zilany et al., (2014) model. The VCN model cell receives both on-CF excitatory input and a delayed inhibitory input from the AN fiber. The IC BE cell stage receives an excitatory input and delayed inhibition from the VCN cell, and lastly the IC BS cell receives excitation from the VCN cell and inhibition from the BS cell. This model configuration with the proper selection of relative strengths and time courses of inhibition and excitation recreates AM sensitivity in BE and BS IC cells. The SFIE model was used to create hypotheses about how AM sensitivity in the IC to neural fluctuations could impact timbre encoding. More information on a modified model that includes AM-tuned broad inhibition is below.

This project investigated how two mechanisms, NF sensitivity and broad inhibition, impact complex sound encoding in the IC. We built upon a previous study that used on-CF tones in a narrowband noise by expanding the stimulus to include on- and off-CF tones in a broadband noise. We initially hypothesized that NF sensitivity would dominate, as it had in the narrowband study; however, results were complex and broad inhibition was investigated as well. This project included recording responses in the central nucleus of the IC in awake Dutch-belted rabbits and extending a current model of the IC to include broad inhibition based on the physiological results. Interestingly, the model that was most accurate featured broad inhibition that was sensitive to AM.

Physiological results did not match the initial hypothesis that neural fluctuations would increase or decrease rate for neurons for an on-CF tone based on MTF type. Instead, rates increased near the tone regardless of MTF type. These results differed from the narrowband tones-in-noise (NB-TIN); BE cells decreased in rate at CF with NB-TIN but increased with WB-TIN. Additionally, we found areas of suppression when the tone was moved off-CF, compared to a noise-alone response. These WB-TIN results were not consistent with NFs and supported that broad inhibition mechanisms influence responses to the broadband stimuli. To further test that hypothesis, off-CF pathways were added to an on-CF model of the IC that featured AM sensitivity. We tried two different types of off-CF inputs: one from the VCN stage of the model without AM sensitivity, and one from the IC BS cell stage, which does have AM sensitivity. The addition of inhibition from the BS cell resulted in a model of the IC that approximated the responses to WB-TIN, replicated many different MTF types depending on input parameters, and correctly

predicted NB-TIN results. The AM-tuning of the broad inhibition model meant that the inhibitory inputs were also taking advantage of NFs in the AN. This result was exciting: the addition of AM-tuned broad inhibition to a model with NF sensitivity improved predictions to the WB-TIN while maintaining NB-TIN results (Fritzinger and Carney, 2025).

Overall, this work highlights the impact that AM-tuned broad inhibition may have on complex sound encoding in the IC, specifically for broadband stimuli. These results set up the hypothesis that spectral peaks may be encoded in the IC due to neural fluctuation sensitivity and AM-tuned broad inhibition.

1.3 How are spectral peaks represented in the IC?

IC neuron responses were recorded to a synthesized stimulus that elicits the 'brightness' timbre percept: a harmonic complex tone with a broad triangular shaped spectrum. This stimulus was chosen for two reasons: 1) it has been used in psychophysical tasks that estimate spectral centroid discrimination (Allen and Oxenham, 2014) and 2) preliminary computational modeling work suggested that neural fluctuation sensitivity in the IC could describe human spectral centroid discrimination (Maxwell et al., 2020).

The psychophysical study used the harmonic complex tones with the triangular shaped stimulus to find a threshold of spectral centroid discrimination (Allen and Oxenham, 2014). In this task the fundamental frequency (F0) remained constant while the spectral centroid, here equivalent to the spectral peak, varied. Discrimination

thresholds for a spectral centroid at 1200 Hz was 4.0% for musicians and 5.0% for non-musicians. Additionally, this study investigated F0/spectral centroid interference by varying F0 while subjects performed the spectral centroid discrimination task, finding that variation in F0 increased spectral centroid discrimination thresholds (Allen and Oxenham, 2014).

To test if IC neuron sensitivity to NFs could account for these results, we modeled the interference task using computational models of the AN and IC (Maxwell et al., 2020). Model responses for a population of BS neurons increased in rate near the spectral peak of the stimulus. This peak in the population response varied as the spectral centroid varied. The model spectral centroid discrimination with no F0 variation was equal to the musicians thresholds of 4.0%, and thresholds increased as F0 interference increased at the same rate as human perceptual thresholds (Maxwell et al., 2020). This model is consistent with the hypothesis that spectral centroid encoding is present in the IC and found that F0-spectral centroid interference may arise early in the auditory system.

IC physiological responses to a harmonic complex with a steep triangular spectrum, created to approximate a vowel formant, can provide initial insight on spectral peak encoding. The key difference in these vowel stimuli and the synthetic timbre stimulus is the steepness of the roll off: the synthetic timbre stimulus has a roll off of 24 dB/octave, whereas the vowel formant stimuli have roll offs around 200 dB/octave. The vowel formant approximation stimulus also results in discrimination thresholds around 3-5% in humans (Lyzenga and Horst, 1995). This stimulus has been modeled in AN fibers and recorded in the IC of budgerigars (Tan and Carney, 2005; Henry et al., 2017). IC

neural discrimination thresholds using rate profiles was sufficient to describe budgerigar behavioral thresholds in quiet (Henry et al., 2017). However, timing information was necessary to describe behavior when a background noise was added. These results suggest that rate and timing representations contain important information for behavior.

The synthetic timbre IC modeling work, IC results for vowels in budgerigar, and the previous investigation into encoding mechanisms set the stage for studying physiological and modeling responses to synthetic timbre stimuli. Similarly to the WB-TIN and NB-TIN results, we hypothesized that spectral peaks are encoded in the IC via rate profiles due to NF sensitivity. We predicted that BE cells would decrease in rate for neurons tuned near the spectral peak, and BS cells would increase in rate for neurons tuned near the spectral peak. Neurons could also potentially encode this stimulus using timing information: IC neurons tuned near the spectral peak would have low phase-locking strength due to the lack of fluctuations, whereas neurons tuned away from the spectral peak may have stronger phase-locking due to fluctuations. Additionally, AM-tuned broad inhibition may also impact these responses and could create an increase in rate response for BE neurons.

Neural responses to the synthetic timbre stimulus from Allen & Oxenham (2014) were recorded using the same methods as Chapter 2. One limitation of the physiological experiments is low yield: the electrodes only sample four different areas of the IC. To get an inferred population response, we shifted the stimulus spectrum to get an estimate of a neural response when the neuron is tuned above, at, or below the spectral peak frequency.

Responses were recorded at several suprathreshold levels as well, to get an estimate of how robust IC responses were over level.

An important result from this work was the finding that the rate representation of spectral peaks in the IC was sufficient to describe human discrimination thresholds (Allen and Oxenham 2014). Performance improved as level improved, and the rate representation was also robust over level. Temporal results were complicated: A group of neurons did support the initial prediction, but more complex patterns were present as well. These results were not consistent with the NF predictions from initial modeling studies. Instead, the model with AM sensitivity and added AM-tuned broad inhibition predicted the average rate results.

This chapter extended IC computational modeling work to physiological results to test neural discrimination thresholds and encoding methods of spectral peaks. We found that rate profiles reliably encode spectral peaks over many sound levels.

1.4 How are pitch and instrument identity represented in the IC?

The previous two projects focused on synthetized sounds to directly investigate encoding mechanisms of spectral peaks in harmonic sounds. These highly controlled stimuli are useful to investigate specific questions about IC encoding. However, a broader goal of this work is to understand how natural musical sounds are represented in the IC. Natural sounds with spectral peaks and timbre, such as vowels or instrument sounds, contain more complexities than the synthetic stimuli. This last project recorded

physiological IC responses to recordings of instruments playing different F0s to study how instrument identity and F0 are represented in the IC.

The majority of psychological studies in the IC have been done using synthetic stimuli. These stimuli allow a systematic investigation of IC encoding mechanisms, but results may not be relevant in natural listening conditions (Theunissen and Elie, 2014). Recent work in the IC has used synthetic stimuli with sound statistics that match natural sounds (Attias and Schreiner, 1997; Escabí et al., 2003; Lesica and Grothe, 2008). Other studies have used vocalizations, providing ecologically relevant stimuli to the IC of models animals such as guinea pig and mouse (Portfors and Felix, 2005; Lyzwa et al., 2016). And lastly, some work has recorded IC responses to spoken vowels (Carney et al., 2015; Mitchell and Carney, 2025). Cortical work uses natural sounds, including instruments and music, much more often in model animals such as ferrets (Landemard et al., 2021). How natural musical sounds are encoded in early auditory areas before cortex is largely unknown.

A comparison of synthetic and natural instrument stimuli highlights differences in sound attributes that may impact timbre encoding of instruments. For most instruments, spectra have a strong downward slope and strong harmonics at low frequencies. This spectral envelope differs from the synthetic stimuli in Chapter 3: if the peak was set to be above around 1200 Hz, the synthetic stimulus had low amplitude harmonics at low frequencies. This difference in energy at low frequencies could impact the IC representation of these sounds. Low-CF AN fibers would increase in rate, or possibly saturate, due to the increase in energy at lower frequencies, and high-CF AN fibers could

also be affected due to the tail at suprathreshold levels. Phase-locking to the fundamental may also be strengthened due to the stronger fundamental. In the IC, neurons often respond at low frequencies due to broad tuning at suprathreshold levels, thus changing rate representations.

A second difference in the instrumental sounds compared to synthetic is the addition of phase delays. The synthetic stimulus consisted of harmonics all added in sine phase, whereas instruments have phase delays at different frequencies based on the resonances of the instrument body (Fletcher, 1999). These phase changes can result in fast frequency sweeps, called chirps, that are present in both vowels and musical sounds. Octopus cells in the cochlear nucleus (Lu et al., 2022) and IC cells (Mitchell et al., 2023) are sensitive to chirps and chirp direction. Thus, these phase delays may have a large impact on IC encoding of timbre.

Along with timbre, this chapter investigates pitch encoding in natural sounds by presenting oboe and bassoon notes of varying fundamental frequencies (F0). These double-reed instruments were chosen because they had a spectral shape more similar to the synthetic stimuli than most other instruments. Studies investigating pitch coding in the IC with flat spectrum harmonic complexes support both a temporal and a rate code for F0s at different ranges of F0 (Su and Delgutte, 2019, 2020). Temporal codes were sufficient for lower F0s, whereas F0s > 900 Hz could be represented by rate due to resolved harmonics in awake rabbit IC. Behaviorally, rabbits can do pitch discrimination tasks over a wide range of F0s (200 – 1600 Hz) and use spectral and temporal cues

(Wagner et al., 2022). This project studied whether these results were consistent for natural sounds with complex spectral envelopes.

A confounding factor when studying F0 discrimination in natural sounds is that the spectrum covaries with F0 (Siedenburg et al., 2021). The spectral centroid increases as a function of F0 for most instruments (Siedenburg et al., 2021). Additionally, as mentioned previously, psychophysics work using synthetic (Allen and Oxenham, 2014) and instrument sounds (McPherson and McDermott, 2023) find pitch and timbre interference while performing pitch or timbre discrimination tasks. The mechanisms behind this interference remain unknown, though modeling work presents the hypothesis that this interference is present in the IC (Maxwell et al., 2020).

This project sought to understand how instrument identity and F0 are represented in the IC by recording IC responses to bassoon and oboe recordings varying in F0. We tested whether energy, STRF, and IC models could predict these responses, but predictions were not above chance for any model. Instead, a variety of decoding models were used to classify instrument identity or F0 using single neurons or the neural population, either using rate or temporal responses. We found that instrument identity was redundantly encoded in the rates of a diverse set of neurons. A population of neurons, using rate or timing, was needed able to reliably discriminate F0. Single neurons were only able to decode lower F0s using timing. Performance dropped when the F0 task included multiple instruments, indicating that the decoding models benefited when only one instrument was included in the task. This study opens up many questions about

pitch/timbre encoding and interference in the IC, but initial results generally support findings from previous studies using synthetic stimuli.

1.5 Thesis Structure Overview

Chapter 1 has introduced the overarching question of this thesis: How is timbre encoded in the IC? This chapter highlighted spectral peaks as an important attribute for timbre perception, motivated this work by examining how timbre perception degrades with hearing loss, and introduced the IC. An overview of AN encoding was included as background on encoding mechanisms that influence encoding in the IC, including NFs and AM-tuned broad inhibition. Background on the simple timbre stimuli and earlier IC vowel work was discussed. Lastly, a comparison of the synthetized timbre stimulus and natural instrument sounds was presented, and F0/timbre encoding was introduced. Chapters 2, 3, and 4 are all presented in the format of a manuscript and a 'roadmap' of the stimuli used in each chapter is presented in Figure 1.2. These three chapters all use single-neuron recordings in the central nucleus of the IC in awake Dutch-belted rabbits along with IC computational models.

Chapter 2 was published in the Journal of Neuroscience (Fritzinger and Carney, 2025) and investigated two mechanisms that may impact the encoding of complex sounds in the IC. These mechanisms are 1) NFs set up in the AN that may influence rate coding in the IC due to AM sensitivity in the IC, and 2) AM-tuned broad inhibition. This chapter describes physiological results to a wideband noise with an added tone that shifts on- and off-CF (Fig 1.2a) and presents a modified IC computational model with AM-tuned broad inhibition and NF sensitivity that explains physiological results.

Chapter 3 is under review at the Journal of Neuroscience. This chapter describes physiological responses to a harmonic tone complex with a triangular-shaped spectrum that elicits a brightness percept in humans (Fig 1.2b) This chapter presents analysis of rate and temporal responses to these stimuli with spectral peaks over a wide range of CFs and levels and investigates spectral peak encoding in the IC. The accuracy of models from Chapter 1 for these stimuli were tested as well.

Chapter 4 is unpublished work that analyzes IC responses to natural instrument recordings of bassoon and oboe over a range of F0s (Fig 1.2c). This chapter uses decoding models to investigate how pitch and instrument identity are represented in IC neurons. Chapter 5 includes a discussion of the results from the previous chapters along with future directions for this work.

Bibliography

Adams JC (1980) Crossed and descending projections to the inferior colliculus. Neurosci Lett 19:1–5.

Adams JC, Mugnaini E (1984) Dorsal nucleus of the lateral lemniscus: A nucleus of GABAergic projection neurons. Brain Res Bull 13:585–590.

Aitkin LM, Phillips SC (1983) Is the inferior colliculus an obligatory relay in the cat auditory system? :6.

Alkhatib A, Biebel UW, Smolders JWT (2006) Inhibitory and excitatory response areas of neurons in the central nucleus of the inferior colliculus in unanesthetized chinchillas. Exp Brain Res 174:124–143.

Allen EJ, Oxenham AJ (2014) Symmetric interactions and interference between pitch and timbre. J Acoust Soc Am 135:1371–1379.

Andoni S, Li N, Pollak GD (2007) Spectrotemporal Receptive Fields in the Inferior Colliculus Revealing Selectivity for Spectral Motion in Conspecific Vocalizations. J Neurosci 27:4882–4893.

ANSI (1994) Am Stand Acoust Terminol R2004.

Attias H, Schreiner CE (1997) Temporal Low-Order Statistics of Natural Sounds. :7.

Carney LH (2018) Supra-Threshold Hearing and Fluctuation Profiles: Implications for Sensorineural and Hidden Hearing Loss. J Assoc Res Otolaryngol 19:331–352.

Carney LH (2024) Neural Fluctuation Contrast as a Code for Complex Sounds: The Role and Control of Peripheral Nonlinearities. Hear Res 443:108966.

Carney LH, Li T, McDonough JM (2015) Speech Coding in the Brain: Representation of Vowel Formants by Midbrain Neurons Tuned to Sound Fluctuations. eneuro 2:ENEURO.0004-15.2015.

Carney LH, McDonough JM (2019) Nonlinear auditory models yield new insights into representations of vowels. Atten Percept Psychophys 81:1034–1046.

Caspary DM, Ling L, Turner JG, Hughes LF (2008) Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. J Exp Biol 211:1781–1791.

Chen G-D, Tanaka C, Henderson D (2008) Relation between outer hair cell loss and hearing loss in rats exposed to styrene. Hear Res 243:28–34.

Chintanpalli A, Ahlstrom JB, Dubno JR (2016) Effects of age and hearing loss on concurrent vowel identification. J Acoust Soc Am 140:4142–4153.

Dicke U, Ewert SD, Dau T, Kollmeier B (2007) A neural circuit transforming temporal periodicity information into a rate-based representation in the mammalian auditory system. J Acoust Soc Am 121:310–326.

Egorova M, Ehret G, Vartanian I, Esser K-H (2001) Frequency response areas of neurons in the mouse inferior colliculus. I. Threshold and tuning characteristics. Exp Brain Res 140:145–161.

Emiroglu S, Kollmeier B (2008) Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. Brain Res 1220:199–207.

Escabí MA, Miller LM, Read HL, Schreiner CE (2003) Naturalistic Auditory Contrast Improves Spectrotemporal Coding in the Cat Inferior Colliculus. J Neurosci 23:11489–11504.

Fan L, Henry KS, Carney LH (2021) Responses to diotic tone-in-noise stimuli in the inferior colliculus: stimulus envelope and neural fluctuation cues. Hear Res 409:108328.

Fletcher NH (1999) The nonlinear physics of musical instruments. Rep Prog Phys 62:723–764.

Fritzinger, J. B., & Carney, L. H. (2025). Mechanisms of Tone-in-Noise Encoding in the Inferior Colliculus. Journal of Neuroscience, 45(23).

Gfeller K, Witt S, Mehr MA, Woodworth G, Knutson J (2002) Effects of Frequency, Instrumental Family, and Cochlear Implant Type on Timbre Recognition and Appraisal. Ann Otol Rhinol Laryngol 111:349–356.

Grey JM (1977) Multidimensional perceptual scaling of musical timbres. J Acoust Soc Am 61:1270–1277.

Grey JM, Gordon JW (1978) Perceptual effects of spectral modifications on musical timbres. J Acoust Soc Am 63:1493–1500.

Henry KS, Abrams KS, Forst J, Mender MJ, Neilans EG, Idrobo F, Carney LH (2017) Midbrain Synchrony to Envelope Structure Supports Behavioral Sensitivity to Single-Formant Vowel-Like Sounds in Noise. J Assoc Res Otolaryngol 18:165–181.

Joris PX, Schreiner CE, Rees A (2004) Neural Processing of Amplitude-Modulated Sounds. Physiol Rev 84:541–577.

Kim DO, Carney L, Kuwada S (2020) Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. J Neurophysiol 124:1198–1215.

Kong Y-Y, Mullangi A, Marozeau J (2012) Timbre and Speech Perception in Bimodal and Bilateral Cochlear-Implant Listeners. Ear Hear 33:645–659.

Kong Y-Y, Mullangi A, Marozeau J, Epstein M (2011) Temporal and Spectral Cues for Musical Timbre Perception in Electric Hearing. J Speech Lang Hear Res 54:981–994.

Krishna BS, Semple MN (2000) Auditory Temporal Processing: Responses to Sinusoidally Amplitude-Modulated Tones in the Inferior Colliculus. J Neurophysiol 84:255–273.

Landemard A, Bimbard C, Demené C, Shamma S, Norman-Haignere S, Boubenec Y (2021) Distinct higher-order representations of natural sounds in human and ferret auditory cortex. eLife 10.

LeBeau FEN, Malmierca MS, Rees A (2001) Iontophoresis In Vivo Demonstrates a Key Role for GABA A and Glycinergic Inhibition in Shaping Frequency Response Areas in the Inferior Colliculus of Guinea Pig. J Neurosci 21:7303–7312.

Lesica NA, Grothe B (2008) Dynamic Spectrotemporal Feature Selectivity in the Auditory Midbrain. J Neurosci 28:5412–5421.

Liberman MC (1978) Auditory-nerve response from cats raised in a low-noise chamber. J Acoust Soc Am 63:442–455.

Looi V, She J (2010) Music perception of cochlear implant users: A questionnaire, and its implications for a music training program. Int J Audiol 49:116–128.

Lu H-W, Smith PH, Joris PX (2022) Mammalian octopus cells are direction selective to frequency sweeps by excitatory synaptic sequence detection. Proc Natl Acad Sci 119.

Lyzenga J, Horst JW (1995) Frequency discrimination of bandlimited harmonic complexes related to vowel formants. J Acoust Soc Am 98:1943–1955.

Lyzwa D, Herrmann JM, Wörgötter F (2016) Natural Vocalizations in the Mammalian Inferior Colliculus are Broadly Encoded by a Small Number of Independent Multi-Units. Front Neural Circuits 9.

Marozeau J, Lamping W (2019) Timbre Perception with Cochlear Implants. In: Timbre: Acoustics, Perception, and Cognition (Siedenburg K, Saitis C, McAdams S, Popper AN, Fay RR, eds), pp 273–293 Springer Handbook of Auditory Research. Cham: Springer International Publishing.

Maxwell BN, Fritzinger JB, Carney LH (2020) Neural Mechanisms for Timbre: Spectral-Centroid Discrimination based on a Model of Midbrain Neurons. In, pp 4.

McAdams S (2019) The Perceptual Representation of Timbre. In: Timbre: Acoustics, Perception, and Cognition (Siedenburg K, Saitis C, McAdams S, Popper AN, Fay RR, eds), pp 23–57 Springer Handbook of Auditory Research. Cham: Springer International Publishing.

McAdams S, Winsberg S, Donnadieu S, De Soete G, Krimphoff J (1995) Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychol Res 58:177–192.

McPherson MJ, McDermott JH (2023) Relative pitch representations and invariance to timbre. Cognition 232:105327.

Mitchell, P. W., & Carney, L. H. (2025). Chirp sensitivity and vowel coding in the inferior colliculus. Hearing Research, 109307.

Mitchell PW, Henry KS, Carney LH (2023) Sensitivity to direction and velocity of fast frequency chirps in the inferior colliculus of awake rabbit. Hear Res 440:108915.

Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J Acoust Soc Am 116:2173–2186.

Oliver DL (2000) Ascending efferent projections of the superior olivary complex. Microsc Res Tech 51:355–363.

Palmer AR, Kuwada S (2005) Binaural and Spatial Coding in the Inferior Colliculus. In: The Inferior Colliculus, pp 377–410. New York: Springer-Verlag. Available at: http://link.springer.com/10.1007/0-387-27083-3_13 [Accessed July 10, 2025].

Palmer AR, Shackleton TM, Sumner CJ, Zobay O, Rees A (2013) Classification of frequency response areas in the inferior colliculus reveals continua not discrete classes. J Physiol 591:4003–4025.

Palombi PS, Caspary DM (1996) Physiology of the aged Fischer 344 rat inferior colliculus: responses to contralateral monaural stimuli. J Neurophysiol 76:3114–3125.

Peeters G, Giordano BL, Susini P, Misdariis N, McAdams S (2011) The Timbre Toolbox: Extracting audio descriptors from musical signals. J Acoust Soc Am 130:2902–2916.

Pollak GD (2011) Discriminating among complex signals: the roles of inhibition for creating response selectivities. J Comp Physiol A 197:625–640.

Portfors CV, Felix RA (2005) Spectral integration in the inferior colliculus of the CBA/CaJ mouse. Neuroscience 136:1159–1170.

Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. J Acoust Soc Am 66:470–479.

Saint-Marie RL, Baker RA (1990) Neurotransmitter-specific uptake and retrograde transport of [3H]glycine from the inferior colliculus by ipsilateral projections of the superior olivary complex and nuclei of the lateral lemniscus. Brain Res 524:244–253.

Siedenburg K, Jacobsen S, Reuter C (2021) Spectral envelope position and shape in sustained musical instrument sounds. J Acoust Soc Am 149:3715–3726.

Su Y, Delgutte B (2019) Pitch of harmonic complex tones: rate and temporal coding of envelope repetition rate in inferior colliculus of unanesthetized rabbits. J Neurophysiol 122:2468–2485.

Su Y, Delgutte B (2020) Robust Rate-Place Coding of Resolved Components in Harmonic and Inharmonic Complex Tones in Auditory Midbrain. J Neurosci 40:2080–2093.

Tan Q, Carney LH (2005) Encoding of vowel-like sounds in the auditory nerve: Model predictions of discrimination performance. J Acoust Soc Am 117:1210–1222.
Theunissen FE, Elie JE (2014) Neural processing of natural sounds. Nat Rev Neurosci 15:355–366.

Wagner JD, Gelman A, Hancock KE, Chung Y, Delgutte B (2022) Rabbits use both spectral and temporal cues to discriminate the fundamental frequency of harmonic complexes with missing fundamentals. J Neurophysiol 127:290–312.

Williams AJ, Fuzessery ZM (2011) Differential roles of GABAergic and glycinergic input on FM selectivity in the inferior colliculus of the pallid bat. J Neurophysiol 106:2523–2535.

Winer JA, Larue DT, Pollak GD (1995) GABA and glycine in the central auditory system of the mustache bat: Structural substrates for inhibitory neuronal organization. J Comp Neurol 355:317–353.

Winer JA, Schreiner C eds. (2005) The inferior colliculus: with 168 illustrations. New York, NY: Springer.

Xie R, Meitzen J, Pollak GD (2005) Differing Roles of Inhibition in Hierarchical Processing of Species-Specific Calls in Auditory Brainstem Nuclei. J Neurophysiol 94:4019–4037.

Yang L, Pollak GD, Resler C (1992) GABAergic circuits sharpen tuning curves and modify response properties in the mustache bat inferior colliculus. J Neurophysiol 68:1760–1774.

Young ED, Sachs MB (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. J Acoust Soc Am 66:1381–1403.

Zilany MSA, Bruce IC (2007) Representation of the vowel /ɛ/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats. J Acoust Soc Am 122:402–417.

Chapter 2: Mechanisms of Tone-in-Noise Encoding in the Inferior Colliculus

Fritzinger JB, Carney LH (2025) Mechanisms of Tone-in-Noise Encoding in the Inferior Colliculus. Journal of Neuroscience 45.

2.1 Abstract

Extracellular single-unit responses to tone-in-noise (TIN) stimuli were recorded in the inferior colliculus (IC) of awake female Dutch-belted rabbits. Stimuli consisted of wideband and narrowband TIN with on-and off-characteristic frequency tones. Neural responses to wideband TIN showed a pattern of rates that increased when the tone matched CF and decreased (with respect to noise-alone responses) when the tone was above or below CF. This result differed from narrowband TIN IC responses that depended on envelope fluctuations in the stimulus, consistent with neural-fluctuation sensitivity. The wideband TIN responses could be fit with a difference-of-gaussians model that had narrow excitation and broad inhibition; responses to TIN could not be predicted by response-maps or spectrotemporal receptive fields. Responses to diotic and contralateral presentations of wideband TIN did not differ due to presentation ear. A single-CF computational model of the IC could not predict responses to wideband TIN. However, adding local off-CF inhibitory inputs to an on-CF IC model improved accuracy. These results suggest that broad inhibition could explain encoding of wideband TIN at suprathreshold signal-to-noise ratios, whereas neural fluctuation sensitivity is more important for narrowband sounds.

2.2 Introduction

In this study we compared mechanisms that influence tone-in-noise (TIN) encoding in the inferior colliculus (IC). TIN stimuli have been used extensively to gain insight on aspects of auditory perception, including energy and envelope cues, binaural perception, and critical bands (Fletcher, 1940; Kidd et al., 1989; Richards, 1992; van de Par and Kohlrausch, 1999). However, neural mechanisms underlying TIN representation in the IC are not well understood. Neural fluctuations (NFs) are hypothesized to provide a cue for tone detection in noise. NFs, or low-frequency changes in amplitude of auditory-nerve (AN) rate functions due to beating between stimulus components, are converted to rate changes by periodicity-tuned neurons in the IC (Carney, 2018, 2024). Addition of a tone to a noise reduces the depth of envelope fluctuations in the stimulus (Richards, 1992) and of NFs (Carney, 2018), providing a cue for tone detection. IC responses are consistent with NF predictions in awake rabbit for an on-characteristic frequency (CF) tone in a narrowband (1/3-octave) noise (Fan et al., 2021, 2022)(Fig. 2.1a).

IC neurons that are enhanced by a band of modulation frequencies, band-enhanced (BE), decrease in rate as SNR increases due to the reduction of NFs at CF (Fig. 2.1c). IC neurons suppressed by modulation, band-suppressed (BS), increase in rate again due to the reduction of NFs at CF (Fig. 2.1d). To further explore how NFs influence physiological responses to TIN, we recorded responses in the IC using a novel paradigm: a wideband noise with an added tone that was stepped past the CF (Fig. 2.1b). The predicted response to WB-TIN was a decrease in rate near- and on-CF, dependent on the periodicity tuning in the IC neuron (Fig. 2.1c,d). Interestingly, the qualitative trends in

responses to wideband TIN (WB-TIN) differed from narrowband TIN (NB-TIN) responses, suggesting that mechanisms in addition to NFs are involved. This study investigated how NFs and inhibition interact to represent WB-TIN in the IC.

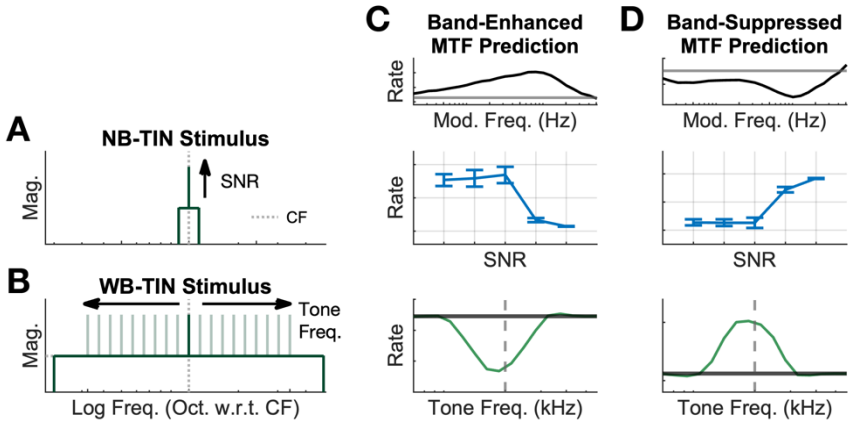

Figure 2.1. Illustration of stimuli and neural fluctuation hypothesis using a computational model of the IC. (A) NB-TIN stimulus with tone increasing in SNR at CF (B) WB-TIN stimulus with tone varying in frequency. (C) Band-enhanced, BE, prediction based on neural fluctuation mechanism. MTF (top), response to increasing SNR NB-TIN (middle), and WB-TIN (bottom). (D) Band-suppressed, BS prediction based on neural fluctuation mechanisms, MTF, NB-TIN, and WB-TIN responses.

TIN stimuli have been widely used in psychophysical and physiological studies to investigate auditory filters in the context of the power-spectrum model, which postulates that tones are encoded by increased energy in the response of peripheral filters tuned to the tone frequency (Moore, 1975; Patterson, 1976). Physiological evidence for energy coding of on-CF tones in broadband noise include increasing rate as a function of signal-to-noise ratio (SNR) in IC neurons (Jiang et al., 1997; Rocchi and Ramachandran, 2018). However, the power-spectrum model is not consistent with psychophysical results, which show little effect on TIN detection performance when energy cues are made unreliable by

random variation of sound level across intervals (Kidd et al., 1989; Richards, 1992). Psychophysical and modeling studies report that envelope-based cues are sufficient to describe TIN detection thresholds and outperform energy-based models (Dau et al., 1996; Kohlrausch et al., 1997; Richards, 1992; Davidson et al., 2009; Mao et al., 2013; Mao and Carney, 2015). These psychophysical and modeling TIN results suggest a role for the IC in representation of envelope cues because most IC neurons are sensitive to amplitude modulation (AM) (Langner and Schreiner, 1988; Krishna and Semple, 2000; Joris et al., 2004; Nelson and Carney, 2007; Kim et al., 2020).

Inhibition also plays an important role in shaping IC cell responses to sound (Casseday et al., 1994; Koch and Grothe, 1998; Caspary et al., 2008; Pollak et al., 2011), though the effect of inhibition on TIN responses has not been fully explored. Two-tone paradigms in the IC have revealed inhibitory response regions that may influence TIN encoding (Egorova et al., 2001; Portfors and Felix, 2005; Alkhatib et al., 2006; Palmer et al., 2013). The frequency range over which neurons are excited broadens when inhibition is blocked or reduced, supporting a role of broad inhibition in shaping responses to tones (Yang et al., 1992; Palombi and Caspary, 1996; LeBeau et al., 2001; Xie et al., 2005). Sideband inhibition is observed in the response maps (RMs) of I-type IC neurons, which are hypothesized to encode tones in background noise levels more robustly than other RM types (Ramachandran et al., 1999, 2000).

This study investigated how NF sensitivity and off-CF inhibition contribute to TIN encoding by characterizing responses in the IC of awake rabbit to TIN stimuli with on- or off-CF tones. Responses of single neurons were compared for on-CF WB-TIN and

NB-TIN. Next, responses to WB-TIN were characterized for a range of spectrum levels and SNRs, for contralateral and diotic stimuli. A difference-of-gaussians (DOG) model (Su and Delgutte, 2020) was used to explore excitatory and inhibitory components in the WB-TIN response. Lastly, computational IC models tested the hypothesis that interactions of NFs and off-CF inhibition influence TIN encoding.

2.3 Methods

2.3.1 Animals & Surgical Procedures

All procedures were approved by the University of Rochester Committee on Animal Resources in compliance with National Institutes of Health Guidelines. Four female Dutch-belted rabbits (*Oryctolagus cuniculus*) ranging from 6 months to 5 years of age were screened for normal hearing periodically using distortion product otoacoustic emissions (Whitehead et al., 1992), and experiments were discontinued in an animal if emission amplitudes decreased by 10 dB.

Surgeries included an initial headbar placement, initial craniotomy and microdrive placement, and microdrive replacements. For all surgeries, rabbits were anesthetized with either 66 mg/kg intramuscular ketamine and 2 mg/kg intramuscular xylazine or 35mg/kg intramuscular ketamine and 0.10-0.15mg/kg intramuscular dexmedetomidine. Animals were given the analgesic meloxicam (0.2mg/kg) subcutaneously for three days following surgery and were monitored for normal behavior.

A custom 3D-printed plastic headbar (ProtoLabs, Maple Plain, Minnesota) was affixed to the dorsal surface of the skull using stainless-steel screws and dental acrylic.

After a one-month recovery period, the initial craniotomy and placement of the microdrive was performed. The microdrive was replaced every 1-6 months and the placement of the tetrodes was varied to sample different locations in the IC. Custom earmolds (Dreve Otoform Ak, Unna, Germany) were cast at the end of the microdrive surgeries while the animal was still under anesthesia.

For daily two-hour physiological recording sessions in a sound-attenuated booth (Acoustic Systems, Austin, Texas), awake rabbits were placed in a custom chair and the head was fixed. Custom earmolds were inserted and audio presentation hardware was calibrated at the beginning of the session using a probe-tube microphone (Etymotic ER10B+ or ER7C, Etymotic Research, Inc., Elk Grove Village, Illinois).

2.3.2 Microdrive and Electrodes

Extracellular recordings were made using a chronically implanted microdrive. The microdrive was 3D printed plastic, modeled after the Neuralynx Five-drive microdrives and supporting EIB-16 Five-Drive board connectors (Neuralynx, Inc., Bozeman, MT). The microdrive positioned four tetrodes, which were made by twisting four 18-μm platinum-iridium wires insulated with epoxy (California Fine Wire Co., Grover Beach, CA). The tetrodes were inserted into polyimide tubes (Polymicro Technologies Inc., Phoenix, AZ) which were passed through four stainless-steel 28-gauge guide tubes (Eagle Stainless, Franklin, MA). Tetrodes were plated with platinum-black plating solution (Neuralynx, Inc., Bozeman, MT) to lower impedances to a range from 0.1-1.5 MΩ.

2.3.3 Spike Sorting & Clustering

Signals were recorded using the Intan RHD recording system, including a 16-channel headstage with initial filtering stages (16-Channel Headstage with RHD2132 chip, Intan Technologies, LLC., Los Angeles, CA). The sampling frequency was set at 30 kHz. These initial filters included a 1st-order, analog, high-pass filter (Fc = 150 Hz), a 3rd-order, analog, Butterworth low-pass filter (Fc = 7.5 kHz), and a digital high-pass filter (Fc = 300 Hz) and saved by the Intan RHD recording system for post-hoc filtering and analysis (Intan Technologies, LLC., Los Angeles, CA).

To isolate neurons for analysis, signals were filtered, thresholded, and then clustered into separate neurons, as described below. First, the signals were filtered with a 4th-order Butterworth bandpass filter (300-3000 Hz). The majority of potential spikes were detected using a threshold of 4 times an estimate of the standard deviation (STD), with the standard deviation estimated by $\sigma_{est} = median(\frac{|x|}{\sqrt{2}\, erf^{-1}(0.5)})$ (Quiroga et al., 2004). The threshold was adjusted higher for a few sessions to improve clustering of neurons that appeared to be separable but were not clustered well using the 4 x STD threshold criterion. For a few sessions, artifacts due to movement were present during the first dataset, and a different dataset was used to calculate the STD of the signal. If two above-threshold events occurred within 0.5 ms, only the event with the larger amplitude was determined to be a potential spike. All potential spike waveforms and times were then saved for clustering. Most neurons were clustered by comparing repolarization slopes; a few neurons were better isolated using principal components analysis. (Schwarz et al., 2012)

Three criteria were required to classify a recording as a well-isolated neuron. First, less than 2% of the inter-spike intervals could be less than 1-ms. Second, spike waveforms for each wire, when aligned by their peaks, were required to have similar shapes. Third, although many recordings could potentially have been clustered into several neurons, clustering was deemed valid only if the quantitative cluster-separation metric (Schwarz et al., 2012) was < 0.1. Cells were classified as unique neurons when the tetrodes were moved and the response properties of the neuron changed.

To verify that tetrode locations were in the central nucleus of the IC, responses to a set of characterizing stimuli were monitored for properties that matched typical IC recordings, e.g., responses to tones and diotic or dichotic noises. CFs were also monitored to check that as tetrode depth increased, CF increased (Schreiner and Langner, 1997). After an animal showed a 10-dB reduction in emission amplitudes or had elevated neural thresholds, the animal was euthanized and histology was completed to confirm tetrode placement.

2.3.4 Basic Stimuli & Analysis

Stimuli were created in MATLAB (Mathworks, Natick, MA) and presented through an audio interface (16A, Mark of the Unicorn, Cambridge, Massachusetts), a digital-to-analog converter (DAC3 HGC, Benchmark Media Systems, Inc., Syracuse, New York), and earphones (Beyerdynamic DT-48, Beyerdynamic GmbH and Co., Heilbronn, Germany or Etymotic ER2, Etymotic Research, Inc., Elk Grove Village, Illinois). The earphones were coupled with the custom earmolds. The sampling frequency for sound presentation was 48,000 Hz.

At the beginning of each session, a set of stimuli was presented to characterize the neuron. These stimuli included tones to construct an RM, sinusoidally modulated noise for a modulation transfer function (MTF), binaural noise to characterize basic binaural response properties, and wideband noise to construct a spectrotemporal receptive field (STRF).

The RM was created by presenting pure-tone stimuli at 10, 30, 50, and 70 dB SPL, with frequencies spanning 250-16000 Hz, with five steps per octave. The stimuli were diotic, 200 ms in duration, with 10-ms $\cos^2$ on/off ramps, repeated 3 times in a random sequence. The CF was estimated by finding the tone frequency that elicited a response at the lowest sound level. If the CF based on the response to pure tones was unclear, the peak of the noise STRF was used to estimate CF. Neurons were categorized as V-type, I-type, O-type, onset, onset/offset, inhibitory, and unusual. Only V-type, I-type, and onset neurons are included in analysis due to small sample size of other categories. Onset neurons were defined as neurons with an onset response but no sustained response to the RM stimulus. I-type neurons had a sustained response and the bandwidth at 70 dB SPL did not exceed 2 times the bandwidth at threshold. Neurons were categorized as V-type if they had a sustained response and the bandwidth at 70 dB SPL was greater than 2 times threshold bandwidth. For some analyses, neurons were separated into three groups based on CF: low (CF $<$ 2 kHz), medium (2 kHz $\leq$ CF $<$ 4 kHz), and high (CF $\geq$ 4 kHz).

The MTF of the neuron was determined by presenting 100% sinusoidally amplitude-modulated broadband (100-10000 Hz) diotic, flat-spectrum, gaussian noise. The amplitude-modulation frequencies tested were 2-600 Hz in increments of three steps

per octave, presented at 33 dB SPL spectrum level, with a duration of 1 s including 50-ms $\cos^2$ ramps, and repeated five times, in a random sequence. Each neuron was categorized into one of four MTF classification types: band-enhanced (BE), band-suppressed (BS), hybrid, and flat (Kim et al., 2020). MTFs were classified as BE if at least two average-rate responses to modulation frequencies were significantly larger than the unmodulated rate (i.e., average rate in response to an unmodulated noise), over a range of modulation frequencies that was not interrupted by a rate significantly lower than the unmodulated rate (unpaired t-test, $p < 0.05$)(Fig. 2.1c). Conversely, MTFs were classified as BS if at least two average-rate responses to modulation frequencies were significantly lower than the unmodulated rate, over a range of modulation frequencies that was not interrupted by a rate significantly above the unmodulated rate (unpaired t-test, $p < 0.05$)(Fig. 2.1d). A cell was classified as hybrid if conditions for both BE and BS categories were met. A cell was classified as flat if neither BE nor BS condition was met. The best modulation frequency (BMF) for BE and hybrid neurons was the modulation frequency that resulted in the highest spline-interpolated, average-rate response. Conversely, the worst modulation frequency (WMF) for BS and hybrid neurons was the modulation frequency that resulted in the lowest spline-interpolated average-rate response. Hybrid neurons were divided into two categories: neurons were classified as H_{BE} if the rate of the neuron in response to modulation frequencies near 100 Hz was greater than the unmodulated response and H_{BS} if responses near 100 Hz were suppressed. The responses near 100-Hz modulation were used in these categorizations because the hybrid neuron response

properties were more similar to the BE or BS MTF-classification type that matched their response to 100-Hz modulation.

Binaural response properties were determined by presenting a wideband (100-15,000 Hz), flat-spectrum, gaussian noise at 0, 10, 20, or 30 dB SPL spectrum level to the ipsilateral, contralateral, or both ears. The stimuli were 1-s duration with 10-ms $\cos^2$ ramps and were presented three times in a random sequence. The average-rate responses were used to identify the contribution of each ear to the binaural response.

The STRF was estimated based on responses to flat-spectrum gaussian noise (100-16,000 Hz) with a 2-s duration, presented at an overall level of 68 dB SPL. Ten repetitions each of 25 different noise tokens were presented in a random sequence, either diotically or to the contralateral ear. The STRFs were created using a 2nd-order Wiener-kernel analysis based on Lewis et al. (2002). Briefly, the 2nd-order reverse correlation was computed between the gaussian-noise stimulus and the response of the neuron. Then, singular-value decomposition was used to decompose the kernel into excitatory and inhibitory components (Lewis et al., 2002).

2.3.5 Tone-in-Noise Stimuli & Analyses

The NB-TIN stimulus consisted of a 1/3 octave flat-spectrum gaussian noise with a tone centered at CF (Fig. 2.1a). The tone frequency was shifted one octave above and one octave below CF, in 6 steps per octave. To create the NB-TIN stimuli, a single frozen broadband, flat-spectrum, gaussian noise was filtered to a 1/3-octave bandwidth centered on each tone frequency, using a 5000-th order finite-impulse-response filter. SNRs tested were –inf (noise alone), 20, 30, and 40 dB tone level with respect to noise spectrum level

in dB SPL. Stimuli had 300-ms duration and 10-ms cos2 on/off ramps and were presented at noise spectrum levels (N_0) of 3, 23, or 43 dB SPL. Twenty repetitions of each NB-TIN stimulus were presented; stimuli were diotic or contralateral. During recording sessions, the CF of the neuron was not precisely known, so shifting the NB-TIN spectrum in frequency ensured that the neural response to an on-CF tone was recorded.

The WB-TIN stimulus consisted of a broadband, flat-spectrum, gaussian noise (3- or 4-octave bandwidth) geometrically centered on CF with an added pure tone (Fig. 2.1b). The tone was varied in frequency over 3 octaves centered on CF, in steps of 6 per octave. SNRs matched NB-TIN stimuli. The SNRs were selected to straddle rabbit WB-TIN behavioral detection thresholds (Zheng et al., 2002). Stimuli were presented in random order, with 300-ms duration and 10-ms $\cos^2$ on/off ramps, and repeated 30 times, with each repetition being a different gaussian-noise token. Diotic and contralateral presentations of the WB-TIN stimulus had noise spectrum levels (N_0) of 3, 23, or 43 dB SPL.

The rate profiles for WB-TIN and NB-TIN were calculated by averaging the firing rate over the duration of the stimulus, excluding a 50-ms onset, and plotted as a function of tone frequency. On-CF responses for NB- and WB-TIN were determined by taking the average rate response of the neuron for the stimulus with the tone nearest to CF. The proportion of predictable variance was quantified by $V_p = \frac{r_{12}}{r_{12}+0.5(1-r_{12})}$, where V_p is the proportion of predictable variance, and r_{12} is the first-half − last-half correlation between responses to repetitions of the stimulus (Ahumada and Lovell, 1971).

WB-TIN rate profiles were compared to linear predictions based on RMs and noise STRFs. Variance explained was calculated between the RM level closest to the overall level of the WB-TIN and the WB-TIN response. To compare WB-TIN responses to the noise STRF of the neuron, the noise STRF was used to predict a neuron's response to the WB-TIN stimulus. First the spectrogram of the sound stimulus was calculated. Then the spectrogram was convolved frequency-by-frequency with the noise STRF. Next, the result was summed over frequency to create a predicted peri-stimulus time histogram (PSTH), which was then averaged over time to compute a predicted average rate. The variance explained by the STRF-model average rate was calculated to determine similarity between STRF and WB-TIN responses.

Neural responses to WB-TIN were fit to a DoG function for further analysis of excitation and inhibition (Su and Delgutte, 2020):

$$W_{DoG}(f) = g_e e^{\frac{-(f-f_e)^2}{2\sigma_e^2}} - g_i e^{\frac{-(f-f_i)^2}{2\sigma_i^2}} \quad ,$$

where f is the frequency range of the tone in the WB-TIN stimulus. The mean of the excitatory and inhibitory gaussian functions, f_e and f_i, respectively, were near the CF of the neuron. All frequencies and bandwidths were fit in logarithmic units and transformed into linear units for plotting. The standard deviations, σ_e and σ_i , represent the excitatory and inhibitory bandwidths. The strength of the excitatory and inhibitory gaussians are g_e and g_i, respectively. The DoG function was fit to the neural data with the noise-alone response subtracted by minimizing the squared error using the MATLAB function fmincon.

2.3.6 Statistical Analyses

Two linear mixed-effect models were fit to detect statistically significant changes in the WB-TIN response due to stimulus parameters and neuron characteristics. MATLAB functions *LME* and *compare* were used for model selection and the statistical software package JASP (JASP Team, 2024, Version 0.18.3, computer software) was used for the remaining analysis. The WB-TIN rates for each neuron were normalized using a noise-referenced z-score approach: $z = \frac{x-k}{\sigma}$, where x was the average rate response, k is the noise-alone rate, and σ is the standard deviation of x. Both models included neuron identity as a random effect, specifically a random intercept, to account for the differences in rate for each neuron.

The first model included the following stimulus parameters as fixed effects: tone frequency, in -1 to 1 octaves above CF, in increments of 0.25 octaves, spectrum level, SNR, and presented ear. All fixed effects were coded as categorical variables. The model chosen for this analysis included two-, three-, and four-way interactions between fixed effects.

The second model included the following stimulus parameters and neural characteristics as fixed effects: tone frequency, spectrum level, CF group (low, medium, high), MTF type, and RM type as categorical variables. The data for this second model was curated to include 40-dB SNR data and only V- and I- RM types due to small sample sizes of other RM types. The binaural/contra variable did not improve model fit and was not included as a parameter. The model chosen for this analysis included two-, three-, and

four- way interactions of the fixed effects. Both model structures were selected using forward stepwise regression based on the Akaike information criterion.

The model fits were assessed by visual examination of residual and quantile-quantile plots in MATLAB. The models were analyzed using type III ANOVAs in JASP using the Satterthwaite approximation. Further analysis of the results was done using contrast testing. Other statistical methods used throughout this study include t-tests, ANOVAs, and comparisons of correlation coefficients and variance explained.

2.3.7 Modeling

All models used the WB-TIN stimulus with a sampling rate of 100 kHz. To compare the neuron responses to WB-TIN to predictions of the energy model, an estimate of energy was computed by passing the stimulus through a 4th-order gammatone filter and taking the RMS of the filter response. Each gammatone filter was centered on the CF of the neuron of interest. The gammatone tau parameter was calculated by finding the cat Q_{10} value for a specific CF, ($Q_{10} = 10^{0.4708*\log_{10}(CF/1000)} + 0.4664$) (Carney and Yin, 1988), converting to $ERB_{CF} = CF/Q_{10}$, then calculating tau, $\tau = \frac{1}{2\pi*1.019*ERB_{CF}}$ (Patterson et al., 1988).

For all remaining models, a phenomenological AN model stage (Zilany et al., 2014) was followed by a same-frequency inhibition-excitation (SFIE) IC model (Nelson and Carney, 2004; Carney and McDonough, 2019). The SFIE model is a phenomenological model that replicates amplitude-modulation sensitivity seen in physiological recordings of IC neurons, including BE (Fig. 2.2a) and BS (Fig. 2.2b)

MTFs (Joris et al., 2004; Nelson and Carney, 2004; Kim et al., 2020). The IC model was expanded to include three SFIE paths with different CFs: one centered on the tone frequency, and two off-CF pathways symmetrically in octaves above and below the on-CF pathway. Six different configurations with off-CF inhibition were tested: ascending off-CF inhibition from cochlear nucleus (CN) to a BE or BS cell, lateral off-CF inhibition from a BS or BE off-CF cell to a target on-CF BE cell, and lateral off-CF inhibition from a BE or BS cell to a target on-CF BS cell (Fig. 2.2c).

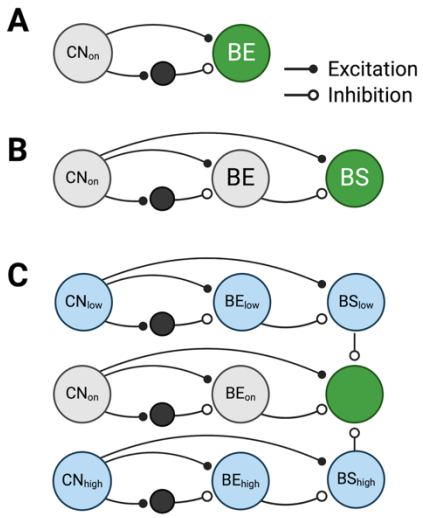

Figure 2.2. Schematics of SFIE IC model and one configuration that added off-CF inhibition. (A) SFIE model configuration with an on-CF CN stage (CN_{on}, grey) and IC BE output cell (green). (B) SFIE model with an output cell that is an IC BS cell (green), on-CF stages are grey. (C) Model configuration with the output model cell as a BS on-CF SFIE BS cell (green) inhibited by off-CF BS cells. On-CF pathways shown in grey, off-CF pathways shown in blue.

All model configurations had on-CF excitation, on-CF inhibition, and two off-CF inhibitions, based on this structure:

$$\mathrm{r_{IC}(t) = g_{ex_{cf}} - g_{inh_{cf}} - g_{inh_{lo}} - g_{inh_{hi}}} \quad ,$$

where $\mathrm{g_{ex_{cf}}}$ was the excitatory on-CF input, $\mathrm{g_{inh_{cf}}}$ was the inhibitory on-CF input, $\mathrm{g_{inh_{lo}}}$ was the inhibitory input from below CF, and $\mathrm{g_{inh_{hi}}}$ was the inhibitory input from above CF. The combined response, $r_{IC}(t)$, which represented the instantaneous firing rate of the IC model neuron, was half-wave rectified. The four model inputs were described as

$$\mathrm{g_{ex_{cf}}(t) = \frac{t}{\tau^2} e^{-t/\tau_{exc}} * r_{CN_{cf}}(t)} \quad ,$$

$$g_{inh_{cf}} = S_{inh} \times \frac{t}{\tau^2} e^{-t/\tau_{inh}} * r_{inh_{cf}}(t - D_{inh}) \quad ,$$

$$g_{inh_{lo}} = S_{lo} \times \frac{t}{\tau^2} e^{-t/\tau_{lo}} * r_{lo}(t - D_{off}) \quad , \qquad \text{and}$$

$$g_{inh_{hi}} = S_{hi} \times \frac{t}{\tau^2} e^{-t/\tau_{hi}} * r_{hi}(t - D_{off}) \quad .$$

Each equation represents an input convolved with an inhibitory or excitatory alpha function, where the alpha function was normalized to an area of 1. The inhibitions were delayed, by *D*, and multiplied by a strength term, *S.* The excitation strength was 1. The input $r_{CN_{cf}}$ is the instantaneous rate of the cochlear nucleus on-CF model neuron, the inputs $r_{inh_{cf}}$, r_{lo}, and r_{hi} , and S_{inh} and D_{inh}, the strength and delay of the on-CF inhibition, respectively, varied across the six model configurations (Table 1). The time constants, τ, were parameters for the alpha functions, where $\tau_{exc} = \frac{1}{10 \times BMF_{on}}$, $\tau_{inh} = 1.5\tau_{exc}$, $\tau_{lo} = \frac{1}{10 \times BMF_{lo}}$, and $\tau_{hi} = \frac{1}{10 \times BMF_{hi}}$.

	$r_{inh_{CF}}$	r_{lo}	r_{hi}	S_{inh}	D_{inh}
BE inhibited by off-CF CN	$r_{CN_{cf}}$	$r_{CN_{lo}}$	$r_{CN_{hi}}$	0.9	$2\tau_{exc}$
BE inhibited by off-CF BE		$r_{BE_{lo}}$	$r_{BE_{hi}}$		

BE inhibited by off-CF BS		$r_{BS_{lo}}$	$r_{BS_{hi}}$		
BS inhibited by off-CF CN	$r_{BE_{cf}}$	$r_{CN_{lo}}$	$r_{CN_{hi}}$	4	$2\tau_{exc}$
BS inhibited by off-CF BS		$r_{BS_{lo}}$	$r_{BS_{hi}}$		
BS inhibited by off-CF BE		$r_{BE_{lo}}$	$r_{BE_{hi}}$		

Table 2.1. Model inputs and fixed parameters for all six lateral inhibition model configurations.

The parameters that were fit to each model configuration were the off-CF values, S_{lo}, S_{hi}, and BMF_{on}, BMF_{lo}, and BMF_{hi}. The delay parameter, D_{off}, was set to 1 ms after an initial grid search revealed that the delay parameter did not impact average-rate results. The stimuli used to fit the model to example neural data included the MTF, an on-CF NB-TIN at 23 dB SPL and -inf, 20, 40 dB SNR, and a WB-TIN stimulus at N_0 = 23 dB and 40 dB SNR. For the fitting procedure, first the AN model was run with three different CFs: CF_{lo}, CF_{on}, and CF_{hi}, where the low and high CF value was varied in increments of 1/6 octave +/- CF up to 1.5 octaves +/- CF. Then, the MATLAB function fmincon with the sequential quadratic programming (SQP) algorithm was used to minimize an objective function to fit the broad inhibition parameters for each CF range. The fit parameters were initialized with random numbers within a bounded range. The range for strength parameters was 0 to 1, and the range for BMFs was 10 to 300 Hz. The objective function was $R_{avg} = -1 * \frac{R_{MTF}+R_{NB}+R_{WB}}{3}$, where R_{avg} is the mean of the correlation between the data and the predicted model for each dataset. After a model was fit to each CF range (12 total), the CF range and parameter set with the highest average correlation coefficient was used as the best fit to the data. CF range was not included in the gradient descent method due to the computational expense of running the AN model for each CF.

Code Accessibility

Custom MATLAB code for clustering data (Schwarz et al., 2012) available at https://www.urmc.rochester.edu/labs/carney/publications-code/spike-sorting-code.aspx. Custom MATLAB code for data analysis and modeling along with data files are available at https://osf.io/p4r82/. Code is also available on GitHub at https://github.com/jfritzinger/FritzingerCarney2025-WBTIN.

2.4 Results

Single-neuron activity was recorded from 229 neurons in the central nucleus of the IC in four awake Dutch-belted rabbits. All neurons had a response to binaural broadband noise greater than 5 sp/s. Responses of 127 neurons to both diotic and contralateral WB-TIN were recorded. Responses of an additional 80 neurons to diotic WB-TIN, and 21 neurons to contralateral WB-TIN were recorded. Responses to NB-TIN were recorded in 106 of the neurons, and noise STRFs were recorded in 170 neurons. Distributions of neuron characteristics matched similar datasets in awake rabbit (Kim et al., 2020; Fan et al., 2021). The CFs ranged from 328 Hz to 10.6 kHz, with a median of 2.8 kHz (Fig. A1b). The MTF distribution was 58 BE (25.3%), 104 BS (45.4%), 43 hybrid (18.8%), and 24 flat (10.5%) (Fig. A1a). The BMF of the BE neurons ranged from 12 Hz to 412 Hz, with a median of 70 Hz (Fig. A1a). The worst modulation frequency (WMF) of BS neurons ranged from 2 Hz to 512 Hz with a median of 102 Hz (Fig. A1c). Hybrid neurons have both a BMF and a WMF; Hybrid BMFs ranged from 5 Hz to 512 Hz with a median of 64 Hz, and WMFs ranged from 2 Hz to 478 Hz with a median of

125 Hz (Fig. A1c). For the purpose of analysis and modeling, the predictable variance, Vp, of the response to WB-TIN was calculated for all conditions to identify reliable responses; conditions with a Vp < 0.4 were excluded from analysis (Fig. A1d).

2.4.1 Single-neuron response to frequency-shifted WB-TIN

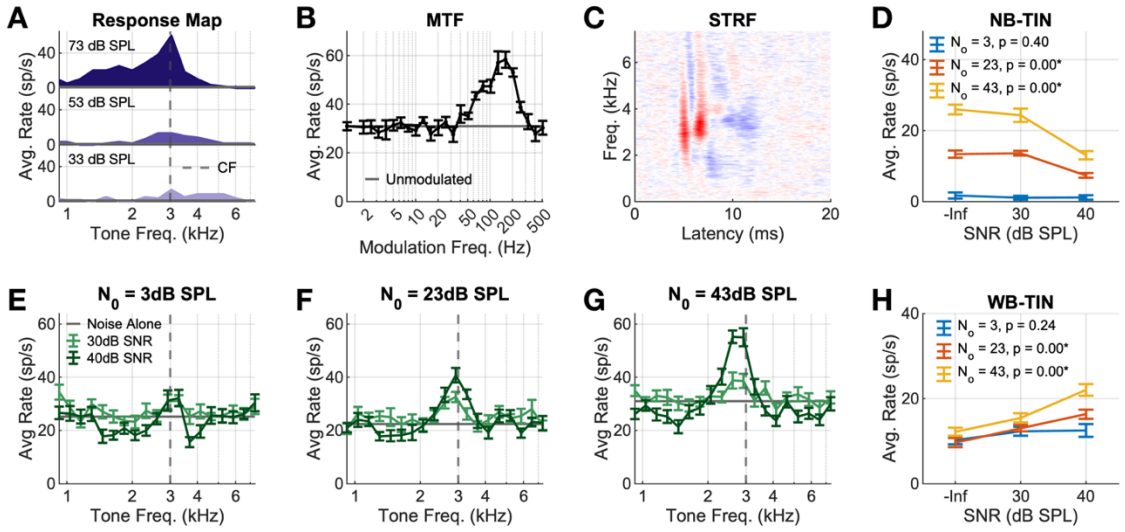

Figure 2.3 Single neuron responses to tones and noise used to characterize the neurons. (A) RM, dashed grey line represents estimated CF. (B) MTF, BE. (C) STRF estimated from responses to wideband noise, excitation in red and inhibition in blue. (D) Response to a NB-TIN centered on CF, for three levels (N_0 = 3, 23, 43 dB SPL, overall level = 31, 51, 71 dB SPL) and 3 SNRs (-inf, 30, 40 dB). (E-G) Response to WB-TIN (N_0 = 3, 23, 43 dB SPL, overall level = 43, 63, 83 dB SPL). Tones were presented at 30 and 40 dB SNR w.r.t. N_0. The grey line is the noise-alone response. (H) Response to WB-TIN with tone at CF, rates estimated from data in (E-G). All error bars represent +/- 1 SEM.

IC neurons were initially characterized by a RM, MTF, and STRF (Fig. 2.3a, b). The noise STRF for the example in Fig. 2.3 featured early and later broad excitation centered on CF and both a lagging off- and on-CF inhibition (Fig. 2.3c). The rate in response to on-CF NB-TIN decreased with increasing SNR for two out of the three

spectrum-level conditions ($p<0.001$ for N0 = 23, 43 dB SPL, $p=0.400$ for N0 = 3 dB SPL, t-test, Fig. 2.3d). This decrease in rate is consistent with a report of BE responses to on-CF tones in narrowband gaussian noise (Fan et al., 2021). In contrast, the response of this neuron to on-CF WB-TIN increased as SNR increased (Fig. 2.3h). Looking over a range of tone frequencies, in comparison to the noise-alone response (Fig. 2.3e-g, solid grey line), rates increased when the tone was near CF, and decreased for tones below or above CF (Fig. 2.3e-g). The excitation and flanking inhibitions were present at all three noise levels tested and excitation increased in magnitude as level increased. The difference between on-CF tones in wideband and narrowband noise was not expected in the neural fluctuation hypothesis, and off-CF rate decreases for WB-TIN were also not predicted.

2.4.2 Responses to on-CF tones in WB and NB noise

To quantify differences in neural responses due to noise bandwidth when a tone was at CF, we compared responses to on-CF WB- and NB-TIN. Due to the previous finding that responses to NB-TIN are dependent on MTF type, neurons were divided into four categories, BE, BS, H_{BE}, and H_{BS}. An on-CF WB-TIN elicited an increase in rate as SNR increased for most neurons, regardless of MTF type, spectrum level, or contralateral or diotic presentation (Fig. 2.4). Rates for diotic presentations of the on-CF WB-TIN increased with SNR (n=57/84 N_0=3, n=67/86 N_0=23, 68/86 N_0=43 dB SPL, Fig. 2.4a-c). In contralateral presentations, rate increased as SNR increased for the majority of neurons in all conditions (n=25/40 N0=3, n=30/42 N0=23, n=35/42 N0=43 dB SPL, Fig. A2). The contralateral 3-dB-SPL N_0 condition was the only condition for which BE and BS cells

had significantly different rate changes in the WB-TIN response (unpaired t-test, t(19)=-2.9263, p=0.0087, Fig. A2a). Overall, five out of six conditions did not show any effect of MTF type on responses to on-CF WB-TIN.

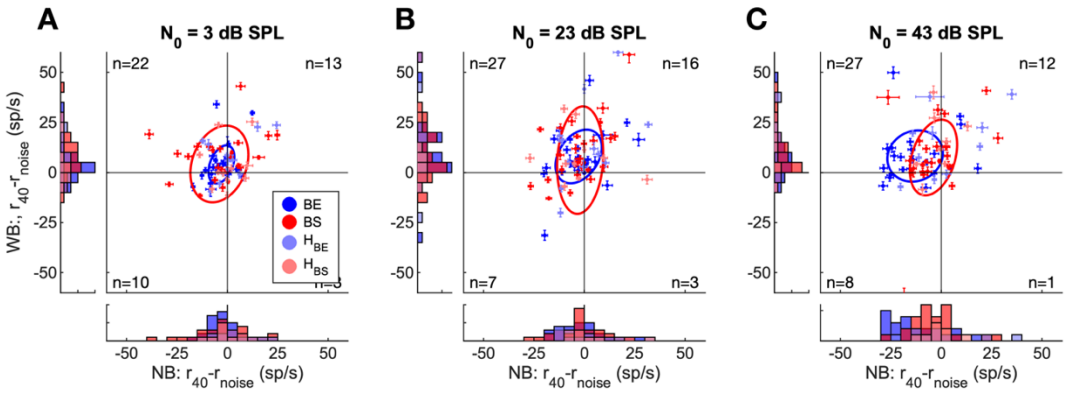

Figure 2.4. Differences between average rates in response to diotic TIN (40 dB SNR) and noise-alone for on-CF tones in wideband vs. narrowband noise. Responses at (A) N_0 = 3 dB SPL, WB-TIN overall level range of 34-48 dB SPL, NB-TIN level range of 28-41 dB SPL, (B) 23 dB SPL, WB-TIN overall level range of 54-69 dB SPL, NB-TIN level range of 48-63 dB SPL, and (C) 43 dB SPL, WB-TIN overall level range of 74-83 dB SPL, NB-TIN level range of 68-83 dB SPL. BE MTFs are blue, BS MTFs are red, H_{BE} are light blue, H_{BS} are pink. Error bars represent +/- 1 SEM. Ellipses represent the standard deviation for BE (blue) and BS (red) neurons. Number of neurons in each quadrant is reported in the corner of each quadrant.

For the NB-TIN condition, responses of BE and BS neurons differed for the contralateral 23-dB-SPL N_0 condition (unpaired t-test, t(25)=-3.1598, p=0.0041, Fig. A2b), and diotic 43-dB-SPL N_0 condition (unpaired t-test, t(46)=-2.4818, p=0.0168, Fig. 2.4c). In both conditions, rates of BE cells decreased when the tone was added. BS cells in the contralateral 23-dB-SPL N_0 condition had increased rates when the tone was added, and BS cells in the diotic 43-dB-SPL N_0 condition had rates that either increased

or decreased when the tone was added. The BE and BS responses were similar in the other four conditions, with distributions centered around 0, indicating that rate increases and decreases occurred in both BE and BS neurons. These NB-TIN on-CF results were only partially consistent with Fan et al. (2021), which reported MTF-type-dependent responses for many levels and SNRs. Potential explanations for the differences between studies are discussed below.

2.4.3 Characterizing responses to frequency-shifted WB-TIN.

The previous analysis found that most neurons responded to WB-TIN with an increase in rate at CF, and our next step was to determine how neurons responded to the rest of the WB-TIN stimulus, including for different sound levels and SNRs. A linear mixed-effects model was implemented to assess how stimulus parameters influence responses to WB-TIN. The mixed model revealed a significant three-way interaction between SNR, N_0, and tone frequency ($F_{(32, 20580)}$ =3.74, p=4.87e-12). SNR increases led to stronger rate increases near CF (+/- 0.25 octaves) and stronger rate decreases for tones away from CF (-0.5 to -1 octaves and above 0.5 octaves) (Fig. 2.5a).

Changes in noise level also affected WB-TIN responses. In the 40-dB-SNR condition, increasing N_0 broadened the excitatory peak and increased rate for tone frequencies ranging from -0.75 to +0.5 octaves relative to CF (Fig. 2.5a, right). In the 30-dB-SNR condition, when N_0 increased, rates increased ¼ octave below CF and decreased ¼ octave above CF (Fig. 2.5a, middle). No changes in rate vs. tone frequency over level were observed for the 20-dB-SNR condition, which is near threshold for TIN detection in rabbits (Zheng et al., 2002) (Fig. 2.5a, left).

The mixed-effects model also identified a significant three-way interaction between SNR, N_0, and contralateral/diotic presentation ($F(4, 20596)=7.77$, $p =2.95e-06$) and a significant two-way interaction between tone frequency and contralateral/diotic presentation in the mixed-effects model ($F_{(8, 20580)}=2.15$, $p = 0.027$); however, there were only significant differences for the tone one octave above CF. Generally, contralateral and diotic presentations of the WB-TIN stimuli resulted in similar responses (Fig. 2.5, Fig. A3). Overall, responses had systematic changes based on SNR and N_0, and only the highest SNR condition revealed rate decreases below the noise-alone rate.

We wondered whether neural characteristics, such as MTF type, CF group, or RM type, changed the responses to WB-TIN. A second linear mixed-effects model was used to analyze the impact of neural characteristics. The mixed model revealed significant three-way interactions between tone frequency, RM type and CF group ($F(16,16586)=3.20$, $p=1.45e-05$, Fig. 2.5d, e). Low-CF I-type neurons had stronger rate decreases (-1, -0.75, 0.25 octaves w.r.t. CF) and stronger, broader rate increases (-0.25 and 0 oct) compared to low-CF V-type neurons (Fig. 2.5b, c, blue). At medium CFs, I-type rate increases were stronger than in V-type (0, 0.25, 0.5 oct, Fig. 2.5b, c, red). For high-CF neurons, V-type neurons had smaller rate decreases at 1 octave above CF (Fig. 2.5b, c, green). Medium-CF I-type neurons had greater rate increases (0, 0.25 oct), and weaker rate decreases (0.5) than low and high CF neurons (Fig. 2.5b). For V-type neurons, low CFs had broader rate increases (0.25 oct) and high CFs often had weaker or no rate decreases (-1, -0.75, 0.5 oct, Fig. 2.5c). A significant three-way interaction was also found between RM type, MTF type, and tone frequency ($F(32,16586)=3.82$,

p=1.96e-12, Fig. A4). There were stronger rate increases and decreases for I-types in BE, flat, and H_{BE} neurons (Fig. A4).

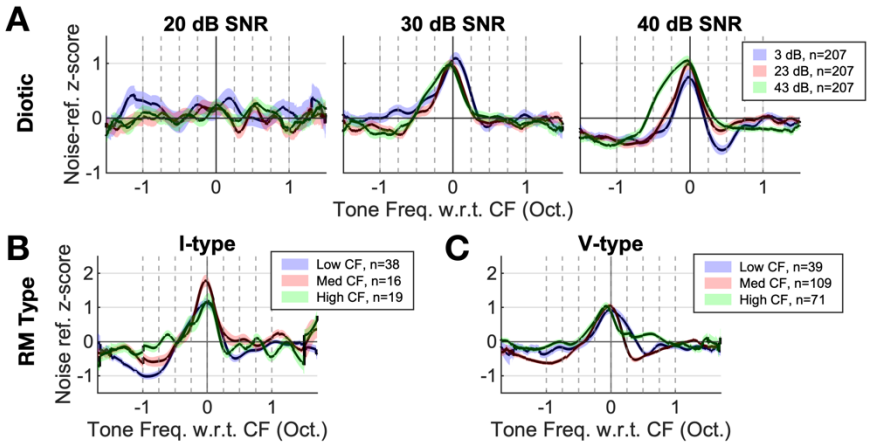

Figure 2.5. Population average responses to WB-TIN stimuli. (A) Diotic condition 20-dB-SNR (left), 30-dB SNR (middle), 40-dB SNR (right), for N0 = 3-dB-SPL (blue), N0 = 23-dB-SPL (pink), N0 = 43-dB-SPL (green), n=207. All data was normalized by noise-referenced z-score. Dotted grey lines represent the tone frequencies tested in the linear mixed model. Colored bands indicate +/-1 SEM for each population average. (D-E). Population average responses to WB-TIN separated by neural characteristics. Responses averaged over contra/diotic presentation and spectrum-level presentation (3, 23, 43 dB SPL) and normalized using a noise-referenced z-score. (B) I-type and (C) V-type responses for low (blue), medium (red), and high (green) CF groups.

Overall, RM type and CF group significantly impacted responses, and I-type neurons had stronger rate increases and decreases that varied with MTF type and CF group. Comparing CF groups, low- and mid-CF neurons had strong rate decreases off CF, but high CFs had weaker rate decreases. MTF type marginally impacted responses, suggesting that changes in the amplitude of NFs in the peripheral responses upon addition of the tone to WB noise had a relatively weak effect on these responses.

Predicting WB-TIN responses using RMs and STRFs

We questioned to what extent the response to pure tones in quiet (RMs) could explain the WB-TIN data. If the response to WB-TIN was driven primarily by the tone, these metrics may be well correlated. To quantify similarity between RM and WB-TIN responses, we calculated the variance explained, R^2, between a neuron's WB-TIN response and the neuron's RM at the sound level most closely matched to the WB-TIN stimulus. Interestingly, most V-type neurons had WB-TIN responses that were poorly explained by their RMs (43-dB-SPL N_0, R^2 mean = 0.24, R^2 median = 0.17, n=121, Fig. 2.6a, Fig. A5a, c). I-type neurons had WB-TIN responses that were overall better explained by RMs, though there was a lot of variation in predictable variance (43-dB-SPL N_0, R^2 mean = 0.36, R^2 median = 0.34, n=45, Fig. 2.6a, Fig. A5b, c). These results are consistent with previous reports that I-type responses to tones are robust when noise is added (Ramachandran et al., 2000). Most onset neurons RMs were not able to explain WB-TIN responses (43-dB-SPL N_0, R^2 mean = 0.16, R^2 median = 0.08, n=20, Fig. A5c). Overall, the I-type RMs were significantly more correlated to the WB-TIN response than the other two types for all sound levels (ANOVA, N_0 = 3 dB SPL, $F(2, 124)= 8.183$, $p=0.00046$; N_0 = 23 dB SPL, $F(2, 174)=17.836$, $p=8.9e-8$; N_0 = 43 dB SPL, $F(2, 181)=9.202$, $p=0.00016$). However, variance explained by the RM alone was generally poor.

Another approach to characterizing neurons is to use an STRF, which is a linear mapping of firing rate to stimulus features, used to describe spectral and temporal tuning properties of IC neurons (Escabí and Schreiner, 2002; Andoni et al., 2007). Many IC

STRFs reveal excitatory and inhibitory regions at varied frequencies and delays, which could be related to the WB-TIN responses. Here we compare WB-TIN responses with a prediction using a noise STRF because both stimuli use gaussian noise. We tested whether STRFs could predict the responses to WB-TIN for the 23-dB-SPL N_0 condition. STRF model predictions of WB-TIN responses were compared to the WB-TIN data using the R^2 goodness-of-fit metric. Results varied; most WB-TIN responses were not well predicted by the STRF, but 25/145 neurons had a $R^2 > 0.5$. One such neuron had an STRF with excitation near CF and delayed on-CF and off-CF inhibition (Fig. 2.6b). The STRF model predicted the neural WB-TIN response well, with the largest errors in the size and frequency range of the inhibition ($R^2 = 0.66$, Fig. 2.6b). A second example neuron that was not well predicted had an STRF with a broad excitation and broad, delayed inhibition (Fig. 2.6c). The STRF prediction did not accurately reflect the inhibition seen in WB-TIN responses (R^2=0.20 Fig. 2.6c). There was no correlation between the RM R^2 and the STRF-model R^2 for the 23-dB-SPL N_0 conditions and no correction for specific RM types (Fig. 2.6d). Overall, STRF and RMs were not sufficient to explain the WB-TIN responses, but STRFs provided the best predictions for neurons that had STRFs with inhibitory sideband regions.

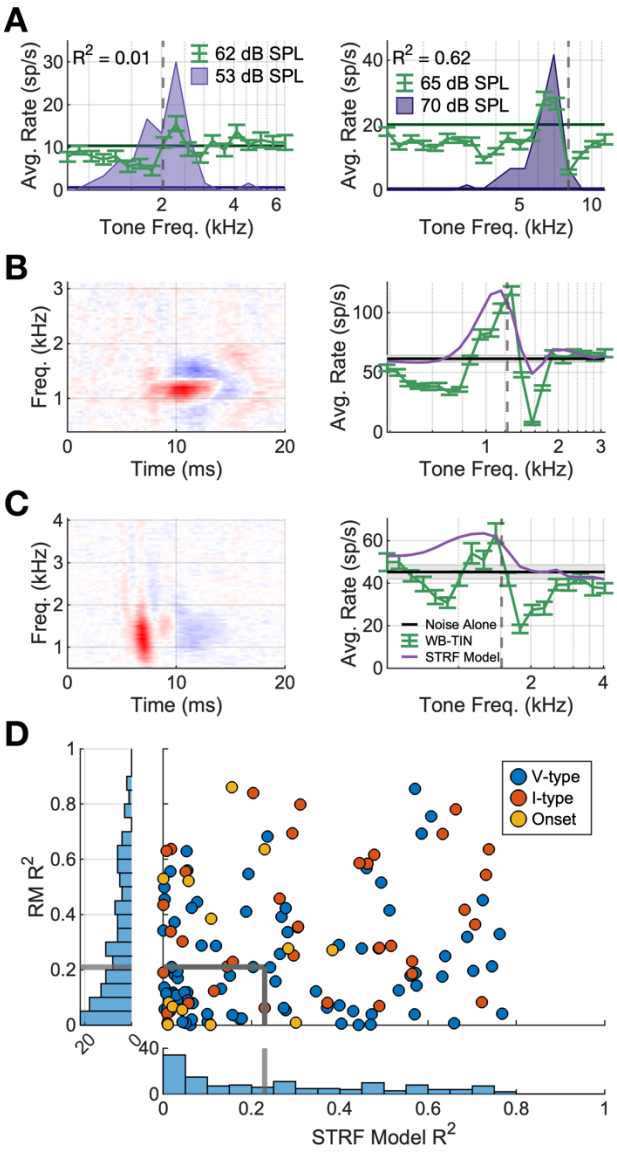

Figure 2.6. Example STRF predictions of WB-TIN and summary of predictions based on RM and STRF. (A) Response map (purple) with WB-TIN response (green) overlayed for two example neurons, one V-type (left) and one I-type (right). (B) Example neuron STRF (left) and STRF predictions of WB-TIN (right, purple) compared to actual WB-TIN response (right, green), R2 = 0.66. (C) Different example neuron for same analysis as (B), R2 = 0.20. (D) Scatter plot of RM R2 and STRF R2 values. Grey lines represent the median R2 values.

Difference of gaussians (DoG) fit to WB-TIN data.

Noting that the STRF predictions with inhibitory sidebands tended to have more accurate WB-TIN predictions, we next asked how broad inhibition factored into these WB-TIN responses. A DoG function was fit to each average-rate WB-TIN profile to quantify patterns in response profiles across neurons for the diotic, N_0= 23 dB SPL, 40-dB-SNR condition (Fig 7). An example fit including the excitatory and inhibitory components is shown in Fig. 2.7a. The DoG fit well (variance explained, $R^2 > 0.5$) to the majority of neurons that had high predictable variances (Fig. 2.7b, blue). The ratio of inhibition to excitation amplitude was generally below one (mean = 0.92, median = 0.81), meaning that inhibition was weaker than excitation in the model fit for most neurons (Fig. 2.7c). The DoG equation had parameters for both the center frequency of the excitatory gaussian (f_{exc}) and the inhibitory gaussian (f_{inh}) to allow fits to asymmetrical off-CF inhibition. The inhibitory and excitatory center frequencies were similar to the CF of the neuron but were often lower than CF (Fig. 2.7d). The difference between inhibitory and excitatory center frequencies had a mean of 0.197 octaves with respect to CF, indicating on average the inhibitory CF skewed to lower frequencies compared to the excitatory CF and created stronger lower inhibitory sidebands. The bandwidths of excitation (σ_{exc}) and inhibition (σ_{inh}) were plotted as a function of CF and fit using a log-log transformed linear model (Fig. 2.7e). Bandwidths for both excitation (Fig. 2.7e, top) and inhibition (Fig. 2.7e, middle) increased as a function of CF ($p<0.001$ for all excitatory conditions, $p=0.002$ for inhibition, 23-dB-SPL N_0, Fig. 2.7e). The slope of the σ_{exc} fit line was 0.78 and the slope of the σ_{inh} was 0.48; inhibition widens slower than

excitation as CF increases. The ratio of CF to excitatory bandwidth is near 1:1, but for inhibitory bandwidth the inhibition gets narrower relative to CF. For many units, inhibitory bandwidths were wider than excitatory bandwidths (n=129/192, Fig. 2.7e, bottom). The same trends were seen in N_0= 3 dB SPL and N_0= 43 dB SPL, 40-db-SNR condition (Fig. A6). Scatter plot shows each model fit as a function of log bandwidth vs. log strength ratio, with example profiles for each quadrant (Fig. 2.7f). Model fits in quadrant one have negative rate profiles due to strong inhibition strength and broad inhibition (n=22/192). Fits in quadrant two can have two profiles: strong, narrow inhibition can fall outside of the excitatory gaussian and thus represent an excitation with a single sideband of inhibition (n=17/192) or can represent a positive rate profile with a dip near CF (n=22/192). Quadrant three, representing weak inhibition and narrow inhibition, can have a profile with a single inhibitory sideband (n=24/192). Lastly, quadrant four, strong excitation with broad inhibition, represents the "classic" difference of gaussians fit with excitation near CF and broad, weaker inhibition off CF. Quadrant four contains 55.7% of the fits and represents the majority of the data (n=107/192). Points near (0,0) represent profiles that were generally flat, where inhibition and excitation were nearly equal in strength and bandwidth. Overall, the majority of neural responses to WB-TIN were successfully fit using a difference of gaussians and the majority of these neurons had broad, weak inhibition, but there was heterogeneity in the data and other rate profiles were seen as well.

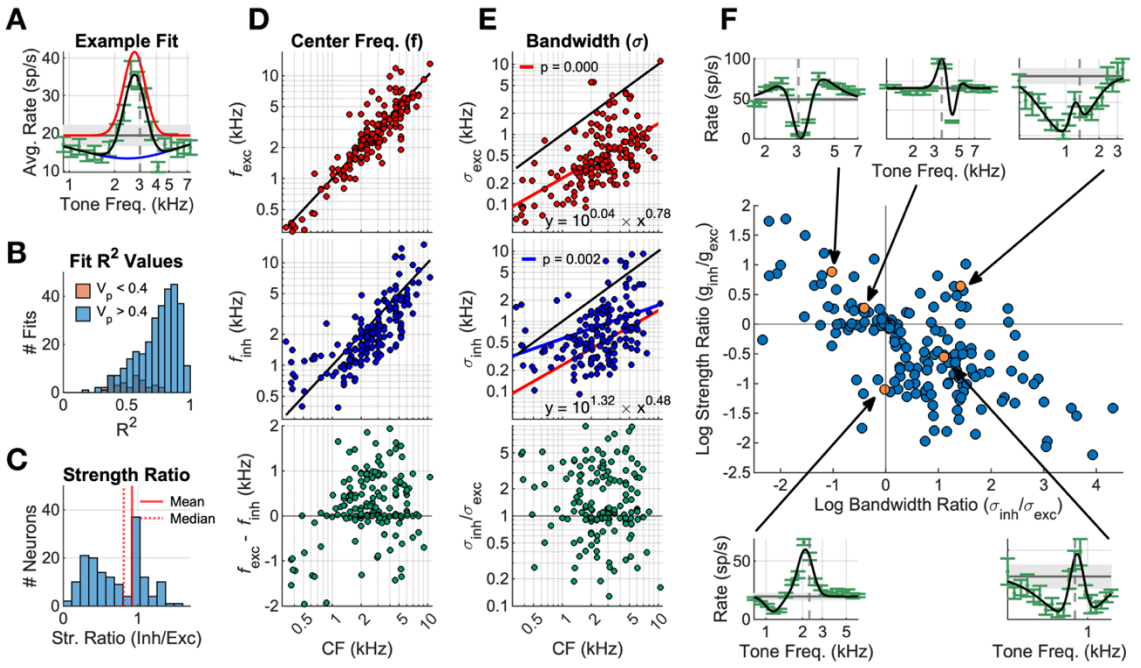

Figure 2.7. Difference of gaussians analysis. (A) DoG fit to an example neuron 40-dB SNR WB-TIN response. Red curve is the excitatory gaussian, blue curve is the inhibitory gaussian. Green bars represent the data and +/-1 SEM. Black curve is the DoG. Horizontal black line indicates the noise-alone rate, and grey shaded area is +/- 1 SEM. (B) Histogram of R^2 values fit to all neurons in the diotic, 23 dB SPL, 40-dB SNR condition. Orange indicates neurons with a predictable variance less than 0.4. (C) Strength ratio (inh/exc) for all fit neurons, including mean (red solid line) and median (red dotted line). (D) Center frequencies for excitatory (red, top) and inhibitory (blue, middle) gaussians as a function of CF. Difference between excitatory and inhibitory center frequencies in green, bottom row. Unity line in black. (E) Bandwidth, σ, for excitatory (red, top) and inhibitory (blue, middle) gaussians. σ_{exc} fit lines in red (top, middle), σ_{inh} fit line in blue (middle). Unity line in black. Bottom row, green, indicates the ratio of inhibitory to excitatory bandwidths. (F) Scatter of the strength ratio vs bandwidth ratio with example neuron fit curves shown for each quadrant. Orange points indicate examples.

Adding off-CF pathways to IC model improved model accuracy.

Lastly, we added off-CF inhibitory pathways to an existing IC SFIE model with an on-CF excitation and a delayed inhibition (Nelson and Carney, 2004; Carney and McDonough, 2019). The goal was to implement the broad inhibition seen in the DoG fits to predict WB-TIN responses in a phenomenological model of the IC, while also retaining amplitude-modulation sensitivity, a key feature of IC neurons. Six model configurations with off-CF inhibitory inputs from different sources (brainstem, BS IC cell, BE IC cell) targeting on-CF IC neurons (BS or BE) were tested to search for model configurations that could replicate physiological results (Fig. 2.2).

Two exemplar neurons were used to fit the model configurations, one BE cell and one BS cell (Fig. 2.8a, Fig 2.9a). Although the WB-TIN results were not affected by amplitude-modulation sensitivity, on-CF NB-TIN results depend on MTF type. Thus, we investigated whether these model configurations could retain accurate MTFs and NB-TIN responses while also fitting the WB-TIN rate profiles.

The model without added inhibition was tested as a baseline. The SFIE BE model had decreased rate in response to on-CF NB-TIN and predicted a dip in rate at CF for the WB-TIN stimulus, which poorly predicted the neural data (r_{MTF} =0.78; r_{NB} = 0.86; r_{WBTIN}=0.03, Fig. 2.8c). The energy model was also tested as a baseline, and as expected, predicted peaks in rate when the tone frequency matched CF but did not predict the rate decreases below and above CF (r_{MTF} = -0.34; r_{NB} = -1.00; r_{WBTIN}=0.56, Fig. 2.8d). All six model configurations were fit to this BE cell; the BS-inhibited-by-BS model configuration (Fig. 2.2c) was able to replicate the MTF, on-CF NB-TIN, and WB-TIN

results and MTF type with the best accuracy (r_{MTF} = 0.81; r_{NB} = 0.99; r_{WBTIN}=0.85, Fig. 2.8b).

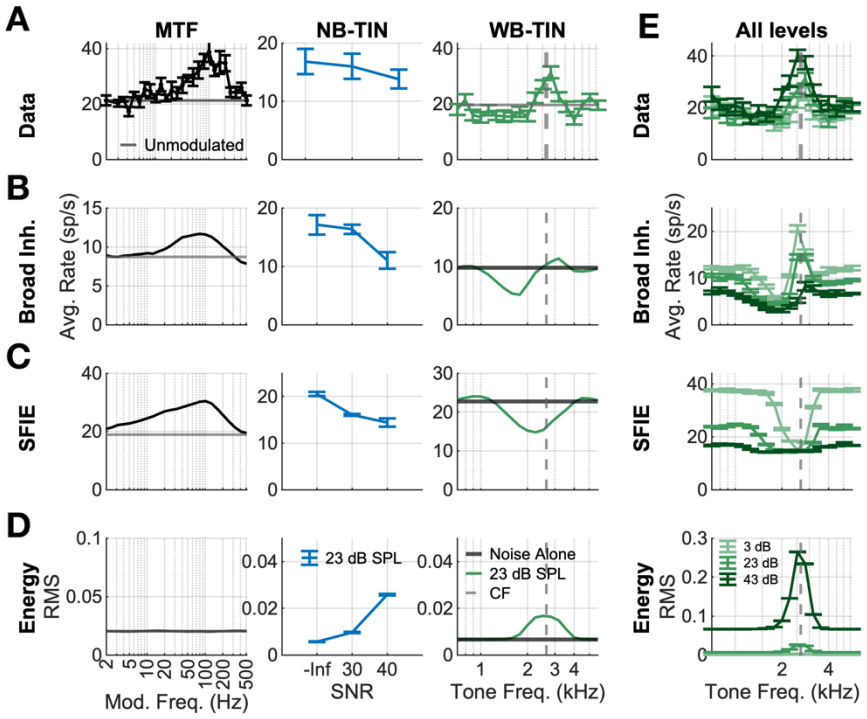

Figure 2.8. Comparison of model responses to data for three models: broad inhibition, SFIE, and energy. (A) Example BE neuron MTF (left), NB-TIN response (middle), and WB-TIN response for the N_0=23 dB SPL, SNR = 40 dB condition (overall level = 50, 62 dB SPL respectively). These three datasets were used for fitting the broad inhibition model. (B) Broad inhibition model responses to MTF, NB-TIN, WB-TIN. S_{low} = 1, S_{high}= 0.28, D_{off} = 1 ms, off-CF range = 0.625 octaves, BMFs = [87 164 111] for [low-CF on-CF high-CF]. (C) SFIE model responses. (D) Energy model responses. (E) Data, broad inhibition, SFIE, and energy model responses for three levels of WB-TIN (3, 23, 43 dB SPL, 40 dB SNR). Error bars represent +/-1 SEM.

All models tested had limitations when predicting responses over a range of sound levels. For the energy model (Fig. 2.8e, bottom), the increase in rate when the tone was off-CF was too large. The response of the SFIE model (Fig. 2.8e) to noise-alone or to a

noise with a tone far from CF decreased as level increased, in stark contrast to the neural data, due to the model sensitivity to neural fluctuations in the AN model responses. As sound level increased, even for a noise, the fluctuations in the AN responses decrease due to IHC saturation, and the SFIE BE cell responded to a lack of fluctuations with a decrease in rate. The response to a tone at CF was also not correct over a wide range of sound levels for the SFIE and broad inhibition models. The tone captured the AN response and decreased fluctuations, thus decreasing the BE SFIE rate, but reached a 'floor' when the AN response was fully captured. The broad inhibition model decreased in rate as level increased near CF, due to the broadening of inhibitory pathways as sound level increased. The broadening inhibition decreased excitation at CF.

The exemplar BS response was fit using a model BS neuron that was inhibited by two off-CF BS neurons (Fig. 2.2c, Fig. 2.9). Again, the SFIE BS cell without lateral inhibition was a baseline. The SFIE BS model increased in rate with increasing SNR for the on-CF NB-TIN stimulus, and there was a peak in rate at CF for the WB-TIN response. This model accurately predicted the on-CF rate increases, but not the off-CF rate decreases seen in data (r_{MTF} = 0.39; r_{NB} = 0.94; r_{WBTIN}=0.39, Fig. 2.9c). The SFIE BS WB-TIN response featured a small amount of inhibition on the high-frequency side due to suppression of the AN responses. However, that inhibition was not strong enough to accurately model the physiological IC responses. With the additional off-CF BS inhibitory inputs to the on-CF BS cell, inhibition was stronger than with the on-CF pathways alone, and this model had the greatest accuracy (r_{MTF} = 0.53; r_{NB} = 0.93; r_{WBTIN}=0.52, Fig. 2.9b). Similar to the BE SFIE and broad-inhibition models, these BS

models also did not accurately predict changes in level (Fig. 2.9d). These inaccuracies were created again by the SFIE model's response to decreases in fluctuations to noise as level increased, causing an increase in rate for noise-alone and off-CF tone responses in the BS models not seen in physiology.

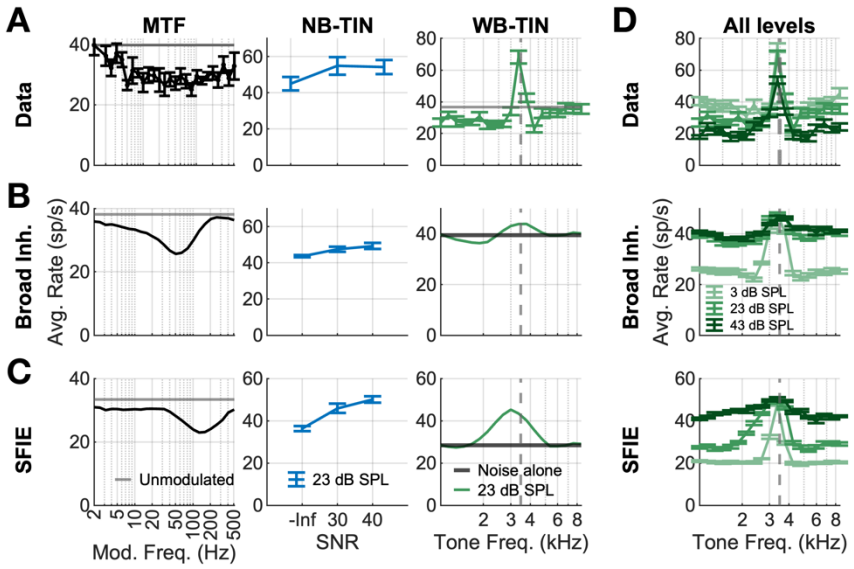

Figure 2.9. Comparison of model responses to data for three models: broad inhibition, SFIE, and energy. (A) Example BS neuron MTF (left), NB-TIN response (middle), and WB-TIN response for the N_0=23 dB SPL, SNR = 40 dB condition (overall level = 50, 62 dB SPL respectively). These three datasets were used for fitting the broad inhibition model. (B) Broad inhibition model responses to MTF, NB-TIN, WB-TIN. S_{low} = 0.19, S_{high}= 0, D_{off} = 1 ms, off-CF range = 0.5 octaves, BMFs = [300 38 10] for [low-CF on-CF high-CF]. octaves (C) SFIE model responses. (D) Data, broad inhibition, SFIE, and energy model responses for three levels of WB-TIN (N_0=3, 23, 43 dB SPL, 40 dB SNR). Error bars represent +/-1 SEM.

The other configurations tested that did not accurately predict WB-TIN were the three models with an on-CF BE target cell and the BS-inhibited-by-off-CF-BE model (not shown). The model configurations that feature an on-CF BE target cell always

decreased in rate near CF due to the underlying on-CF BE cell, which responded to WB-TIN with a dip near CF. Adding off-CF inhibition to the BE-model configuration caused a larger decrease in rate. As the strength of inhibition was increased, the rates further decreased to zero. The BS-inhibited-by-off-CF-BE model configuration also did not replicate inhibition off CF in the IC responses in any combination of model parameters tested. The off-CF BE cell responded with a decrease in rate when the tone was near the CF, which lead to the on-CF BS cell being inhibited by a weak off-CF response. This model configuration resulted in a lack of inhibition off CF in the WB-TIN response. Lastly, the BS-inhibited-by-off-CF-CN model did not predict the off-CF inhibition observed in the IC responses.

An interesting aspect of the broad-inhibition models is the way BE and BS cells were created from only a combination of BS cells (Fig. 2.10), largely due to BMF and inhibitory strength parameters. To illustrate these MTFs, the on- and off-CF BS cell responses were plotted individually (Fig. 2.10, top row). The on-CF BS cell for the BE model fit has a shallow BS MTF, and the off-CF MTFs have more salient BS MTFs (Fig. 2.10b, top). When combined, the on-CF response is strongly inhibited at low and high modulation frequencies, and less inhibited at the WMF of the off-CF response (approx. 100 Hz), resulting in a broad-inhibition model with a BE MTF profile (Fig 10b, bottom). The broad-inhibition BS cell had a stronger BS on-CF MTF that was weakly inhibited by the off-CF BS cells, which were flatter due to changes in WMF parameters (Fig. 2.10c).

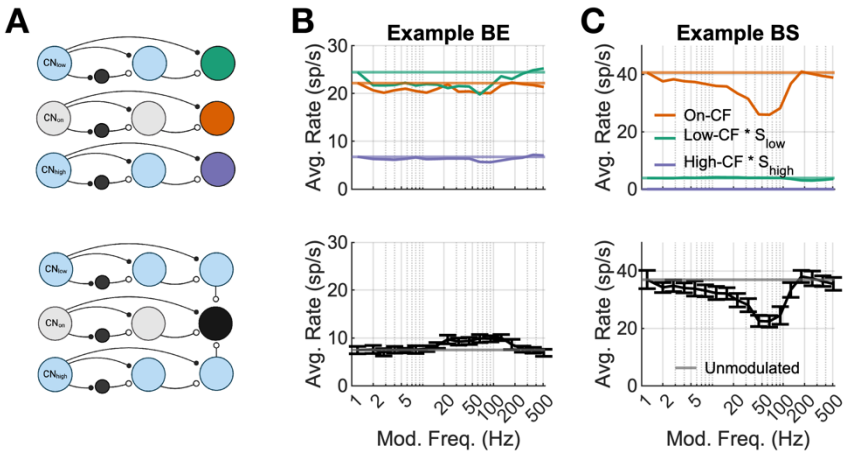

Figure 2.10. Model MTF responses for intermediate components of the broad inhibition model. (A) Model schematics, top row has on-CF (orange), low-CF (green), and high-CF (purple) cells highlighted. Bottom row shows full broad inhibition model cell (target cell in black). (B) Example BE broad inhibition model response, top row has on-CF, low-CF, and high-CF MTFs plotted separately with the low- and high-CF responses scaled by inhibition strength, same parameters as Fig 8. Bottom row is the aggregate broad inhibition response. (C) Same plots for an example BS broad inhibition cell, for Fig 9 model cell.

We further investigated how changing the parameters in the model would influence model responses by varying the strengths, delays, on- and off-CF parameters. Varying the strength (S_{off}), or weight of the inhibitory inputs, resulted in changing MTF types and changing overall rate responses to WB-TIN (Fig. 2.11a). At $S_{off} = 0.1$ the MTF is BS, at 0.3 the MTF is hybrid, and at 0.5 the MTF is BE (Fig. 2.11a). As S_{off} increases, the overall rate response to WB-TIN decreased and inhibitory regions became more salient (Fig. 2.11a). Next, we changed the CF of the off-CF inputs and found that, as expected, the further away in frequency the off-CF inputs were, the broader the inhibitory regions of the response were (Fig. 2.11b). The MTF was not largely impacted by

changing the CF-range parameter (Fig. 2.11b). We also tested off-CF delays, D_{off}, of 0, 1, and 2 ms and found this parameter had a slight impact on the MTF of the neuron but had no effect on the WB-TIN response (not shown).

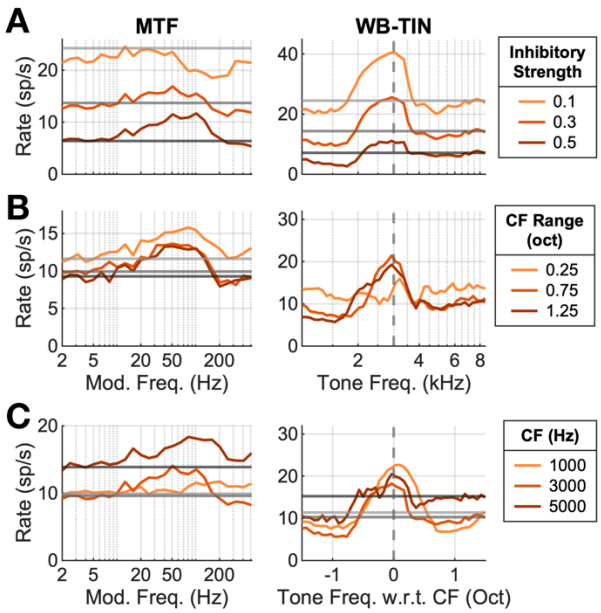

Figure 2.11. Varying lateral-inhibition model (see Fig. 2.4H) parameters changes model responses. (A) Lateral SFIE model response for a 3 kHz model neuron with S_{off} parameter set to 0.1 (light orange), 0.3 (orange), or 0.5 (brown). D_{off} = 0 ms, off-CF range = 1 octave. Model responses to MTF and WB-TIN with N_0 = 23 dB SPL. Noise-alone responses for WB-TIN shown in light grey (0.1), grey (0.3) and dark grey (0.5). (B) Responses of a 3-kHz model neuron with off-CF range parameter set to 0.25 (light orange), 0.75 (orange), or 1.25 (brown) octaves away from CF. S_{off} = 0.4, D_{off} = 0 ms. Noise-alone response not shown for clarity. (C) Responses of a 1-kHz (light orange), 3-kHz (orange), and 5-kHz (brown) model neurons with S_{off} = 0.4, D_{off} = 0 ms, off-CF range = 1 octave. All models have BMF = 100 Hz.

Lastly, the CF of the on-CF model neuron was varied to compare WB-TIN and MTF responses across a range of CFs (Fig. 2.11c). The 1000-Hz model neuron with S_{off} =

0.4, D_{off} = 0 ms, and off-CF range = 1 octave created a mostly flat MTF with a WB-TIN response with excitation and flanking inhibition (Fig. 2.11c). The 3-kHz model neuron, which was considered a 'mid' CF range, with the same model parameters, resulted in a hybrid MTF and WB-TIN responses that were consistent with physiological responses. The high-CF 5-kHz model neuron had a BE MTF, and the WB-TIN responses did not have as clear excitation and inhibition (Fig. 2.11c). Overall, adding an off-CF input from local BS model cells to the on-CF BS cell (Fig. 2.4c) created responses that retained MTF sensitivity by replicating BE, BS, hybrid, and flat MTF types and accurately predicted WB-TIN responses seen in physiological recordings.

2.5 Discussion

Our study revealed that average rates in response to WB-TIN increased for tones near CF and decreased for tones away from CF, compared to responses to noise-alone stimuli. This pattern differed from responses to narrowband noise, suggesting that mechanisms in addition to NF sensitivity impact TIN encoding. This pattern also often differed from what would be expected based on neurons' tone RMs and noise STRFs. The responses to WB-TIN featured off-CF rate decreases that occurred for a range of sound levels and at suprathreshold SNRs, supporting the hypothesis that broad inhibition shaped responses. We developed a model that incorporated on-CF excitation and inhibition with two off-CF inhibitory inputs from local BS model cells to approximate broad inhibition. This model accurately predicted WB-TIN responses while maintaining

MTF sensitivity. Overall, these results highlight broad inhibition as a mechanism influencing complex-sound encoding in the IC.

Our approach of shifting the tone frequency with respect to CF can also be used to infer the population response of IC neurons. In this inferred population, the neurons tuned near the tone frequency are excited and the neurons tuned away from the tone frequency are inhibited with respect to their noise-alone response. This combination of excitation and inhibition would create a rate profile across a population of neurons with enhanced representation of a spectral peak in noise.

Comparisons to TIN IC studies

Our findings are consistent with studies that observed rates that increased with SNR for an on-CF tone in wideband noise (Rees and Palmer, 1988; Jiang et al., 1997; Rocchi and Ramachandran, 2018). For on-CF NB-TIN, BE IC neural responses were largely consistent with Fan et al. (2021), showing decreased rates with increasing SNR. However, BS cell responses here were more variable and did not consistently show increased rates with increasing SNR (Fig. 2.4). A possible explanation could be methodological: we used diotic MTFs instead of contralateral MTFs and did not rely on identification of CF during the experiment because we sampled many tone frequencies. We observed that IC responses to NB-TIN often changed from a rate increase to a rate decrease at different tone frequencies near CF, highlighting the impact of the choice of CF value to test (not shown).

IC responses to off-CF tones have not been widely studied, though one group found that responses to off-CF tones helped to explain behavioral TIN-detection

thresholds in budgerigar (Wang et al., 2021). That finding, combined with our results that off-CF rates contain information in the form of inhibition, supports the idea that off-CF information could be important in TIN detection.

We originally hypothesized that ipsilateral inhibition in the binaural response may strengthen off-CF inhibitory bands in responses, but overall did not see changes in inhibitory sidebands due to presentation ear in our population. However, differences in diotic and contralateral responses on an individual neuron level could be studied in tandem with interaural response properties in the future.

Broad Inhibition

Broad inhibition may arise from various sources, including GABAergic projections from the contralateral and ipsilateral dorsal nucleus of the lateral lemniscus (DNLL) (Adams and Mugnaini, 1984), glycinergic input from the ipsilateral lateral superior olive and ventral nucleus of the lateral lemniscus (Saint-Marie et al., 1989; Saint-Marie and Baker, 1990; Winer et al., 1995), or local or commissural connections within the IC itself (Saldaña and Merchań, 1992). There is also evidence of lateral inhibition or suppression in earlier auditory areas such as ventral and dorsal cochlear nucleus (Martin and Dickson, 1983; Rhode and Greenberg, 1994).

Lateral inhibition is also postulated to enhance the rate representation of spectral information at high sound levels (Shamma, 1985). Our physiological results showed that inhibition was more pronounced at high SNRs (40 dB SNR w.r.t. No) than low SNRs (30 dB SNR), consistent with reports that STRFs at higher sound levels have increased areas of inhibition (Lesica and Grothe, 2008). These results suggests that strength of inhibition

may be level or SNR dependent, which could contribute to sound processing in the IC at high levels. Future physiological work understanding the origin and function of broad inhibition in the subcortical auditory system is necessary to understand the impact of broad inhibition.

Modeling

Our broad-inhibition computational model is phenomenological and thus replicates IC responses, not physiological structure. However, we do think the model is physiologically possible. Firstly, in our model fitting process we found that local inhibition of BS cells was the only configuration that could replicate IC responses, consistent with physiological results that found blocking local IC inhibitory receptors reduced off-CF inhibition in STRFs (Andoni et al., 2007).

Secondly, the model relies on two subpopulations of BE cells, because the on-CF BS cell was created from an on-CF BE cell, which is a limitation currently (Fig 2c). However, we hypothesize that the initial BE stage may represent neurons in other regions, such as the DNLL. The DNLL receives input from the ventral CN and provides strong inhibitory input to IC cells. Additionally, some neurons in the DNLL have band-pass rate MTFs (Yang and Pollak, 1997). Therefore, the first BE cell stage can represent the DNLL.

Lastly, we found that as the WB-TIN level was increased, the overall excitation decreased in model BE cells, contrary to physiological responses. We explored adding efferent feedback (Farhadi et al., 2023), but this model, which was designed based on narrowband stimuli, did not improve the predictions of our wideband responses. The

incorrect responses over level were due to saturated on- and off-CF HSR AN fibers that provided the ascending inputs to the model IC neuron. A new IC model is needed to overcome this issue, possibly by including wideband efferent feedback pathways.

Alternative Explanations

Suppression and efferent feedback are alternative mechanisms that could appear as off-CF inhibition in WB-TIN responses. Suppression occurs when a tone outside of the tuning curve of an AN fiber decreases the rate of the fiber in response to a tone at CF (Abbas and Sachs, 1976; Delgutte, 1990). In response to WB-TIN, AN suppression could result in a rate profile in the IC that appears to have inhibitory sidebands. However, the AN model used replicates suppression and the BS SFIE model response did not have decreased rates for tones below CF and had smaller decreases above CF as compared to physiological data (Fig. 2.9c), suggesting that suppression alone cannot account for the changes in rate in WB-TIN responses.

Subcortical efferent feedback also has the potential to alter responses in the IC. Responses in the IC to AM stimuli increase or decrease over a timescale consistent with the medial olivocochlear (MOC) efferent system, dependent on MTF type, indicative that efferent feedback may change responses in the IC (Farhadi et al., 2023). Many neurons adapted over time in response to WB-TIN, and some neurons increased in rate over duration for specific tone frequencies (not shown), possibly due to efferent effects. Additionally, some MOC neurons project to outer hair cells across many frequency channels (Brown, 2014). This wide-bandwidth efferent feedback, which could impact IC responses to off-CF tones, is a topic for future study.

Visual retinal ganglion cell & DoG models

The excitatory region and inhibitory sidebands seen in the WB-TIN data are reminiscent of retinal ganglion cell (RGC) responses in the visual system. RGCs have center-surround organization, with excitation (or inhibition, for off-center RGCs) when light is within a receptive field, and inhibition (or excitation) when light is presented in the surrounding area (Rodieck, 1965). This mechanism arises from lateral inhibition and enhances edge and contrast detection in a visual scene (Boycott and Wässle, 1999). Our IC results show that a tone is presented at a frequency near CF excited neurons, but tones away from CF neuron inhibited or suppressed responses. We hypothesize that the excitation-inhibition pattern in IC cell responses to TIN serves to increase contrast, making the tone response a more salient feature in the rate code.

DoG models, previously used for RGCs (Rodieck and Stone, 1965; Enrorh-Cugell and Robson, 1984) were used to fit WB-TIN responses. Additionally, DoG spectral receptive field models fit to IC rate responses to harmonic complex tones predicted responses more accurately than a gaussian due to the added inhibition (Su and Delgutte, 2020). Our results had similar parameter fit trends to the Su and Delgutte study, with one exception: we observed a strong correlation between inhibitory bandwidth and CF, which they did not report.

Future Experiments / Summary

The comparison in this study between NF sensitivity and broad inhibition raises the following question: at what bandwidth does inhibition become more prevalent than NF sensitivity? Future studies could include a band-widening experiment to test the

bandwidth-dependence of NF-sensitivity or inhibition mechanisms. The shifting tone frequency in a bed of noise is also similar to the two-tone paradigm, which also reveals inhibitory regions (Egorova et al., 2001; Palmer et al., 2013). It would be interesting to know whether a response map created using the two-tone paradigm more accurately reflects WB-TIN responses. Lastly, an STRF created using ripples may result in better predictions than noise STRFs, as these ripple STRFs perform better for higher auditory areas (Andoni et al., 2007; Versnel et al., 2009).

In conclusion, this study revealed distinct response patterns of rate increases to on-CF tones and rate decreases to off-CF tones added to broadband noise in the IC. These results suggest that broad inhibition may have a critical role in complex-sound encoding by shaping rate profiles of sounds in a noisy background. A computational model of the IC including both broad inhibition and AM sensitivity could explain both narrowband and wideband TIN results, highlighting the complexity in encoding different TIN stimuli. These findings could impact our understanding of how natural sounds with broad bandwidths and amplitude modulation, such as speech and music, are represented in the rate profile of IC neurons.

Acknowledgements

This study was funded by NIH-R01-DC010813 and NIH-F31-DC020630-03. Thanks to Paul W. Mitchell and Swapna Agarwalla for assisting in data collection, and to Douglas Schwarz for assistance with hardware and software. Thanks to Kris Abrams for helping with the rabbit surgery and care.

Bibliography

Abbas PJ, Sachs MB (1976) Two-tone suppression in auditory-nerve fibers: Extension of a stimulus-response relationship. J Acoust Soc Am 59:112–122.

Adams JC, Mugnaini E (1984) Dorsal nucleus of the lateral lemniscus: A nucleus of GABAergic projection neurons. Brain Res Bull 13:585–590.

Ahumada A, Lovell J (1971) Stimulus Features in Signal Detection. J Acoust Soc Am 49:1751–1756.

Alkhatib A, Biebel UW, Smolders JWT (2006) Inhibitory and excitatory response areas of neurons in the central nucleus of the inferior colliculus in unanesthetized chinchillas. Exp Brain Res 174:124–143.

Andoni S, Li N, Pollak GD (2007) Spectrotemporal Receptive Fields in the Inferior Colliculus Revealing Selectivity for Spectral Motion in Conspecific Vocalizations. J Neurosci 27:4882–4893.

Boycott B, Wässle H (1999) Parallel Processing in the Mammalian Retina: The Proctor Lecture. Invest Ophthalmol Vis Sci 40:1313–1327.

Brown MC (2014) Single-unit labeling of medial olivocochlear neurons: the cochlear frequency map for efferent axons. J Neurophysiol 111:2177–2186.

Carney LH (2018) Supra-Threshold Hearing and Fluctuation Profiles: Implications for Sensorineural and Hidden Hearing Loss. JARO 19:331–352.

Carney LH (2024) Neural Fluctuation Contrast as a Code for Complex Sounds: The Role and Control of Peripheral Nonlinearities. Hear Res 443:108966.

Carney LH, McDonough JM (2019) Nonlinear auditory models yield new insights into representations of vowels. Atten Percept Psychophys 81:1034–1046.

Carney LH, Yin TC (1988) Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. J Neurophysiol 60:1653–1677.

Caspary DM, Ling L, Turner JG, Hughes LF (2008) Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. J Exp Biol 211:1781–1791.

Casseday JH, Ehrlich D, Covey E (1994) Neural Tuning for Sound Duration: Role of Inhibitory Mechanisms in the Inferior Colliculus. Science 264:847–850.

Dau T, Püschel D, Kohlrausch A (1996) A quantitative model of the ‘“effective”’ signal processing in the auditory system. II. Simulations and measurements. J Acoust Soc Am 99:3623–3631.

Davidson SA, Gilkey RH, Colburn HS, Carney LH (2009) An evaluation of models for diotic and dichotic detection in reproducible noises. J Acoust Soc Am 126:1906.

Delgutte B (1990) Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level. Hear Res 49:225–246.

Egorova M, Ehret G, Vartanian I, Esser K-H (2001) Frequency response areas of neurons in the mouse inferior colliculus. I. Threshold and tuning characteristics. Exp Brain Res 140:145–161.

Enrorh-Cugell C, Robson JG (1984) Functional Characteristics and Diversity of Cat Retinal Ganglion Cells. Invest Ophthalmol Vis Sci 23:250–267.

Escabí MA, Schreiner CE (2002) Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain. J Neurosci 22:4114–4131.

Fan L, Henry KS, Carney LH (2021) Responses to diotic tone-in-noise stimuli in the inferior colliculus: stimulus envelope and neural fluctuation cues. Hear Res 409:108328.

Fan L, Henry KS, Carney LH (2022) Responses to dichotic tone-in-noise stimuli in the inferior colliculus. Front Neurosci 16:997656.

Farhadi A, Jennings SG, Strickland EA, Carney LH (2023) Subcortical auditory model including efferent dynamic gain control with inputs from cochlear nucleus and inferior colliculus. The Journal of the Acoustical Society of America 154:3644–3659.

Fletcher H (1940) Auditory Patterns. Rev Mod Phys 12:47–65.

Jiang D, McAlpine D, Palmer AR (1997) Responses of Neurons in the Inferior Colliculus to Binaural Masking Level Difference Stimuli Measured by Rate-Versus-Level Functions. J Neurophysiol 77:3085–3106.

Joris PX, Schreiner CE, Rees A (2004) Neural Processing of Amplitude-Modulated Sounds. Physiol Rev 84:541–577.

Kidd G, Mason CR, Brantley MA, Owen GA (1989) Roving-level tone-in-noise detection. J Acoust Soc Am 86:1310–1317.

Kim DO, Carney L, Kuwada S (2020) Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. J Neurophysiol 124:1198–1215.

Koch U, Grothe B (1998) GABAergic and Glycinergic Inhibition Sharpens Tuning for Frequency Modulations in the Inferior Colliculus of the Big Brown Bat. J Neurophysiol 80:71–82.

Kohlrausch A, Fassel R, Ptischel D, Gottingen A (1997) Detection of Tones in Low-noise Noise: Further Evidence for the Role of Envelope Fluctuations. Acta Acust 83:659–669.

Krishna BS, Semple MN (2000) Auditory Temporal Processing: Responses to Sinusoidally Amplitude-Modulated Tones in the Inferior Colliculus. J Neurophysiol 84:255–273.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799–1822.

LeBeau FEN, Malmierca MS, Rees A (2001) Iontophoresis In Vivo Demonstrates a Key Role for GABA A and Glycinergic Inhibition in Shaping Frequency Response Areas in the Inferior Colliculus of Guinea Pig. J Neurosci 21:7303–7312.

Lesica NA, Grothe B (2008) Dynamic Spectrotemporal Feature Selectivity in the Auditory Midbrain. J Neurosci 28:5412–5421.

Lewis ER, Henry KR, Yamada WM (2002) Tuning and timing of excitation and inhibition in primary auditory nerve fibers. Hear Res 171:13–31.

Mao J, Carney LH (2015) Tone-in-Noise Detection Using Envelope Cues: Comparison of Signal-Processing-Based and Physiological Models. JARO 16:121–133.

Mao J, Vosoughi A, Carney LH (2013) Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues. J Acoust Soc Am 134:396–406.

Martin MR, Dickson JW (1983) Lateral inhibition in the anteroventral cochlear nucleus of the cat: A microiontophoretic study. Hear Res 9:35–41.

Moore BCJ (1975) Mechanisms of masking. J Acoust Soc Am 57:391–399.

Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J Acoust Soc Am 116:2173–2186.

Nelson PC, Carney LH (2007) Neural Rate and Timing Cues for Detection and Discrimination of Amplitude-Modulated Tones in the Awake Rabbit Inferior Colliculus. J Neurophysiol 97:522–539.

Palmer AR, Shackleton TM, Sumner CJ, Zobay O, Rees A (2013) Classification of frequency response areas in the inferior colliculus reveals continua not discrete classes. The Journal of Physiology 591:4003–4025.

Palombi PS, Caspary DM (1996) Physiology of the aged Fischer 344 rat inferior colliculus: responses to contralateral monaural stimuli. J Neurophysiol 76:3114–3125.

Patterson R, Nimmo-Smith I, Holdsworth J, Rice P (1988) Implementing a GammaTone Filter Bank. SVOS.

Patterson RD (1976) Auditory filter shapes derived with noise stimuli. J Acoust Soc Am 59:640–654.

Pollak GD, Xie R, Gittelman JX, Andoni S, Li N (2011) The dominance of inhibition in the inferior colliculus. Hear Res 274:27–39.

Portfors CV, Felix RA (2005) Spectral integration in the inferior colliculus of the CBA/CaJ mouse. Neuroscience 136:1159–1170.

Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Neural Comput 16:1661–1687.

Ramachandran R, Davis KA, May BJ (1999) Single-Unit Responses in the Inferior Colliculus of Decerebrate Cats I. Classification Based on Frequency Response Maps. J Neurophysiol 82:152–163.

Ramachandran R, Davis KA, May BJ (2000) Rate Representation of Tones in Noise in the Inferior Colliculus of Decerebrate Cats. JARO 1:144–160.

Rees A, Palmer AR (1988) Rate-intensity functions and their modification by broadband noise for neurons in the guinea pig inferior colliculus. J Acoust Soc Am 83:1488–1498.

Rhode WS, Greenberg S (1994) Lateral suppression and inhibition in the cochlear nucleus of the cat. J Neurophysiol 71:493–514.

Richards VM (1992) The detectability of a tone added to narrow bands of equal-energy noise. J Acoust Soc Am 91:3424–3435.

Rocchi F, Ramachandran R (2018) Neuronal adaptation to sound statistics in the inferior colliculus of behaving macaques does not reduce the effectiveness of the masking noise. J Neurophysiol 120:2819–2833.

Rodieck RW (1965) Quantitative analysis of cat retinal ganglion cell response to visual stimuli. Vision Res 5:583–601.

Rodieck RW, Stone J (1965) Analysis of Receptive Fields of Cat Retinal Ganglion Cells. J Neurophysiol 28:833–849.

Saint-Marie RL, Baker RA (1990) Neurotransmitter-specific uptake and retrograde transport of [3H]glycine from the inferior colliculus by ipsilateral projections of the superior olivary complex and nuclei of the lateral lemniscus. Brain Res 524:244–253.

Saint-Marie RL, Ostapoff E -Michael, Morest DK, Wenthold RJ (1989) Glycine-immunoreactive projection of the cat lateral superior olive: Possible role in midbrain ear dominance. J Comp Neurol 279:382–396.

Saldaña E, Merchań MA (1992) Intrinsic and commissural connections of the rat inferior colliculus. J Comp Neurol 319:417–437.

Schreiner CE, Langner G (1997) Laminar fine structure of frequency organization in auditory midbrain. Nature 388:383–386.

Schwarz DM, Zilany MSA, Skevington M, Huang NJ, Flynn BC, Carney LH (2012) Semi-supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. J Neurosci Methods 206:120–131.

Shamma SA (1985) Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. J Acoust Soc Am 78:1622–1632.

Su Y, Delgutte B (2020) Robust Rate-Place Coding of Resolved Components in Harmonic and Inharmonic Complex Tones in Auditory Midbrain. J Neurosci 40:2080–2093.

van de Par S, Kohlrausch A (1999) Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters. J Acoust Soc Am 106:1940–1947.

Versnel H, Zwiers MP, Van Opstal AJ (2009) Spectrotemporal Response Properties of Inferior Colliculus Neurons in Alert Monkey. J Neurosci 29:9725–9739.

Wang Y, Abrams KS, Carney LH, Henry KS (2021) Midbrain-Level Neural Correlates of Behavioral Tone-in-Noise Detection: Dependence on Energy and Envelope Cues. J Neurosci 41:7206–7223.

Whitehead ML, Lonsbury-Martin BL, Martin GK (1992) Evidence for two discrete sources of 2 f1−f2 distortion-product otoacoustic emission in rabbit: I. Differential dependence on stimulus parameters. J Acoust Soc Am 91:1587–1607.

Winer JA, Larue DT, Pollak GD (1995) GABA and glycine in the central auditory system of the mustache bat: Structural substrates for inhibitory neuronal organization. J Comp Neurol 355:317–353.

Xie R, Meitzen J, Pollak GD (2005) Differing Roles of Inhibition in Hierarchical Processing of Species-Specific Calls in Auditory Brainstem Nuclei. J Neurophysiol 94:4019–4037.

Yang L, Pollak GD (1997) Differential Response Properties to Amplitude Modulated Signals in the Dorsal Nucleus of the Lateral Lemniscus of the Mustache Bat and the Roles of GABAergic Inhibition. J Neurophysiol 77:324–340.

Yang L, Pollak GD, Resler C (1992) GABAergic circuits sharpen tuning curves and modify response properties in the mustache bat inferior colliculus. J Neurophysiol 68:1760–1774.

Zheng L, Early SJ, Mason CR, Idrobo F, Harrison JM, Carney LH (2002) Binaural detection with narrowband and wideband reproducible noise maskers: II. Results for rabbit. J Acoust Soc Am 111:346–356.

Zilany MSA, Bruce IC, Carney LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. J Acoust Soc Am 135:283–286.

Chapter 3: Timbre Encoding in the Inferior Colliculus

This chapter is in revision at Journal of Neuroscience.

3.1 Abstract

Timbre, or the quality of a sound, is a critical component in speech and music. One percept of timbre, brightness, is correlated with the spectral centroid and the spectral envelope of harmonic sounds. Little is known about how this aspect of timbre is encoded in the subcortical auditory system. We used physiological and computational modeling methods to investigate the representation of spectral peaks in a harmonic complex tone with a broad, triangular-shaped spectrum. Extracellular single-neuron recordings were made in the central nucleus of the inferior colliculus in awake, female Dutch-belted rabbits. A population response to the timbre stimulus was inferred by shifting the stimulus spectrum above and below the characteristic frequency of each neuron. Spectral peaks in the stimulus were encoded in peaks in the average-rate profiles of the majority of neurons, and this representation was robust over a range of suprathreshold levels. Neural discrimination thresholds were also sufficient to describe human behavioral thresholds. Temporal responses were complex and often exhibited mode-locking. Computational models that included neural-fluctuation sensitivity and broad inhibition captured the major trends in physiological results. These findings demonstrate that multiple mechanisms may influence robust spectral-peak encoding in inferior colliculus neurons.

3.2 Introduction

Timbre is the sound quality that allows listeners to distinguish between complex sounds that are identical in pitch, duration, and loudness (ANSI, 1994). The percept of timbre contributes to speech intelligibility, music enjoyment, and sound-source recognition (Review: McAdams 2019). Timbre can be decomposed into spectral features (e.g., spectral centroid, slope, and variation) and temporal features (e.g., attack time) (Grey and Gordon, 1978; Iverson and Krumhansl, 1993; McAdams et al., 1995). Timbre perception has been studied psychophysically; however, subcortical encoding of timbre remains poorly understood. This study tested the hypothesis that spectral peaks, related to the spectral centroid, are robustly encoded in inferior colliculus (IC) rate and temporal profiles (Fig. 3.1). The hypothesis was tested using harmonic complex tones with triangularly shaped spectra that elicit a timbral 'brightness' percept (from Allen and Oxenham, 2014) (Fig. 3.1a).

Although psychophysical timbre dimensions have not been studied in the IC, vowel-encoding research gives insight into spectral-peak encoding (Perez et al., 2013; Carney et al., 2015). Studies of vowel-like sounds have used harmonic complexes with steeply sloping, triangular spectral peaks, steeper but otherwise similar to the stimulus used in this study (Lyzenga and Horst, 1995; Tan and Carney, 2005; Henry et al., 2017). Model and physiological responses to vowels support a transformation from temporal to rate coding in the IC (Carney et al., 2015; Carney, 2018; Carney and McDonough, 2019). However, although IC rate coding of single-formant stimuli in budgerigar can explain spectral-peak discrimination in quiet, temporal coding was needed to explain behavioral

thresholds in noise (Henry et al., 2017). Timing information in the IC is excellent for strongly periodic sounds, such as vowels or amplitude modulated (AM) stimuli with low-frequency periodicities (Rees and Langner, 2005; Kim et al., 2020). Most IC neurons are sensitive to AM (Langner and Schreiner, 1988; Krishna and Semple, 2000; Joris et al., 2004; Nelson and Carney, 2007; Kim et al., 2020), a key feature of harmonic sounds.

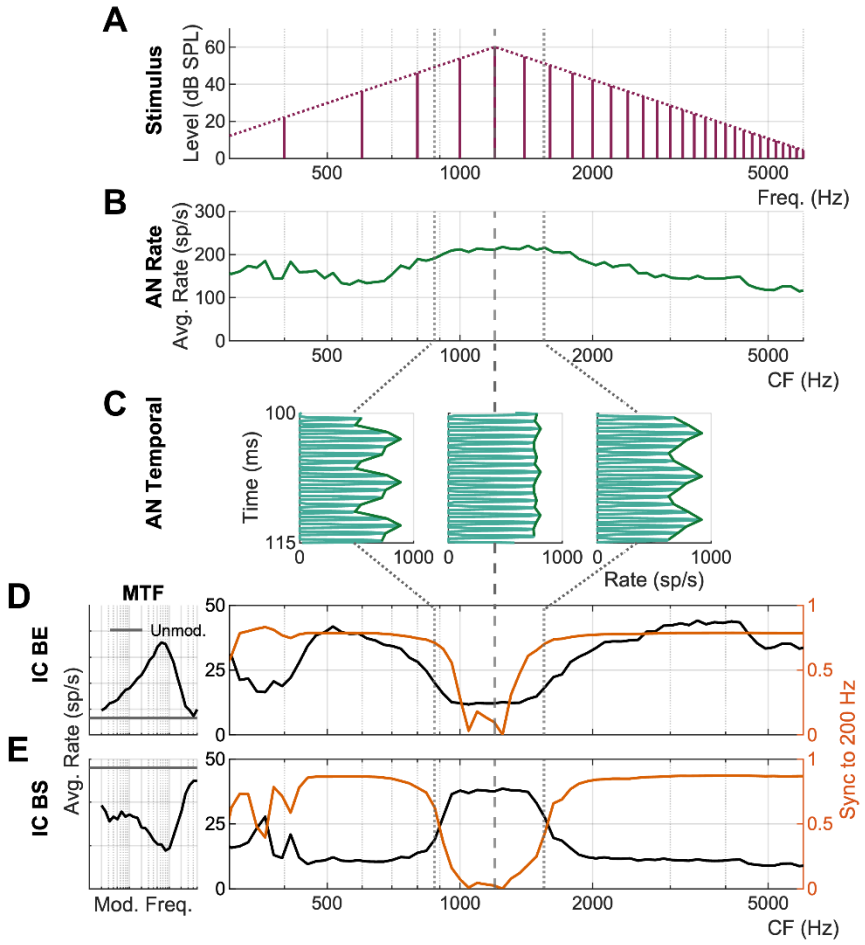

Figure 3.1. Illustration of rate and temporal coding of a spectral peak based on model AN and IC neurons. (A) Synthetic timbre stimulus with spectral peak at 1200 Hz (dashed line) and F0 of 200 Hz. (B) AN average rate profile for a population of fibers with CFs spanning the stimulus spectrum. (C) Temporal responses for AN fibers tuned to 875, 1200, and 1550 Hz (dotted and dashed lines). (D) Population response for same-

frequency inhibition-excitation (SFIE) model IC-BE cells (MTF, left) (Nelson & Carney, 2004). Rate response (black) and synchronization coefficient to F0 = 200 Hz (orange) (E) Population response for model IC-BS cells (MTF, left). Rate response (black) and synchronization coefficient to F0 = 200 Hz (orange).

We hypothesize that AM sensitivity in the IC could contribute to the encoding spectral peaks through IC responses to neural-fluctuation (NF) profiles present in auditory-nerve (AN) fibers. NFs are low-frequency temporal changes in firing rate that are shaped by nonlinear inner-hair-cell (IHC) transduction (Carney, 2018, 2024). AN fibers tuned near a spectral peak are "captured," and the NFs are reduced, due to inner-hair-cell saturation (Fig 3.1c, middle). AN fibers tuned away from the spectral peak, where IHCs are not saturated, have NFs at the fundamental frequency (F0) (Fig. 3.1c, left, right). The IC could respond to NF profiles in at least two ways: 1. Neurons tuned away from spectral peaks could phase-lock to F0 more strongly than neurons tuned at the spectral peak, thus representing the spectral peak by reduced phase locking to F0 (Fig. 3.1d, e); 2. Neurons that are band-enhanced (BE, Fig. 3.1d) or band-suppressed (BS, Fig. 3.1e) by fluctuations, in comparison with unmodulated stimuli, could encode spectral peaks in firing-rate changes corresponding to NF-depth variations (Mao and Carney, 2015; Carney, 2018, 2024; Carney and McDonough, 2019; Maxwell et al., 2020). Broad inhibition, from frequencies spanning characteristic frequency (CF), could also sharpen the rate profile at a spectral peak (Fritzinger and Carney, 2025). Inhibition strongly impacts IC representations of complex sounds; blocking inhibition broadens tuning curves, removes frequency-sweep direction selectivity, and alters binaural responses (Andoni et al., 2007; Pollak, 2011; Williams and Fuzessery, 2011).

This study investigated IC encoding of spectral peaks in harmonic stimuli by recording single-neuron responses in awake Dutch-belted rabbits and shifting the stimulus spectrum with respect to CF to infer population responses. Temporal patterns were more complex than predicted, exhibiting mode locking (Aihara et al., 1984) in addition to phase-locking to F0. Rate profiles robustly encoded spectral peaks at suprathreshold levels. Rate-based neural timbre-discrimination thresholds were sufficient to explain human thresholds. Lastly, we used computational IC models and receptive-field models to investigate how energy, neural fluctuations, and broad inhibition impact rate profiles.

3.3 Methods

3.3.1 Surgical Procedures

All procedures were approved by the University of Rochester Committee on Animal Resources in compliance with National Institutes of Health Guidelines. Four female Dutch-belted rabbits (Oryctolagus cuniculus) ranging from six months to five years of age were studied. All animals had normal hearing, monitored using distortion product otoacoustic emissions (Whitehead et al., 1992).

Animals were anesthetized intramuscularly with 66 mg/kg ketamine and 2 mg/kg xylazine, or 35mg/kg ketamine and 0.10-0.15mg/kg dexmedetomidine, for headbar placement, initial craniotomy and microdrive placement, and microdrive replacements. The headbar placement surgery consisted of affixing a custom 3D-printed headbar (ProtoLabs, Maple Plain, MN) to the skull of the animal using screws and dental acrylic.

The headbar was designed with a chamber to hold a microdrive. After recovery, a craniotomy was performed and a custom microdrive containing four tetrodes, modeled after the Neuralynx Five-drive microdrive (Neuralynx, Inc., Bozeman, MT), was inserted into the headbar. The microdrive consisted of a 3D-printed body with stainless-steel guide tubes that extended into the craniotomy. Microdrive replacement surgeries were performed, on average, every six months.

Experiments were discontinued if emissions dropped by 10 dB or neural thresholds increased. Animals were euthanized and the IC was sectioned and stained to verify tetrode location in the central nucleus.

3.3.2 Stimulus Presentation & Neural Recording Procedure

Each recording session was two hours in duration inside a sound-attenuated booth (Acoustic Systems, Austin, Texas, USA). Animals were head fixed using the headbar and placed in a custom chair and wrapped in a towel to limit movement. Custom earmolds (Hal-Hen Company, Inc., Garden City Park, NY, USA OR Dreve Otoform Ak, Unna, Germany) were inserted for closed-field stimulus presentation.

Stimuli were generated using MATLAB, sent to an audio interface (16A, Mark of the Unicorn, Cambridge, Massachusetts, USA) and converted from digital to analog (DAC3 HGC, Benchmark Media Systems, Inc., Syracuse, New York, USA). Stimuli were then presented using earphones (Beyerdynamic DT-48, Beyerdynamic GmbH and Co., Heilbronn, Germany or Etymotic ER2, Etymotic Research, Inc., Elk Grove Village, Illinois). At the beginning of each session, a calibration curve was calculated by presenting 50- to 20-kHz tones and recording the speaker output using a probe-tube

microphone (Etymotic ER10B+ or ER7C, Etymotic Research, Inc., Elk Grove Village, Illinois). The stimuli were filtered by the calibration curve to compensate for the frequency response of the acoustic system.

Neural recordings were made with tetrodes consisting of four twisted 18-µm epoxy-coated platinum-iridium wires (California Fine Wire Co., Grover Beach, CA) plated with a platinum-black solution (Neuralynx, Inc., Bozeman, MT) to lower impedances to approximately 0.1-1.5 MOhms. Neural signals from the tetrodes were recorded using a 16-channel headstage, amplifier, and software from the Intan RHD Recording system (Intan Technologies, LLC., Los Angeles, CA, USA). Tetrodes were advanced and retracted at the beginning or end of a session to sample different areas of the IC. The location of the recordings was determined to be the central nucleus of the IC if tuning increased with tetrode depth.

After recording, waveforms were filtered and sorted into single-unit neurons as described below. First, waveforms were filtered using a 4^{th}-order Butterworth bandpass filter with passband from 300 to 3000 Hz. Second, a threshold of four times the standard deviation (STD) of the waveforms was set to determine spike times (Quiroga et al., 2004). Lastly, the waveforms were sorted into neurons using custom spike-sorting software (Schwarz et al., 2012). A waveform had to meet two criteria to be considered a single-unit neuron: less than 2% of spikes occurred with intervals shorter than 1 ms, and the cluster entanglement metric was less than 0.1 (Schwarz et al., 2012).

3.3.4 Stimuli & Analysis

At the beginning of each recording session, a set of standard stimuli were presented to characterize each neuron. This characterization included estimating the best frequency of a neuron over many sound levels by presenting pure-tone stimuli with varying frequencies and sound levels. The pure tones were presented from 250 – 16000 Hz in 5 steps per octave at 10, 30, 50, and 70 dB SPL. The stimuli were 200 ms in duration with 10-ms raised-cosine on/off ramps and were presented diotically for 3 repetitions in random order, with 400-ms interstimulus intervals. The CF was estimated as the frequency that elicited the largest average-rate response at the lowest sound level.

Modulation transfer functions, to assess AM sensitivity, were measured by presenting broadband Gaussian noise (100-10 kHz) sinusoidally modulated from 2-600 Hz, with 3 steps per octave. Stimuli were presented at 33 dB SPL spectrum level, were 1 s in duration with 50-ms raised-cosine on/off ramps, with 500-ms interstimulus intervals. Stimuli were presented five times, diotically, in a random sequence. MTFs were classified into four categories based on the average-rate responses to modulated noises compared to an unmodulated noise baseline (Fritzinger & Carney, 2025). Neurons were classified as band-enhanced (BE) if two modulation-frequency rates were significantly higher than the unmodulated rate, without a rate at an intermediate modulation frequency that was significantly below the unmodulated rate. Categorization for BS neurons required two rates that were significantly below the unmodulated rate, without an intermediate response that exceeded the unmodulated response. Hybrid neurons satisfied

both BE and BS criteria, at different modulation frequencies, and flat neurons did not satisfy either criteria.

The synthetic-timbre reference stimulus was based on Allen & Oxenham (2014) and consisted of a harmonic complex tone with 200-Hz F0 and a triangular-shaped spectrum with a roll-off of 24 dB/octave. The peak harmonic in the reference stimulus was the multiple of F0 nearest to the CF of the neuron. All harmonic components were added in sine phase. The upper frequency limit of the stimulus was 10 kHz. The stimulus was 300-ms duration with 20-ms raised-cosine on/off ramps. The entire spectrum of the reference stimulus was shifted in 35-50 Hz increments to span six harmonics above and below the reference (Fig. 3.2). All stimuli were presented in random order, with 30 repetitions of each stimulus. The stimuli were presented at 43, 63, 73, or 83 dB SPL to either the contralateral ear or diotically.

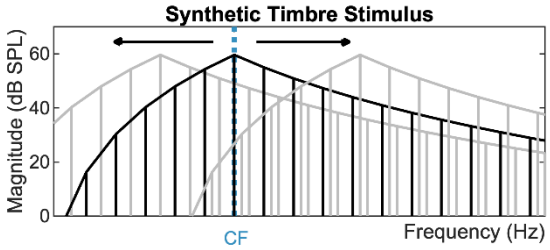

Figure 3.2. Schematic of shifted synthetic-timbre stimulus with triangular spectral envelope. The peak frequency of the reference stimulus (black) was the multiple of 200 Hz closest to the CF of the neuron.

Rate profiles in response to the synthetic-timbre stimuli were calculated by averaging the firing rate over the duration of the stimulus, excluding a 50-ms onset. Rates were plotted as a function of spectral-peak frequency. Temporal analysis included calculating period histograms and vector strength at the 200-Hz F0.

Rate profiles were separated into three categories that described the rate-profile shape: peak, dip, and sloping (Fig. 3.3). First, average-rate profiles were smoothed and converted to z-scores using the mean and standard deviation from the smoothed rate profile. Next, we identified peak profiles as those with a peak in the rate profile within +/- one octave around CF using the MATLAB function 'findpeaks' with a minimum peak prominence set at 0.25. Dip profiles were identified by inverting the rate profile (multiplying by -1) and applying the same criteria as for the peak analysis. Many neurons had both peaks and dips in the rate profiles; those neurons were categorized by which feature (peak or dip) was nearest to CF (Fig. 3.3b). Sloping neurons were categorized as having no peaks or dips within +/- 1 octave around CF. A salience metric, Q, was calculated for rate profiles that were categorized as peak or dip:

$$Q = \frac{f_p}{BW_{0.75}},$$

where f_p refers to the frequency at the peak or dip, and $BW_{0.75}$ refers to the bandwidth of the response at 0.75 z-score below the peak rate (or 0.75 z-score above the dip).

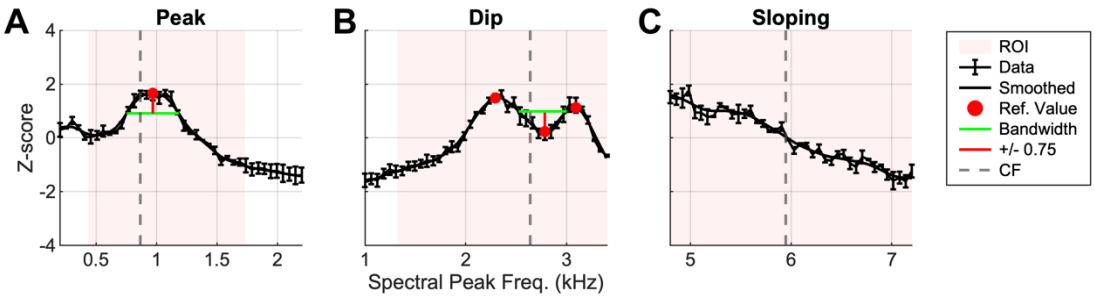

Figure 3.3. Examples of peak, dip, and sloping rate profiles, with quantification of peak and dip profiles. (A) Example rate profile with a peak near CF. Pink shaded area represents region of interest (+/- one octave around CF). Red dots indicate peaks/dips. The peak or dip (red circles) that was closest to CF (dotted grey line) was chosen as the

reference value for determining bandwidth. Red vertical line indicates +/-0.75 from reference value, depending on peak or dip. Green line indicates bandwidth measurement. (B) Example rate profile with a dip. (C) Example of a sloping rate profile with no peaks or dips.

Neurons were also classified as sharpening, broadening, or unchanging over sound level, depending on changes in Q as level increased. A slope was determined using linear regression for Q-values at 43, 63, and 83 dB SPL. If the slope was greater than 0.03/dB, the neuron was classified as sharpening, if the slope was less than -0.03/dB the neuron was classified as broadening. A slope within -0.03 and 0.03/dB was determined to be unchanging.

A neural discrimination threshold was calculated for each IC neuron. First, rates and standard deviations of rates were linearly interpolated. Then, the steepest slope between two consecutive rates (n, n+1) was found, and the d' was calculated for those two rates. If the d' was greater or equal to 1, then threshold was calculated using those two rates. However, if d' was less than one, the steepest slope between (n, n+2) was then found. The process continued, sampling (n, n+i) where i = 1, 2, 3, …. 100, until either a d' of one was found or no threshold could be calculated. Lastly, the frequency discrimination threshold, in percent of frequency, was calculated as: $\Theta = \frac{\Delta f}{f_{mid}} \times 100$, where Δf is the difference in spectral peak frequency between the two rates and f_{mid} is the frequency at the midpoint of the two spectral peak frequencies.

A rate analysis was performed to test how the rate representation changed over the duration of the stimulus. The average-rate profiles were computed over two time

windows, 50-150 ms and 200-300 ms. The salience metric, Q, was used to quantify changes in peaks or dips in the rate profiles over the duration of the stimulus.

3.3.5 Models

Linear, spectral receptive-field models were used to predict neural rate profiles based on the synthetic-timbre stimulus spectra. Two models were tested, a gaussian and a difference of gaussians (DoG). The procedure and equations used to fit both models are described in Su & Delgutte (2020). Briefly, model predictions were calculated as the sum of stimulus spectral components, each weighted by the corresponding model magnitude, with the spontaneous firing rate of the neuron added as a constant. For the model with inhibitory weights, the prediction was half-wave rectified. Model parameters were determined using the MATLAB function fmincon to minimize the mean squared error between the rate profile and the prediction. The result was a gaussian or DoG model that provided the best model prediction fit to the neural rate profiles. An adjusted goodness-of-fit R^2 metric (Su & Delgutte 2020) was used to evaluate model performance while accounting for differences in the number of parameters in each model. Single-sided F tests were used to evaluate if the DoG model fit better than the gaussian model, with the null hypothesis that both models fit the data equally well.

Three IC models (energy, same-frequency inhibition-excitation (SFIE), and broad-inhibition) were evaluated against synthetic-timbre rate profiles. Model predictions were made for each neuron in the dataset. Detailed implementations of these models are described in Fritzinger & Carney (2025). The energy model consisted of a 4^{th}-order gammatone filter with a bandwidth based on cat Q_{10} values (Carney & Yin, 1988)

centered on the CF of the neuron. The SFIE and broad-inhibition models both used an auditory-nerve model with cat tuning parameters (Zilany et al., 2014). The SFIE model is a phenomenological model of BE and BS MTF IC neurons that accurately predicts AM sensitivity in IC neurons (Nelson and Carney, 2004; Carney and McDonough, 2019). This model includes same-frequency inhibition and excitation, where the time course and amplitude of inhibition compared to excitation creates AM sensitivity. The broad-inhibition model expands on the SFIE model by adding two parallel off-CF BS pathways that inhibit the on-CF BS cell. The three inhibitory pathways (one below CF, one at CF, and one above CF) constitute the broad inhibition. Depending on the inhibition strength, BMFs, and off-CF frequencies chosen, BE, BS, hybrid, and flat MTFs can be simulated using this model configuration. The parameters used in the BE and BS broad-inhibition model were from Fritzinger & Carney (2025). Model fits to data were evaluated using the variance-explained goodness-of-fit metric, R^2. The Q value and thresholds calculated from neural rate profiles were also applied to the model rate profiles to compare trends in the model responses to those in the data.

3.3.6 Statistics

A linear mixed-effects model was used to detect statistically significant changes in salience, log(Q), due to stimulus parameters and neuron characteristics. Q was log transformed for normality. Model selection was done using MATLAB functions LME and compare, then JASP (JASP Team, 2024, Version 265 0.18.3, computer software) was used for the remaining analyses. The fixed effects included MTF type, level (dB SPL), CF group, and presentation ear (contralateral or diotic). Neuron identity was added as a

random effect, implemented as a random intercept, to account for heterogeneity in the neuron population. The model chosen included two-, three- and four-way interactions between the four fixed effects. Model fit was assessed using visual examination of residuals and quantile-quantile plots in MATLAB. In JASP, models were analyzed using type III ANOVAs using the Satterthwaite approximation. Other statistical methods used throughout this study include t-tests, the Kruskal-Wallis test for non-parametric data, and comparisons of correlation coefficients and variance explained.

3.4 Results

We recorded responses of 188 neurons in the central nucleus of the IC in four awake Dutch-belted rabbits to synthetic-timbre stimuli. Responses to diotic synthetic-timbre presentations at 43, 63, 73, and 83 dB SPL were recorded in 153, 163, 82, and 152 of the 188 neurons, respectively. Responses to contralateral synthetic-timbre presentations at 43, 63, 73, and 83 dB SPL were recorded in 66, 70, 13, and 66 of the 188 neurons, respectively. The CFs of the dataset ranged from 320 to 9236 Hz, with a median of 2313 Hz (Fig. B1a). The MTF distribution was 45 BE (23.9%), 103 BS (54.8%), 24 hybrid (12.8%), and 16 flat (8.5%, Fig. B1b). Best modulation frequencies (BMFs) of BE neurons ranged from 12 to 324 Hz with a median of 67 Hz, and worst modulation frequencies (WMFs) of BS neurons ranged from 2 to 435 Hz with a median of 78 Hz (Fig. B1c, d).

3.4.1. Rate and timing properties of responses to synthetic-timbre stimuli

The hypothesis that the spectral peaks are encoded by peaks or dips in the rate profiles, dependent on MTF type, was tested by analyzing average-rate responses to the synthetic-timbre stimuli (Fig. 3.1). Each rate profile represents an inferred population response for neurons tuned at, above, or below the spectral-peak frequency, and rate profiles were expected to be similar but reflected around CF compared to the population-model predictions (Fig. 3.1).

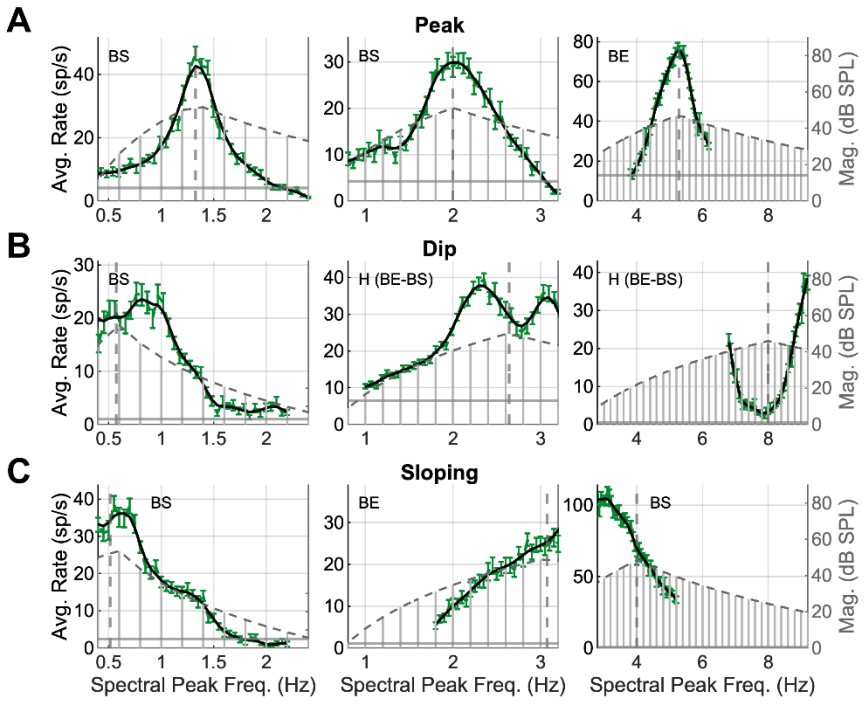

Figure 3.4. Nine example neuron responses to synthetic timbre stimuli, with low, medium, and high CFs (left to right). (A) Three examples of responses with a peak near CF. Labels of MTF type in the top left corner of each plot. Data +/- standard error of the mean (SEM) (green), smoothed data (black), estimated CF (dashed grey line), and spontaneous rate (dark grey horizontal line). The stimulus spectrum is overlayed in grey with a grey dotted line outlining the spectral envelope. Synthetic timbre stimulus spans

2400 Hz for all examples. (B) Three responses with a dip near CF. (C) Three neurons with sloping responses.

Neural responses varied, and rate profiles were classified into three groups: peaks, dips, and sloping responses (Fig. 3.4). Peaked responses occurred in neurons at low (<2 kHz), mid (2-4 kHz) and high (4+ kHz) CFs (Fig. 3.4a, left, middle, right). Similarly, responses with dips or slopes occurred in neurons with a range of CFs (Fig. 3.4b, c). The response category did not depend on the MTF type of the neuron. Peaks, dips, and sloping responses were present in all MTF types (Fig. 3.4), inconsistent with the hypothesis that MTF type would determine the shape of the rate profile. Rate profiles were sharper than the stimulus envelope in all classifications.

Vector strength to the 200-Hz F0 was calculated for each spectral-peak frequency to test the hypothesis that the vector strength decreases near the spectral peak (Fig. 3.1), thus providing information about the spectral-peak location. Vector strength varied in individual neurons (Fig. 3.5a), and overall, many neurons had a dip in vector strength near the CF of the neuron (Fig. 3.5b). Period histograms (for a 200-Hz F0) were plotted across the spectral peak frequencies to further scrutinize temporal response features (Fig. 3.5c-g). Temporal responses were highly variable, and neurons were classified into five groups (Fig. 3.5C-G). First, 27% of neurons matched the initial hypothesis that vector strength decreased near CF (n=49/184, Fig. 3.5c), often exhibiting mode-locking. In fact, 61% of neurons exhibited mode locking (n=112/184, Fig. 5c-f). Sixteen % of responses were consistent over the range of spectral-peak frequencies tested (n=30/184, Fig. 3.5d), whereas 11% of responses were more phase-locked or mode-locked near the CF

(n=20/184, Fig. 3.5e). Interestingly, 17% of neurons, often with high CFs, had 3-4 peaks in the period histogram in each stimulus period (n=32/184, Fig. 3.5f). Lastly, 29% of IC neurons had little to no phase locking to the F0 (Fig. 3.5g). The large diversity in temporal responses and the mode locking in these responses complicated the analysis; vector strength is typically low in neurons that mode lock. Though some neurons were consistent with the initial hypothesis, more analysis is needed to understand the mode-locked responses. Due to the complexity of the period histograms and the result that rate profiles provided robust information about the spectral peak, the rest of the study focused on rate profiles.

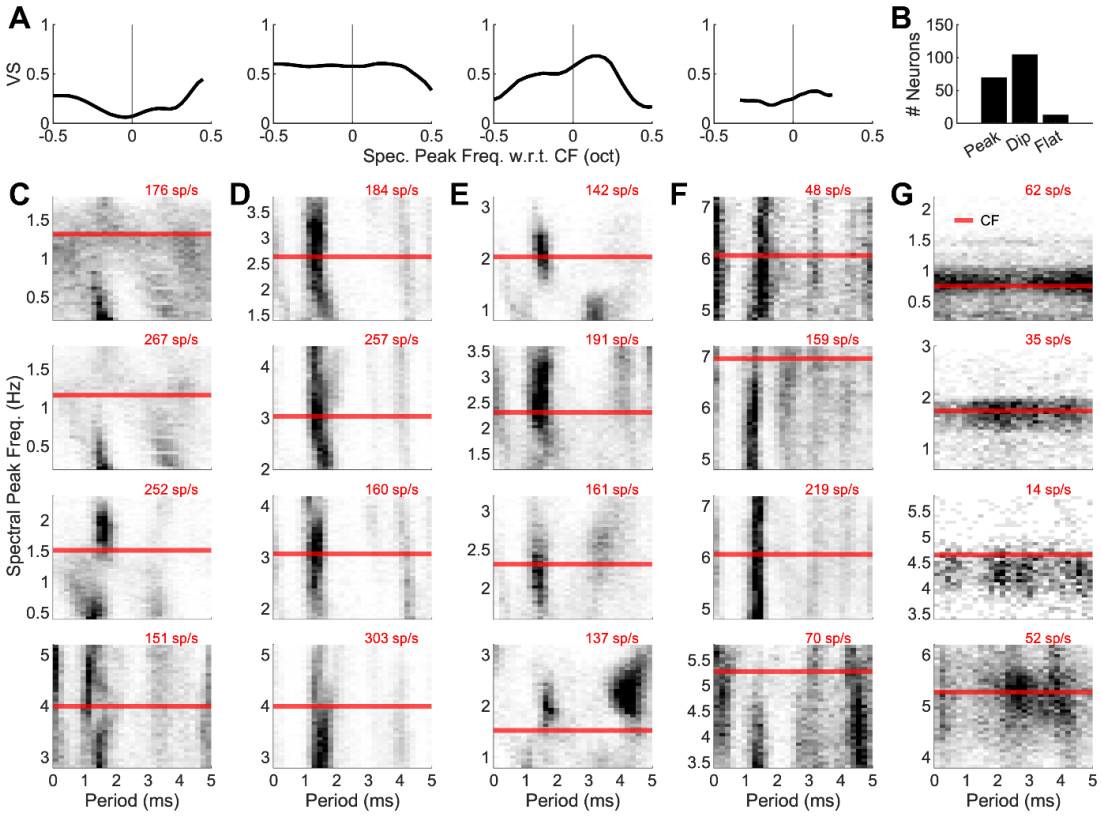

Figure 3.5. Temporal responses were influenced by mode locking. (A) Example vector strength calculations for the top example neuron in each column (C-F). Vector strength

for G is near 0 and is not shown. (B) Histogram of vector strength profiles characterized as peak, dip, or flat. (C-G) Normalized period histograms for a range of peak frequencies for 20 example neurons to synthetic timbre, diotic, at 63 dB SPL. Red line indicates CF, and numbers in the upper-right corner of each plot are the maximum rate across all spectral peaks. Example neurons were grouped as follows: (C) decrease in phase locking near CF, (D) strong mode followed by 1-2 weaker modes, (E) mode(s) that decrease away from CF, (F) many modes, and (G) little or no phase locking.

3.4.2. Rate profiles peaked at the harmonic nearest to spectral peak over a wide range of CFs

Neural rate profiles featured robust peaks, dips, and sloping responses. For the 63-dB SPL, diotic presentation, 81% (132/163) of rate profiles featured peaks near CF (Fig 3.6a), 10% (17/163) featured dips near CF (Fig 3.6b), and 9% (14/163) featured sloping responses (Fig 3.6c). The majority of responses had strong peaks, regardless of MTF type (Fig. 3.6d).

Hybrid neurons were generally expected to have a rate profiles similar to those of BE or BS types, depending on the MTF shape at 200 Hz. Flat neurons were hypothesized to have flat rate profiles in response to synthetic-timbre stimuli. However, the majority of both hybrid and flat neurons also exhibited peaks in rate profiles. Overall, BS neurons were consistent with the hypothesis that responses would have peaks in the response due to NF profiles (Fig. 3.1); however, BE cells did not support the hypothesis that rates would decrease for spectral peaks near CF.

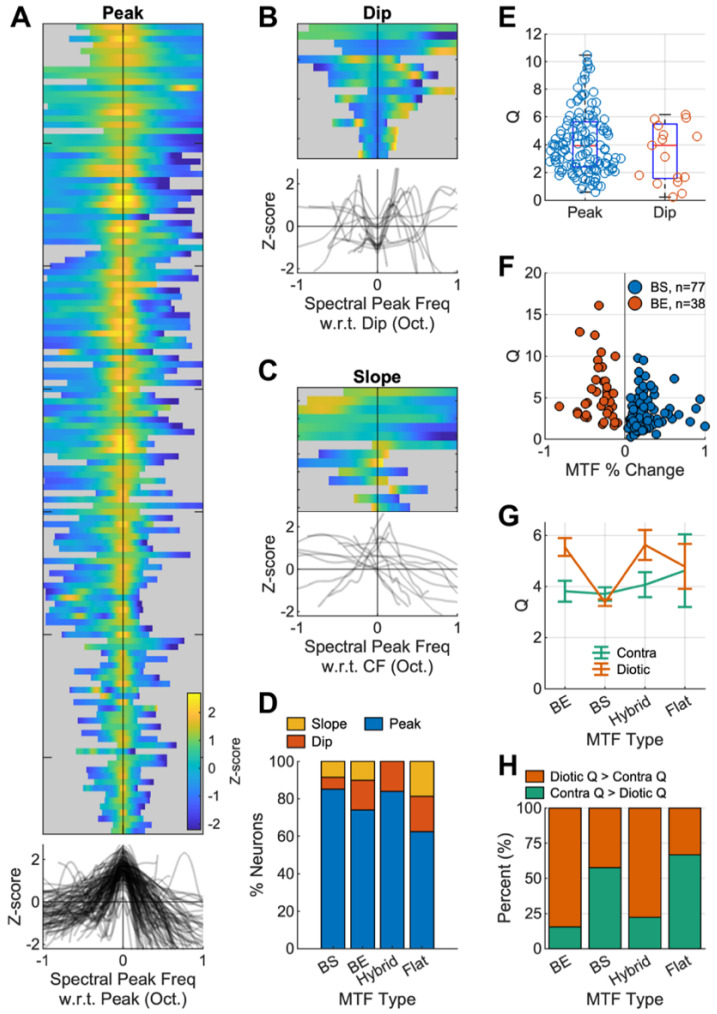

Figure 3.6. Population analysis to test the hypothesis that increases and decreases in rate (with respect to MTF type) capture the spectral peak. (A) Neuron responses to 63 dB SPL synthetic timbre that are categorized with a peak near CF, ordered by CF with low CFs at the top and high CFs at the bottom. Grey traces (bottom) are all response profiles overlaid. (B) Neuron responses that feature a dip near CF. (C) Neuron responses that feature sloping profiles. (D) Proportions of peak, dip, and sloping responses for BE, BS, hybrid, and flat MTFs. (E) Rate-profile salience, Q, for neurons with peaks and dips, no significant difference between groups. (F) Q value as a function of % MTF rate changes ($\frac{r_{unmod}-r_{max}}{r_{unmod}+r_{max}}$) for BE (red) and BS (blue). (G) Significant interaction between MTF and contralateral/diotic presentation, Q-value plotted as a function of MTF type, separated

into binaural and diotic presentations. (H) Percentage of neurons that increase (orange) or decrease (green) in Q from diotic to contra stimulus presentation.

Next, the salience metric, Q, was calculated for every response with a peak or a dip (Fig. 3.3) to compare the sharpness of the rate profiles across neurons and sound levels. There was no significant difference in salience (Q) between responses with peaks or dips ($p = 0.5077$, Welch's t-test, Fig. 3.6e). Salience was plotted as a function of MTF % change (i.e. the MTF rate for unmodulated noise minus the maximum or minimum rate, for BE or BS, respectively). There was not a significant linear relationship between MTF strength and salience (BE, $p=0.7221$; BS, $p=0.6853$, linear regression). Interestingly, BE and hybrid neurons had significantly sharper salience in the diotic condition compared to the contralateral condition (Fig. 3.6g, $F(4, 565) =3.33$, $p=0.010$). A larger proportion of neurons with BE and hybrid MTFs had sharper responses to diotic compared to contralateral presentations, whereas BS and flat MTF responses were more evenly divided (Fig. 3.6h). The involvement of the ipsilateral pathway, which is often inhibitory, in the diotic presentation sharpened responses to the stimulus, but only for neurons with BE and hybrid MTFs. More experiments are required to clarify how the ipsilateral pathway impacts responses in these neurons.

3.4.3. Rate profiles were robust across a range of suprathreshold levels

We next tested the robustness of rate profiles in response to synthetic-timbre stimuli over suprathreshold levels. Salience was calculated for all neurons with responses to diotic stimuli at 43, 63, and 83 dB SPL (n=97). When pooling across all neurons, there

was no significant change in Q as level increased (p=0.4208, Kruskal-Wallis). However, individual neurons had heterogeneous responses and many increased or decreased in Q.

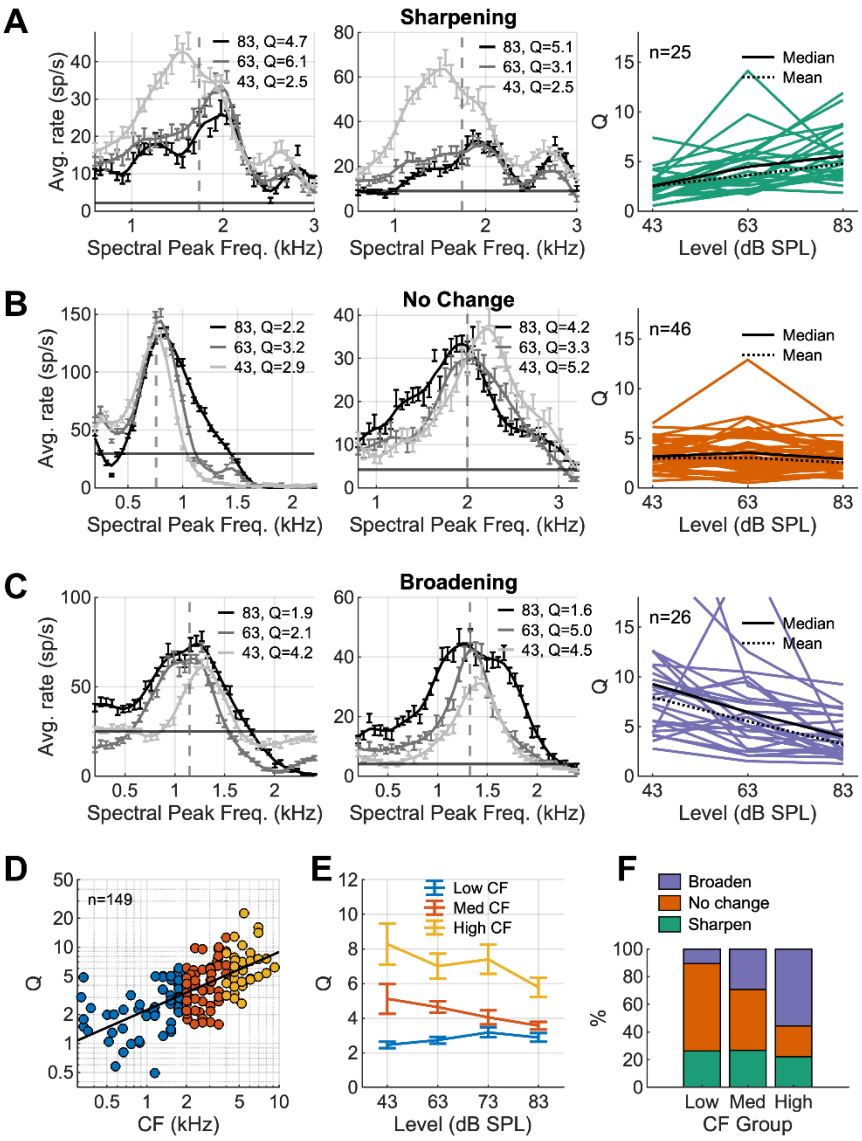

Figure 3.7. Analysis of neural responses over level. (A) Two example neuron responses (left, middle), and Q vs level for neurons that sharpen in Q as level increases. (B) Same as A, for neurons that do not change over level. (C) Same as A, for neurons that broaden in Q as level increases. (D) Q values as a function of CF for the 63 dB SPL condition, split into three CF groups: low (blue), med (orange), and high (yellow). Trendline in

black. (E) Significant interaction between level and CF grouping in LMM, Q-value plotted as a function of level for three different CF groups. (F) Histogram showing percent of neurons that broaden, have no change, or sharpen in Q value split into CF groups. Low CF group has a higher number of neurons that are unchanging, whereas the high CF group shows more units that decrease (broaden) in Q.

To quantify these changes, neurons were divided into three categories: sharpening, no change, and broadening (Fig. 3.7). Sharpening as level increased occurred in 26% of neurons, often with an overall rate decrease as levels increased ($p=1.2245e-5$, Kruskal-Wallis, Fig. 3.7a). Neurons without large changes in Q made up 47% of neurons, and these rate profiles were mostly static as level increased ($p=0.4196$, Kruskal-Wallis, Fig. 3.7b). Lastly, 27% of neurons broadened as level increased ($p=8.3599e-5$, Kruskal-Wallis, Fig. 3.7c). The rate profile for synthetic timbre was robust over many suprathreshold levels, with almost half of the neurons having little change in Q as level increased.

Salience increased (sharpened) as a function of CF for all three sound levels tested (Fig. 3.7d). When neurons were divided into low, mid, and high CFs, there was a significant decrease in Q as level increased for medium and high CFs, whereas Q did not change over level for low CFs (Fig. 7e, $F(1, 555) =3.192$, $p=0.026$). When the proportions of broadening, sharpening, and unchanging neurons were divided into three categories based on CF, low CFs had a greater proportion of neurons that did not change in salience as level increased (Fig. 3.7f).

3.4.4. Neural thresholds based on rate changes explained human timbre-discrimination thresholds

Next, neural discrimination thresholds for changes in the peak frequency of synthetic-timbre stimuli were compared to human timbre-discrimination thresholds from Allen and Oxenham (2014). Timbre-discrimination thresholds in humans for a 200-Hz F0 and a 1200-Hz peak were 4% of the peak frequency for musicians and 5% for non-musicians (Allen and Oxenham, 2014). An example neuron with a CF = 1326 Hz had a discrimination threshold of 3.00%, lower than the behavioral thresholds (Fig. 3.8a). Across all neural threshold estimates in response to 63 dB SPL, diotic stimuli, 22.1% of neurons had thresholds lower than 4% (Fig. 3.8b). Neurons with a wide range of CFs (758-8000 Hz) had thresholds lower than 4% (Fig. 3.8c).

Next, changes in neural discrimination thresholds as sound level increased were calculated. The behavioral timbre-discrimination task has not been tested over level, but behavioral thresholds generally improve with level over a range of suprathreshold sound levels (Pickett, 1956; Studebaker et al., 1999; Heinz et al., 2001). Overall, there was no significant difference in thresholds at 43, 63, or 83 dB SPL ($p=0.0798$, Kruskal-Wallis, Fig 3.8d). Neurons were divided into three categories: Neurons with improved thresholds made up 46% of the population ($n=57/124$, $p=1.5792e-5$, Kruskal-Wallis), whereas neurons with worsening thresholds made up 34% ($n=42/124$, $p=2.8456e-7$, Kruskal-Wallis) and neurons with unchanging thresholds made up 20% ($n=25/140$, Fig. 3.8e, $p=0.9819$, Kruskal-Wallis). Many IC neurons had improved thresholds based on the rate profiles as levels increased, generally consistent with behavioral trends.

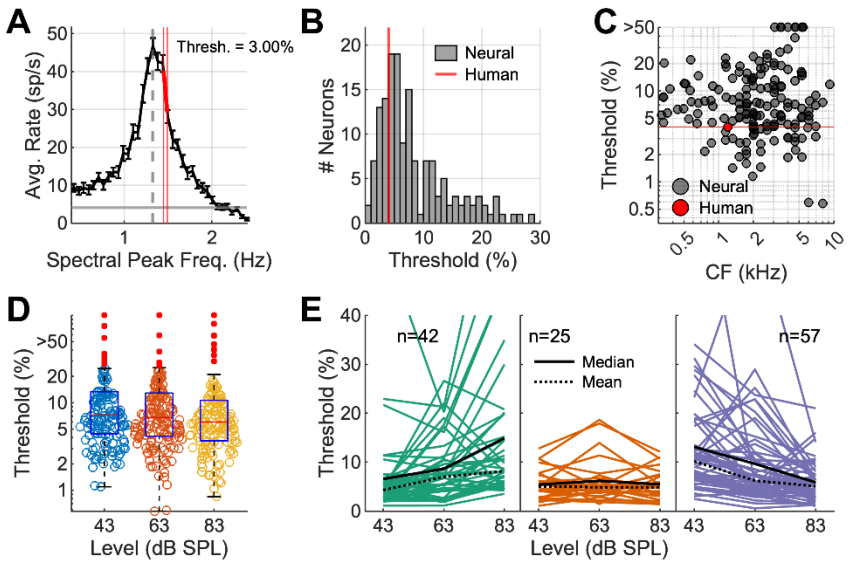

Figure 3.8. Thresholds for each neuron, for all sound levels and for diotic presentation. (A) Example neuron response, thick red line indicates the steepest slope that has a d'>= 1. Thin red vertical lines indicate the spectral peak frequencies used for threshold calculation. $Threshold = \frac{\Delta f}{f_{mid}} \times 100$, where Δf is the smallest difference in spectral peak frequency for which d' = 1 based on a difference in rate. (B) Histogram of thresholds, where red line indicates human threshold at 1200 Hz for musicians (Allen & Oxenham 2014). (C) Thresholds as a function of CF. Human threshold (red circle) shown for reference. (D) Plot of all thresholds for diotic presentation as a function of level. No significant differences between groups. (E) For the subset of neurons for which responses to all three levels (43-, 63-, and 83-dB SPL) were available, split into three categories: increasing (worsening) threshold, steady thresholds, and decreasing (improving) thresholds.

3.4.5. IC models predicted rate profiles at one level, but failed to capturing trends over level

Three IC models were used to predict the rate profiles in responses to synthetic timbre: energy, SFIE, and broad-inhibition models. These models were run for each

neuron with a BE or BS MTF for the 63-dB-SPL, diotic condition. For 28% of neurons, usually with BS MTFs, all three models had an R^2 (calculated between each model and neuron response) above 0.4 (n=36/127, Fig. 3.9a). Alternatively, for 35% of neurons, often with BS MTFs and higher CFs, all models had R^2 values below 0.4 (n=45/127, Fig. 3.9b). The broad-inhibition model predictions were better than the other models for 10% of neurons (n=13/127 Fig. 3.9c). For neurons with complex responses, broad-inhibition and SFIE models outperformed the energy predictions (Fig. 3.9d). The SFIE model incorrectly predicted a 'dip' in rate near CF for BE neurons (Fig. 3.9e, f). Energy and broad-inhibition models predicted rate profiles well for BE neurons with peaked responses (Fig. 3.9e), but all models failed to predict sloping responses (Fig. 3.9f). Variance explained for each model and each neuron example is detailed in Table 1. There was a large heterogeneity across neural responses, which could not be fully predicted by any of the IC models tested.

The correlation, R, of a model prediction and the neural data for each neuron was used to compare model performance. Neurons were separated into BE and BS MTFs, and predictions were compared for SFIE and energy models for the 63-dB-SPL condition (Fig. 3.9g). SFIE and energy predictions for BS neurons were relatively accurate; however, the SFIE model failed to predict BE responses (Fig. 3.9g). Next, the broad-inhibition and energy models were compared (Fig. 3.9h). Both of these models performed well, with energy having a slightly higher R overall (mean $R_{\text{broad inhibition}} = 0.6008$, mean $R_{\text{energy}} = 0.4847$). However, the broad-inhibition model also predicted MTFs and

responses to tones in narrowband and wideband noise, unlike the energy model (Fritzinger & Carney, 2025).

Period histograms from the energy, SFIE, and broad-inhibition models were also compared to neural data (Fig. 3.5, Fig. B2). SFIE-model period histograms exhibited phase-locking away from the spectral peak, and a lack of phase-locking at the spectral peak (Fig. B2a). The broad-inhibition (Fig. B2b) and energy (Fig. B2c) models predicted trends in rate profiles that were similar trends to those of the SFIE model. None of these models had the complexity in temporal responses seen in the physiological data.

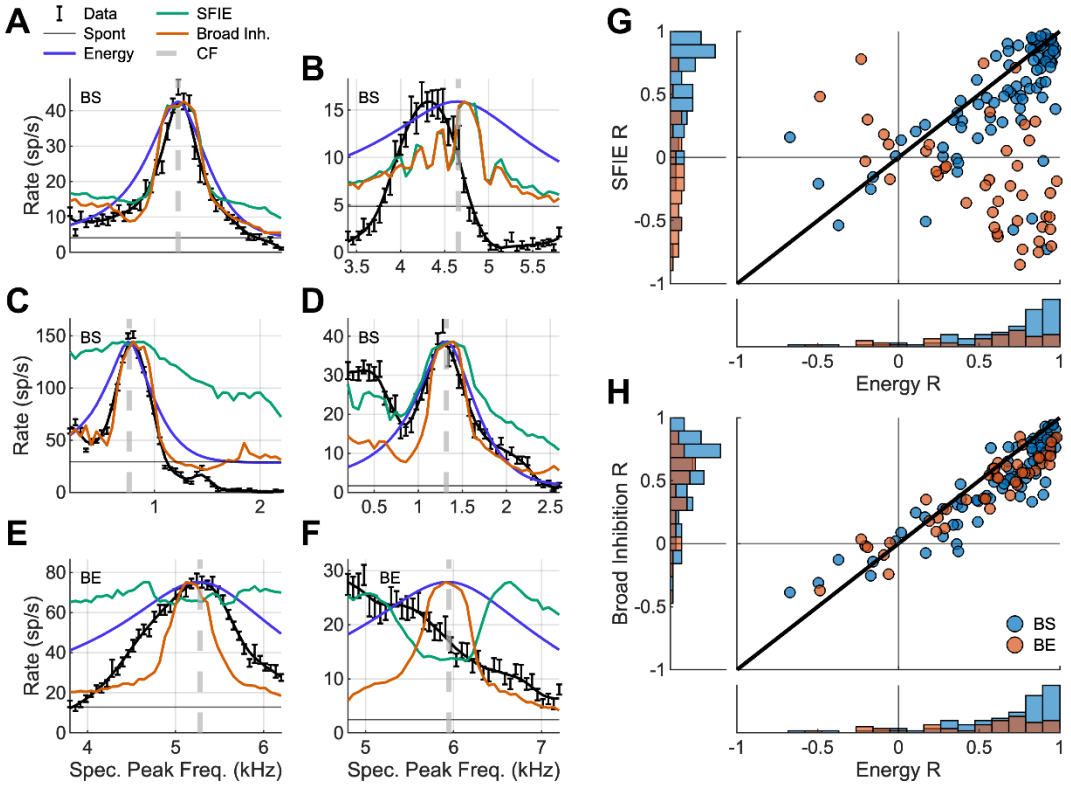

Figure 3.9. Rate profiles for six example neurons at 63-dB SPL and model predictions for energy, SFIE, and broad-inhibition models. (A) BS neuron with responses that are predicted well by models. (B) BS neuron with poor model predictions. (C) BS neuron with inhibitory area consistent with broad-inhibition model. (D) BS neuron with 'dip' in

rate off-CF, predicted by SFIE and broad inhibition models. (E) BE neuron with acceptable predictions from energy and broad inhibition. (F) BE neuron with poor model predictions. (G) Energy and SFIE model correlations to neural data for BE (red) and BS (blue) neurons, with histograms showing distribution for BE and BS neurons. (H) Same as G, but for broad-inhibition and energy model comparison.

Neuron	SFIE R^2	Energy R^2	Broad Inhibition R^2
A	0.89	0.90	0.89
B	0.19	0.35	0.22
C	0.71	0.92	0.89
D	0.71	0.47	0.59
E	0.04	0.96	0.72
F	0.02	0.06	0.04

Table 3.1. Variance explained by the three model types for the six neuron examples in Figure 8.

Changes in Q as sound level increased were calculated for model predictions of each neuron and were compared with trends in the neural data. First, the relationship of Q to CF in the neural data (Fig. 3.7d) and model predictions were compared (Fig. 3.10a). The SFIE and broad-inhibition model predictions had steeper slopes than the neural data and energy model (Fig. 3.10a). Model responses for each neuron were classified into broadening, sharpening, or unchanging categories (Fig. 3.10b). The energy model predicted no change in Q vs. level, whereas some SFIE and broad-inhibition model neurons predicted broadened rate-profile peaks as level increased. No model matched the distribution of sharpening, broadening, and unchanging neurons in the data: the energy model overestimated the number of unchanged neurons, and the SFIE and broad-inhibition models overestimated the number of broadened neurons.

Last, we calculated the changes in neural discrimination threshold as level increased for each of the models. The majority of all three models fit to the neurons had unchanging thresholds vs. level, whereas the data had more neurons that decreased (improved) in threshold (Fig. 3.10c). These results emphasize that although rates can be predicted at one sound level with good accuracy, changes over level in the model responses did not match the trends in the data, and improvements to the models are needed.

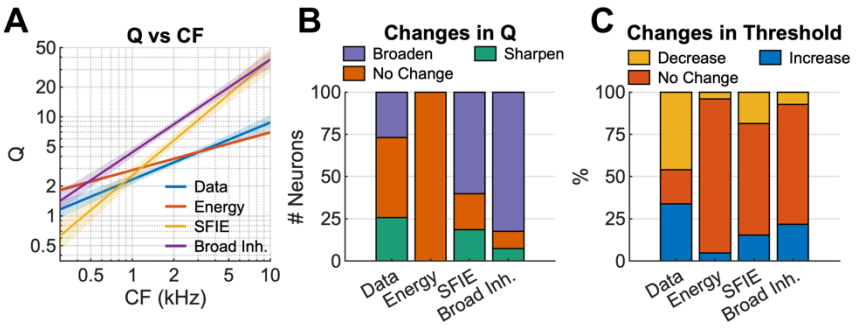

Figure 3.10. Evaluations of Q vs. CF and trends in Q and threshold over sound level for energy, SFIE, and broad inhibition models. (A) Salience, Q, trendlines for data and models as a function of CF. (B) Broadening, sharpening, and unchanging Q calculated as level increases for data and three models. (C) Changes in neural discrimination thresholds for data and three models as level increases.

3.4.6. Spectral-receptive-field models with broad inhibition predicted rate profiles

Spectral-receptive-field models, based on a gaussian or a DoG, were fit to neural rate profiles. The DoG model's addition of an inhibitory gaussian led to better predictions of rate profiles (Fig. 3.11). An example neuron and the DoG and gaussian fits to the rate

profile shows how the added inhibitory term in the DoG model allowed predictions below the spontaneous rate of the neuron, increasing model accuracy (Fig. 3.11a, b).

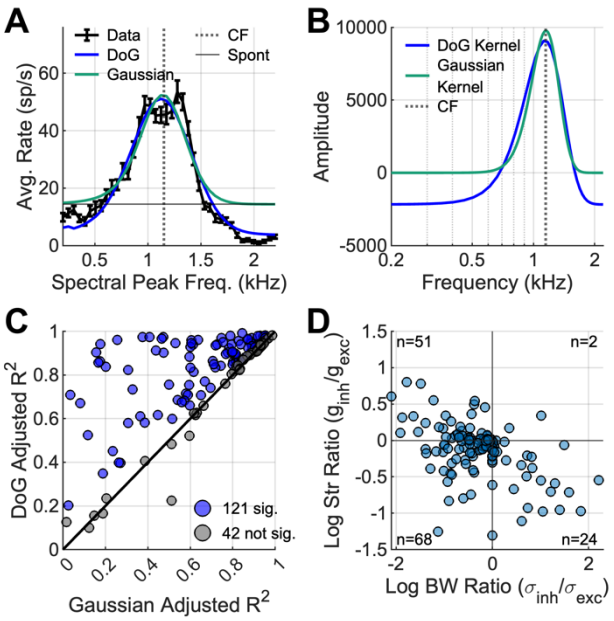

Figure 3.11. Gaussian vs difference of gaussian spectral receptive field model results to investigate importance of adding broad inhibition. (A) Example DoG and gaussian fit to an example neuron. (B) DoG and gaussian receptive field model kernel for example shown in A, kernel is multiplied and summed with the stimulus spectrum to obtain neural predictions seen in A. (C) Comparison of adjusted R^2 for gaussian and DoG, significant increases in DoG R^2 in blue. (D) Scatter plot of strength ratio vs. bandwidth ratio for fitted DoG parameters.

Seventy-four % of model fits were significantly improved by adding the inhibitory term (Fig. 3.11c). The shapes of the DoG model fits were analyzed by comparing the ratio of inhibitory and excitatory strength to the inhibitory and excitatory bandwidth parameters (Fig. 3.11d, Fritzinger & Carney, 2025). Fits in quadrant 4 (n=24) represent models with broad, but weak, inhibition (Fig. 3.11b). Fits in quadrant 2 and 3 (n=51, n=68, respectively) represent models with narrow inhibition, that either have a

single inhibitory sideband or have double peaks with a dip in the middle of the response (examples: Fig. 3.4b, Fig. 3.9d). Though 17.9% of DoG models featured broad inhibition (n=26/145), narrow inhibition improved fits for neurons with multiple peaks and single sidebands (n=119/145), emphasizing the heterogeneity in rate profiles and the general importance of inhibition.

3.5 Discussion

Spectral-peak encoding was investigated by recording single-neuron responses in the central nucleus of the IC in awake rabbits. Using a novel frequency-shifting paradigm, a population response was inferred from responses of a single neuron to a harmonic stimulus with a triangular spectral peak. Results support the hypothesis that spectral peaks in the rate profiles are sufficient to explain human thresholds in a timbre-discrimination task over a wide range of peak frequencies. Many neurons had robust rate profiles over several suprathreshold sound levels, and neural discrimination thresholds improved as level increased. An IC model with broad inhibition predicted rate responses, and the addition of inhibition to a spectral-receptive-field model improved rate predictions, highlighting the potential importance of inhibition for the robustness of responses across sound level.

Temporal responses to the synthetic stimuli were more complicated than initially predicted, leaving an open question about temporal coding of these stimuli in the IC. Interestingly, IC responses to a flat harmonic complex with F0 = 200 Hz had a mean vector strength to F0 of 0.5 (Su and Delgutte, 2019). The mean vector strength to F0 for

the synthetic timbre stimuli was approximately 0.25. This result suggests a possible difference in temporal code between the flat and triangular-shaped harmonic complexes. The mode-locked responses in the data also leave open questions about how and why mode locking arises in these IC responses. Mode-locked responses to periodic stimuli, such as AM tones, have been noted in the ventral cochlear nucleus (Mitchell and Carney, 2025), but are relatively understudied in the IC (Farokhniaee, 2016).

Another impact on these IC rate profiles may be subcortical efferent-system feedback. Medial olivocochlear neurons receive input from cochlear nucleus and IC neurons and project to outer hair cells across frequency channels (Brown, 2014). The rate representation in the IC to these timbre stimuli may be sharpened (increased Q) by efferent feedback (Carney, 2024). We found that the Q of rate profiles in response to synthetic-timbre stimuli increased, decreased, or remained unchanged over the time course of the stimuli in about equal proportions, and this effect did not depend on MTF type (Fig. B3). However, this result could be due to the 300-ms duration of the stimulus; a longer duration stimulus would be predicted to have larger changes in rate due to efferent effects.

Comparisons to vowel and timbre studies

An important consideration when studying the human percept of spectral timbre perception is whether model animals can perform spectral discrimination tasks. Though there have not been behavioral studies on timbre discrimination in rabbits, behavioral tasks discriminating vowel sounds have been successful in other model species (Bizley et

al., 2013, ferrets; Henry et al., 2017, budgerigar). A difference between humans and most laboratory animal models is the difference in cochlear-tuning bandwidths: humans have sharper cochlear tuning than rabbits, but interestingly the rabbit rate profiles had timbre-discrimination thresholds similar to those in humans, despite the increased tuning bandwidth.

Timbre results can be compared to IC neural responses to harmonic stimuli with triangular spectra meant to approximate spectral peaks, or formants, in vowels. These stimuli differ in the steepness of the spectral slope: often the vowel approximations use slopes of 200 dB/oct, whereas the slope in our stimuli was 24 dB/oct. Neural timbre-discrimination thresholds based on rate profiles were generally consistent with neural vowel-discrimination thresholds in quiet (Henry et al., 2017). Despite the broader bandwidth of timbre stimuli compared to the vowel approximations, spectral peaks in synthetic-timbre stimuli were still encoded via IC rate profiles in quiet. Henry et al. (2017) found neural discrimination thresholds based on timing were needed to explain behavioral thresholds in noise, and while we did not test timbre stimuli in noise, the timing information in the IC for these stimuli were rich and varied (Fig. 3.5).

Modeling

The broad-inhibition model was an improvement over the SFIE model and predicted rate profiles in neurons with both BE and BS MTFs (Fig 3.9). The spectral-receptive-field model with broad inhibition in the form of the difference of gaussians also predicted responses more accurately than an excitatory-only gaussian model (Fig. 3.11).

This finding is consistent with an implementation of a DoG spectral-receptive-field model for flat harmonic tone complexes (Su and Delgutte, 2020). Together, these modeling results suggest that broad inhibition may have a large impact on rate coding for harmonic stimuli.

Limitations of the current IC models include inaccurate period histograms and trends in rate responses over sound level. Many neurons mode-lock to F0, but the models do not exhibit mode locking except at very low fundamental frequencies (not shown). Twenty-seven percent of neurons (n=49/184) had decreased vector strengths and mode-locking when the peak harmonic was near CF, consistent with the models, but 73% of neurons (n=135/184) had responses that the models did not predict. Changes across sound levels in the models also did not match the trends in the data. For the SFIE and broad-inhibition models, most predictions had Qs that broadened as level increased, in part due to AN saturation occurring more broadly across a range of CFs as level increases. For all three models studied here, the majority of responses for models fit to neurons did not have level-dependent thresholds, whereas many neurons had thresholds that improved as level increased.

A new IC model is needed to accurately predict response timing and changes with level, as the limitations in the current models are not easily fixed. Mode-locking, for example, can be modeled with biological Hodgkin-Huxley models (Lee and Kim, 2006), leaky integrate-and-fire models (Keener et al., 1981; Laudanski et al., 2010), and Izhikevich models (Farokhniaee and Large, 2017). Another option for a new IC model is a conductance-based model (Gerstner, 2001). Current SFIE and broad-inhibition models

use two alpha functions to model excitation and inhibition, whereas conductance-based models could provide more flexibility in model structure.

Differences in synthetic and natural harmonic sounds with timbre

The synthetic-timbre stimulus is simplified compared to natural sounds with timbre, and encoding natural instrument spectra may be more complex. For example, the amount of energy at low harmonics differs: most instruments have strong fundamentals and low harmonics (for examples, see Siedenburg et al., 2021), whereas the synthetic-timbre stimulus used here had a weak or missing fundamental. This difference in energy at low frequencies may affect both temporal and rate coding, specifically for suprathreshold sounds. For example, AN fibers with higher CFs respond to low frequencies in the tail of the tuning curve at higher sound levels, which could alter the stimulus representation in these AN fibers and change the neural-fluctuation patterns (Carney, 2018). IC neurons, similarly, have response maps that are often V-shaped and broad at high sound levels, so low-frequency harmonics would alter IC firing rates. The phase-locking in the IC to the 200-Hz F0 could also be strengthened due to the higher amplitude fundamental in natural stimuli.

Another key difference between synthetic and natural stimuli is the phase of the harmonics. The synthetic-timbre stimuli had harmonics added in sine phase. Vowels and instrument sounds have harmonic components with phase delays based on the resonances in the vocal tract or instrument body (Wolfe et al., 2009). These phase changes result in fast frequency sweeps, called chirps. Neurons in the IC are selective for upward or

downward chirps, and thus chirps in natural sounds would also alter IC responses (Mitchell et al., 2023).

Implications of hearing loss

A motivation for studying spectral-peak and timbre encoding is hearing loss: the perception of the sound spectrum is altered by hearing loss and in cochlear-implant users (Gfeller et al., 2002; Emiroglu and Kollmeier, 2008; Kong et al., 2011, 2012; Chintanpalli et al., 2016). Perceptual studies in which listeners identify just-noticeable-differences between morphed instruments or identify vowels in concurrent-vowel experiments report deficits in instrument and vowel identification in listeners with hearing loss (Emiroglu and Kollmeier, 2008; Chintanpalli et al., 2016). Possible mechanisms related to perceptual deficits include loss of cochlear sensitivity due to outer-hair-cell damage and loss of inhibition due to aging (Caspary et al., 2008). Loss of cochlear sensitivity, for example, could lead to decreases in peak rates at the spectral peak of IC cells due to reduced contrast in NF profiles (Carney, 2018), whereas loss of inhibition could alter AM sensitivity and reduce broad-inhibition mechanisms for AM sensitivity. Future work studying IC encoding of spectral peaks after hearing loss is critical to understand how IC representations are degraded in hearing loss.

Future Directions / Conclusions

This study investigated IC encoding of a specific dimension of timbre, the spectral envelope, which is correlated to the brightness percept. More work is needed to understand spectral-peak encoding with additional complexities, such as those in natural

harmonic stimuli, and to understand how the auditory midbrain encodes other features of timbre, such as attack time. Overall, this study fills a gap in understanding timbre encoding in the auditory midbrain. Results illustrate that rate coding of spectral peaks in IC neurons can explain timbre-discrimination thresholds in humans.

Acknowledgements

This study was funded by NIH-R01-DC010813 and NIH-F31-DC020630-03. Thanks to Paul W. Mitchell and Swapna Agarwalla for assistance in data collection, and to Douglas Schwarz for assistance with hardware and software. Thanks to Braden Maxwell for initial modeling studies and feedback. Thanks to Kris Abrams for helping with the rabbit surgery and care.

Bibliography

Aihara K, Matsumoto G, Ikegaya Y (1984) Periodic and non-periodic responses of a periodically forced Hodgkin-Huxley oscillator. J Theor Biol 109:249–269.

Allen EJ, Oxenham AJ (2014) Symmetric interactions and interference between pitch and timbre. J Acoust Soc Am 135:1371–1379.

Andoni S, Li N, Pollak GD (2007) Spectrotemporal Receptive Fields in the Inferior Colliculus Revealing Selectivity for Spectral Motion in Conspecific Vocalizations. J Neurosci 27:4882–4893.

ANSI (1994) Am Stand Acoust Terminol R2004.

Bizley JK, Walker KMM, Nodal FR, King AJ, Schnupp JWH (2013) Auditory Cortex Represents Both Pitch Judgments and the Corresponding Acoustic Cues. Curr Biol 23:620–625.

Brown MC (2014) Single-unit labeling of medial olivocochlear neurons: the cochlear frequency map for efferent axons. J Neurophysiol 111:2177–2186.

Carney LH (2018) Supra-Threshold Hearing and Fluctuation Profiles: Implications for Sensorineural and Hidden Hearing Loss. J Assoc Res Otolaryngol 19:331–352.

Carney LH (2024) Neural Fluctuation Contrast as a Code for Complex Sounds: The Role and Control of Peripheral Nonlinearities. Hear Res 443:108966.

Carney LH, Li T, McDonough JM (2015) Speech Coding in the Brain: Representation of Vowel Formants by Midbrain Neurons Tuned to Sound Fluctuations. eneuro 2:ENEURO.0004-15.2015.

Carney LH, McDonough JM (2019) Nonlinear auditory models yield new insights into representations of vowels. Atten Percept Psychophys 81:1034–1046.

Caspary DM, Ling L, Turner JG, Hughes LF (2008) Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. J Exp Biol 211:1781–1791.

Chintanpalli A, Ahlstrom JB, Dubno JR (2016) Effects of age and hearing loss on concurrent vowel identification. J Acoust Soc Am 140:4142–4153.

Emiroglu S, Kollmeier B (2008) Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. Brain Res 1220:199–207.

Farokhniaee A (2016) Simulation and Analysis of Gradient Frequency Neural Networks. Farokhniaee A, Large EW (2017) Mode-locking behavior of Izhikevich neurons under periodic external forcing. Phys Rev E 95:062414.

Fritzinger, J. B., & Carney, L. H. (2025). Mechanisms of Tone-in-Noise Encoding in the Inferior Colliculus. Journal of Neuroscience, 45(23).

Gerstner W (2001) A framework for spiking neuron models: The spike response model. In: Handbook of Biological Physics (Moss F, Gielen S, eds), pp 469–516 Neuro-Informatics and Neural Modelling. North-Holland.

Gfeller K, Witt S, Mehr MA, Woodworth G, Knutson J (2002) Effects of Frequency, Instrumental Family, and Cochlear Implant Type on Timbre Recognition and Appraisal. Ann Otol Rhinol Laryngol 111:349–356.

Grey JM, Gordon JW (1978) Perceptual effects of spectral modifications on musical timbres. J Acoust Soc Am 63:1493–1500.

Heinz MG, Colburn HS, Carney LH (2001) Evaluating Auditory Performance Limits: I. One-Parameter Discrimination Using a Computational Model for the Auditory Nerve. Neural Comput 13:2273–2316.

Henry KS, Abrams KS, Forst J, Mender MJ, Neilans EG, Idrobo F, Carney LH (2017) Midbrain Synchrony to Envelope Structure Supports Behavioral Sensitivity to Single-Formant Vowel-Like Sounds in Noise. J Assoc Res Otolaryngol 18:165–181.

Iverson P, Krumhansl CL (1993) Isolating the dynamic attributes of musical timbrea). J Acoust Soc Am 94:2595–2603.

Keener JP, Hoppensteadt FC, Rinzel J (1981) Integrate-and-Fire Models of Nerve Membrane Response to Oscillatory Input. SIAM J Appl Math 41:503–517.

Kim DO, Carney L, Kuwada S (2020) Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. J Neurophysiol 124:1198–1215.

Kong Y-Y, Mullangi A, Marozeau J (2012) Timbre and Speech Perception in Bimodal and Bilateral Cochlear-Implant Listeners. Ear Hear 33:645–659.

Kong Y-Y, Mullangi A, Marozeau J, Epstein M (2011) Temporal and Spectral Cues for Musical Timbre Perception in Electric Hearing. J Speech Lang Hear Res 54:981–994.

Laudanski J, Coombes S, Palmer AR, Sumner CJ (2010) Mode-Locked Spike Trains in Responses of Ventral Cochlear Nucleus Chopper and Onset Neurons to Periodic Stimuli. J Neurophysiol 103:1226–1237.

Lee S-G, Kim S (2006) Bifurcation analysis of mode-locking structure in a Hodgkin-Huxley neuron under sinusoidal current. Phys Rev E 73:041924.

Lyzenga J, Horst JW (1995) Frequency discrimination of bandlimited harmonic complexes related to vowel formants. J Acoust Soc Am 98:1943–1955.

Mao J, Carney LH (2015) Tone-in-Noise Detection Using Envelope Cues: Comparison of Signal-Processing-Based and Physiological Models. J Assoc Res Otolaryngol 16:121–133.

Maxwell BN, Fritzinger JB, Carney LH (2020) Neural Mechanisms for Timbre: Spectral-Centroid Discrimination based on a Model of Midbrain Neurons. In, pp 4.

McAdams S, Winsberg S, Donnadieu S, De Soete G, Krimphoff J (1995) Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychol Res 58:177–192.

Mitchell, P. W., & Carney, L. H. (2025). Chirp sensitivity and vowel coding in the inferior colliculus. Hearing Research, 109307.

Mitchell PW, Henry KS, Carney LH (2023) Sensitivity to direction and velocity of fast frequency chirps in the inferior colliculus of awake rabbit. Hear Res 440:108915.

Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J Acoust Soc Am 116:2173–2186.

Perez CA, Engineer CT, Jakkamsetti V, Carraway RS, Perry MS, Kilgard MP (2013) Different Timescales for the Neural Coding of Consonant and Vowel Sounds. Cereb Cortex 23:670–683.

Pickett JM (1956) Effects of Vocal Force on the Intelligibility of Speech Sounds. J Acoust Soc Am 28:902–905.

Pollak GD (2011) Discriminating among complex signals: the roles of inhibition for creating response selectivities. J Comp Physiol A 197:625–640.

Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Neural Comput 16:1661–1687.

Rees A, Langner G (2005) Temporal Coding in the Auditory Midbrain. In: The Inferior Colliculus (Winer JA, Schreiner CE, eds), pp 346–376. New York, NY: Springer New York. Available at: https://link.springer.com/10.1007/0-387-27083-3_12 [Accessed April 22, 2025].

Schwarz DM, Zilany MSA, Skevington M, Huang NJ, Flynn BC, Carney LH (2012) Semi-supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. J Neurosci Methods 206:120–131.

Siedenburg K, Jacobsen S, Reuter C (2021) Spectral envelope position and shape in sustained musical instrument sounds. J Acoust Soc Am 149:3715–3726.

Studebaker GA, Sherbecoe RL, McDaniel DM, Gwaltney CA (1999) Monosyllabic word recognition at higher-than-normal speech and noise levels. J Acoust Soc Am 105:2431–2444.

Su Y, Delgutte B (2019) Pitch of harmonic complex tones: rate and temporal coding of envelope repetition rate in inferior colliculus of unanesthetized rabbits. J Neurophysiol 122:2468–2485.

Su Y, Delgutte B (2020) Robust Rate-Place Coding of Resolved Components in Harmonic and Inharmonic Complex Tones in Auditory Midbrain. J Neurosci 40:2080–2093.

Tan Q, Carney LH (2005) Encoding of vowel-like sounds in the auditory nerve: Model predictions of discrimination performance. J Acoust Soc Am 117:1210–1222.

Whitehead ML, Lonsbury-Martin BL, Martin GK (1992) Evidence for two discrete sources of 2 f1−f2 distortion-product otoacoustic emission in rabbit: I. Differential dependence on stimulus parameters. J Acoust Soc Am 91:1587–1607.

Williams AJ, Fuzessery ZM (2011) Differential roles of GABAergic and glycinergic input on FM selectivity in the inferior colliculus of the pallid bat. J Neurophysiol 106:2523–2535.

Wolfe J, Garnier M, Smith J (2009) Vocal tract resonances in speech, singing, and playing musical instruments. HFSP J 3:6–23.

Zilany MSA, Bruce IC, Carney LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. J Acoust Soc Am 135:283–286

Chapter 4: Representations of Pitch and Timbre of Instrument Sounds in the Inferior Colliculus

4.1 Abstract

The neural encoding of pitch and timbre in complex sounds has previously been studied using synthetic, controlled stimuli to investigate mechanisms of encoding. These studies provide information about how single attributes of sound are represented in the inferior colliculus (IC). However, there is a gap in understanding how natural sounds with both pitch and timbre attributes, such as instrument sounds, are represented in the IC. In this study, extracellular recordings were made of single-neuron responses to natural instrument stimuli varying in fundamental frequency (F0) to determine how instrument identity (timbre) and F0 (pitch) are represented in IC neurons. We found that instrument identity was encoded in the responses of a small population of neurons with diverse rate and timing characteristics, and that information was redundant. Predictions of F0 using single-neuron rate responses was poor; but the population of rate responses contained enough information to decode F0 reliably. F0 information was also encoded in single-neuron temporal responses up to 196 Hz. Decoding F0 from a population of temporal responses was accurate up to approximately 900 Hz, but accuracy decreased at high F0s. For the task where F0 was identified in both oboe and bassoon in the overlapping F0 conditions, performance in decreased compared to single-instrument F0 decoding. This result supports the hypothesis that pitch/timbre information are encoded together in the

IC, possibly explaining why changes in pitch interfere with timbre discrimination tasks in psychophysics, and vice versa.

4.2 Introduction

Pitch and timbre are critical aspects of speech and music. Previous studies used synthetic sounds to explore how neurons in the inferior colliculus (IC) encode these acoustic features. However, how IC neurons respond to natural instrument sounds, which have more spectral peaks and phase differences than the synthetic stimuli used in previous studies, is unknown. This study aimed to fill this gap by recording IC responses in awake rabbit to bassoon and oboe stimuli with varying fundamental frequencies (F0). Bassoon and oboe were chosen due to the shape of the spectra; the spectral envelope was more similar to the triangular envelope used in previous synthetic studies than many other instruments. This study determined how neurons in the IC encode F0, instrument identity, and their interactions.

Studies using synthetic sounds to investigate one characteristic, F0 or timbre, provide the basis for initial hypotheses about natural instrument encoding. F0 encoding has been studied in the IC of awake rabbits using flat-spectrum, harmonic-tone complexes (HTCs) (Su and Delgutte, 2019, 2020). Behaviorally, rabbits discriminate F0 using temporal envelope cues at low F0s and spectral cues of resolved harmonics at high F0s (Wagner et al., 2022). Analysis of IC responses to these stimuli support encoding of F0 based on average rate (for F0s > 800 Hz), phase-locking to envelope (F0s < 900 Hz),

and, for neurons excited by amplitude modulations, by band-pass rate tuning to envelope repetition rate (200 – 1600 Hz) (Su and Delgutte, 2019, 2020).

Timbre- and vowel-encoding studies in the IC have also used synthetic stimuli to investigate how spectral peaks are represented in the IC. HTCs with broad triangular spectral peaks and zero-phase spectra are encoded in the average discharge rates of neurons (Fritzinger and Carney, in review). Similarly, formant discrimination based on average-rate information in budgerigar IC is sufficient to describe behavioral formant discrimination in quiet, though timing information was necessary when background noise was added (Henry et al., 2017). Timing information was also necessary for single-neuron identification of spoken vowels, and vowel accuracy was correlated with neural sensitivity to low-velocity chirps (Mitchell and Carney, 2025). Another study found that IC models with amplitude-modulation sensitivity accurately predicted some neural responses to spoken vowels (Carney et al., 2015). These studies were limited to F0s near the voice-pitch range; higher F0s may be encoded by different mechanisms.

By using natural instrument sounds of varying F0s, this study was also designed to explore an open question about pitch and timbre interactions. Psychophysical studies using synthetic stimuli (Allen and Oxenham, 2014) and natural vowel and instrument stimuli (McPherson and McDermott, 2023) have shown that pitch variation impacts timbre discrimination performance, and vice versa. Modeling work has proposed that this interaction may be present in the IC (Maxwell et al., 2020). The spectral centroid, a measure of timbral brightness, varies with F0 within a single instrument (Siedenburg et al., 2021), thus complicating pitch and timbre separability in natural sounds.

In this study, we hypothesized that rate information is sufficient to describe instrument identity in a population of IC neurons, based on previous timbre-discrimination results in quiet (Fritzinger and Carney, in review). For F0 identification, we hypothesized that timing information is necessary for F0s < 900 Hz, but that rate-based decoders would be accurate at high F0s, in part due to changes in spectral centroid for different F0s. We hypothesized that neurons with sensitivity to amplitude modulation could encode F0 and instrument identity. Previous cortical studies have shown that pitch and timbre information of vowels are often represented in the same neurons (Bizley et al., 2009). It is unknown whether this interaction occurs in the IC, and our study aimed to investigate that question as well.

Single-neuron responses were recorded in the central nucleus of the IC in awake rabbits to bassoon and oboe sounds with F0s ranging from 58 to 1661 Hz. We started with an analysis of the variability in these stimuli. Then we investigated whether a single neuron, or a population of neurons, could decode instrument identity and/or F0 using rate or timing information. We used classification and linear regression models to study information present in the IC, determine which IC neuron types are necessary for identification, and investigate overlap in neurons that discriminate pitch and instrument identity. We found that a small population of neurons with diverse characteristics was sufficient to encode instrument identity, and both rate and timing information in populations of IC units could encode F0. Often, single neurons that contained information about lower F0s also contained information about instrument identity.

4.3 Methods

Animal Care and Procedures

Extracellular, single-unit recordings were made in the central nucleus of the IC in four female Dutch-belted rabbits from 5 months to 5 years of age. All methods were approved by the University of Rochester Committee on Animal Resources. Surgical procedures consisted of an initial headbar placement, craniotomy and microdrive placement, and subsequent microdrive replacements. Animals were anesthetized intramuscularly with 66 mg/kg ketamine and 2 mg/kg xylazine, or 35mg/kg ketamine and 0.10-0.15mg/kg dexmedetomidine for all procedures. An initial surgery was performed to affix a custom, plastic, 3D-printed headbar (ProtoLabs, Maple Plain, MN) to the skull using screws and dental acrylic. One month after the initial surgery, another surgery was performed to create the craniotomy and insert the microdrive into the headbar. Subsequent surgeries were performed every 3-6 months to replace the microdrive and change position to sample different areas of the IC. Normal hearing was monitored using distortion-product otoacoustic emissions (Whitehead et al., 1992). When emissions dropped by 10 dB, experiments were discontinued, and animals were perfused to extract the brain. The IC tissue was stained with Nissl to confirm tetrode location in the IC.

Physiological Recordings

Rabbits were trained to sit in a custom chair with their heads fixed using the headbar. They were placed inside a sound-attenuated booth (Acoustic Systems, Austin, Texas, USA). Daily recording sessions lasted up to two hours.

Calibration tones and stimuli were created in MATLAB and send to an audio interface (16A, Mark of the Unicorn, Cambridge, Massachusetts, USA). The stimuli were converted from digital to analog (DAC3 HGC, Benchmark Media Systems, Inc., Syracuse, New York, USA) and presented through a speaker (Beyerdynamic DT-48, Beyerdynamic GmbH and Co., Heilbronn, Germany or Etymotic ER2, Etymotic Research, Inc., Elk Grove Village, Illinois).

At the beginning of each session, custom earmolds (Hal-Hen Company, Inc., Garden City Park, NY, USA or Dreve Otoform Ak, Unna, Germany) were inserted and the frequency response of the acoustic system was calibrated. Tones of 50 Hz-20 kHz were presented and a probe-tube microphone (Etymotic ER10B+ or ER7C, Etymotic Research, Inc., Elk Grove Village, Illinois) was used to record in the ear canal. All stimuli were filtered to compensate for the frequency response of the system.

Neural signals were recorded using a microdrive that contained four tetrodes. Tetrodes consisted of four twisted 18-μm epoxy-coated platinum-iridium wires (California Fine Wire Co., Grover Beach, CA) plated with a platinum-black solution (Neuralynx, Inc., Bozeman, MT) to lower impedances to approximately 0.1-1.5 MOhms. The tetrodes were advanced or retracted at the end the sessions using the microdrive to sample different areas of the IC. Signals from the tetrodes were amplified (RHD2216 16-channel amplifier chip) and recorded using the Intan RHD recording system (Intan Technologies, LLC., Los Angeles, CA, USA) and accompanying software. Signals were recorded at 30-kHz sampling rate and filtered with a 4th-order Butterworth 300 Hz- 3 kHz filter.

After each recording session, signals were sorted using custom MATLAB code to identify single neurons. The thresholds for each wire were set at 4 times the standard deviation of the waveform (Quiroga et al., 2004). Waveforms were sorted into individual neurons using custom software (Schwarz et al., 2012). Waveforms had to meet 2 criteria to be considered single-neuron recordings: 1) the percent of spikes that occurred with intervals less than 1 ms was less than 2%, and a cluster entanglement metric had to be less than 0.1 (Schwarz et al., 2012). Spike times were extracted and saved.

Stimuli & Analysis

Several basic stimuli were included to characterize each neuron. First, pure tones from 250 – 16,000 Hz were presented in 5 steps per octave to classify the CF and response map of each neuron. Tones were presented at 10, 30, 50 and 70 dB SPL with a 200-ms duration and 10-ms raised-cosine on/off ramps. Stimuli were presented diotically in random order for three repetitions with 400-ms interstimulus intervals. The CF of the neuron was defined as the frequency that elicited an increase in average rate at the lowest sound level.

Modulation transfer functions (MTFs) in response to amplitude-modulated (AM) stimuli were characterized by presenting sinusoidally amplitude-modulated gaussian noise. The noise bandwidth was 100 Hz – 10 kHz, modulation frequencies spanned 2-600 Hz in 3 steps per octave. AM stimuli were presented at 33 dB SPL spectrum level, 1-s duration with 50-ms raised-cosine on/off ramps, and 500-ms interstimulus intervals. The stimuli were repeated 5 times and were presented diotically in a random order. Average-rate responses were calculated, excluding a 50-ms onset response. Neurons were

classified into 5 types based on the comparison of responses to modulated and unmodulated noises: band-enhanced (BE), band-suppressed (BS), hybrid (BE-BS), hybrid (BS-BE), and flat. BE neurons had at least 2 modulation-frequency responses with rates significantly higher than the unmodulated rate (two-sample t-test, $p<0.05$), without a response at an intermediate modulation frequency that was significantly below the unmodulated rate. BS neurons had at least 2 modulation-frequency response rates significantly below the unmodulated rate, again without a significantly higher response rate at an intermediate modulation frequency. Hybrid neurons met both BE and BS criteria. H (BE-BS) neurons were BE at low modulation frequencies and BS at higher modulation frequencies, whereas H (BS-BE) were the opposite. Flat neurons did not meet either BE or BS criteria. The best modulation frequency (BMF) and worse modulation frequency (WMF) was calculated from the MTFs of BE and BS neurons, respectively. The BMF/WMF was the modulation frequency with the maximum/minimum of the spline-interpolated average rate response.

Neuron sensitivity to fast frequency sweeps, known as chirps, was characterized by rate-velocity functions (RVFs, Mitchell et al., 2023; Mitchell and Carney, 2025) based on stimuli that consisted of chirps with velocities of 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 4, 5, 6, 7, 8, 9 kHz/ms in both the positive and negative directions. Stimuli were normalized by energy and included raised-cosine on/off ramps that were 10% of the duration of each chirp. Chirp duration varied based on velocity. Chirps were presented in random order, diotically, with a random interstimulus interval of 40-60 ms. Each chirp was presented 80 times. Response rates were based on sums over a 15-ms time window

that started at an estimate of neural latency, defined as the latency for tones at CF at 73 dB SPL. Principal components analysis was used to extract prominent features of the chirp responses, with the first three principal components matching previous work (Mitchell and Carney, 2025).

The natural stimuli consisted of samples of musical instrument sounds (University of Iowa Electronic Music Studios, Iowa City, IA). A 300-ms duration, steady-state segment of the sound was extracted and 20-ms raised cosine ramps were added to the beginning and end of the stimulus. Stimuli were presented diotically at 73 dB SPL, in random order, for 20 repetitions. The bassoon stimulus set included all semitones from Bb1 (58 Hz) to D5 (587 Hz), and the oboe stimulus set included all semitones from Bb3 (233 Hz) to Ab6 (1661 Hz). Sixteen of the bassoon stimuli had F0s that overlapped with the oboe.

The spectral centroid of each stimulus was calculated to quantify changes over F0. To calculate the spectral centroid of a stimulus, the fast-Fourier transform was computed with a resolution of 3.33 Hz. The spectral centroid in Hz was calculated as:

$$Centroid = 10^{\frac{\sum log(f_i)x_i}{\sum x_i}},$$

where x_i is the log amplitude of the frequency spectrum at bin i, and f_i is the logarithmically transformed frequency of that bin. The F0 values of the stimuli were also calculated from the stimulus spectrum, as the labeled F0 from the recordings were not 100% accurate. F0 was estimated first by finding the frequencies of the stimulus harmonics using Welch's power spectral density estimate and finding the peaks of the spectrum. Then, an initial estimate of F0 was defined as the first spectral peak. The

estimate was then refined iteratively by taking the difference between the next harmonic and previous harmonic and averaging that difference with the current F0 estimate. This process continued for harmonics up to 10 kHz.

Neural responses to natural-timbre stimuli were analyzed by calculating rate and timing metrics. Average-rate responses were computed over the duration of the stimulus, averaged over all repetitions, excluding a 50-ms onset response. Additionally, classification analyses required calculating average rate per repetition of the stimulus. Temporal analyses included calculating peri-stimulus time histograms (PSTHs), period histograms, inter-spike intervals, vector strength, and reliability. Vector strength (VS) was calculated as:

$$VS = \left| \sum_{k=1}^{N} e^{2\pi i \frac{t_k \bmod T}{T}} \right| \quad ,$$

where t_k are spike times for all repetitions, excluding the 50 ms onset, and T is the period of the harmonic of interest. The time of each spike within a period is by calculating spike times modulo the period. Vector strength to F0 and individual harmonics were calculated. Significance of vector strength was determined using a standard Rayleigh test ($p<0.01$).

A reliability metric was used to determine consistency of spiking activity of a neuron over many repetitions (Gai and Carney, 2006). Reliability was defined as the correlation coefficient between the PSTHs of even and odd repetitions of the stimulus.

Classification Tasks

Three different decoding tasks were performed using various classification analyses and decoding models. The first task was identifying instrument identity, the

second task was identifying F0 of either oboe or bassoon, and the third task involved identifying both instrument identity and F0 of the stimulus. The instrument-identification task included both oboe and bassoon stimuli that overlapped in F0, which accounted for 16 stimuli with F0s ranging from 233 to 587 Hz. The goal of the analysis was to correctly identify the instrument identity regardless of F0. The F0-identification task was split into identifying F0 based on responses to the bassoon stimuli (40 stimuli), identifying F0 based on responses to the oboe stimuli (35 stimuli), or identifying F0 in the overlapping F0 range using responses to both bassoon and oboe (16 stimuli). Two other F0 identification tasks were performed: 1) training on a subset of F0s and predicting the held-out F0s, and 2) training on 16 overlapped oboe stimuli and testing on the 16 bassoon stimuli, and vise versa. Lastly, the instrument-identity-and-F0-discrimination task used responses to all 75 stimuli, and the goal was to determine if neurons or populations of neurons could correctly identify both instrument and F0. All tasks used the same types of analysis and models, presented below.

Single-Neuron Classification Analyses

First, a simple rate metric was used to determine how well a single-neuron rate response could classify either instrument identity, F0, or both (Mitchell and Carney, 2025). For the F0 identification, the average-rate response for a single repetition of one F0 stimulus was calculated. Then, the average-rate response across the other 19 repetitions of the stimulus and across the 20 repetitions of all other F0 stimuli provided an array of overall-average rates for all F0 stimuli. The single repetition response was compared to the overall average rates. The F0 classification of the single average rate

response was determined as the stimulus with an overall-average rate closest to the single-repetition rate. This analysis was performed for all repetitions and all conditions and for all neurons. Instrument identity and F0/instrument identity rate discriminations were performed in the same manner. A confusion matrix was calculated based on the classifications. Overall accuracy was defined as the sum of the diagonal of the confusion matrix (correct responses) divided by the total number of responses. Accuracy was calculated for all classification tasks.

Next, classification models were used to test how well timing information from a neuron could classify instrument identity and/or F0. Model features included binned spike times from each trial, where the bin width was varied from 0.1 ms to 300 ms (full rate response) and a bin width was chosen that resulted in the highest accuracy. All models were trained on 80% of the data and tested on the other 20%. For these tasks, the training data was chosen as 16 random repetitions out of the 20 stimulus repetitions to ensure instruments were equally represented in the training data. Support vector machines (SVM) were used as the base learning algorithm. Five-fold cross validation was performed on the training data to provide robust estimates of model performance. The regularization strength, lambda, was chosen by systematically varying lambda and using the lambda value that resulted in the lowest mean squared error in the final model. Ridge regularization was used to prevent overfitting and improve model generalization. The convergence criterion was set to 0.0001 for stable model convergence. The instrument identification task used the MATLAB function fitclinear, the F0 task used fitcecoc.

Population Classification Analyses

Population analyses included decoders that used the average-rate responses or PSTH information to test rate vs timing predictions in a population of IC neurons. For all tasks, the models were trained on 80% of the data and tested on the remaining 20%, again stratified such that 16 reps of each stimulus were used in training. All decoders used the MATLAB function fitcecoc. The learning algorithm was SVM, using a multi-class strategy of error-correcting output codes to predict the classifications. Ridge regularization was used to prevent overfitting and improve generalization, and the optimal regularization strength (lambda) was optimized as mentioned previously. Five-fold cross validation was performed on the training data. An analysis of feature weights, called beta weights, was performed to investigate the contribution of each neuron to the model. Timing information in a population of neurons was used for the F0-decoding task and F0- and-instrument-identification task, using PSTHs from 1 – 40 neurons with a bin size of 1 ms to keep computation time down. This model was not used for the instrument identity task because rates were sufficient to accurately predict instrument. Permutation testing was performed to determine the importance of each neuron in the model.

Regression Analyses for the F0-Indentification Task

Lastly, linear regression models were tested for the F0-identification task. Regression models were fit to the single-neuron PSTH data (bin size of 0.25 ms), the population rate data, and the population PSTH data (bin size of 1 ms). All models were fit using the MATLAB function fitrlinear. An 80/20 train/test split was used for model evaluation. For the basic F0 discrimination tasks, the training/testing sets were stratified

such that for each set of 20 repetitions, 16 (80%) repetitions were used in training and 4 (20%) were used in testing. Least squares regression was selected as the learning algorithm. Ridge regression (L2) was applied to prevent the models from overfitting. The regularization strength, lambda, was chosen by systematically varying lambda and using the lambda value that resulted in the lowest mean squared error in the final model. The limited-memory Broyden-Fletcher-Goldfarb-Shanno solver was used for parameter estimation. A convergence criterion was set to 0.001 for the beta coefficients to maintain efficiency. Two other training/testing configurations were performed. The first task trained the linear regression models on 80% of the F0s and tested the model using the remaining 20% of the F0s. The F0s used for training/testing were randomly selected. The second task trained on the 16 overlapping F0s for one instrument (50% training data) and tested on the 16 overlapping F0s for the other instrument (50% testing data). The model was fit using the above methods.

IC Computational Models

Three computational models of IC neurons were evaluated to determine accuracy of each model. Detailed implementations of all three models can be found in (Fritzinger and Carney, 2025). The first model was an energy model consisting of a 4th order gammatone filterbank with cat Q10 tuning values. The other two models used an auditory-nerve model stage (Zilany et al., 2014). The second model was the same-frequency excitation-inhibition (SFIE) model, which consists of a VCN stage and IC BE and IC BS cell stage that accurately predict IC sensitivity to modulation (Nelson and Carney, 2004; Carney and McDonough, 2019). The last model is a modified SFIE model

that adds inhibition from two off-CF BS cells (Fritzinger and Carney, 2025) to simulate an AM-sensitive broad inhibition. This model also predicts BE and BS MTFs.

For the energy, SFIE, and broad inhibition+SFIE model, model predictions were made for each recorded neuron, with parameters matched for CF and MTF type. Model average rate responses were compared to data using a variance explained metric, R2. The amount of neural variance explained by the model rates was above chance if the R2 value was significantly above an R2 value where model rates were shuffled. For the SFIE model, a non-homogenous Poisson process was used as a spiking generator and spiking activity was calculated for each model neuron. SFIE spiking activity outputs were used as inputs to the F0 and instrument identification decoding models to analyze differences between model and data.

Statistics

Statistical tests included linear regression to determine relationships between model accuracies or model weights and features such as CF, second principal component of the RVF, phase-locking strengths, and spiking reliability. Kruskal-Wallis for non-parametric data and Mann-Whitney U-tests were also used to determine significant differences in accuracies due to PSTH bins or MTF groups. Variance explained, R2 was used in linear regression models to determine how well model predictions fit to data.

4.4 Results

We determined that bassoon and oboe stimuli had different spectral-peak locations, and the oboe had more energy in the higher harmonics (Fig. 4.1a) for all

overlapping F0s. We calculated the spectral centroid of each stimulus; for both oboe and bassoon, spectral centroid increased as a function of F0 (Fig. 4.1b). Oboe stimuli also had a higher spectral centroids compared to bassoon, eliciting a brighter timbre (Fig. 4.1b), as expected.

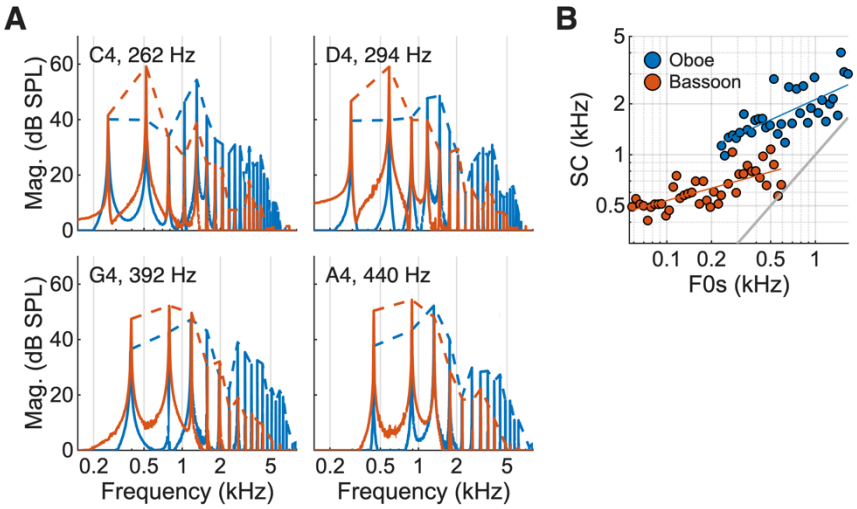

Figure 4.1. Characteristics of the natural-timbre stimuli. (A) Four examples of oboe and bassoon spectra with different F0s. (B) Spectral centroid plotted as a function of F0 for oboe and bassoon stimuli.

Single-neuron responses were recorded from a total of 298 neurons in the central nucleus of the IC. Responses to both bassoon and oboe stimuli were collected for 249 neurons. Bassoon responses only were recorded for an addition 41 neurons and oboe responses only were recorded for an additional 7 neurons. Response maps and MTFs were collected for all neurons, and RVFs were collected for 188 neurons. Neural CFs ranged from 332 Hz – 13.9 kHz (Fig. 4.2a). MTFs were analyzed for the population of responses: 22% (n=67) were BE, 49% (n=146) were BS, 18% (n=54) were hybrid, and

11% (n=31) were flat (Fig. 4.2b). The median BMF for BE neurons was 83 Hz, and the median WMF for BS neurons was 76 Hz (Fig. 4.2c, d).

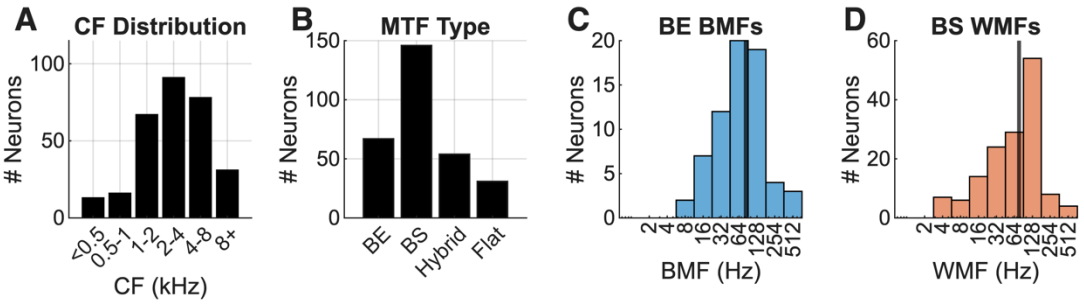

Figure 4.2. Data distribution, n = 298 neurons. (A) Distribution of CFs and (B) MTF types. (C) Distribution of best modulation frequencies (BMFs) for BE neurons, n=67. (D) Distribution of worst modulation frequencies (WMFs) for BS neurons, n=146.

Instrument Identification

Identification of instrument using individual-neuron rate and timing information.

The first task was decoding instrument identity over the range of overlapping F0s, 233 to 587 Hz, using rate or timing information from single neurons. Instrument-identification accuracy based on rate responses from single neurons had a mean average of 59% accuracy and ranged from 34 to 88%. The decoding accuracy of 16 neurons was below chance (50%) due to an unreliable classification boundary because of high variance in the data (Fig. 4.3D). Rate responses from 12 example neurons revealed that neurons with the highest accuracy had consistent rate differences over the range of overlapping F0s (Fig. 4.3a, b, c); in contrast, neurons exhibiting chance accuracy showed overlapping response rate (Fig. 4.3d). Identification accuracy did not correlate with CF or differ significantly across MTF type (Fig. 4.3e).

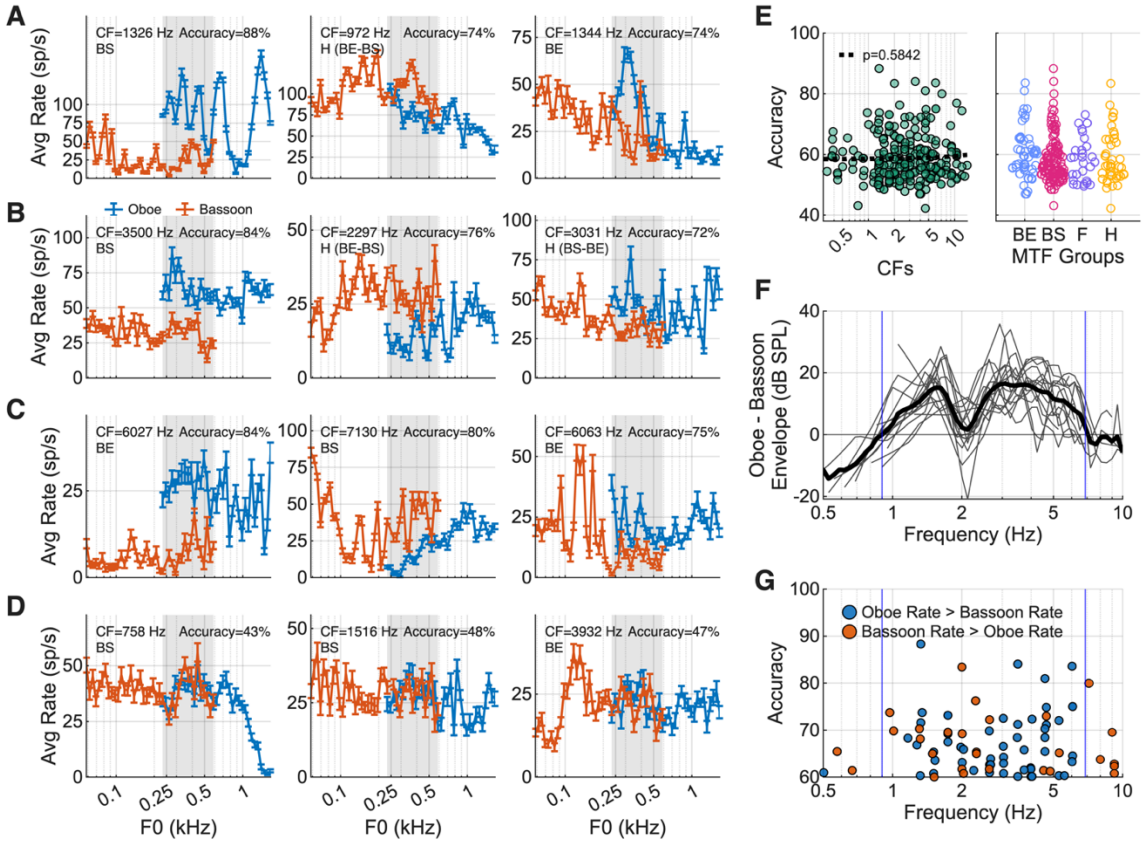

Figure 4.3. Example single-neuron average rate responses to oboe and bassoon and analysis of rate predictions of instrument identity. (A) Low CF, (B) medium CF, and (C) high CF neuron examples of responses to oboe and bassoon. Grey bar indicates the range of overlapping F0s between oboe and bassoon.. (D) Examples with poor instrument identification. (E) Accuracy of instrument identity classification for each neuron as a function of CF or classified by MTF type. (F) Differences between oboe and bassoon spectral envelopes, oboe – bassoon with zero crossings (blue). (G) Prediction accuracy as a function of CF split into two groups, oboe rate greater than bassoon (blue), and bassoon rate greater than oboe (orange).

We calculated the difference in the spectral envelopes for bassoon and oboe over the 16 stimuli with overlapping F0 values to ask whether high identification accuracy in neurons could be explained by energy differences between the stimuli. Oboe stimuli had

higher spectral magnitudes than bassoon from 900 to 6875 Hz with a dip near 2 kHz (Fig. 4.3f). We then calculated which instrument elicited a larger rate response and plotted accuracy as a function of CF (Fig. 4.3g). If energy were the sole driver of the neural response rates, then average rates in response to oboe- would have been greater than for bassoon from 900 to 6875 Hz. However, the results were mixed, and rates in response to bassoon were often greater than for oboe in the frequency region where oboe spectral envelopes had larger magnitudes, indicating that energy was not the sole driver of these responses.

Next, the instrument-identification task was performed on responses of single neurons using timing information from PSTHs and an SVM classification model. The PSTH bin size that resulted in the highest median accuracy was 20 ms, though 0.75 – 150 ms bins were not significantly different (ANOVA, Fig. 4.4a). The 20-ms bin width indicates that no phase-locking information about F0 was present, and instead the differences in onset and sustained activity were used to classify the two instruments (Fig. 4.4b). The median neural accuracy for this task was 61%, with a range of 43 to 91%. For 70.0% of neurons, classification based on timing information was better than for average rates (Fig. 4.4c). Accuracy was correlated with CF ($p=0.0463$, linear regression) but did not differ across MTF groups ($p=0.1431$, Kruskal-Wallis). We investigated how much temporal information mattered by shuffling the PSTH bins within each trial. Shuffling the PSTHs reduced the accuracy ($p=1.1425e-9$, Mann-Whitney U-test), indicating that temporal structure contributed to these results (Fig. 4.4d). Although the temporal

information in the PSTH carried information, it is possible that the F0 information contained in the PSTH added noise when discriminating across F0s.

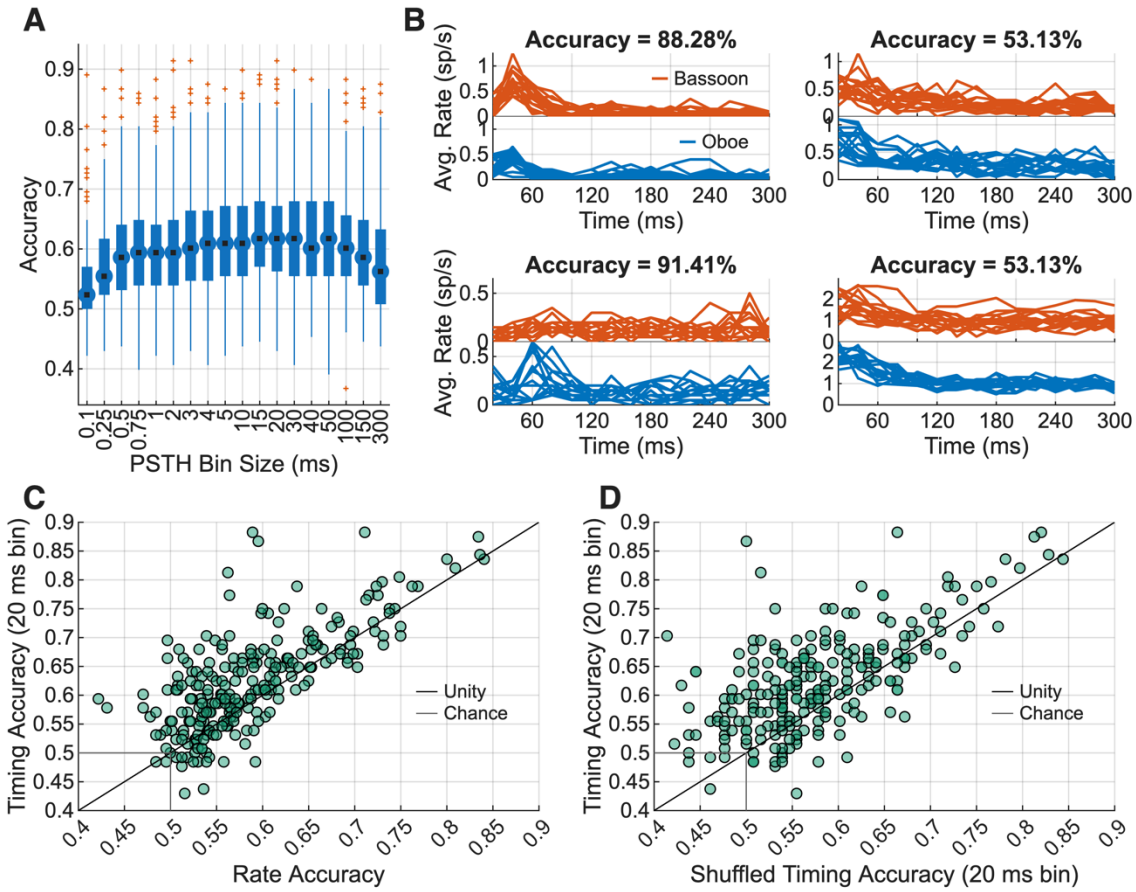

Figure 4.4. Single-unit identification of instrument using timing information. (A) Accuracy for all neurons using different PSTH bin sizes, outliers (red) and median (black) marked. (B) Example PSTHs using 20-ms bins for four example neurons, oboe (blue) and bassoon (orange) responses for each repetition. Top left: CF = 5278 Hz, BE. Bottom left: CF = 10556 Hz, Hybrid. Top right: CF = 2297 Hz, BS. Bottom right: CF = 2633 Hz, BS. (C) Timing accuracy plotted for each neuron as a function of rate accuracy. (D) Comparison of timing results with a model using shuffled timing with a 20-ms bin width, n=237 neurons.

Instrument identification using rate and timing information across a population.

Average-rate responses for a population of neurons were used in an SVM classifier to determine how a population of IC neurons encoded instrument identity. The population of neurons spanned 332 to 13929 Hz, covering the range of the stimulus spectrum (Fig. 4.5a). When comparing driven rates to an example F0, overall population responses were similar but had small differences (Fig. 4.5a). Decoding accuracy using that average rates of all neurons was 100%, with shuffled accuracy at chance (Fig. 4.5b). To investigate the contributions of each neuron in the model, feature weights, called beta weights, were plotted as a function of CF ($p=0.4855$), or classified by MTF group ($p=0.0744$, Kruskal-Wallis, Fig. 4.5c, d). A correlation between beta weights and single-unit rate accuracy was found ($p=1.3952e-32$, regression, Fig. 4.5e). A permutation test, in which a single neuron was taken out of the model, was used to test the impact of a single neuron on performance. The test revealed that each individual neuron had no contribution on model results (not shown), indicating that instrument identification information was encoded redundantly in many neurons. To further investigate contributions of CF and MTF groups, models were trained using 1 to 100 neurons, sorted by beta weights. The model using the highest weighted neurons reached 99% accuracy with 5 neurons and 100% accuracy with 20 neurons (Fig. 4.5f, g, green). When, neurons were split by CF group and the model was re-trained, the low-CF and medium-CF groups reached 99% accuracy when 80 or 75 neurons were included, respectively (Fig. 4.5g). Similarly, when only BS neurons were included, a model with 50 neurons predicted instrument identity with 99% accuracy (Fig. 4.5g). Overall, instrument identity was redundantly encoded in many neurons, but a diverse set of neurons was needed to decode instruments using a

small number of neurons. Timing information of a population of neurons was also used as an input to the instrument-identification SVM classifier. These results were similar to the rate responses: 10 neurons diverse in CF and MTF type yielded 99% accuracy (Fig. C1). Overall, results are consistent with previous spectral-discrimination tasks in quiet using synthetic stimuli: rate and timing responses encoded spectral information about the instrument identity.

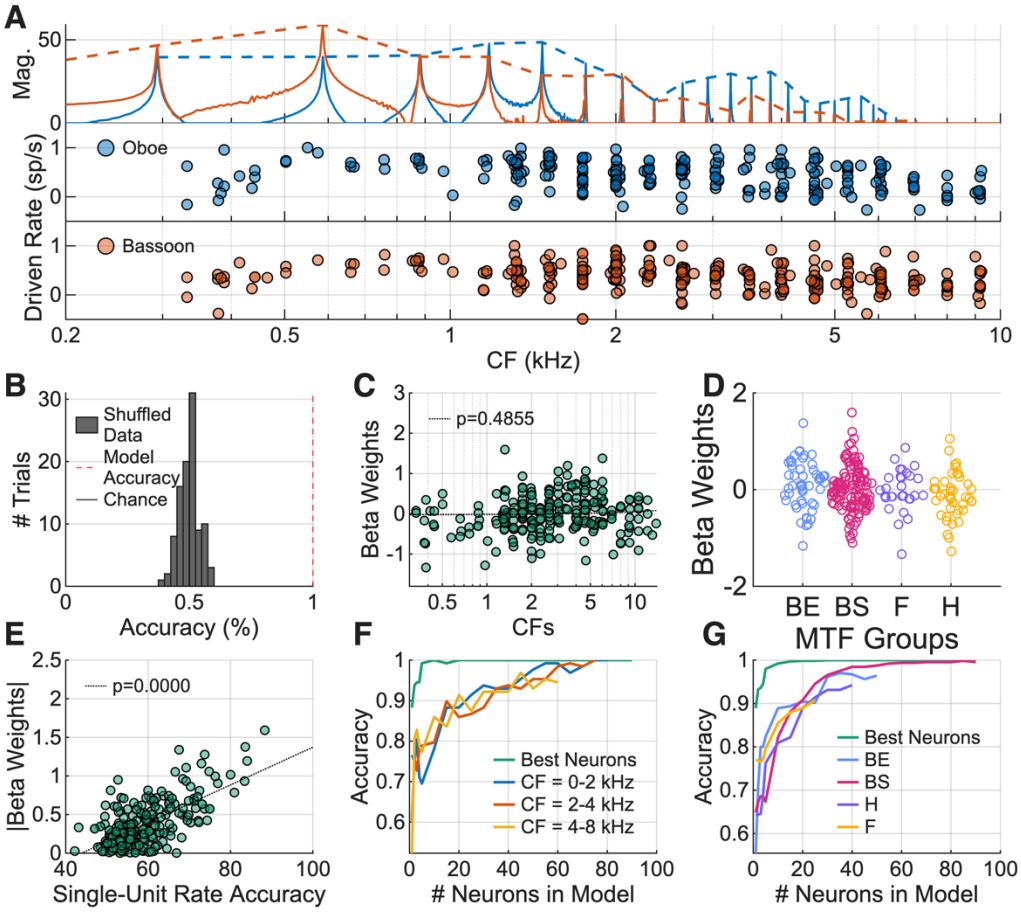

Figure 4.5. Instrument identification task using a population of neural rates. (A) Example population driven rate responses to a single F0 in either bassoon or oboe. (B) Decoding model performance and shuffled performance. (C) Neuron weights plotted as a function of CF. (D) Neuron weights plotted by MTF group. (E) Single neuron weights in the

model as a function of the single-unit rate accuracy. (F) Model performance when using the highest weighted neurons for all neurons, low, medium, or high CF neurons. (G) Same as G, but with MTF types.

F0 Identification

F0 identification for oboe and bassoon using rate or timing information in single neurons.

Next, the stimulus F0 of bassoon, oboe, or both, was discriminated in single neuron responses to learn how F0 information was represented by single neuron responses. Rates of a single neurons was a poor predictor of F0 for both bassoon and oboe, with an average decoding accuracy of 6.00% and 6.30%, respectively, and highest accuracy of 17.12% and 15.57% (Fig. 4.6a, b). Decoding models using timing information were more accurate: accuracy for decoding bassoon F0s had an average of 14.48% and maximum of 61.12%, and accuracy for decoding oboe had an average of 5.60% and a maximum of 16.71% (Fig. 4.6a, b). F0-identification results of single neurons for responses to bassoon and oboe were correlated: neurons with responses that allowed accurate identification of F0 in bassoon stimuli also contained information about oboe F0 ($p=7.2618e-56$, regression, Fig. 4.6g). We further investigated the timing information present in the F0- decoding task for responses to bassoon. There was no correlation between accuracy and CF ($p=0.3597$, regression). BE neurons were significantly more accurate than other MTFs ($p=0.0019$, Kruskal-Wallis), and there was a correlation between RVF PCA 2 score and accuracy ($p= 0.0094$, regression, Fig. 4.6c, d, e). The highest decoding accuracy occurred at low F0s; decoding accuracy from the top

15 neurons was over 50% for F0s 58 to 196 Hz (Fig. 4.6f). The reliability of spike timing was highly correlated with the F0-identification accuracy for responses to bassoon, but vector strength was not (Fig. 4.6f, h, i). These results suggest that the temporal responses of neurons contained information about lower F0s. Reliability was also a better predictor than vector strength to F0, suggesting that though vector strength may be low, precise timing in IC neurons can be decoded.

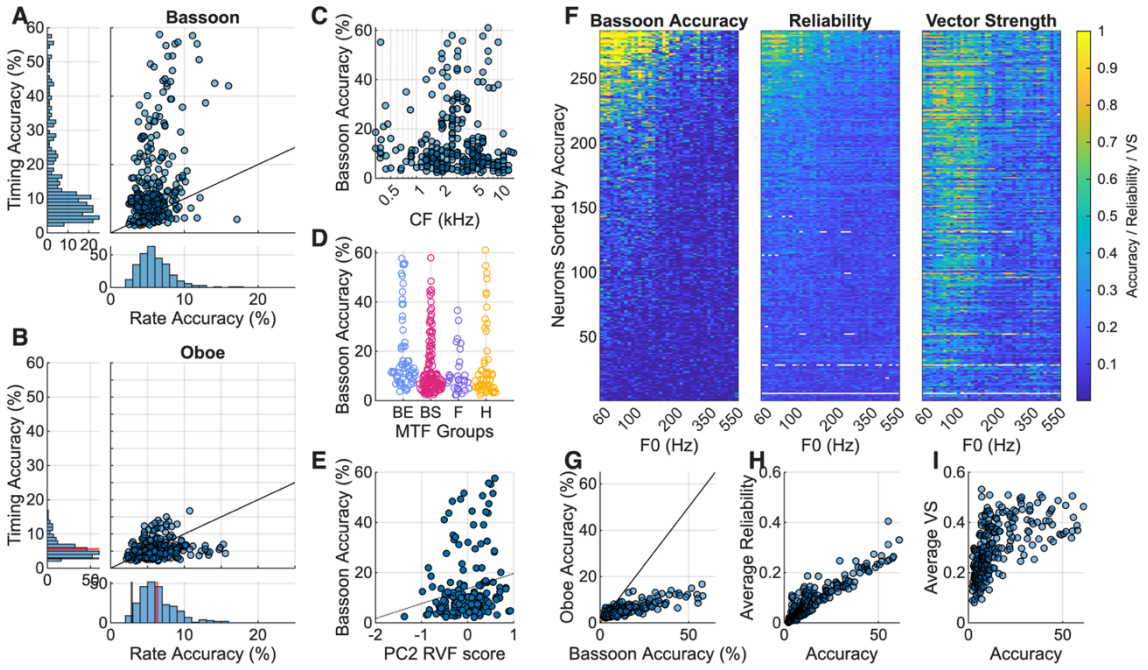

Figure 4.6. Single neuron rate and timing decoding F0.(A) Decoding accuracy using timing for every neuron as a function of accuracy using rate for bassoon stimuli and (B) oboe stimuli. (C) Neuron timing accuracy for bassoon as a function of CF, (D) MTF, or (E) PCA 2 from the RVF of the responses. (F) Accurate predictions for each bassoon F0 for all neurons, sorted by highest accuracy (left). Reliability (middle) and vector strength (right) calculations for each F0 and each neuron also sorted by highest accuracy. (G) Oboe timing accuracy vs bassoon accuracy. (H) Average reliability as a function of

accuracy for each neuron. (I) Average vector strength as a function of accuracy for each neuron.

Neurons with accurate F0 identification contain reliable timing information.

Neurons with reliable timing information resulted in more accurate F0 classifications, but reliability was only weakly correlated with vector strength. One example neuron with precise timing had a CF = 1529 Hz, BS MTF, and was sensitive to positive frequency sweeps (Fig. 4.7a, b, c). This neuron's response was significantly phase locked to the majority of the bassoon stimuli (Fig. 4.7d). Period histograms revealed that, for most bassoon stimuli, the neuron fired at the same phase of the stimulus, but there were often two spikes per period (Fig. 4.7e).

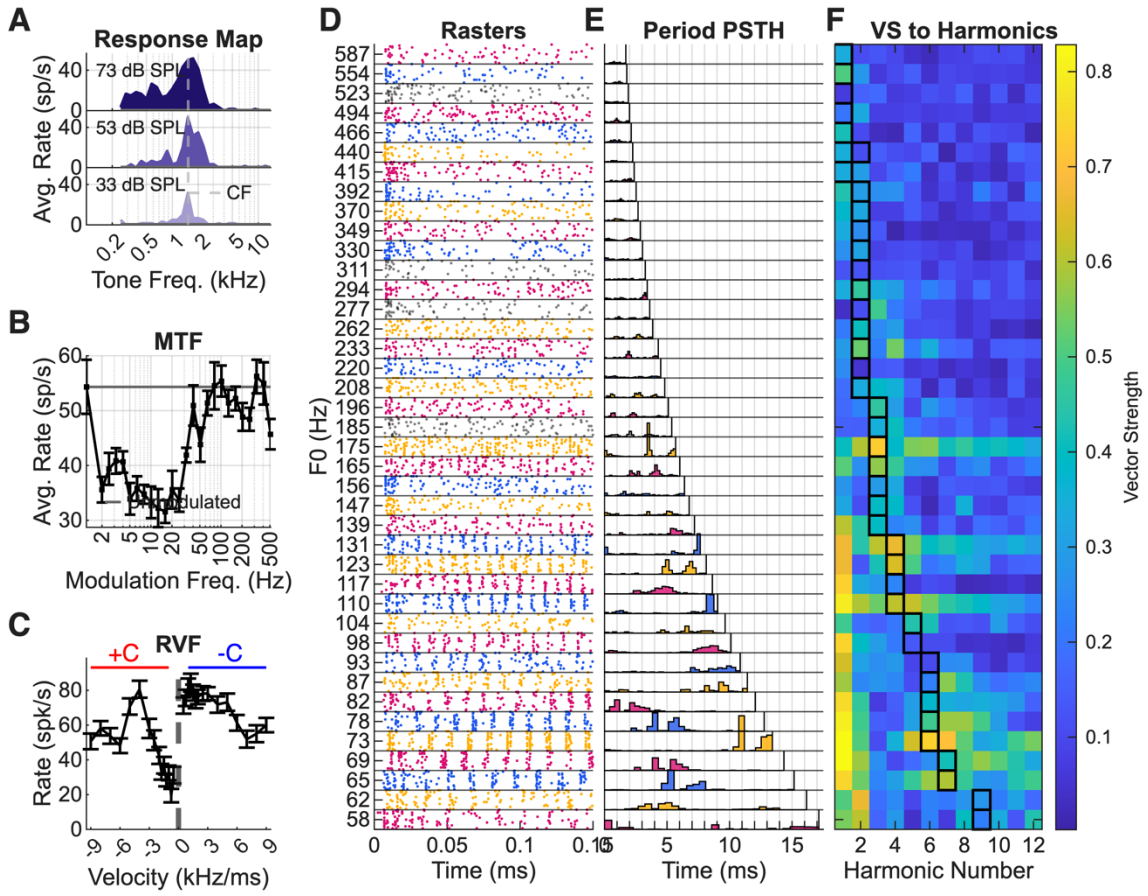

Figure 4.7. Neuron with the highest prediction accuracy. Response map, (B) MTF, and (C) RVF of the neuron. (D) PSTH and (E) period histogram of the neural responses to different bassoon F0s reveals complex responses. (F) Vector strength plotted for each harmonic, with the peak harmonic of each stimulus outlined in black.

Calculating vector strength to each harmonic revealed that this neuron had strong phase-locking responses to the fundamental, but also exhibited phase locking near or at the spectral peak harmonic, which was approximately 500 Hz for all stimuli (Fig. 4.7f). This was found in multiple neurons with high accuracy; neurons that had multiple spikes per period often phase locked to peak harmonics (Fig. 4.7, Fig. C2).

F0 identification for oboe and bassoon using rate or timing information in a population.

SVM and linear regression models were used to predict F0 based on rate and timing information from populations of neurons. First, an SVM decoder was used to predict bassoon or oboe F0 based on the rate responses of the entire population of recordings and reached 100% and 99% accuracy respectively (Fig. 4.8a). We also tested whether a simpler, linear regression model could predict F0s. Linear regression models performed well, with R^2 between the predicted F0 and actual F0 of 0.97 and 0.96 for both tasks (Fig. 4.8b). F0 was predicted within 5 semitones for bassoon and oboe (Fig. 4.8b). Beta weights varied with CF for the oboe (p=1.9698e-06). We split up neurons into CF groups and MTF groups and tested model accuracy using those groups against using all neurons, sorted by using highest beta weight neurons in the model first (Fig. 4.8c, d). Similar to results in the instrument identification task, responses of a small number of diverse neurons yielded accurate F0 predictions (Fig. 4.8c, d). Models based on neurons for a single CF or MTF group had worse performance. Interestingly, a linear readout of F0 using rate responses may exist, but it was necessary that the neurons included in the regression model were diverse in CF and MTF group.

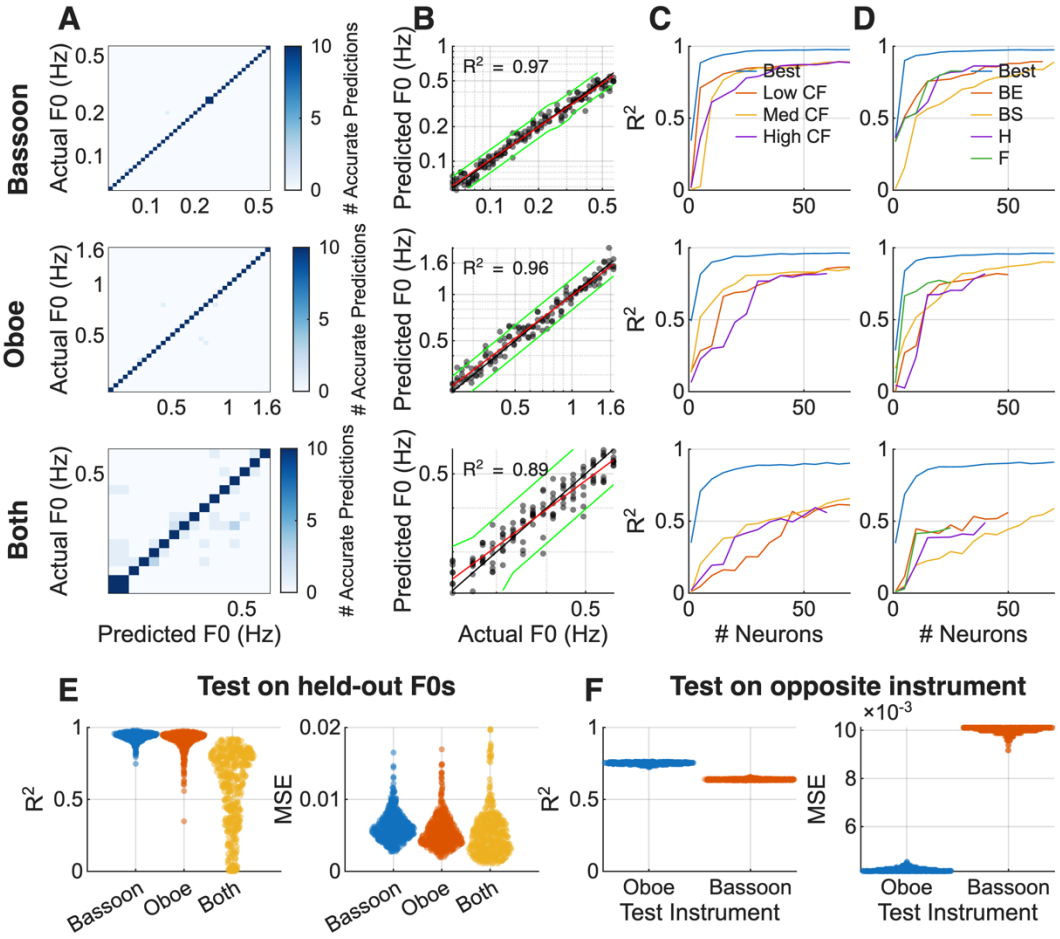

Figure 4.8. F0 decoding using rate responses for a population of neurons. Confusion matrices for classification model. (B) Linear regression model predictions, fit line (red) and lines indicating 5 semitones above/below F0 (green). (C) Accuracy of a linear regression model using a subset of highest beta weight neurons split into CF groups or (D) MTF groups. (E) Results of training on 80% of F0s and testing on 20% of F0s for bassoon, oboe, and both conditions, R^2 and mean squared error results, 500 trials. (F) Results when linear regression model was trained on one instrument and tested on the other instrument, 500 trials.

Though this task was framed as an F0-identification task, spectral centroid varied as a function of F0 (Fig. 4.1). Thus, these models could be using information present in the neurons about F0 and/or spectral centroid. To specifically investigate F0 information

in these neurons, F0 was predicted regardless of instrument identity, for F0s from 233 to 587 Hz, the F0 range over which stimuli for both instruments were available. SVM performance reached 95% accuracy, and the R2 for the linear regression model was 0.89 (Fig. 4.8a, b). Beta weights were correlated with CF ($p=0.0014$). Linear regression models were run with a subset of neurons sorted by beta weights and split into CF groups or MTF types (Fig. 4.8c, d). Using all MTF and CF groups, models reached a maximum R^2 of 0.90 using 50 neurons (Fig. 4.8c, d). However, performance decreased when the model included only neurons from single MTF or CF groups. The overall drop in performance for this task compared to the bassoon or oboe tasks indicated that the neural responses to the instrument spectrum had an impact on F0 identification.

Next, we tested how well linear regression models generalized though two tasks: 1) training on a subset of F0s and using the model to predict new F0s (Fig. 4.8e), and 2) training F0 discrimination using data from one instrument and testing the other instrument (Fig. 4.8f). Testing on held-out F0s resulted in R^2 values near the original F0 discrimination tasks (Fig. 4.8b), where bassoon F0 discrimination accuracy had a mean R^2 of 0.94 oboe F0 discrimination accuracy had a mean of 0.91, and discrimination F0s from both instruments had a mean R^2 of 0.58 (Fig. 4.8e). For the second task, the mean R^2 for models trained on bassoon data and tested on oboe F0s was 0.75 and interestingly the models trained on oboe F0s and tested on bassoon F0s had an R^2 of 0.64 (Fig. 4.8f). These results suggest that the models generalize to different F0s, but there are limitations in generalizing across different instruments.

Next, SVM models were used to predict F0 based on timing information from a population of neurons. Again, F0 was identified for bassoon and oboe stimuli separately, and then F0s were identified for oboe and bassoon for overlapping F0s. Models were run using 1 to 200 neurons in the population, sorted by individual-neuron timing accuracy. Bassoon F0 predictions rose to 99% accuracy with 30 neurons (Fig 4.9a, top). Oboe F0 predictions had similar trends, though it took 100 neurons to reach 88% accuracy (Fig. 4.9a, middle). For both tasks, higher F0s were more accurately classified as more neurons were added to the model (Fig. 4.9b, top, middle). These results indicated that F0 and spectral changes with F0 are also encoded in neural timing information in lower F0 regions.

The F0 identification model using timing from bassoon and oboe overlapping F0s was created to test for F0 decoding and avoid differences due to spectral centroid. The SVM model using this information reached 88% accuracy using 200 neurons (Fig. 4.9a, bottom). Predictions were less accurate over all F0s when spectrum information was mixed (Fig. 4.9b, bottom). Overall, spectral changes due to instrument identity across F0 values aided in prediction accuracy in this neural population for analyses based on both timing and rate information.

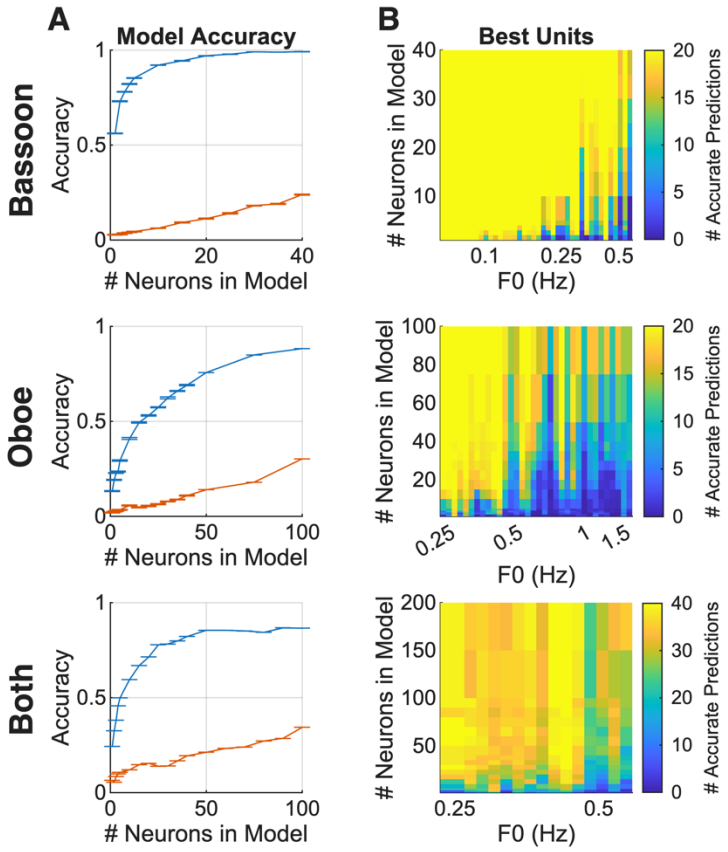

Figure 4.9. F0-identification accuracy using multiple neuron PSTHs for bassoon, oboe, and both bassoon and oboe. (A) F0 discrimination accuracy as a function of number of neurons in the model, with neurons sorted by best/worst based on individual neuron PSTH F0 accuracy. (B) Number of accurate predictions for each F0 as number of neurons in the model increases.

Instrument/F0 Identification

The result that spectral information improved F0 identification in our models led us to investigate whether information about both F0 and instrument identity was encoded by the same neurons. The bassoon F0-identification and instrument-identification accuracy using timing information had overall highest accuracies in single neurons were positively correlated (bassoon rate p=4.6377e-04, bassoon timing p=1.1148e-07, oboe

rate p=4.3173e-09, oboe timing= 4.5046e-11, Fig. 4.10A, bottom left). In fact, bassoon or oboe F0-identification were correlated with instrument-identification accuracies for both rate and timing, though accuracies were overall poor (Fig. 4.10a).

Next, we used an SVM to decode both instrument identity and F0 for all 75 possible stimuli (Fig. 4.10b, c, d). This 'token' decoding task asked whether rate or timing were sufficient to decode specific instrument F0s in a population of responses. SVM accuracy using rate was 98.8% using all 247 neurons, whereas timing was 85.9% using 40 neurons (Fig. 4.10b, c). The model using timing information again struggled with high F0s, confusing instrument identity and F0s at higher F0s (Fig. 4.10d). Within this population of neurons, enough information exists to decode both instrument identity and F0. Decoding using rate, specifically, did not confuse instrument identity, whereas decoding using timing was less accurate and did confuse instrument identity. Over this population of IC neurons, rate responses were more robust than timing.

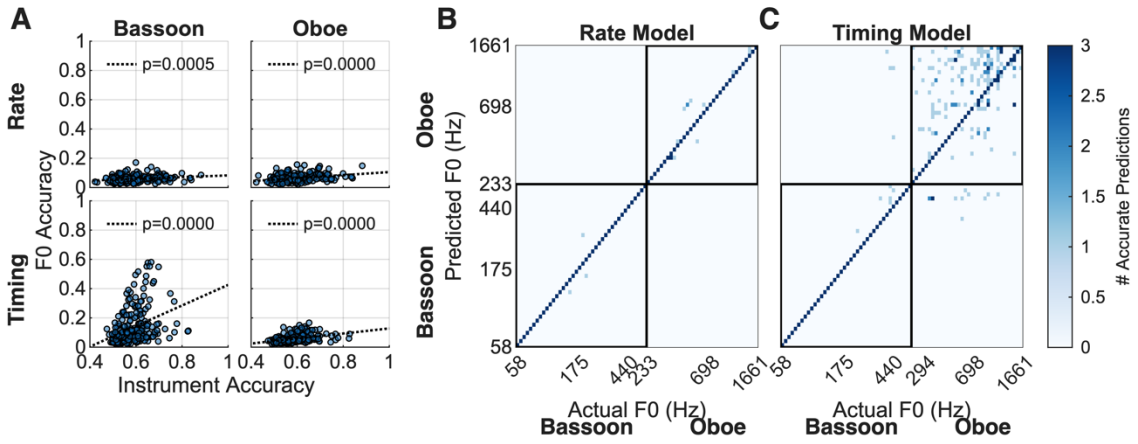

Figure 4.10. Investigating pitch/timbre interference. (A) Scatter plots of accuracy using rate or timing for bassoon and oboe F0s. (B) Confusion matrix for rate population model

identifying both instrument identity and F0. (C) Confusion matrix using timing information for 40 neurons.

Why Current IC Models Fail

For each BE and BS neuron in the neural population, we ran an SFIE model with the CF, MTF shape, and best/worst modulation frequency of each neuron. Response map, MTF, bassoon, and oboe stimuli were used as input to the SFIE model. Model responses were also calculated using the broad inhibition model and an energy model. We analyzed average rate responses to bassoon and oboe stimuli and compared the model rates with the neural rate responses and found that the variance of the data explained by the model was at chance for the majority of the data (Fig. 4.11a). IC models could not accurately predict rate changes as F0 increased in bassoon or oboe.

Next, we used the decoding models to investigate differences between SFIE model and neural responses. The spike times and average rates from the SFIE model used in the same decoding tasks as the data. Decoding F0 or instrument identity in single neurons often had similar distributions of accuracies as the data (Fig. 4.11b). However, there was no correlation between data accuracy and SFIE accuracy for single neurons except in the oboe and bassoon F0 identification with rate information task ($p=0.0150$, $p=0.0126$, Fig. 4.11b). However, we questioned whether the response maps and MTFs of the SFIE model accurately reflected neural response maps and MTFs. The correlation between neural response map and SFIE model response map and the correlation between neural MTF and SFIE model MTF were calculated. The correlations were squared and

then averaged to get the neural variance explained by the SFIE model. Model and data accuracies have no relationship.

To further investigate differences between the SFIE model responses and the neural data, we compared temporal responses to bassoon and oboe in the SFIE model and data. Both example neurons have MTFs and response map SFIE model responses with R^2 to data above 0.5. Despite MTF and response map similarity, temporal patterns for bassoon stimuli differed from neural data in both examples (Fig. 4.11d, e). The first example had an SFIE model response that could be decoded similarly to the neural response, but there are clear differences in onsets and temporal patterns (Fig. 4.11d). The second example features an SFIE model response with little phase locking to stimulus features and poor decoding accuracy (Fig. 4.11e).

Population instrument and F0 identification accuracies were similar between the SFIE model and neural data (not shown). The instrument identification population SFIE model also relied on a diverse set of CFs and MTF types to encode instrument, but in general contained redundant information. The F0 identification decoding accuracy using rate SFIE information was also similar to neural decoders, with bassoon identification 100% accurate, oboe identification 95% accurate, and both oboe and bassoon F0 identification 94% accurate. Next, F0 identification using SFIE timing information was similar for the bassoon F0 task, with SFIE accuracy of 96% with 30 neurons. Interestingly, one difference between neural decoding and SFIE decoding was found in decoding oboe F0 identity. In the SFIE model, decoding accuracy increased to 50% with 25 neurons in the model but additional model neurons did not improve accuracy, unlike

in the data (Fig. 4c). Specifically, prediction accuracy sharply fell off at F0s greater than 370 Hz (Fig. 4c).

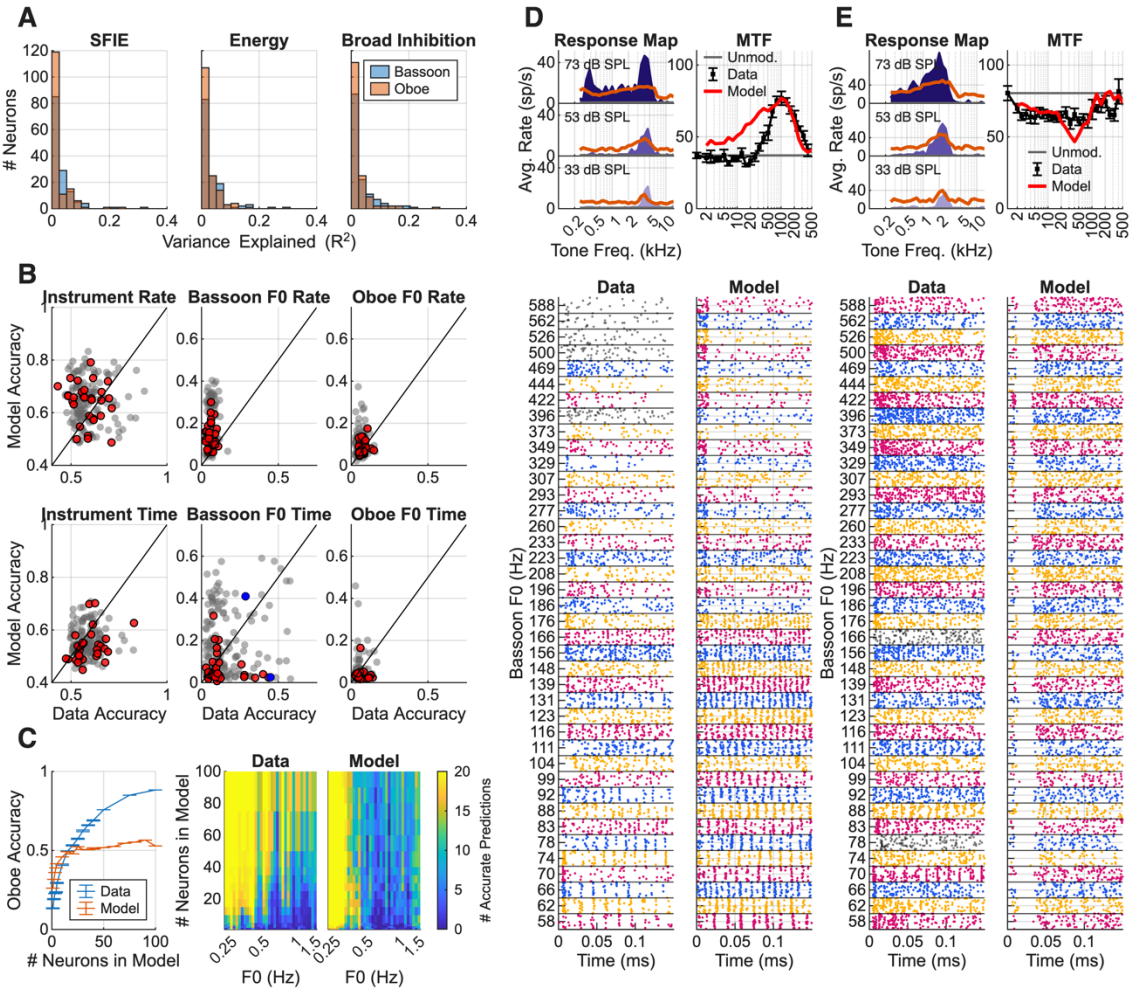

Figure 4.11. Model predictions of instrument and F0 identity and comparisons to neural data. (A) Variance of the neural rate responses to bassoon and oboe explained by the model responses, SFIE, energy, and broad inhibition for each neuron. (B) SFIE model accuracy in instrument and F0 identification tasks plotted as a function of neural accuracy for each single neuron. Unity line (black). Grey dots indicate models with response map and MTF R^2 to data less than 0.5, red indicates R^2 between data and model MTFs and response maps greater than 0.5. Blue dots indicate examples in (D) and (E). (C) Population decoding oboe F0 for data and SFIE model, including overall accuracy as

more neurons are added to the decoding model (left), and accuracies for each F0 prediction (middle data, right SFIE model). (D) Response map, MTF, and bassoon response rasters for data and model BE neuron and (E) BS neuron.

4.5 Discussion

We recorded extracellular responses in the IC in awake Dutch-belted rabbits to recorded bassoon and oboe sounds over a range of F0. Decoding models were used to characterize how instrument and F0 information were represented in single neurons and population responses. Information contained in rate responses of a population was sufficient to decode F0 and/or instrument identify, and models had the best performance when neurons were diverse in CF and MTF type. Additionally, decoding models could use timing information to decode F0 and instruments. Models used information from single-neuron timing responses to decode lower F0s. Population timing information also predicted F0s, though accuracy decreased as F0 increased.

Rate responses are sufficient for timbre encoding in quiet.

Single neurons contained rate and timing information sufficient to identify instruments with up to 88% accuracy. This result differed from a study using recorded vowels, in which rate-based vowel identification was poor, but timing-based vowel identification was accurate (Mitchell and Carney, 2025). This difference could be due to the variation of F0 information in the instrument-identification task and the higher F0s used, whereas the previous study decoded sets of vowels at a single F0, ranging from 95-202 Hz (Mitchell and Carney, 2025). It is likely that more precise timing in neurons at lower F0s contributes to higher decoding accuracy using timing in the vowel study.

Rate or timing information from a population of IC neurons could also be used to decode instrument identity. Previously, behavioral spectral peak discrimination thresholds in quiet could be explained by rate responses in budgerigar IC, and human spectral peak discrimination thresholds were similar to neural discrimination thresholds in rabbit IC rate responses (Henry et al., 2017; Fritzinger and Carney, in review). Interestingly, temporal information was necessary to discriminate spectral peaks in noise in budgerigar IC, providing a possible role of redundant encoding of spectral information using timing (Henry et al., 2017). We also found that neurons that were diverse in CF and MTF types yielded the best model performance with the least number of neurons.

A critical area of timbre investigation is identifying which stimulus attributes are essential for instrument identification. The presence of frequency sweeps in natural stimuli may affect encoding: a vowel study in the IC correlated decoding accuracy using timing with the second principal component of the RVF (Mitchell and Carney, 2025). Psychophysical studies have correlated perceptual attributes of timbre to specific sound characteristics (Review: McAdams, 2019) and it is unknown how most attributes are encoded in the IC. Studies have found that spectral peaks, which are related to the spectral centroid and the percept of brightness, are encoded in the IC (Fritzinger and Carney, in review). However other attributes most likely also contribute to timbre encoding, for example, attack time could largely impact temporal responses, enhancing temporal information about instrument identity. Though the significance of these attributes could not be tested in the present study, which focused on steady-state sounds,

future studies could systematically vary these attributes to further investigate this question.

Both rate and timing contain information about F0.

In responses of single neurons, decoding models were able to use timing information but not rate information. One study using HTCs found that rate information could reliably encode F0s > 800 Hz (Su and Delgutte, 2020), but we found no evidence of an increase in accuracy for high F0s using rate responses (Fig. S3). Phase locking was present up to 900 Hz in HTCs (Su and Delgutte, 2019). Neurons in our study significantly phase locked to F0 up to approximately 600 Hz, but generally, decoding models could only use temporal information up to 193 Hz. Differences between the HTCs and instrument results could be due to differences in the amplitude of harmonics, but more work needs to be done to examine differences between encoding F0 in flat and non-flat spectrums.

Decoding models using population information were much more accurate than single neurons, as expected. F0 identification using rate information from a population was accurate using either an SVM model or linear regression. The linear-regression model relied on a small subset of neurons with diverse CFs and MTF types, and rate responses in each of these neurons had a relatively linear relationship with F0. The linear model was chosen over the SVM to test the simpler hypothesis that F0 can be decoded linearly from rate. Model results were constant as a function of F0, supporting the idea that rate can encode higher F0s where responses may be resolved (Su and Delgutte, 2019), but interestingly lower F0s were encoded using rate as well. Note, these studies

were all performed in rabbit, changes in tuning with different animal species may impact results.

F0 identification using timing information for a population of neurons also had high accuracy. However, at higher F0s, temporal decoding models struggled, even as more neurons were added to the model. These results support a limit of temporal information, generally consistent with previous work.

Pitch-timbre interference may be present in the IC.

This study aimed to investigate whether pitch and timbre interference was present in the IC. Interference was investigated by using decoding models that 1) predicted F0 across a single instrument or 2) predicted F0 across both instruments for the F0s where oboe and bassoon overlapped. Decoding F0 in both instruments decreased model performance using both rate and timing population responses. This decrease in performance indicates that the decoding model was taking advantage of spectral changes due to F0 in the single-instrument F0 identification task. Additionally, single neuron accuracy of F0 identification was highly correlated with instrument identification accuracy, indicating that neurons often encode both pitch and timbre. These results support the hypothesis that pitch and timbre information are encoded together in the IC, and that variation in timbre disrupts pitch discrimination in IC neurons (Maxwell et al., 2020).

Benefits and limitations of computational modeling approaches.

Decoding models can be a useful way to reveal the format in which information is represented in a brain region, but results must be interpreted carefully (Review: (Kriegeskorte and Douglas, 2019). First, though decoding models are useful for finding information in the IC, how that information is used in the brain is unknown. Decoding models are not biologically plausible, in general; the medial geniculate nucleus does not necessarily integrate information from the IC like the decoding model. Second, weights of decoders are difficult to interpret, due to the nonlinear combinations in the SVMs. Another approach to determine neuron impact on the model is to remove a neuron from the model, run the model several times, and determine if model accuracy drops when that neuron is removed. In our data, removing one neuron did not alter performance, generally due to the redundant information present in the IC, and thus the importance of each neuron was zero. Instead, we trained models on subsets of neurons to determine what types of neurons carried important pitch or timbre information. Lastly, overfitting can lead to high accuracies even when information is not present in the neural data. We used 5-fold cross validation to prevent overfitting, but also analyzed neural responses used in the models to validate the presence of information.

Other computational IC models have been used to investigate specific encoding mechanisms in the IC. SFIE, broad inhibition, and energy models were poor predictors of bassoon and oboe rate responses over many F0s. However, other work has found that the SFIE-model predictions, with sensitivity to amplitude modulation and neural fluctuations, significantly correlated with a majority of neural responses to different vowels (Carney et al., 2015). Building upon that work, a model that was sensitive to amplitude modulation

and frequency-sweep direction/velocity better predicted vowel identification using timing information (Mitchell and Carney, 2025). Separately, the SFIE model with the addition of broad inhibition more accurately modeled IC responses at spectral peaks of a harmonic stimulus (Fritzinger and Carney, in review). Though the SFIE model was a poor predictor of average rate in response to bassoon and oboe F0s, we used decoding models with SFIE model inputs to examine how the information in the model differs from neural responses. We found large differences in temporal patterns of responses, and thus average rate responses as well for the model compared with neural data. More work needs to be done to investigate these differences in depth.

Human vs Animal Pitch/Timbre Perception

It is important to consider differences between human and animal perception when studying pitch and timbre. It has been hypothesized that humans generally use rate information, because humans are better at pitch discrimination when resolved harmonics are present (Shackleton and Carlyon, 1994) and cortical responses are more driven by resolved harmonics (Norman-Haignere et al., 2013). Most animal models, however, with broad tuning, are better at pitch tasks when harmonics are unresolved (Walker et al., 2019). Previous behavioral work has found that ferrets, gerbils, and chinchillas may be more reliant on temporal cues, whereas marmosets rely on resolved harmonics (Shofner, 2002; Klinge and Klump, 2009; Song et al., 2016; Walker et al., 2019). However, recent behavioral studies suggest that rabbits can discriminate F0s without the fundamental in the ranges 200 – 400 Hz, 400 – 800 Hz, and 800 – 1600Hz using resolved harmonics at high F0s and timing at lower F0s (Wagner et al., 2022).

The analyses here focused on information present in IC neurons that may be useful in instrument and F0 identification, or both, but remains limited by the stimuli used. A more diverse set of stimuli, including responses from instruments with strong F0s (such as trombone), unusual spectral envelopes (such as clarinets), and time-varying sounds (such as piano) would be useful to tease apart differences in neuronal coding of these aspects of sound. Another stimulus to add to this experiment would be the addition of simplified, synthetized instrument sounds. For example, the impact of phase on encoding could be determined by playing a natural instrument sound followed by a synthesized sound with harmonics in sine phase. To investigate how timbre interferes with pitch, a flat spectrum with phase information could be played along with the natural stimulus. Such a paradigm would help answer questions about encoding that this initial study leaves open.

Acknowledgements

This study was funded by NIH-R01-DC010813 and NIH-F31-DC020630-03. Thanks to Douglas Schwarz for assistance with hardware and software. Thanks to Braden Maxwell for feedback on analysis. Thanks to Kris Abrams for helping with the rabbit surgery and care. Thanks to Daniel Guest for feedback on machine-learning models.

Bibliography

Allen EJ, Oxenham AJ (2014) Symmetric interactions and interference between pitch and timbre. J Acoust Soc Am 135:1371–1379.

Attias H, Schreiner CE (1997) Temporal Low-Order Statistics of Natural Sounds. :7.

Bizley JK, Walker KMM, Silverman BW, King AJ, Schnupp JWH (2009) Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. J Neurosci 29:2064–2075.

Carney LH, Li T, McDonough JM (2015) Speech Coding in the Brain: Representation of Vowel Formants by Midbrain Neurons Tuned to Sound Fluctuations. eneuro 2:ENEURO.0004-15.2015.

Carney LH, McDonough JM (2019) Nonlinear auditory models yield new insights into representations of vowels. Atten Percept Psychophys 81:1034–1046.

Escabí MA, Miller LM, Read HL, Schreiner CE (2003) Naturalistic Auditory Contrast Improves Spectrotemporal Coding in the Cat Inferior Colliculus. J Neurosci 23:11489–11504.

Fritzinger, J. B., & Carney, L. H. (2025). Mechanisms of Tone-in-Noise Encoding in the Inferior Colliculus. Journal of Neuroscience, 45(23).

Gai Y, Carney LH (2006) Temporal Measures and Neural Strategies for Detection of Tones in Noise Based on Responses in Anteroventral Cochlear Nucleus. J Neurophysiol 96:2451–2464.

Henry KS, Abrams KS, Forst J, Mender MJ, Neilans EG, Idrobo F, Carney LH (2017) Midbrain Synchrony to Envelope Structure Supports Behavioral Sensitivity to Single-Formant Vowel-Like Sounds in Noise. J Assoc Res Otolaryngol 18:165–181.

Klinge A, Klump GM (2009) Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans. J Acoust Soc Am 125:304–314.

Kriegeskorte N, Douglas PK (2019) Interpreting encoding and decoding models. Curr Opin Neurobiol 55:167–179.

Landemard A, Bimbard C, Demené C, Shamma S, Norman-Haignere S, Boubenec Y (2021) Distinct higher-order representations of natural sounds in human and ferret auditory cortex. eLife 10 Available at: https://elifesciences.org/articles/65566 [Accessed July 9, 2025].

Lesica NA, Grothe B (2008) Dynamic Spectrotemporal Feature Selectivity in the Auditory Midbrain. J Neurosci 28:5412–5421.

Lyzwa D, Herrmann JM, Wörgötter F (2016) Natural Vocalizations in the Mammalian Inferior Colliculus are Broadly Encoded by a Small Number of Independent Multi-Units.

Front Neural Circuits 9 Available at: http://journal.frontiersin.org/article/10.3389/fncir.2015.00091 [Accessed July 9, 2025].

Maxwell BN, Fritzinger JB, Carney LH (2020) Neural Mechanisms for Timbre: Spectral-Centroid Discrimination based on a Model of Midbrain Neurons. In, pp 4.

McAdams S (2019) The Perceptual Representation of Timbre. In: Timbre: Acoustics, Perception, and Cognition (Siedenburg K, Saitis C, McAdams S, Popper AN, Fay RR, eds), pp 23–57 Springer Handbook of Auditory Research. Cham: Springer International Publishing. Available at: http://link.springer.com/10.1007/978-3-030-14832-4_2 [Accessed September 9, 2021].

McPherson MJ, McDermott JH (2023) Relative pitch representations and invariance to timbre. Cognition 232:105327.

Mitchell, P. W., & Carney, L. H. (2025). Chirp sensitivity and vowel coding in the inferior colliculus. Hearing Research, 109307.

Mitchell PW, Henry KS, Carney LH (2023) Sensitivity to direction and velocity of fast frequency chirps in the inferior colliculus of awake rabbit. Hear Res 440:108915.

Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J Acoust Soc Am 116:2173–2186.

Norman-Haignere S, Kanwisher N, McDermott JH (2013) Cortical Pitch Regions in Humans Respond Primarily to Resolved Harmonics and Are Located in Specific Tonotopic Regions of Anterior Auditory Cortex. J Neurosci 33:19451–19469.

Portfors CV, Roberts PD, Jonson K (2009) Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. Neuroscience 162:486–500.

Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Neural Comput 16:1661–1687.

Schwarz DM, Zilany MSA, Skevington M, Huang NJ, Flynn BC, Carney LH (2012) Semi-supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. J Neurosci Methods 206:120–131.

Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J Acoust Soc Am 95:3529–3540.

Shofner WP (2002) Perception of the periodicity strength of complex sounds by the chinchilla. Hear Res 173:69–81.

Siedenburg K, Jacobsen S, Reuter C (2021) Spectral envelope position and shape in sustained musical instrument sounds. J Acoust Soc Am 149:3715–3726.

Song X, Osmanski MS, Guo Y, Wang X (2016) Complex pitch perception mechanisms are shared by humans and a New World monkey. Proc Natl Acad Sci 113:781–786.

Su Y, Delgutte B (2019) Pitch of harmonic complex tones: rate and temporal coding of envelope repetition rate in inferior colliculus of unanesthetized rabbits. J Neurophysiol 122:2468–2485.

Su Y, Delgutte B (2020) Robust Rate-Place Coding of Resolved Components in Harmonic and Inharmonic Complex Tones in Auditory Midbrain. J Neurosci 40:2080–2093.

Theunissen FE, Elie JE (2014) Neural processing of natural sounds. Nat Rev Neurosci 15:355–366.

Wagner JD, Gelman A, Hancock KE, Chung Y, Delgutte B (2022) Rabbits use both spectral and temporal cues to discriminate the fundamental frequency of harmonic complexes with missing fundamentals. J Neurophysiol 127:290–312.

Walker KM, Gonzalez R, Kang JZ, McDermott JH, King AJ (2019) Across-species differences in pitch perception are consistent with differences in cochlear filtering. eLife 8:e41626.

Whitehead ML, Lonsbury-Martin BL, Martin GK (1992) Evidence for two discrete sources of 2 f1−f2 distortion-product otoacoustic emission in rabbit: I. Differential dependence on stimulus parameters. J Acoust Soc Am 91:1587–1607.

Zilany MSA, Bruce IC, Carney LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. J Acoust Soc Am 135:283–286.

Chapter 5: Summary and Discussion

5.1 Summary and Novel Results

In Chapter 2, an AM-tuned broad inhibition was introduced as a mechanism for encoding complex sounds in tandem with IC sensitivity to NFs. Though off-CF inhibition has been quantified in the past with two-tone paradigms (Egorova et al., 2001; Portfors and Felix, 2005; Alkhatib et al., 2006; Palmer et al., 2013) and blocking inhibition is known to change response properties (Yang et al., 1992; LeBeau et al., 2001; Xie et al., 2005; Andoni et al., 2007), there has been a lack of computational models to investigate how broad inhibition impacts encoding of complex sounds in the IC. A novel result of Chapter 2 is that the addition of an AM-tuned broad-inhibition mechanism to an IC model improved predictions of responses to tones in noise. The AM-tuned aspect of this broad inhibition means that the off-CF responses in the model are also sensitive to neural fluctuations. This AM-tuning was necessary to accurately predict WB-TIN and MTFs, a model with a simpler broad inhibition could not replicate results.

The theory that IC mechanisms are impacted by stimulus bandwidth can also help reconcile differing results in previous studies. For example, IC responses in macaques increase with SNR for tones in wideband noise (Rocchi and Ramachandran, 2018), whereas IC responses in budgerigar decrease with SNR for tones-in-narrowband noise (BE MTFs, Henry et al., 2017). Rabbit IC responses increase and decrease, dependent on MTF type, with SNR for narrowband TIN (Fan et al., 2021). These previous studies fit

under a framework that NFs impact responses to narrowband TIN, but AM-tuned broad inhibition has a more important in encoding wideband TIN.

In Chapter 3, we investigated how spectral peaks in a synthetic stimulus were represented in the IC, characterized responses to the stimuli, and used computational models to understand the mechanisms that impact timbre encoding. BS cells encoded spectral peak as an increase in rate which supported the hypothesis that NFs could be transformed into rate changes due to AM sensitivity of IC neurons. However, BE cells did not support the hypothesis. Instead, BE cells increased in rate near the spectral peak. The computational model with both AM-tuned broad inhibition and AM sensitivity correctly predicted physiological results, supporting the hypothesis that multiple mechanisms influence encoding. Another novel result was that neural-discrimination thresholds were similar to psychophysical thresholds, and generally neural-discrimination thresholds improved as sound level increased. Additionally, temporal responses contained rich information about the stimulus.

In the last project, Chapter 4, we presented musical instrument sounds to IC neurons. The goal of this work was to determine how well computational models predicted responses to real sounds. However, the computational IC models failed to predict rate responses to natural sounds, and instead we used decoding models to investigate how F0 and timbre information were represented in the IC. Instrument identity and rate responses were redundantly encoded using a population of IC rate responses, supporting the theory of IC transforming a temporal code into a rate code. Interestingly, timing information was incredibly rich; single neurons encoded F0s in

voice pitch range, and decoders using timing information from a population could identify instruments and F0s, though accuracy decreased at high F0s.

5.2 Future Work

We used physiological recordings and computational modeling approaches to investigate encoding of complex sounds, specifically focused on the percept of timbre. Many future directions and open questions about timbre encoding in the IC remain, here split into four main sections:

1. Understanding encoding mechanisms in the IC.
2. Improving computational models of the IC.
3. Investigating timbre and pitch/timbre interactions in the IC.
4. Advancing human hearing health.

5.2.1. Understanding encoding mechanisms in the IC.

Chapter 2 introduced a theory that narrowband sounds are primarily encoded in the IC due to sensitivity to NFs in the AN, whereas an AM-tuned broad-inhibition mechanism helps encode broadband sounds. Additional experiments need to be done to investigate these mechanisms. A band-widening experiment could be performed with tones in noise to investigate at what bandwidths the on-CF NF encoding dominates AM-tuned broad inhibition. The switch from NFs as the dominant encoding mechanism to AM-tuned broad inhibition would be evident in BE neurons that decrease in rate for a tone at CF in 1/3rd octave noise but increase for a tone at CF in 4 octave noise.

Next, is it true that the IC is where a timing code is transformed into a rate code? Chapters 3 and 4 demonstrated that in quiet environments, rate codes are sufficient to describe human thresholds of timbre discrimination, and IC population rates contain enough information to decode instrument identity and F0. However, Henry et al., (2017) demonstrated that temporal information was necessary in budgerigar to describe behavioral timbre discrimination thresholds in noise. We also found that although phase-locking limits to temporal fine structure near CF decrease in the IC, precise spike timing, related to envelope and phase locking to harmonics, carries important information about natural stimuli. We can speculate as to the purpose of both timing and rate carrying pitch and timbre information: 1) It is possible that the redundancy is necessary to carry information when rate and timing information may be degraded in noisy environments and 2) These dual encoding strategies may allow neurons to encode information about several different sound attributes through different spiking characteristics, known as multiplexing. These questions and ideas could be investigated by recording IC responses to complex sounds in noise, and by studying how neurons in the medial geniculate body (MGB) respond to complex sounds. Neurons in MGB are sensitive to AM, with similar but sometimes more complex MTFs than IC neurons (Kim et al., 2020). AN sensitivity could be inherited from IC neurons and could contribute to the transformation from a temporal code to rate code. However, little is known about how MGB encodes complex sounds, and more work will need to be done to understand the importance of rate or timing inputs from the IC to MGB.

The projects in Chapters 2, 3, and 4 have approached IC encoding from a population or a single neuron. Chapter 2 investigates AM-tuned broad inhibition as a possible mechanism for single neuron rate encoding. However, an alternative framework for these responses was briefly mentioned: the rate profile could be interpreted as an inferred population response. In this framework, a tone excites neurons tuned near the tone frequency, but neurons tuned to frequencies within the broadband noise are inhibited due to the presence of the tone. Thus, broad inhibition could sharpen rate representations of spectral peaks across a population. In Chapter 3, we infer a population response while recording from one neuron, which imposes limitations on the population code. This neural population contains neurons with different frequency responses but assumes the MTF sensitivity is the same throughout the population. Lastly, Chapter 4 studies both single neuron and population decoding, finding robust and redundant rate codes in neural populations made up of many CFs, MTF types, and frequency-sweep selectivity. However, again, little is known about encoding strategies in the MGB, and thus there are open questions about how the MGB integrates information from a population of IC neurons. We propose studying MGB responses to complex sounds and using encoding and decoding methods to investigate possible IC inputs.

Alternatively, genetically identifiable neuron types have been identified (Goyer et al., 2019; Silveira et al., 2020; Kreeger et al., 2021) that could be used to explore mechanisms of encoding in the IC with more specificity. For example, how AM sensitivity arises in the IC is still unknown, and investigating AM sensitivity would improve computational models of the IC and enhance understanding of AM coding in the

IC. Different computational models of the IC have proposed AM sensitivity as arising from either coincidence detection (Langner, 1981; Hewitt and Meddis, 1994) or inhibition (Nelson and Carney, 2004; Dicke et al., 2007). Coincidence detection AM sensitivity relies on oscillator neurons in the CN, dependent on either VCN chopper cells (Hewitt and Meddis, 1994) or an oscillator/integrator circuit (Langner, 1981), though support of coincidence detection in the IC remains low (Burger and Pollak, 1998). AM sensitivity can also be modeled through precise timing of excitation followed by delayed inhibition in a two-stage circuit (Nelson and Carney, 2004) or through inhibition from a lowpass IC neuron stage to create a bandpass IC neuron (Dicke et al., 2007). Genetic approaches have begun to investigate AM sensitivity with more specificity, finding that intrinsic properties of IC neurons contribute to AM sensitivity (Drotos et al., 2023). Blocking NMDA receptors in vivo in a specific subtype of IC neurons, NPY neurons, decrease sensitivity to AM and thus may facilitate temporal summation (Drotos et al., 2023). Broad inhibition mechanisms could also be studied more thoroughly using these techniques, and there is much to explore to understand how ascending or recurrent connections in the IC impact encoding complex sounds.

5.2.2. Improving computational models of the IC.

Modeling IC neuron responses to complex sounds has been a challenge and there are several different models of the auditory periphery and midbrain that have not been reconciled. First, there is the AM-tuned broad-inhibition/AM model from Chapter 2, built upon the SFIE model (Nelson and Carney, 2004). A limitation of the current AM-tuned broad inhibition model is the single-neuron nature of the model. A true population version

of this model would have recurrent connections across CF. Second, there is a model that captures both AM sensitivity and chirp-direction sensitivity in IC neurons using the Krips and Furst framework and an octopus cell stage (Mitchell and Carney, 2024). Third, great progress has been made on an AN model that uses efferent feedback from IC model neurons (Farhadi et al., 2022, 2023). These models intend to replicate midbrain responses to different complex sounds or use IC neural activity to modulate feedback mechanisms. However, no current model can accurately predict responses to both synthetic timbre and to chirps. And no model can accurately replicate IC responses to vowel (Mitchell and Carney, 2025) or musical instrument sounds.

Additionally, the SFIE and AM-tuned broad inhibition models have an issue: neither model accurately predicts responses over a wide range of suprathreshold levels. This behavior occurs due to saturated on- and off-CF high-spontaneous-rate AN fibers that provide ascending inputs to the model IC neuron. The AN model responds appropriately to complex sounds and has been rigorously tested against physiological data (Carney, 1993; Zilany et al., 2009, 2014), thus the IC model must be improved. Unfortunately, using the AN model with narrowband efferent feedback did not improve predictions over level in IC model responses, but version of the AN model with wideband feedback is under development and could impact IC predictions over level.

These shortcomings with the current models illustrate a need for a new IC model, perhaps with an entirely new framework. One option is the Krips and Furst framework, used in the chirp sensitivity model, where model cells are described as coincident detectors that combine inhibitory and excitatory inputs with strength and timing

parameters (Krips and Furst, 2009a, 2009b). Conductance-based modeling approaches may also be better suited for modeling IC neurons: a conductance-based model fit using extracellular spike train data that maps synaptic conductances that govern sub-threshold membrane dynamics has been able to model retinal-ganglion-cell activity (Gerstner, 2001; Latimer et al., 2019). Intracellular recordings and circuit mapping using genetic approaches could inform these types of models as well, limiting the number of parameters that would need to be fit with extracellular data.

A drawback to all IC models is the lack of comprehensive models of nuclei in the auditory brainstem. The SFIE uses a model VCN cell, other models use octopus cells, chopper cells, LSO, MSO, or model a stage in the lateral lemniscus (Review of IC models: Davis et al., 2010). To avoid limitations from earlier model stages, IC models could be built specifically to explore circuit mechanisms within the IC, momentarily disregarding earlier computations. Inhibition-stabilized networks, a recurrent circuit found in other sensory areas (Sanzeni et al., 2020; Sadeh and Clopath, 2021), could be modeled in the IC to test hypotheses about computations in the IC.

5.2.3 Investigating timbre and pitch/timbre interactions in the IC.

Chapter 3 focused on a specific aspect of timbre, related to the spectral peak and spectral centroid of a musical stimulus. However, as detailed in Chapter 1, many other sound attributes contribute to timbre perception. More specific experiments could be done to investigate timbre encoding in the IC by manipulating other attributes of timbre. For example, reducing phase contrast in vowels makes vowel discrimination more difficult (Molis et al., 2013). Neurons in the IC are impacted by phase, due to sensitivity

to frequency sweeps, but Chapter 4 was not able to determine the impact of phase in musical sounds on IC encoding. Future work may manipulate phase of natural or synthetic sounds to understand how phase impacts IC responses. Temporal aspects of timbre were also disregarded; the attack of a musical sound was not included in the natural stimuli. Attack and decay times offer information about instrument identity, string instruments have longer attack times than percussion, and piano sounds decay over time. These attributes also have not yet been studied in the IC. Additionally, there are also more questions about IC encoding of spectral attributes, such as the intensity of even and odd harmonics, called spectral irregularity. Spectral irregularity impacts our perception, making instruments like clarinets sound hollow (McAdams, 2019). Computational models and physiological experiments could be used to further explore spectral encoding related to spectral irregularity.

5.2.4. Advancing human hearing health.

Hearing loss is a public health concern that affected 14% (27.7 million) people in the US between the ages of 20-69 in 2012 (NIDCD, Hoffman et al., 2017). As mentioned in the introduction, timbre and vowel perception degrades in older listeners, listeners with hearing loss, and in cochlear implant patients (Gfeller et al., 2002; Emiroglu and Kollmeier, 2008; Kong et al., 2011, 2012; Chintanpalli et al., 2016). There is a large interest in advancing hearing aids and cochlear implants to better public health. Mechanisms that lead to perceptual degradation could include the loss of OHCs degrade NF profiles or loss of inhibition changes IC encoding of these sounds (Caspary et al., 2008). From the results of Chapters 2 and 3, it is possible that the loss of inhibition

impacts timbre perception more than degradation of NF profiles. Studying these mechanisms in older or noise exposed animals could further explore the impact of loss of inhibition on spectral peak representations in the IC. Computational models including broad inhibition could also alter inhibition strengths to model the impact of loss of inhibition on IC representations of timbre.

Bibliography

Alkhatib A, Biebel UW, Smolders JWT (2006) Inhibitory and excitatory response areas of neurons in the central nucleus of the inferior colliculus in unanesthetized chinchillas. Exp Brain Res 174:124–143.

Andoni S, Li N, Pollak GD (2007) Spectrotemporal Receptive Fields in the Inferior Colliculus Revealing Selectivity for Spectral Motion in Conspecific Vocalizations. J Neurosci 27:4882–4893.

Burger RM, Pollak GD (1998) Analysis of the Role of Inhibition in Shaping Responses to Sinusoidally Amplitude-Modulated Signals in the Inferior Colliculus. J Neurophysiol 80:1686–1701.

Carney LH (1993) A model for the responses of low-frequency auditory-nerve fibers in cat. J Acoust Soc Am 93:401–417.

Caspary DM, Ling L, Turner JG, Hughes LF (2008) Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. J Exp Biol 211:1781–1791.

Chintanpalli A, Ahlstrom JB, Dubno JR (2016) Effects of age and hearing loss on concurrent vowel identification. J Acoust Soc Am 140:4142–4153.

Davis KA, Hancock KE, Delgutte B (2010) Computational Models of Inferior Colliculus Neurons. In: Computational Models of the Auditory System (Meddis R, Lopez-Poveda EA, Fay RR, Popper AN, eds), pp 129–176 Springer Handbook of Auditory Research. Boston, MA: Springer US. Available at: http://link.springer.com/10.1007/978-1-4419-5934-8_6 [Accessed April 12, 2021].

Dicke U, Ewert SD, Dau T, Kollmeier B (2007) A neural circuit transforming temporal periodicity information into a rate-based representation in the mammalian auditory system. J Acoust Soc Am 121:310–326.

Drotos AC, Zarb RL, Booth V, Roberts MT (2023) GluN2C/D-containing NMDA receptors enhance temporal summation and increase sound-evoked and spontaneous firing in the inferior colliculus. Available at: http://biorxiv.org/lookup/doi/10.1101/2023.04.27.538607 [Accessed August 30, 2024].

Egorova M, Ehret G, Vartanian I, Esser K-H (2001) Frequency response areas of neurons in the mouse inferior colliculus. I. Threshold and tuning characteristics. Exp Brain Res 140:145–161.

Emiroglu S, Kollmeier B (2008) Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. Brain Res 1220:199–207.

Fan L, Henry KS, Carney LH (2021) Responses to diotic tone-in-noise stimuli in the inferior colliculus: stimulus envelope and neural fluctuation cues. Hear Res 409:108328.

Farhadi A, Jennings SG, Strickland EA, Carney LH (2022) Auditory-nerve Model including Efferent Dynamic Gain Control with Inputs from Cochlear Nucleus and Inferior Colliculus. Neuroscience. Available at: http://biorxiv.org/lookup/doi/10.1101/2022.10.25.513794 [Accessed February 6, 2023].

Farhadi A, Jennings SG, Strickland EA, Carney LH (2023) Subcortical auditory model including efferent dynamic gain control with inputs from cochlear nucleus and inferior colliculus. J Acoust Soc Am 154:3644–3659.

Gerstner W (2001) A framework for spiking neuron models: The spike response model. In: Handbook of Biological Physics (Moss F, Gielen S, eds), pp 469–516 Neuro-Informatics and Neural Modelling. North-Holland. Available at: https://www.sciencedirect.com/science/article/pii/S1383812101800154 [Accessed September 14, 2021].

Gfeller K, Witt S, Mehr MA, Woodworth G, Knutson J (2002) Effects of Frequency, Instrumental Family, and Cochlear Implant Type on Timbre Recognition and Appraisal. Ann Otol Rhinol Laryngol 111:349–356.

Goyer D, Silveira MA, George AP, Beebe NL, Edelbrock RM, Malinski PT, Schofield BR, Roberts MT (2019) A novel class of inferior colliculus principal neurons labeled in vasoactive intestinal peptide-Cre mice. :32.

Henry KS, Abrams KS, Forst J, Mender MJ, Neilans EG, Idrobo F, Carney LH (2017) Midbrain Synchrony to Envelope Structure Supports Behavioral Sensitivity to Single-Formant Vowel-Like Sounds in Noise. J Assoc Res Otolaryngol 18:165–181.

Hewitt MJ, Meddis R (1994) A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus. J Acoust Soc Am 95:2145–2159.

Hoffman HJ, Dobie RA, Losonczy KG, Themann CL, Flamme GA (2017) Declining Prevalence of Hearing Loss in US Adults Aged 20 to 69 Years. JAMA Otolaryngol Neck Surg 143:274.

Kim DO, Carney L, Kuwada S (2020) Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. J Neurophysiol 124:1198–1215.

Kong Y-Y, Mullangi A, Marozeau J (2012) Timbre and Speech Perception in Bimodal and Bilateral Cochlear-Implant Listeners. Ear Hear 33:645–659.

Kong Y-Y, Mullangi A, Marozeau J, Epstein M (2011) Temporal and Spectral Cues for Musical Timbre Perception in Electric Hearing. J Speech Lang Hear Res 54:981–994.

Kreeger LJ, Connelly CJ, Mehta P, Zemelman BV, Golding NL (2021) Excitatory cholecystokinin neurons of the midbrain integrate diverse temporal responses and drive auditory thalamic subdomains. Proc Natl Acad Sci 118:e2007724118.

Krips R, Furst M (2009a) Stochastic Properties of Coincidence-Detector Neural Cells. Neural Comput 21:2524–2553.

Krips R, Furst M (2009b) Stochastic properties of auditory brainstem coincidence detectors in binaural perception. J Acoust Soc Am 125:1567–1583.

Langner G (1981) Neuronal mechanisms for pitch analysis in the time domain. Exp Brain Res 44 Available at: http://link.springer.com/10.1007/BF00238840 [Accessed July 6, 2025].

Latimer KW, Rieke F, Pillow JW (2019) Inferring synaptic inputs from spikes with a conductance-based neural encoding model. eLife 8:e47012.

LeBeau FEN, Malmierca MS, Rees A (2001) Iontophoresis In Vivo Demonstrates a Key Role for GABA A and Glycinergic Inhibition in Shaping Frequency Response Areas in the Inferior Colliculus of Guinea Pig. J Neurosci 21:7303–7312.

McAdams S (2019) The Perceptual Representation of Timbre. In: Timbre: Acoustics, Perception, and Cognition (Siedenburg K, Saitis C, McAdams S, Popper AN, Fay RR, eds), pp 23–57 Springer Handbook of Auditory Research. Cham: Springer International Publishing. Available at: http://link.springer.com/10.1007/978-3-030-14832-4_2 [Accessed September 9, 2021].

Mitchell P, Carney L (2024) A Computational Model of Auditory Chirp-Velocity Sensitivity and Amplitude-Modulation Tuning in Inferior Colliculus Neurons. J Comput Neurosci 52:285–302.

Mitchell, P. W., & Carney, L. H. (2025). Chirp sensitivity and vowel coding in the inferior colliculus. Hearing Research, 109307.

Molis MR, Diedesch A, Gallun F, Leek MR (2013) Vowel Identification by Amplitude and Phase Contrast. J Assoc Res Otolaryngol 14:125–137.

Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J Acoust Soc Am 116:2173–2186.

Palmer AR, Shackleton TM, Sumner CJ, Zobay O, Rees A (2013) Classification of frequency response areas in the inferior colliculus reveals continua not discrete classes. J Physiol 591:4003–4025.

Portfors CV, Felix RA (2005) Spectral integration in the inferior colliculus of the CBA/CaJ mouse. Neuroscience 136:1159–1170.

Rocchi F, Ramachandran R (2018) Neuronal adaptation to sound statistics in the inferior colliculus of behaving macaques does not reduce the effectiveness of the masking noise. J Neurophysiol 120:2819–2833.

Sadeh S, Clopath C (2021) Inhibitory stabilization and cortical computation. Nat Rev Neurosci 22:21–37.

Sanzeni A, Akitake B, Goldbach HC, Leedy CE, Brunel N, Histed MH (2020) Inhibition stabilization is a widespread property of cortical networks. eLife 9:e54875.

Silveira MA, Anair JD, Beebe NL, Mirjalili P, Schofield BR, Roberts MT (2020) Neuropeptide Y Expression Defines a Novel Class of GABAergic Projection Neuron in the Inferior Colliculus. J Neurosci 40:4685–4699.

Xie R, Meitzen J, Pollak GD (2005) Differing Roles of Inhibition in Hierarchical Processing of Species-Specific Calls in Auditory Brainstem Nuclei. J Neurophysiol 94:4019–4037.

Yang L, Pollak GD, Resler C (1992) GABAergic circuits sharpen tuning curves and modify response properties in the mustache bat inferior colliculus. J Neurophysiol 68:1760–1774.

Zilany MSA, Bruce IC, Carney LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. J Acoust Soc Am 135:283–286.

Zilany MSA, Bruce IC, Nelson PC, Carney LH (2009) A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. J Acoust Soc Am 126:2390–2412.

Appendix A: Chapter 2 Supplemental Material

Supplemental material for Chapter 2 as published in the Journal of Neuroscience.

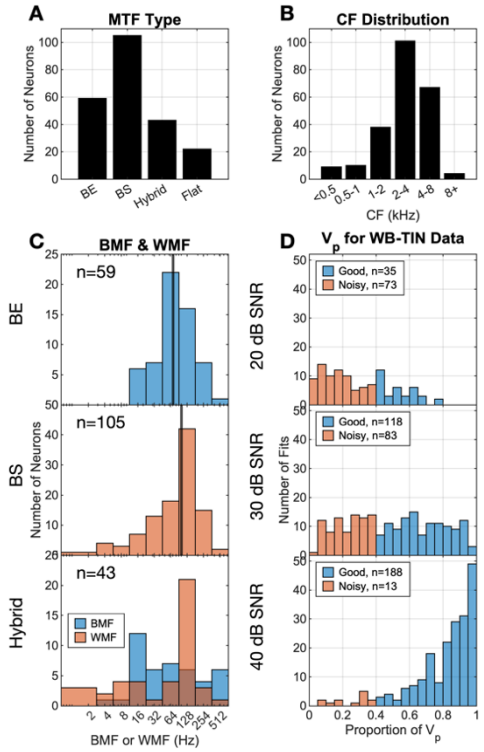

Figure A.1. CF and MTF distributions of neurons, n = 229. (A) Distributions for each MTF type. (B) CF distribution of dataset, in octaves. (C) Distribution of BMFs for BE neurons, median = 70 Hz. Distribution of WMFs for BS neurons, median = 102 Hz. Distributions of BMFs and WMFs for hybrid neurons, BMF median = 64 Hz, WMF median = 125 Hz. (D) Predictable variance distributions for 20-, 30-, and 40-dB SNR for binaural N_0= 23dB SPL dataset.

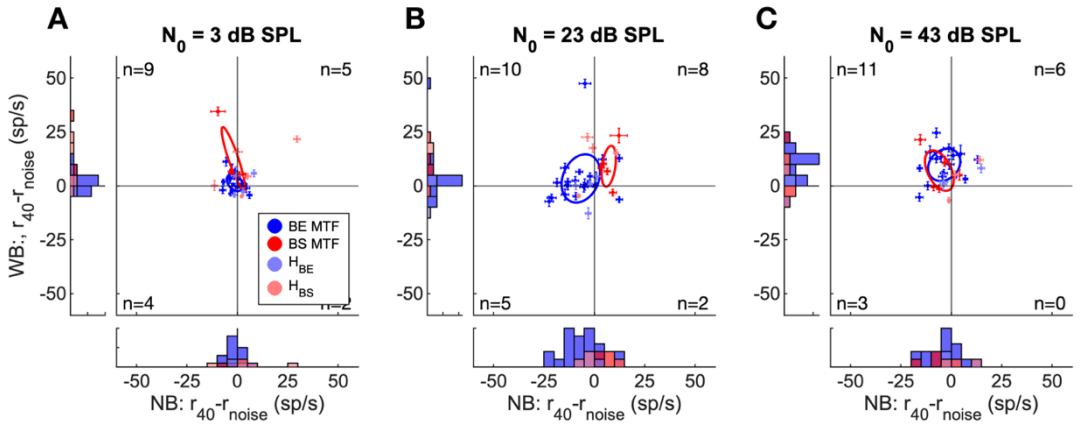

Figure A.2. Differences between average rates in response to contralateral TIN (40 dB SNR) and noise-alone for on-CF tones in wideband vs. narrowband noise. Responses at (A) N0 = 3 dB SPL, WB-TIN overall level range of 34-48 dB SPL, NB-TIN level range of 28-41 dB SPL, (B) 23 dB SPL, WB-TIN overall level range of 54-69 dB SPL, NB-TIN level range of 48-63 dB SPL, and (C) 43 dB SPL, WB-TIN overall level range of 74-83 dB SPL, NB-TIN level range of 68-83 dB SPL. BE MTFs are blue, BS MTFs are red, HBE are light blue, HBS are pink. Error bars represent +/- 1 SEM. Ellipses represent the standard deviation for BE (blue) and BS (red) neurons. Number of neurons in each quadrant is reported in the corner of each quadrant.

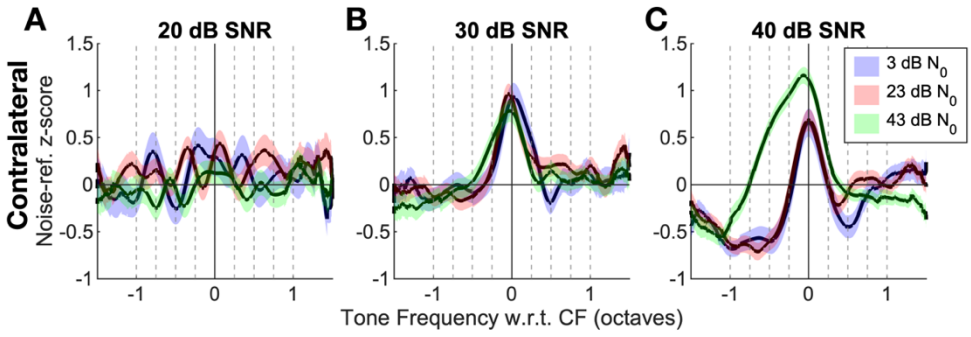

Figure A.3. Population average responses to contralateral WB-TIN stimuli. (A) 20-dB-SNR condition, (B) 30-dB-SNR, (C) 40-dB-SNR, blue curve is the N_0 = 3-dB-SPL condition, pink curve is the N_0 = 23-dB-SPL condition, and green is the N_0 = 43-dB-SPL condition. All data was normalized by noise-referenced z-score. Dotted grey lines

represent the tone frequencies tested in the linear mixed model. Colored bands indicate +/-1 SEM for each population average.

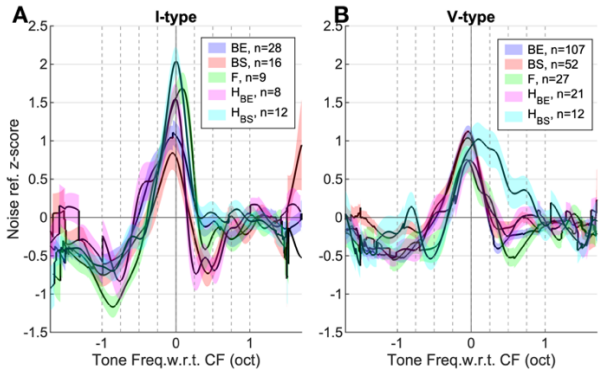

Figure A.4. Population average responses to WB-TIN separated by neural characteristics. Responses averaged over contra/diotic presentation and spectrum-level presentation (3, 23, 43 dB SPL, 40-dB-SNR condition) and normalized using a noise-referenced z-score. (A) I-type and (B) V-type responses for BE (blue), BS (red), flat (green), HBS (pink), and HBE (cyan). Dotted grey lines represent the tone frequencies tested in the linear mixed model. Colored bands indicate +/-1 SEM for each population average.

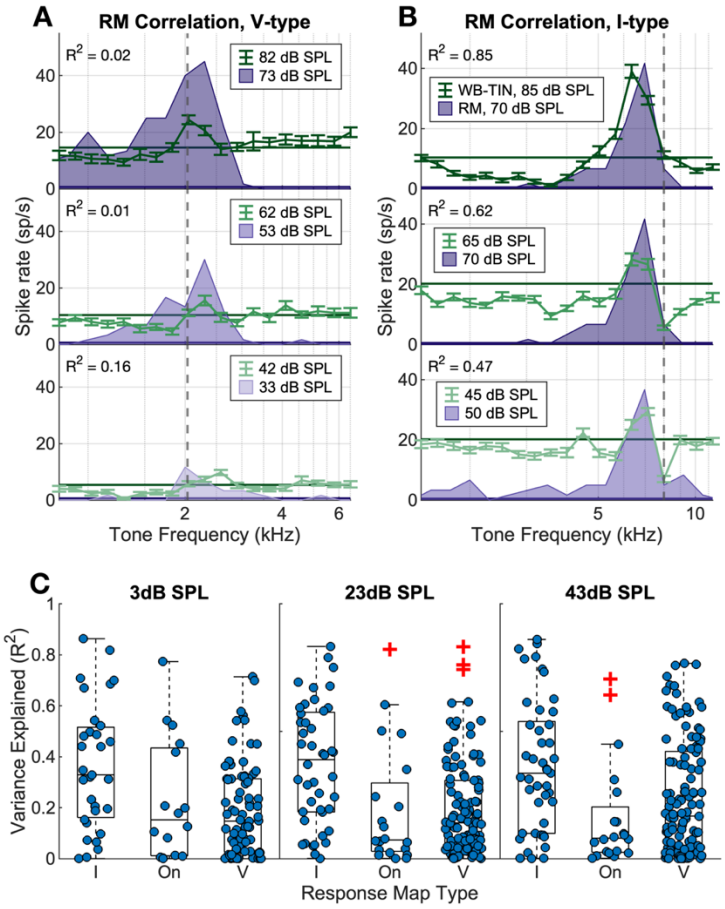

Figure A.5. Variance explained in the WB-TIN responses by neuron responses to pure tones for diotic stimuli. (A) Example V-type unit, CF = 2036 Hz, pure tone responses at 33, 53, and 73 dB SPL in purple. Response to WB-TIN at three levels, N_0 = 3, 23, 43 dB SPL in green. (B) Example I-type neuron, presentation, CF = 8000 Hz. (C-E) Variance explained for all units, condition. Neurons were split into three groups: V-type, I-type, and onset.

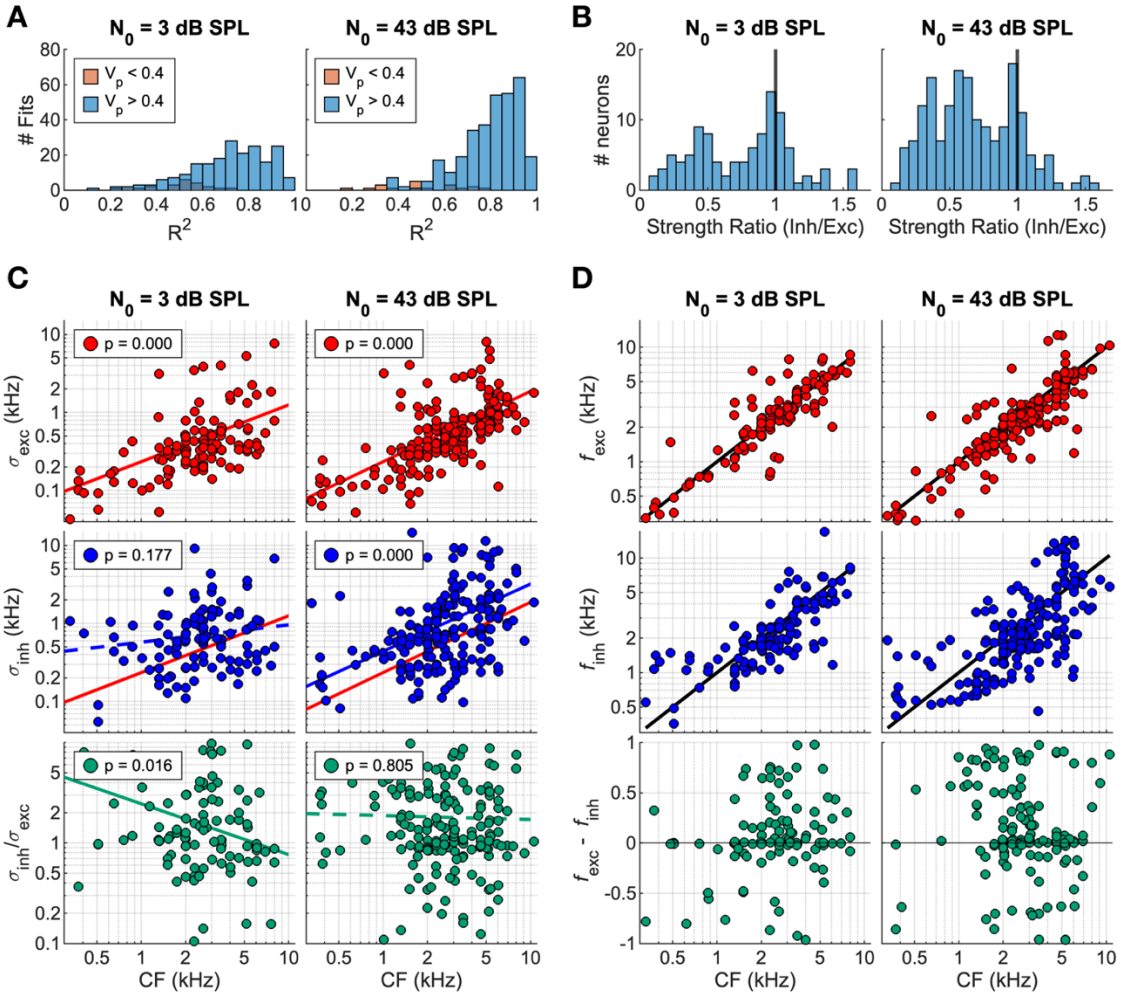

Figure A.6. Difference of gaussians analysis for 3- and 43-dB SPL. (A) Histogram of R^2 values fit to all neurons in the diotic, 40-dB SNR, 3 dB SPL (left) and 23 dB SPL (right) condition. Orange indicates neurons with a predictable variance less than 0.4. (B) Strength ratio (inh/exc) for all fit neurons. (C) Bandwidth, σ, for excitatory (red, top) and inhibitory (blue, middle) gaussians. σ_{exc} fit lines in red (top, middle), σ_{inh} fit line in blue (middle). Bottom row, green, indicates the ratio of inhibitory to excitatory bandwidths. (D) Center frequencies for excitatory (red, top) and inhibitory (blue, middle) gaussians as a function of CF. Difference between excitatory and inhibitory center frequencies in green, bottom row.

Appendix B: Chapter 3 Supplemental Material

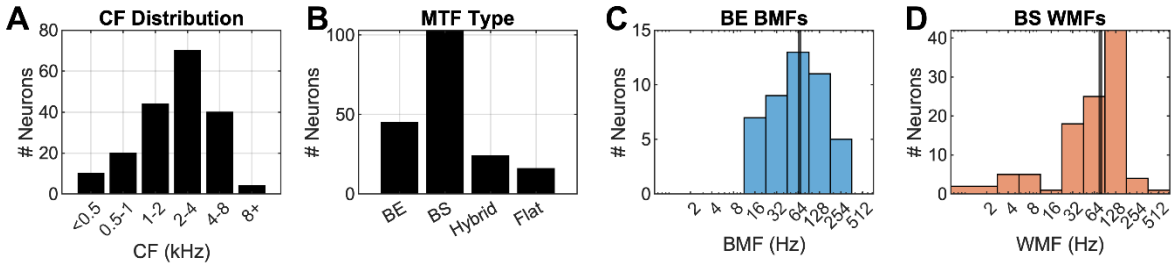

Figure B.1. CF and MTF distribution of neurons. (A) CF distribution. (B) MTF type distribution. (C) Best modulation frequencies for BE neurons, black line indicates median. (D) Worst modulation frequency for BS neurons, black line indicates median.

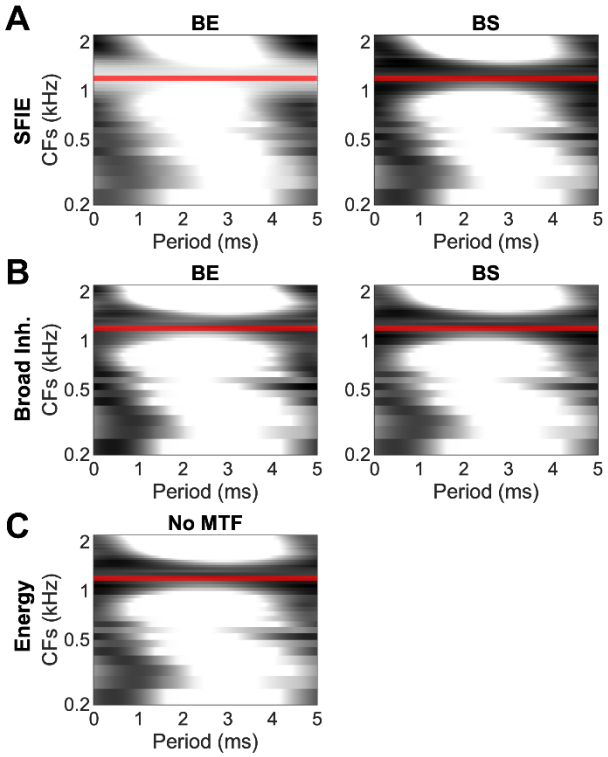

Figure B.2. Model period histograms for (A) SFIE model, BE (left) and BS (right) example neurons with a CF = 1200 Hz and a stimulus level of 63 dB SPL. (B) Broad inhibition model, BE and BS. (C) Energy model, only one plotted due to lack of MTFs in energy model.

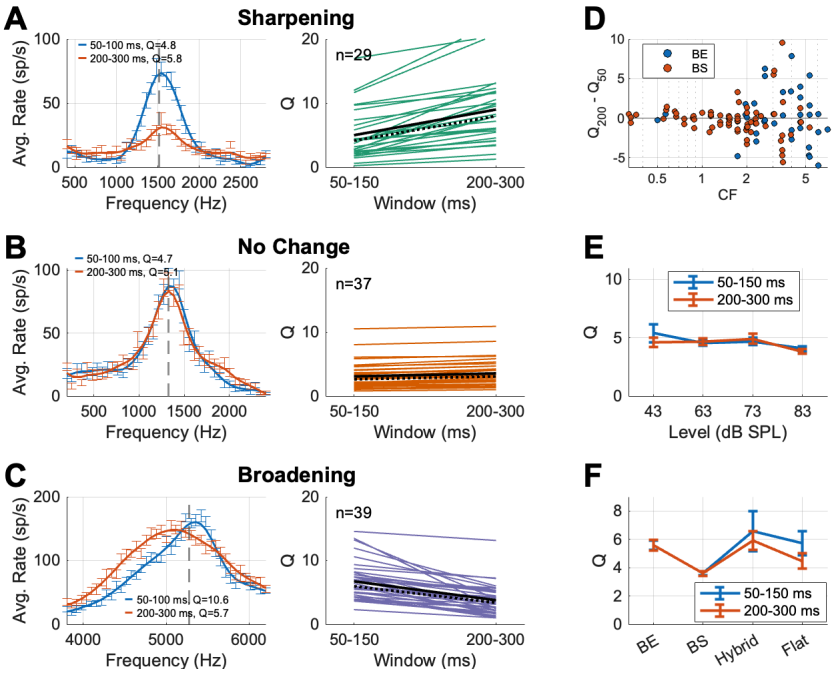

Figure B.3. Changes in average rate over the duration of the synthetic timbre stimulus. (A) Neurons that sharpen in Q value over time, example neuron (left) and population (right). Criteria for sharpening is Q must increase by 1. (B) Neurons that do not change in Q over time, example neuron (left) and population Q (right). (C) Neurons that broaden over time, example neuron (left) and population Q (right). (D) Change in Q plotted as a function of CF for BE (blue) and BS (red) neurons. (E) Q as a function of level for two windows (50-100 ms and 200-300 ms), no significant differences. (F) Q as a function of MTF type again for two windows, no significant differences in Q for windows.

Appendix C: Chapter 4 Supplemental Material

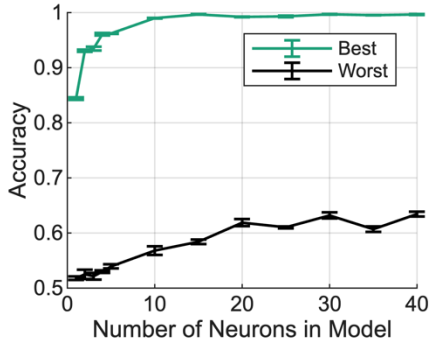

Figure C.1. Decoding instrument identity using timing information in a population of neurons. Neurons were added to the model from best to worst (green) or worst to best (grey) sorted by single neuron accuracy in the decoding task using timing.

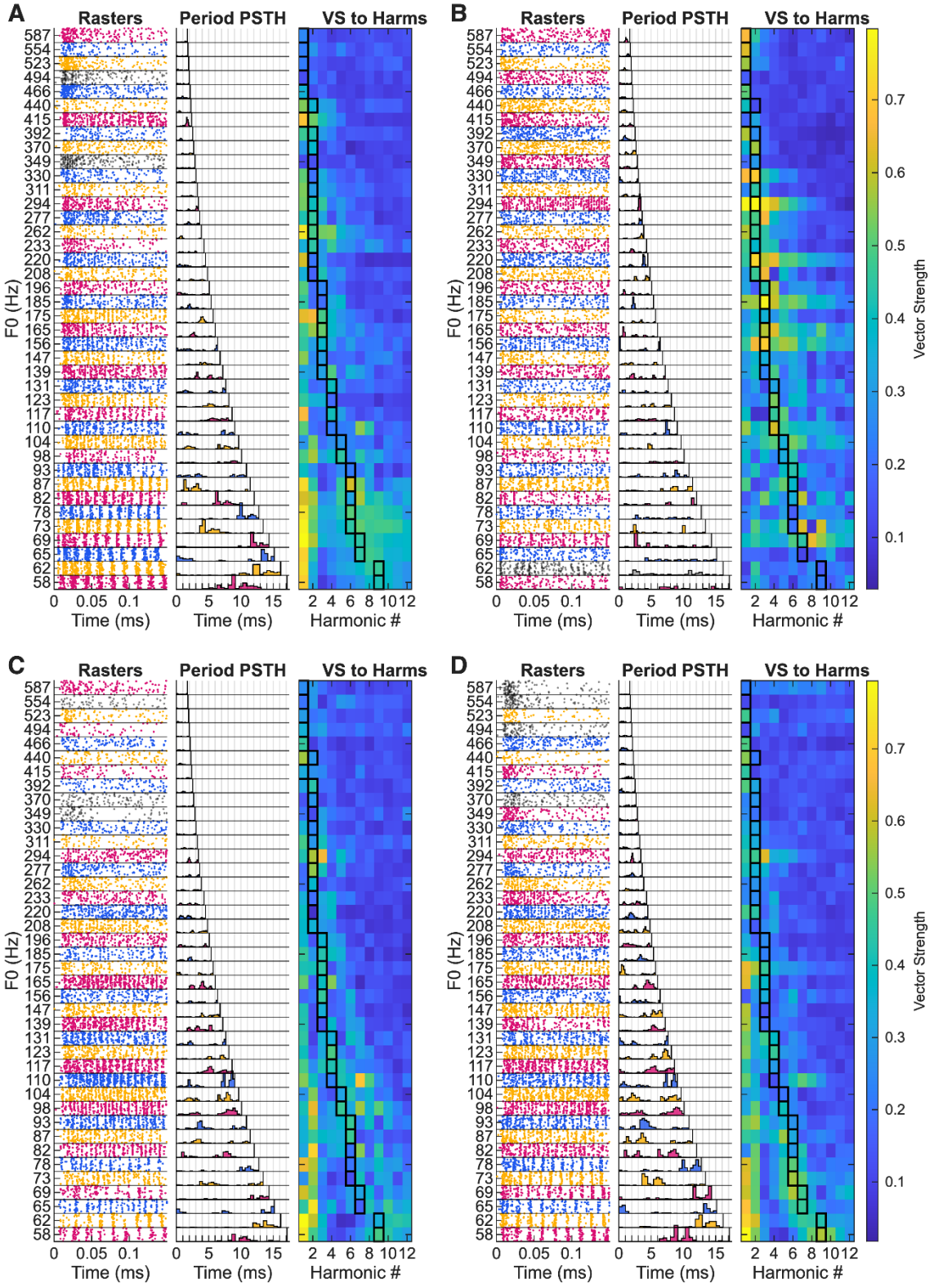
A
B
C
D
Rasters
Period PSTH
VS to Harms
F0 (Hz)
Time (ms)
Harmonic #
Vector Strength

Figure C.2. Four neuron temporal responses to bassoon stimuli. Dot rasters (left), period histogram (middle) and vector strength to the first 12 harmonics (right), grey rasters indicate non-significant phase locking to the F0. Black squares indicate peak harmonic of each stimulus. (A) CF = 6966 Hz, Hybrid. (B) CF = 2629 Hz, BS. (C) 6063 Hz, BE. (D) 3482 Hz, BE.

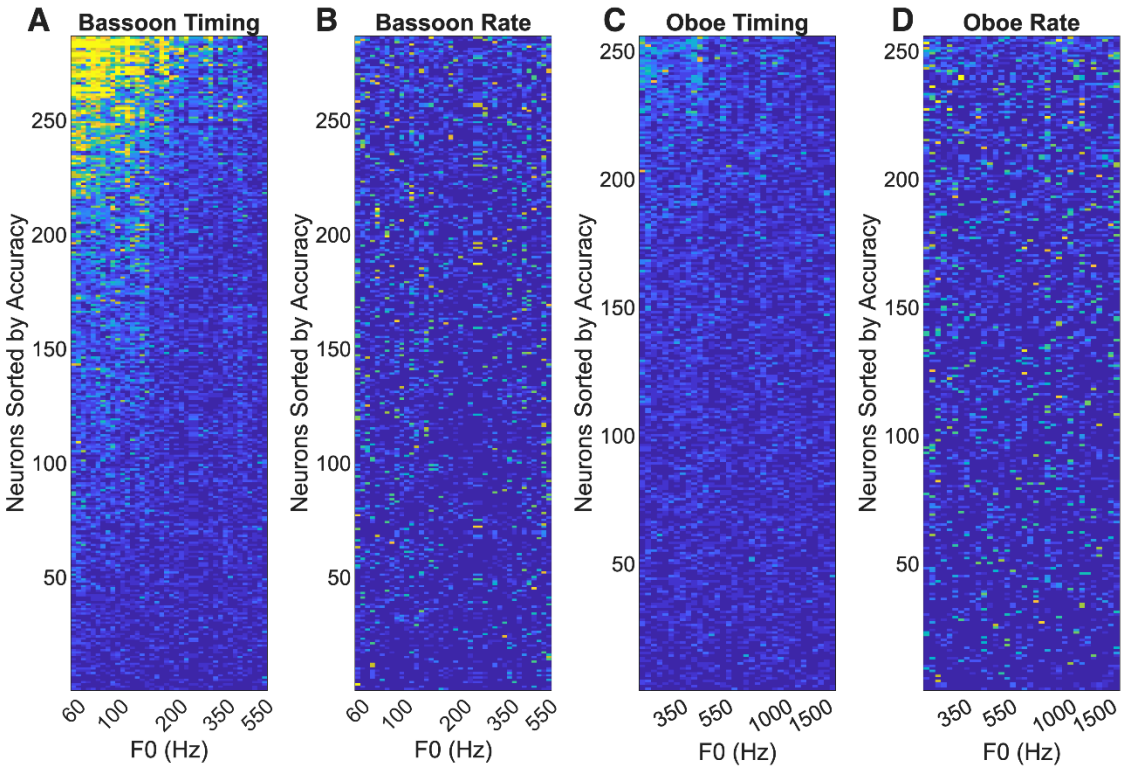

Figure C.3. Decoding accuracy for each F0 for all neurons. Sorted so that the most accurate decoding of F0 is at the top. (A) Decoding results for bassoon F0 identification using timing. (B) Decoding results for bassoon F0 identification using rate. (C) Decoding results for oboe F0 identification using timing. (D) Decoding results for oboe F0 identification using rate.