NEURAL REPRESENTATIONS OF NATURAL SPEECH IN A CHINCHILLA MODEL OF NOISE-INDUCED HEARING LOSS

by

Satyabrata Parida

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Weldon School of Biomedical Engineering

West Lafayette, Indiana

December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Michael G. Heinz, Chair

Speech, Language, and Hearing Sciences, Weldon School of Biomedical Engineering

Dr. Edward L. Bartlett

Department of Biological Sciences, Weldon School of Biomedical Engineering

Dr. Hari M. Bharadwaj

Speech, Language, and Hearing Sciences, Weldon School of Biomedical Engineering

Dr. Keith R. Kluender

Speech, Language, and Hearing Sciences

Approved by:

Dr. George R Wodicka

Head of the Graduate Program

To baba, mommy, and bhai

ACKNOWLEDGMENTS

Working towards this dissertation has been one of the most enjoyable, challenging, and fulfilling experiences of my life, and for that I owe gratitude to many.

First and foremost, I would like to thank my advisor, Dr. Michael Heinz, for his guidance, patience, and constant encouragement throughout my PhD. The unrestricted independence he granted me to shape my projects, in addition to his unbounded enthusiasm for every single aspect of these projects, propelled me to push my limits as a researcher. Over the years, he has offered much essential advice often through amusing anecdotes. Beginning my first semester when I took his "Auditory Periphery" course, he instilled in me the importance of the ability as an engineer/scientist to communicate with the general audience. He has helped me appreciate the power of objectivity in the scientific process. Some of the phrases he liked to use were "the data are the data" and "the data speak for themselves". His influence extends beyond just academic aspects as I strive towards the level of passion, kindness, and work-life balance he has managed to set.

I am extremely fortunate to have a committee of knowledgeable, passionate, and affable mentors. Hari Bharadwaj had a consistent impact on my work and direction. He helped me understand the nature and origins of evoked responses, and introduced me to numerous quantitative tools for evoked-responses as well as for single-unit data. Ed Bartlett has always shown a keen interest in my work, has provided many critical insights at several stages of these projects, and has encouraged me to think about both peripheral and central aspects of sound encoding. I have thoroughly enjoyed my discussions with Keith Kluender, starting from his "Speech Perception" course early on in my program to the many impromptu scribbling sessions we had (on any writable wall that was nearby). Our discussions have helped me appreciate the complexities of speech even more.

Ravi Krishnan has been a kind and influential mentor to me. I learnt a great deal about the auditory system by taking the courses he offered ("Neural basis of hearing" and "Neural Codes: From Auditory Nerve to Cortex"). In addition, I thank him for always being so welcoming, showing excitement to discuss my data, and entertaining all my scientific curiosities; I benefited immensely from his insights.

Mark Sayles helped me by providing constructive feedback, surgical tips, and neurophysiological insights early on in my PhD. Josh Alexander offered many psychoacoustics and modeling insights, which expanded my understanding of our neurophysiological data and the clinical implications of this work. I would also like to thank Beth Strickland for her support and feedback at various stages during my PhD.

Mike Walls and Vijay Prakash patiently trained me to perform the not-so-trivial auditory-nerve experiments, and for that I am grateful. I will cherish the moments we spent in the deep despair of a failed experiment and the immense joy of hearing periodic bursts from that first auditory-nerve-fiber recording after a long exhausting day. I would also like to thank Amanda, who carried our lab for the past two years, for all her help with my training. Mike, Vijay, and Amanda have also been amazing friends.

Early on in my PhD, I benefited from our collaboration with the Hearing Systems group at Technical University of Denmark (DTU). It was an amazing experience to host Christoph Scheidiger and Suyash Joshi here at Purdue, who introduced me to several subfields of hearing research. Even more beneficial was my month-long visit to DTU, which exposed me to a wide variety of hearing research. I would like to thank Christoph, Suyash, Torsten Dau, and Johannes Zaar for hosting me and giving me the opportunity to discuss my research.

I am grateful for the many excellent individuals who have been a part of the Auditory Neurophysiology and Modeling Lab and made my lab experience truly wonderful. I would like to thank Dave, Jessica, Shannon, Vibha, Hannah, Caitlin, Rav, Ivy, Francois, Andrew, and Jim for all their help and feedback, and more importantly, for their friendship. Chandan, Will, and Rav need special thanks for all the late-nighters we pulled together working at Lyles. I would also like to thank Anusha, Varsha, Agudemu, and Homeira for making working at Lyles so much enjoyable.

I am thankful to Kavya, Shruthi, Shan, Sid, Aayush, Priya, Ayush, and Abhinav for being amazing friends and human beings, and making my stay at West Lafayette quite memorable.

I would like to express my special thanks to my father, Gadadhar Parida, and my mother, Sukanti Parida, for supporting me throughout this long but gratifying journey. Finally, I would like to thank my brother, Priyabrata Parida, for convincing me to pursue a PhD and for always being there to support and guide me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ABR	Auditory brainstem responses
AN	Auditory nerve
ANF	Auditory nerve fiber
CAP	Compound action potential
CF	Characteristic frequency
CN	Cochlear nucleus
DPOAE	Distortion product otoacoustic emissions
DT	Distorted tonotopy
ENV	(Temporal) envelope
FFR	Frequency-following responses
FTC	Frequency tuning curve
HI	Hearing impaired
IHC	Inner hair cell
NH	Normal hearing
NIHL	Noise-induced hearing loss
OHC	Outer hair cell
PSD	Power spectral density
PDF	Probability density function
SI	Speech intelligibility
SNHL	Sensorineural hearing loss
SR	Spontaneous rate
TC	Tuning curve
TFS	Temporal fine structure

ABSTRACT

Hearing loss hinders the communication ability of many individuals despite state-of-the-art interventions. Animal models of different hearing-loss etiologies can help improve the clinical outcomes of these interventions; however, several gaps exist. First, translational aspects of animal models are currently limited because anatomically and physiologically specific data obtained from animals are analyzed differently compared to noninvasive evoked responses that can be recorded from humans. Second, we lack a comprehensive understanding of the neural representation of everyday sounds (e.g., naturally spoken speech) in real-life settings (e.g., in background noise). This is even true at the level of the auditory nerve, which is the first bottleneck of auditory information flow to the brain and the first neural site to exhibit crucial effects of hearing-loss.

To address these gaps, we developed a unifying framework that allows direct comparison of invasive spike-train data and noninvasive far-field data in response to stationary and nonstationary sounds. We applied this framework to recordings from single auditory-nerve fibers and frequency-following responses from the scalp of anesthetized chinchillas with either normal hearing or noise-induced mild-moderate hearing loss in response to a speech sentence in noise. Key results for speech coding following hearing loss include: (1) coding deficits for voiced speech manifest as tonotopic distortions without a significant change in driven rate or spike-time precision, (2) linear amplification aimed at countering audiometric threshold shift is insufficient to restore neural activity for low-intensity consonants, (3) susceptibility to background noise increases as a direct result of distorted tonotopic mapping following acoustic trauma, and (4) temporal-place representation of pitch is also degraded. Finally, we developed a noninvasive metric to potentially diagnose distorted tonotopy in humans. These findings help explain the neural origins of common perceptual difficulties that listeners with hearing impairment experience, offer several insights to make hearing-aids more individualized, and highlight the importance of better clinical diagnostics and noise-reduction algorithms.

1. INTRODUCTION

1.1 Hearing loss is a pervasive and complicated problem

Our auditory system possesses exquisite spectrotemporal processing abilities that facilitate seamless communication even in extremely adverse listening environments. Although intricate neural circuitries are dedicated to achieving this feat, these abilities originate in the inner ear, where specialized sensors are housed in a small structure called the cochlea, embedded within the densest bone in the body. These sensors encode information that allows the auditory system to encode sounds spanning a 120 dB intensity range and a 20 kHz frequency bandwidth at microsecond precision (Klumpp and Eady, 1956; Robles and Ruggero, 2001). However, the inner ear, which is inarguably a bioengineering marvel, is a delicate structure and susceptible to injury in numerous ways. Exposure to loud sounds (e.g., frequently attending loud concerts), aging, and ototoxic medications can cause significant damage to the inner ear. The most common consequence of such damage is a reduced hearing sensitivity, i.e., a reduced ability to hear soft sounds (i.e., hearing loss). For example, compared to the sound amplitude threshold for normal-hearing listeners, a sound has to be amplified by a factor of 30 (mild hearing loss), 100 (moderate hearing loss), or even 1000000 (profound hearing loss) for it to be audible by listeners with hearing impairment (Clark, 1981). Hearing loss often accompanies a reduced audible-frequency range of hearing and a narrower intensity dynamic range (Moore, 2007).

Hearing loss can affect family, social, and/or professional lives of those afflicted, with disabling effects in 6% of the world's population (WHO, 2019; estimated to be 10% by 2050). In the US, hearing loss directly affects 23% of the population over the age of 12 (Goman and Lin, 2016). Hearing loss substantially impairs the communication ability of many patients, which can lead to social isolation, depression, and other neurological disorders (Arlinger, 2003; Dawes et al., 2015; Gopinath et al., 2009; Mener et al., 2013), and even increased mortality risk (Karpa et al., 2010). The recent trends paint a sobering picture in the aging population of this ever-noisier planet, and therefore hearing loss has been identified as a serious public health issue (Davis and Hoffman, 2019).

The best help for hearing loss is using a hearing aid, which amplifies sounds to improve audibility for listeners with hearing loss, although it does not restore hearing to normal. Hearing aids or other hearing prostheses aim to restore sufficient communication ability so that listeners can continue their daily life. However, only one out of six people who would benefit from a hearing aid use one (World Health Organization, 2020). Part of the reason is the limited benefit hearing aids provide at present (McCormack and Fortnum, 2013). In particular, listeners with HI particularly struggle in noisy environments, where the target is not clear and background noise can be distracting to the point it is annoying and stressful (Hicks and Tharpe, 2002; McCormack and Fortnum, 2013; Pichora-Fuller et al., 2016). All this goes to show that hearing loss is a complicated problem; there is no “one glass fits all” solution, and we have to improve our understanding of the various effects of hearing loss.

1.2 Audiograms do not predict everyday speech perception difficulties

A key factor contributing to the limited success of hearing aids is that they are largely fit based on audiograms. Audiograms are clinically used to measure hearing sensitivity, where listeners are asked to detect pure tones in a quiet sound booth. However, audiograms do not predict speech perception of abilities of many individuals. In fact, listeners with similar audiograms can demonstrate widely varying speech understanding ability (Middelweerd et al., 1990). Unsurprisingly, amplification strategies aimed at compensating for audiogram-based hearing threshold loss for tones do not restore perception of complex signals, like speech (Dubno et al., 1982; Plomp, 1994). Therefore, we need to understand the effect of hearing loss on how the auditory system processes suprathreshold sounds, especially in noisy environments.

1.3 Animal models can be used to thoroughly characterize sensorineural hearing loss anatomically, physiologically, and behaviorally

Animal models provide an excellent window into obtaining physiological and anatomical data following hearing loss, which is not possible in humans. Anatomical studies can identify the most vulnerable structures in the inner ear, which can ultimately help with pharmacological interventions using precision medicine (Kujawa and Liberman, 2009; Liberman and Dodds, 1984b). Physiological insight can guide the development of optimal hearing-aid strategies to enhance the most informative speech features (Heinz, 2015; Sachs et al., 2002). Additionally, animal models of well-specified sensorineural hearing loss phenotypes can be induced in the controlled settings of a laboratory; this helps with minimizing individual variability that is often seen in human subjects as their exposure history is often unknown (Wible et al., 2005). Finally, we can study the peripheral representation as well as the central neural circuitry that extract behaviorally crucial features in the same animals. In this regard, chinchillas serve as an excellent model because their frequency hearing range and behavioral thresholds are similar to humans and extensive literature exists on the properties of nuclei along the auditory pathway (Martin, 2012; Trevino et al., 2019).

A key benefit of using animal models pursued in this dissertation is to leverage them to understand speech intelligibility variability in human listeners with sensorineural hearing loss (Heinz, 2015). It is likely that peripheral differences contribute significantly to this variability. For example, listeners with similar audiograms, based on their age and noise-/drug-exposure history, may have different degrees of inner- versus outer-hair-cell loss, accompanied with different degrees of endocochlear potential loss and synaptic degeneration (Dubno et al., 1982; Parthasarathy et al., 2020; Wu et al., 2019). The effects of these different sources are evident at the level of the auditory nerve (AN), which is the bottleneck of auditory information available to the brain. Therefore, studying the neural representation of speech in the AN in systems with normal and impaired functions will be crucial in furthering our understanding of speech-perception deficits with hearing loss.

1.4 Overview of perceptually important speech features

Speech is a highly complex yet structured stimulus, which has information spread out over multiple temporal and spectral scales. In linguistics, the smallest speech units are called phonemes. For example, the word "put" consists of the phonemes /p/, /ʊ/, and /t/). The English language consists of 44 phonemes, which are categorized as either vowels (e.g., /ʊ/) or consonants (e.g., /p/ and /t/). Groups of phonemes form syllables, which in turn form words, and words form sentences. This hierarchical division highlights the multiresolution scales of temporal information in speech.

Spectral properties of speech are readily understood using an articulatory treatment of phonemes. Speech production can be modeled using the source-filter theory, which is a two-stage process (Diehl, 2008; Fant, 1970). The first stage, the source, which includes the larynx, controls the generation of sound. Based on whether the vocal folds in the larynx are vibrating periodically, speech can be either periodic (voiced) or aperiodic (unvoiced). Voiced speech (periodic) is marked by a spectrum that has harmonic structure as voiced speech contains energy only at multiples of the fundamental frequency (inverse of the pitch period or voicing period). In contrast, unvoiced speech uses noise as the source, which lacks any discernible temporal structure and does not have spectral harmonicity. The second stage, the filter, which includes the vocal tract, shapes the source spectrum to add resonances (i.e., formants) and anti-resonances. These resonances and anti-resonances carry critical cues for speech perception (Diehl, 2008; Klatt, 1980). In speech, vowels are usually voiced (e.g., /a/), except when whispered. In contrast, consonants can be either voiced (e.g., /p/ and /l/) or unvoiced (e.g., /b/ and /s/).

Spectral and temporal information in a signal is usually divided into two distinct components: (1) a slowly varying temporal envelope (ENV) and (2) a rapidly varying temporal fine structure (TFS). ENV conveys information regarding the signal's slow intensity fluctuation, which usually reflects word and phonemic structures. TFS conveys information about formants (for voiced speech), and in general reflects the harmonic structure of the stimulus. While ENV is sufficient to facilitate robust speech perception in quiet environments (Shannon et al., 1995; Smith et al., 2002), recent studies have emphasized the role of TFS in adverse listening conditions (Ding et al., 2014; Viswanathan et al., 2019). Furthermore, pitch cues, which convey prosodic information such as emotion and gender of the speaker and help with source segregation, are encoded via TFS (Micheyl and Oxenham, 2010; Moore and Carlyon, 2005). It should be highlighted that features like formant and intensity are dynamic in nature, which means these features are not stationary and change over time (e.g., Fig 3.1). In fact, studies have shown that these dynamic properties aid in speech understanding (Hillenbrand and Nearey, 1999; Nearey and Assmann, 1986). Therefore, we should study how these dynamic features are encoded by the normal auditory systems and how this code is degraded with hearing loss.

1.5 Existing literature on physiological changes in the AN following NIHL

The scope of this thesis is limited to the physiological consequences of noise-induced hearing loss on natural speech coding. Therefore, we begin with a brief review of relevant existing literature on the physiological effects of NIHL. The scope of this review is rather selective. For a detailed review of physiological changes following NIHL, see Sayles and Heinz, 2017.

1.5.1 For tones

The cochlea in a normal-hearing system can be thought of as a bank of narrowband filters that are tonotopically arranged. In other words, different places along the cochlea are sensitive to a narrow range of frequencies. From base to apex, the cochlear locations are progressively tuned from higher to lower frequencies. This tonotopicity is preserved in higher auditory nuclei along the auditory pathway, starting with the auditory nerve. Frequency selectivity of AN fibers is usually measured using frequency tuning curve (FTCs), where each data point denotes the intensity required at that frequency to drive the AN fiber's activity above baseline. As previously mentioned, AN fibers are quite frequency selective (as reflected by a sharp FTC) as well as remarkably sensitive (low threshold at characteristic frequency or CF, the frequency at which the AN fiber is the most sensitive). As a result, information in a broadband stimulus can be robustly encoded by a set of independent frequency channels. However, following NIHL, AN fibers become less sensitive (higher threshold) near CF, FTCs become broader, and AN fibers often become more sensitive at frequencies below CF (Liberman, 1984; Liberman and Dodds, 1984b). These changes are hypothesized to reduce the information capacity of the AN (as fewer independent frequency channels), which can lead to degraded sound coding and perceptual deficits.

1.5.2 For stationary synthesized vowels

Speech coding at the AN level has been predominantly carried out using short stationary synthesized vowels and with normal-hearing animals (Delgutte, 1980; Delgutte and Kiang, 1984a; Palmer, 1990; Sachs and Young, 1979; Sinex and Geisler, 1983; Young and Sachs, 1979). A consistent finding from these studies is that formants, which are important for vowel perception, dominate the responses of AN fibers with CF near that formant frequency. In particular, the temporal aspect of the response of AN fiber population conveys robust tonotopic information regarding vowels. The handful of studies that have evaluated speech coding in the AN following NIHL have found a degradation of highly informative high-frequency formants due to broader tuning at the expense of overrepresented low-frequency stimulus energy (Miller et al., 1999; Miller et al., 1997).

1.5.3 For White Gaussian noise

A number of studies have also investigated general physiological changes, such as the coding of ENV and TFS, following NIHL using broadband White Gaussian noise as stimulus (Henry and Heinz, 2013; Henry et al., 2016). White Gaussian noise, because of its random nature, serves as an excellent stimulus to characterize ENV and TFS coding in a systems-identification framework. There is an enhanced representation of the ENV, although this enhancement is deemed pathological as it is hypothesized to be detrimental for speech perception (Füllgrabe et al., 2003; Moore et al., 1996). Similarly, overall TFS coding "strength" in broadband responses is also enhanced following NIHL (Henry and Heinz, 2013; Henry et al., 2016). However, there is severe tonotopic distortion in the spectral profile of AN fiber responses. This effect of distorted tonotopy is consistent with a decreased sensitivity of FTCs near CF and increased sensitivity at lower (<CF) frequencies. Similar to enhanced envelope, distorted tonotopy is also hypothesized to degrade the neural representation of speech (Henry et al., 2016; Sayles and Heinz, 2017).

1.6 General gaps

This thesis includes experimental results that address the following three broad gaps in the literature regarding the effects of NIHL on the neural coding of speech.

- Our understanding of speech coding in the AN primarily comes from studies that have used short-duration stationary synthesized vowels. However, as described earlier, naturally spoken speech is nonstationary and is characterized by its time-varying amplitude fluctuations and spectral (both formant and fundamental frequency) trajectories, including both vowels and consonants. In fact, these dynamics improve speech perception (Hillenbrand and Nearey, 1999; Nearey and Assmann, 1986). The few rare studies that have used natural speech sentences to record from AN fibers have yielded limited datasets and only used normal-hearing animals (Delgutte et al., 1998; Young, 2008). No study has evaluated consonant coding deficits following NIHL. Therefore, we lack a clear understanding of the neural representation of time-varying statistics naturally present in spoken speech, including vowels and consonants.

- Another major limitation of the existing studies is that stimuli have primarily been presented without background noise, particularly for studies involving impaired animals. While these studies have provided invaluable insight into neural representation of speech for animals with normal and impaired hearing, these experimental settings are not representative of our everyday environments. In fact, the number one complaint in the clinic is regarding difficulties listening in noisy environments, rather than the listening in quiet conditions most often studied (Chung, 2004; Souza, 2016). Therefore, insights from studying the neural representation of speech mixed with noise will help improve the perceptual outcomes in everyday noisy environments.
- Finally, studies have rarely demonstrated similar neural deficits using both invasive and noninvasive data from the same animal cohort. Establishing relations between invasive and noninvasive data is beneficial as we can only record noninvasive far-field evoked data, such as electroencephalograms and frequency following responses, from humans. These far-field responses can be used to infer anatomical and physiological changes in the auditory system following NIHL. Ultimately, the ability to noninvasively assess neural processes underlying human communication will enable pharmacological interventions as well as to objectively guide hearing aid fitting and optimizing stimulation strategies.

1.7 Overview of the dissertation document

In this study, both invasive and noninvasive data were collected from anesthetized chinchillas with either normal hearing or with mild-moderate noise-induced hearing loss, the most common hearing-loss phenotype in the US (Goman and Lin, 2016). Invasive data comprise spike-train responses recorded from single AN fibers. Noninvasive data are based on scalp-recorded frequency following responses. Both data types were collected from the same cohort of animals. A naturally spoken speech sentence was used as the stimulus in most experiments. The speech sentence was presented at conversational sound levels in varying degrees of background noise, including in quiet. Noise was speech shaped; noise was either stationary or was sinusoidally amplitude modulated at 8 Hz. General methods employed in this dissertation are described in Chapter 2).

In Chapter 3, a unifying quantitative framework is described where the same spectrotemporal analyses can be applied to both invasive and noninvasive data, and therefore direct comparison between these data types can be possible. In addition, advanced signal processing tools can be applied to neural spike-train data so that the neural coding of natural (nonstationary) speech can be studied with the same spectral specificity as previously reported studies on stationary speech coding.

In Chapter 4, quantitative tools described in Chapter 3 were applied to spike-train data and frequency following responses to investigate speech coding deficits following NIHL. Specifically, coding changes for vowels and consonants were considered. Four physiological changes following NIHL that potentially contribute to speech perception deficits were investigated. These factors are distorted tonotopy, broader bandwidth, reduced temporal precision, and auditory-nerve-fiber (ANF) threshold compression. Our results suggest that deficits in neural coding primarily stem from distorted tonotopy and, surprisingly, from ANF threshold *expansion*. Broader bandwidths likely contributed but this was a minor contribution compared to distorted tonotopy. Temporal precision of spiking activity was intact following NIHL contrary to common beliefs (Moore, 2008, 2014).

In Chapter 5, the effect of hearing loss on envelope coding of speech and noise mixtures is evaluated. It was hypothesized that, following NIHL, speech-related envelope would be degraded, and noise-related envelope would be enhanced. These hypotheses were tested for stationary unmodulated noise and 8-Hz sinusoidally amplitude modulated noise. Modulation in noise is known to help with speech perception for normal-hearing listeners but not for listeners with hearing impairment. The results supported the hypotheses and provided the neural bases for why adding modulation to noise does not improve speech understanding following NIHL. Distorted tonotopy provided the best explanation for these results.

In Chapter 6, degradation in voice-pitch coding following NIHL is investigated. In western languages, pitch conveys non-semantic cues, such as gender, emotion, and intonation (Ladd, 2008). In other languages, such as Mandarin Chinese, pitch can provide semantic cues (Stagray et al., 1992). While perceptual deficits in pitch processing following NIHL are well documented (Moore and Carlyon, 2005), neural bases underlying these deficits are still unknown (Kale et al., 2013). Here, we quantified the neural representation of pitch using two leading theories- temporal and spectral. Results suggested enhanced temporal and degraded spectral representations of pitch. Furthermore, enhanced temporal representation accompanied complex response patterns, which may adversely affect pitch perception. These results offer new insight into pitch perception deficits that listeners with NIHL experience.

In Chapter 7, a noninvasive measure is presented to diagnose distorted tonotopy using frequency following responses. DT plays an important role in complex sound coding for the NIHL population. Moreover, DT has been shown to vary across hearing-loss etiologies (Henry et al., 2019). Therefore, DT is hypothesized to contribute to individual variability in human speech perception in the hearing-impaired population. The noninvasive diagnostic DT metric will allow us to evaluate this hypothesis in humans.

Finally, Chapter 8 first summarizes the results from the preceding chapters with particular emphasis on how these results relate to common perceptual deficits. Next, implications for precision clinical diagnostics, hearing aid stimulation strategies, and the importance of noise reduction algorithms are described. The chapter ends with a discussion of the limitations of the experiments presented, key unanswered questions, and future research directions to mitigate these limitations and answer these questions.

2. GENERAL METHODS

2.1 Animals preparation

Young chinchillas (6 to 12 months old) weighing between 400 and 650 g were used in all experiments. Experiments were of two types: (1) recovery experiments (to collect screening data or frequency-following responses or to induce noise-induced hearing loss) and (2) acute auditory-nerve experiments. In both recovery and acute procedures, anesthesia was induced with xylazine (2 to 3 mg/kg, subcutaneous), followed a few minutes later with ketamine (30 to 40 mg/kg, intraperitoneal). For recovery procedures, xylazine reversal agent atipamezole (0.4 to 0.5 mg/kg, intraperitoneal) was used after procedures for faster recovery. For acute procedures, anesthesia was maintained with sodium pentobarbital (∼ 15 mg/kg/2 h, intraperitoneal). Eyes were lubricated with eye ointment. Animals' vital signs were monitored throughout procedures with a pulse oximeter (Nonin 8600V, Plymouth, MN). An oxygen tube was placed near animals' nostrils. Animals were provided Lactated Ringer's solution both before and after recovery procedures (each 6 cc, subcutaneous) and throughout acute procedures (∼1 cc/h, subcutaneous). Body temperature was maintained near 37 °C using a closed-loop heating pad with a rectal probe (50-7220F, Harvard Apparatus). All procedures were carried out in a double-walled, electrically shielded, sound-attenuating booth (Acoustic Systems, Austin, TX, USA). All procedures were approved by Purdue Animal Care and Use Committee (PACUC Protocol #: 1111000123).

2.2 Noise exposure

Animals were exposed to 116 dB SPL octave-wide noise centered at 500 Hz, for two hours, using an enclosed woofer (Selenium 10PW3, Harman) facing downwards ∼ 25 cm above the animal's heads. Noise intensity was measured near the animal's pinnae every 20 minutes using a sound level meter to confirm steady exposure levels (Simpson model 886-2, Elgin, IL, USA). Animals were anesthetized during the exposure (as previously described). Two weeks of recovery were allowed after noise exposure to minimize temporary hearing-loss effects.

2.3 Surgical preparation

Following anesthesia induction, tracheotomy was performed to facilitate low-resistance airway and to minimize breathing artifacts. Skin and muscles were transected following a dorsal midline incision, external ear canals and bullae were exposed. Hollow ear bars were inserted into ear canals, which allowed sound delivery to the right ear from a dynamic loudspeaker (DT48, Beyerdynamic). The right bulla was vented with a 30 cm long polyethylene tubing (Guinan and Peake, 1967). A silver electrode was placed near the round window through a small opening, later sealed with dental cement, in the right ventral bulla to monitor compound action potential (CAP) during nerve experiments; single unit experiments were terminated when CAP thresholds and single unit frequency tuning curve thresholds elevated by more than 10 dB. Craniotomy was performed using posterior fossa approach; cerebellum was partially aspirated until cochlear nucleus (CN) was visible. Small cotton pellets were used to push the CN away from the lateral wall to expose the auditory nerve bundle. Glass micropipettes ($> 10M\Omega$) electrodes were placed close to the internal auditory meatus for extra-axonal recording to avoid recordings from CN cells. Monopolar action potential waveforms and spike latency confirmed recordings were made from auditory nerve fibers.

2.4 Speech stimulus

One speech sentence from the Danish speech intelligibility test (CLUE, Nielsen and Dau, 2009) material was used (practice list #1, sentence #3) in many acute and FFR experiments. The sentence was shortened to 1.3 s duration by truncating it at silence intervals to allow more stimulus repetitions (Fig 2.1). Experiment-specific stimuli are discussed in individual chapters.

2.5 Data recordings for recovery procedures

A microphone-transducer pair (Etymotic ER-10B, Etymotic ER-2, Etymotic Research, Elk Grove Village, IL, USA) was used for calibration and sound presentation, with a foam ear-tip inserted into the ear canal. The transducer speaker was calibrated at the beginning

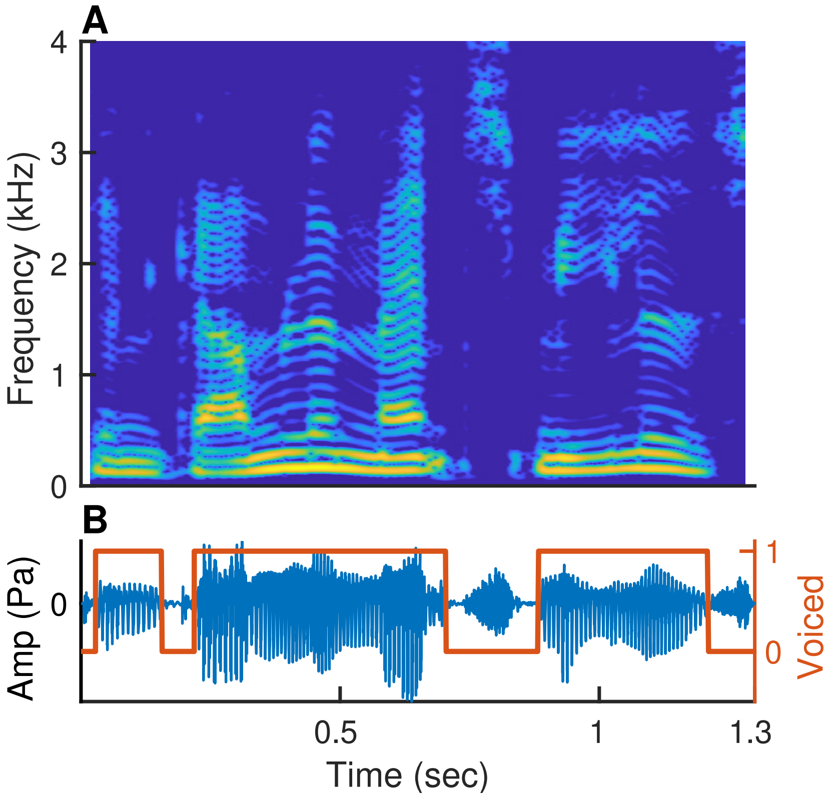

Figure 2.1. **A** The spectrogram of a speech sentence used in this dissertation, with warm (cool) colors representing regions of high (low) power, **B** time-domain waveform (blue), with voiced segments in the stimulus marked in red.

of all experiments. Calibration data showed frequency response was flat (within ± 5 dB) until 10 kHz. In later experiments, a 256-tap digital FIR filter (Tucker-Davis Technologies, Alachua, FL) was used to ensure calibration was even flatter (within ± 2 dB until 10 kHz). Auditory brainstem responses (ABR) and distortion product otoacoustic emissions (DPOAE) were used as screening procedures. ABRs were recorded to short tone bursts at .5, 1, 2, 4, and 8 kHz, at 0 dB SPL to 80 dB SPL in 10 dB steps. Another intensity near threshold (odd multiple of 5 dB) was used to fine-tune threshold estimate. ABR threshold was estimated as the intensity at which linear regression of waveform correlation of ABRs at intensities below 60 dB SPL with a template ABR (response waveform at 60 dB SPL) was three standard deviations above the correlation noise floor. Distribution of the correlation noise floor was constructed from the spurious peaks in the cross-correlation function between the template and physiological noise in no-response regions of the recordings (Henry et al., 2011).

DPOAEs were recorded in response to sets of two tones, ranging in frequency from .5 kHz to 16 kHz, with a frequency ratio of 1.2, played at 75 (tone with lower frequency) and 65 dB SPL.

ABRs and FFRs were recorded using subdermal needle electrodes in a vertical montage (mastoid to vertex, differential mode, with the common ground placed near the nose). Both ABR and FFR signals were filtered through ISO-80 (5 Hz-10 kHz, 100 gain; World Precision Instruments, Sarasota, FL), Dagan (open filter, 200 gain; 2400A, Minneapolis, MN). ABRs were further filtered in the band 0.3-3 kHz (Krohn-Hite, 3550, Brockton, MA). FFRs were bandpass filtered offline between 70-500 Hz.

2.6 Data recordings for acute procedure

Sound was presented using dynamic loudspeaker (DT48, Beyerdynamic) for acute experiments as mentioned above. CAP signal was bandpass filtered and amplified (0.1-10 kHz; 1000 gain; World Precision Instruments ISO-80, Sarasota, FL) before storing on PC. Glass micropipettes were advanced using a hydraulic microdrive (640, Kopf Instruments). Neuronal recordings were amplified (Dagan 2400A, Minneapolis, MN), and bandpass filtered between 0.3-6 kHz (Krohn-Hite 3550, Brockton, MA)). Spike timings were identified using a time-amplitude window discriminator (BAK Electronics) and counter boards (National Instruments, Austin, TX, USA). Individual auditory nerve fibers were isolated using a wide-band noise burst and listening for driven activity. An automated tuning curve algorithm was used to characterize fibers' frequency selectivity by estimating threshold at different frequencies (Chintanpalli and Heinz, 2007; Kiang et al., 1970). The criteria for threshold at a frequency was the minimum tone intensity that elicits a difference of at least 1 spike between 50 ms of the tone and 50 ms of silence. Spontaneous activity was recorded for 30 seconds, followed by a rate-level curve to tones at characteristic frequency at intensities of 0 dB till rate saturation in 1 dB steps. Then speech, noise, and noisy-speech mixtures were played in an interleaved fashion.

In both recovery and acute procedures, stimulus presentation and data acquisition were controlled using custom and commercial hardware (Tucker-Davis Technologies, Alachua, FL) and custom-made GUI written in MATLAB (The MathWorks, Natick, MA).

3. SPECTRALLY SPECIFIC TEMPORAL ANALYSES OF SPIKE-TRAIN RESPONSES TO COMPLEX SOUNDS: A UNIFYING FRAMEWORK

SUMMARY[1]

Significant scientific and translational questions remain in auditory neuroscience surrounding the neural correlates of perception. Relating perceptual and neural data collected from humans can be useful; however, human-based neural data are typically limited to evoked far-field responses, which lack anatomical and physiological specificity. Laboratory-controlled preclinical animal models offer the advantage of comparing single-unit and evoked responses from the same animals. This ability provides opportunities to develop invaluable insight into proper interpretations of evoked responses, which benefits both basic-science studies of neural mechanisms and translational applications, e.g., diagnostic development. However, these comparisons have been limited by a disconnect between the types of spectrotemporal analyses used with single-unit spike trains and evoked responses, which results because these response types are fundamentally different (point-process versus continuous-valued signals) even though the responses themselves are related. Here, we describe a unifying framework to study temporal coding of complex sounds that allows spike-train and evoked-response data to be analyzed and compared using the same advanced signal-processing techniques. The framework uses alternating-polarity peristimulus-time histograms computed from single-unit spike trains to allow advanced spectral analyses of both slow (envelope) and rapid (temporal fine structure) response components. Demonstrated benefits include: (1) generalization beyond classic metrics of temporal coding, e.g., vector strength and correlograms, (2) novel spectrally specific temporal-coding measures that are less corrupted by distortions due to hair-cell transduction, synaptic rectification, and neural stochasticity compared to previous metrics, e.g., the correlogram peak-height, (3) spectrally specific analyses of spike-train modulation coding that can be directly compared to perceptually based models of speech intelligibility, and (4) superior spectral resolution in analyzing the neural representation of nonstationary sounds, such as speech and music. This unifying framework significantly

[1] Manuscript in revision (PLOS Computational Biology)

expands the potential of preclinical animal models to advance our understanding of the physiological correlates of perceptual deficits in real-world listening following sensorineural hearing loss.

3.1 Introduction

Normal-hearing listeners demonstrate excellent acuity while communicating in complex environments. In contrast, hearing-impaired listeners often struggle in noisy situations, even with state-of-the-art intervention strategies (e.g., digital hearing aids). In addition to improving our understanding of the auditory system, the clinical outcomes of these strategies can be improved by studying how the neural representation of complex sounds relates to perception in normal and impaired hearing. Numerous electrophysiological studies have explored the neural representation of perceptually relevant sounds in humans using evoked far-field recordings, such as frequency following responses (FFRs) and electroencephalograms (Clinard et al., 2010; Kraus et al., 2017; Tremblay et al., 2006). Note that we use *electrophysiology* and *neurophysiology* to refer to evoked far-field responses and single-unit responses, respectively (See Table 3.1 for glossary). While these evoked responses are attractive because of their clinical viability, they lack anatomical and physiological specificity. Moreover, the underlying sensorineural hearing loss pathophysiology is typically uncertain in humans. In contrast, laboratory-controlled animal models of various pathologies can provide specific neural correlates of perceptual deficits that humans experience, and thus hold great scientific and translational (e.g., pharmacological) potential. In order to synergize the benefits of both these approaches to advance basic-science and translational applications to real-world listening, two major limitations need to be addressed.

First, there exists a significant gap in relating spike-train data recorded invasively from animals and evoked noninvasive far-field recordings feasible in humans (and animals) because the two signals are fundamentally different in form (i.e., binary-valued point-process data versus continuous-valued signals). While the continuous nature of the evoked-response amplitude allows for any of the advanced signal-processing techniques developed for continuous-valued signals to be applied (e.g., multitaper approaches to robust spectral estimation (Thomson, 1982), spike-train analyses have been much more limited (e.g., in their application to real-world signals as reviewed in the next section). This is a critical gap because most perceptual deficits and limits in machine hearing occur for speech in noise rather than for speech in quiet (Moore, 2007; Scharenborg, 2007). For example, classic neurophysiological studies have quantified the temporal coding of stationary and periodic stimuli using metrics such as vector strength (VS, (Goldberg and Brown, 1969; Joris and Yin, 1992; Rees and Palmer, 1989), whereas more recent correlogram analyses have provided temporal-coding metrics for nonperiodic stimuli, such as noise (Joris et al., 2006; Louage et al., 2004). However, these metrics can be influenced by distortions from nonlinear cochlear processes (Heinz and Swaminathan, 2009; Young and Sachs, 1979, and often ignore response phase information that is likely to be perceptually relevant for simple tasks (Colburn et al., 2003 as well as for speech intelligibility (Paliwal and Alsteris, 2003; Relaño-Iborra et al., 2016).

A second important gap exists because current spectrotemporal tools to evaluate temporal coding in the auditory system are largely directed at processing of stationary signals by linear and time-invariant systems. However, the auditory system exhibits an array of nonlinear (e.g., two-tone suppression, compressive gain, and rectification) and time-varying (e.g., adaptation and efferent feedback) mechanisms (Heil and Peterson, 2015; Sayles and Heinz, 2017). These mechanisms interact with nonstationary stimulus features (e.g., frequency transitions and time-varying intensity fluctuations, Figs 3.1A and 3.1B) to shape the neural coding and perception of these signals (Delgutte, 1997; Hillenbrand and Nearey, 1999; Nearey and Assmann, 1986). In fact, the response of an auditory-nerve (AN) fiber to even a simple stationary tone shows nonstationary features, such as a sharp onset and adaptation (Fig 3.1C), illustrating the need for nonstationary analyses of temporal coding. However, the extensive single-unit speech coding studies using classic spike-train metrics have typically been limited to synthesized and stationary speech tokens, which has deferred the study of the rich kinematics present in natural speech (Delgutte, 1980; Sinex and Geisler, 1983; Young and Sachs, 1979). Some windowing-based approaches have been used to study time-varying stimuli and responses (Cariani and Delgutte, 1996a; Sayles and Winter, 2008), but the approaches used have imposed a limit on the temporal and spectral resolution with which dynamics of the auditory system can be studied.

The present study focuses on developing spectrotemporal tools to characterize the neural representation of kinematics naturally present in real-world signals, speech in particular, that are appropriate for the nonlinear and time-varying auditory system. We describe a unifying framework to study temporal coding in the auditory system, which allows direct comparison of single-unit spike-train responses with evoked far-field recordings. In particular, we demonstrate the unifying merit of using alternating-polarity peristimulus time histograms (*apPSTHs*, Table 3.2), a collection of PSTHs obtained from responses to both positive and negative polarities of the stimulus. By using both polarities, neural coding of natural sounds can be studied using the common temporal dichotomy between the slowly varying envelope (ENV) and rapidly varying temporal fine structure (TFS) (Figs 3.1E and 3.1F), which has been especially relevant for speech-perception studies (Shannon et al., 1995; Smith et al., 2002). Here, we first review some of the existing tools that have been used to quantify temporal coding in auditory neurophysiology. We derive explicit relations between existing metrics, namely VS and correlograms, and *apPSTHs* to show that no information is lost by using *apPSTHs*. In fact, the use of *apPSTHs* is computationally more efficient, provides more precise spectral estimators, and opens up new avenues for perceptually relevant analyses that are otherwise not possible. Next, an *apPSTH*-based ENV/TFS taxonomy is presented, including existing and new metrics. This taxonomy allows for spectrally specific analyses that avoid distortions due to inner-hair-cell transduction and synaptic rectification processes, resulting in more accurate characterizations of temporal coding than with previous metrics. Finally, these methods are extended in novel ways to include the study of nonstationary signals at superior spectrotemporal resolution compared to conventional windowing-based approaches, like the spectrogram.

3.2 Materials and Methods

3.2.1 Experimental procedures and neuro/electrophysiological recordings

Spike trains were recorded from single AN fibers of anesthetized chinchillas using standard procedures in our laboratory (Henry et al., 2019; Kale and Heinz, 2010). All procedures followed NIH-issued guidelines and were approved by Purdue Animal Care and

Use Committee (Protocol No: 1111000123). Anesthesia was induced with xylazine (2 to 3 mg/kg, subcutaneous) and ketamine (30 to 40 mg/kg, intraperitoneal), and supplemented with sodium pentobarbital ($\sim$7.5 mg/kg/hour, intraperitoneal). FFRs were recorded using subdermal electrodes in a vertical montage (mastoid to vertex with common ground near the nose) under the same ketamine/xylazine anesthesia induction protocol described above using standard procedures in our laboratory (Zhong et al., 2014). Spike times were stored with 10-μs resolution. FFRs were stored with 48-kHz sampling rate. Stimulus presentation and data acquisition were controlled by custom MATLAB-based (The MathWorks, Natick, MA) software that interfaced with hardware modules from Tucker-Davis Technologies (TDT, Alachua, FL) and National Instruments (NI, Austin, TX).

3.2.2 Speech stimuli

The following four stimuli were used in these experiments. (s_1) Stationary vowel, $\wedge$ (as in c$\underline{\text{u}}$p): F_0 was 100 Hz. The first three formants were placed at $F_1 = 600$, $F_2 = 1200$, and $F_3 = 2500$ Hz. The vowel was 188 ms in duration. (s_2) Nonstationary vowel, $\wedge$: F_0 increased linearly from 100 to 120 Hz over its 188-ms duration. The first two formants moved as well (F_1: 630 $\rightarrow$ 570 Hz; F_2: 1200 $\rightarrow$ 1500 Hz; see S2 Fig). F_3 was fixed at 2500 Hz. The formant frequencies for both s_1 and s_2 were chosen based on natural formant contours of the vowel $\wedge$ in American English (Hillenbrand et al., 1995; Hillenbrand and Nearey, 1999). s_1 and s_2 were synthesized using a MATLAB instantiation of the Klatt synthesizer (courtesy of Dr. Michael Kiefte, Dalhousie University, Canada). (s_3) A naturally uttered Danish sentence (list #1, sentence #3 in the CLUE Danish speech intelligibility test, Nielsen and Dau, 2009). (s_4) A naturally uttered English sentence (Sentence #2, List #1 in the Harvard Corpus, (Rothauser, 1969). All speech and speech-like stimuli were played at an overall intensity of 60 to 65 dB SPL.

Table 3.1. List of terms and definitions.

Term	Definition
Electrophysiology	Studies that record and analyze far-field (gross) potentials, e.g., electroencephalography
Neurophysiology	Studies that record and analyze spike-train data from neurons, e.g., AN fiber spike trains
Stationarity	A signal is stationary when the signal parameters do not change over time. For example, a stochastic signal like white Gaussian noise is stationary if the amplitude probability density function is constant across time. Similarly, a deterministic pure tone can be considered an example of a stationary sinusoidal process with a particular amplitude, frequency, and initial phase.
Second-order stationarity	A stochastic signal is second-order stationary if its mean and autocorrelation function do not change over time. Second-order stationarity is also referred to as wide-sense stationarity.
Linearity	A system is linear if it obeys the rules of superposition. For example, consider a system for which inputs x_1 and x_2 evoke responses y_1 and y_2, respectively. Then, the system is linear if the response to input $ax_1 + bx_2$ is $ay_1 + by_2$. An auditory corollary of linearity is that a linear system (e.g., the ear canal) processes sound in the same way at soft and loud sound levels, which means that for every dB increase in the input, the output is increased by the same dB.
Time invariance	A system is time invariant if its parameters (e.g., gain at all frequencies) do not change over time
Periodic signal	A perfectly repeating signal, e.g., a tone, or a synthetic vowel with constant pitch
Aperiodic signal	A signal that does not repeat, e.g., white Gaussian noise
Polarity-tolerant response	Response component that does not depend on stimulus polarity, e.g., the onset response
Polarity-sensitive response	Response component that depends on stimulus polarity, e.g., phase-locked spike trains in response to a low-frequency tone
Even sequence	$x[n]$ is even if $x[n] = x[-n]$
Odd sequence	$x[n]$ is odd if $x[n] = -x[-n]$

3.2.3 Power along a spectrotemporal trajectory

Consider a known frequency trajectory, $f_{traj}(t)$, along which we need to estimate power in a signal, $x(t)$. The phase trajectory, $\Phi traj(t)$, can be computed as the integration of $f_{traj}(t)$

$$\Phi_{traj}(t) = \int_0^t f_{traj}(\tau) d\tau. \tag{3.1}$$

For discrete-time signals, the phase trajectory can be estimated as

$$\Phi_{traj}[n] = \frac{1}{f_s} \sum_{m=1}^{n} f_{traj}[m]. \tag{3.2}$$

The phase trajectory can be demodulated from $x(t)$ by multiplying a complex exponential with phase $= -\Phi_{traj}(t)$ (Olhede and Walden, 2005)

$$x_{demod}(t) = x(t)\ \mathrm{e}^{-\jmath 2\pi \Phi_{traj}(t)}. \tag{3.3}$$

The power along $f_{traj}(t)$ in $x(t)$ can be estimated as the power in $x_{demod}(t)$ within the spectral resolution bandwidth (W) near 0 Hz in the spectral estimate, $P_{x_{demod}}(f)$, of $x_{demod}(t)$.

$$P_{traj} = 2 \int_{-W/2}^{W/2} P_{x_{demod}}(f) df \tag{3.4}$$

The scaling factor 2 is required because the integral in Eq 3.4 only represents the original positive-frequency band of the real signal, $x(t)$; the equal amount of power within the original negative-frequency band, which is shifted further away from 0 Hz by $\Phi_{traj}(t)$, should also be included (see Fig 3.11).

3.2.4 The harmonicgram

Consider a harmonic complex, $x(t)$, with a time-varying (instantaneous) fundamental frequency, $F_0(t)$. For a well-behaved and smooth $F_0(t)$, energy in $x(t)$ will be concentrated at multiples of the instantaneous fundamental frequency, i.e., $kF_0(t)$. Thus, $x(t)$ can be represented by the energy distributed across the harmonics of the fundamental. The time-varying power along the k-th harmonic of $F_0(t)$ can be estimated by first demodulating $x(t)$ with the $kF_0(t)$ trajectory using Eq 3.3, and then using an appropriate low-pass filter to limit energy near 0 Hz (say within $\pm W/2$). We define the *harmonicgram* as the matrix of time-varying power along all harmonics of the F_0. Thus, the harmonicgram is

$$harmonicgram(k,t) = \mathcal{LPF}_{[-W/2,W/2]}\{x(t)\ \mathrm{e}^{-\jmath 2\pi k F_0(t)}\}. \tag{3.5}$$

3.3 Classic metrics for quantifying temporal coding in the auditory system

Various approaches and metrics have been developed to quantify auditory temporal coding in neurophysiological responses. In this section, we motivate the need for a unified framework for auditory temporal coding by briefly reviewing these classic metrics and discussing their benefits and limitations.

3.3.1 Period-histogram based metrics

The ability of AN fibers to follow the temporal structure of an acoustic stimulus has been known for a long time (Galambos and Davis, 1943). Using tones as stimuli, Kiang and colleagues showed that AN fibers prefer to discharge spikes around a particular phase of the stimulus cycle (Kiang et al., 1965). Their analysis was qualitative and involved the period histogram, which is constructed as the histogram of spike times modulo the period of one stimulus cycle (e.g., Fig 3.1D). Rose and colleagues used the period histogram to quantify the preference of neurons to fire during one half-cycle of a periodic stimulus (Rose et al., 1967). They introduced a metric, called the coefficient of synchronization, which is defined as the ratio of the spike count in the most effective half-cycle to the spike count during the whole stimulus cycle. The coefficient of synchronization ranges from 0.5 (for a flat period histogram) to 1.0 (for all spikes within one half-cycle). The coefficient of synchronization does not truly quantify the strength of phase locking to the stimulus cycle as it does not

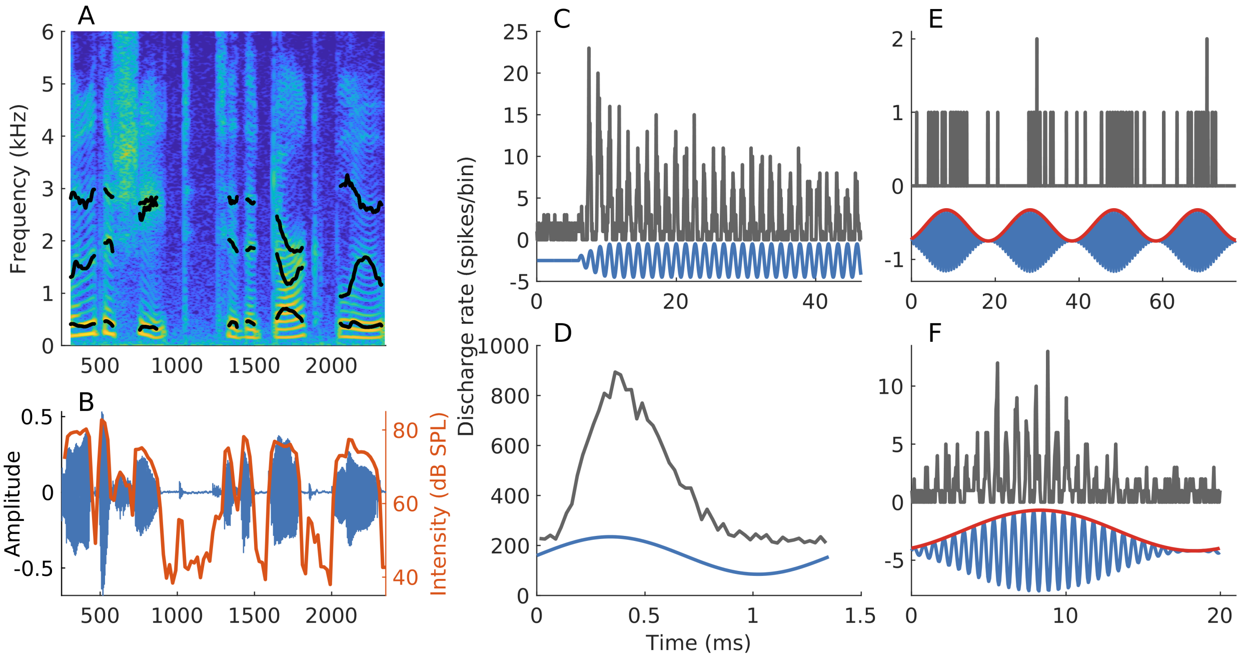

Figure 3.1. Neural responses of AN fibers are invariably nonstationary, even when the stimulus is not. (A, B) Spectrogram and waveform of a speech segment (s_4 described in *Materials and Methods*). Formant trajectories (black lines in panel A) and short-term intensity (red line in panel B, computed over 20-ms windows with 80% overlap) vary with time, highlighting two nonstationary aspects of speech stimuli. (C) PSTH constructed using spike trains in response to a tone at the AN-fiber's characteristic frequency (CF, most-sensitive frequency; fiber had CF=730 Hz, and was high spontaneous rate or SR, Liberman, 1978). Tone intensity = 40 dB SPL. Even though the stimulus is stationary, the response is nonstationary (i.e., sharp onset followed by adaptation). (D) Period histogram, constructed from the data used in C, demonstrates the phase-locking ability of neurons to individual stimulus cycles. (E) PSTH constructed using spike trains in response to a sinusoidally amplitude-modulated (SAM) CF-tone (50-Hz modulation frequency, 0-dB modulation depth, 35 dB SPL) from an AN fiber (CF = 1.4 kHz, medium SR). (F) Period histogram (for one modulation period) constructed from the data used in E. The response to the SAM tone follows both the modulator (envelope, red, panels E and F) as well as the carrier (temporal fine structure), the rapid fluctuations in the signal (blue, panel F). Bin width = 0.5 ms for histograms in C-F. Number of stimulus repetitions for C and E were 300 and 16, respectively.

consider the spread of the period histogram. For example, two period histograms, one where all spikes occur at the peak of the stimulus cycle (strong phase locking), and the other where all spikes are uniformly distributed across one stimulus half-cycle (weak phase locking), will yield the same coefficient of synchronization of 1.0.

A more sensitive measure of phase locking derived from the period histogram is the vector strength (VS, Goldberg and Brown, 1969; Greenwood and Durand, 1955), which is identical to the synchronization index metric described by Johnson (Johnson, 1980). VS has been used extensively to quantify phase-locking strength in spike-train recordings in response to periodic stimuli (Joris et al., 2004; Palmer and Russell, 1986), including stationary speech (Young and Sachs, 1979). In this framework, each spike is treated as a complex vector that has a magnitude of 1 and an angle that is defined by the spike phase relative to the stimulus phase; VS is defined as the magnitude of the average of all such vectors for spikes pooled across all stimulus repetitions (Sec 3.8.1). VS is a biased estimator of the "true" vector strength (Mardia, 1972) and can reach spuriously high values at low spike counts (Yin et al., 2010). This problem is avoided by using a modification of the vector strength, called the phase-projected vector strength (VS_{pp}, Yin et al., 2010). Similar approaches have been used in electrophysiological studies (Vinck et al., 2011). VS_{pp} differs from VS in that trial-to-trial phase consistency is also considered in computing VS_{pp} (Sec 3.8.1).

Overall, the period histogram and metrics derived from it (VS and VS_{pp}) work well for applications involving stationary signals with periodic TFS (e.g., tones), ENV (e.g., sinusoidally amplitude-modulated noise), or both (e.g., sinusoidally amplitude-modulated tones). However, the period histogram ignores nonstationary features in the response that arise from the auditory system. For example, spikes in the first few stimulus cycles are often ignored while constructing the period histogram to avoid the nonstationary onset response. Similarly, since spikes corresponding to different stimulus cycles are wrapped onto a single cycle, effects of adaptation are not captured in the period histogram. Moreover, its application to nonstationary or aperiodic stimuli (e.g., natural speech) is not straightforward.

3.3.2 Peristimulus-time-histogram (PSTH) based metrics

The single-polarity PSTH, $p(t)$, is constructed as the histogram of spike times pooled across all stimulus repetitions at a certain bin width (e.g., Fig 3.1C). As the PSTH shows the rate variation along the course of the stimulus, it captures the onset as well as adaption effects in the response (Kiang et al., 1965; Westerman and Smith, 1988). $p(t)$ has been applied to analyze spike trains recorded in response to periodic signals, both in the temporal and spectral domains (Delgutte, 1980; Palmer et al., 1986; Young and Sachs, 1979). A limitation of the *p(t)*-spectrum is that it is corrupted by harmonic distortions due to the rectified nature of the PSTH response (Young and Sachs, 1979). For example, the spectrum of a PSTH constructed using spike trains recorded from an AN fiber in response to a tone (F_c) can show energy at F_c as well as $2F_c$ even though the stimulus itself does not have energy at $2F_c$. These issues related to rectifier distortion can be minimized by using both polarities of the stimulus as we describe in a later section. Similar to the period histogram, the PSTH can also be used to derive phase-locking metrics, such as VS and VS_{pp}. These synchrony-based metrics have been recently overshadowed by correlogram-based metrics, which are described next, since the synchrony-based metrics are limited to periodic signals. In contrast, correlogram-based approaches offer more general metrics to evaluate temporal coding of both periodic and aperiodic stimuli in the ENV/TFS dichotomy.

3.3.3 Interspike-interval (ISI) based approaches

Interspike interval histogram analyses were developed to quantify the correlation between two spike trains, either from the same neuron or from different neurons (Hagiwara, 1954; Perkel et al., 1967a, 1967b; Rodieck et al., 1962). Interspike intervals between adjacent spikes (also called first-order intervals) within a stimulus trial are used to construct per-trial estimates of the ISI histogram, which are then averaged across trials to form the final first-order ISI histogram (Fig 3.2C). An alternative to the first-order ISI histogram, called the all-order ISI histogram (or the autocorrelogram), can be estimated in a similar way with the only difference being the inclusion of intervals between all spikes within a trial (not only adjacent spikes) to construct the histogram (Fig 3.2E, Møller, 1970; Rodieck, 1967). The autocorrelogram has been used to study the temporal representation of stationary as well as nonstationary stimuli (Bourk, 1976; Cariani and Delgutte, 1996a, 1996b; Sinex and Geisler, 1981). While the autocorrelogram is attractive for its simplicity, it is confounded by refractory effects (Figs 3.2E and 3.2F). In particular, since successive spikes within a single trial cannot occur within the refractory period, the autocorrelogram shows an artifactual absence of intervals for delays less than the refractory period (Fig 3.2E). As a result, the autocorrelogram spectrum is partly corrupted.

Joris and colleagues extended these ISI-based analyses to remove the confounds of the refractory effects by including all-order interspike intervals *across* stimulus trials to compute a shuffled correlogram (Louage et al., 2004). A shuffled correlogram computed using spike trains in response to multiple repetitions of a single stimulus from a single neuron is called the shuffled autocorrelogram (or the *SAC*, Fig 3.2G). Similarly, a shuffled correlogram computed using spike trains from different neurons, or for different stimuli, is called the shuffled cross-correlogram (or the *SCC*). The use of across-trial all-order ISIs provides substantially more smoothing than simple all-order ISIs because many more intervals are included in the histogram (compare Fig 3.2E with Fig 3.2G, and Fig 3.2F with Fig 3.2H).

In addition, both polarities of the stimulus can be used to separate out ENV and TFS components from the response. Stimuli with alternating polarities share the same envelope, but their phases (TFS) differ by a half-cycle at all frequencies. By averaging shuffled autocorrelograms for both stimulus polarities and shuffled cross-correlograms for opposite stimulus polarities, the *polarity-tolerant* (ENV) correlogram (called the *sumcor*) is obtained (Louage et al., 2004, Sec 3.8.4). Similarly, the *polarity-sensitive* (TFS) correlogram, the *difcor*, is estimated as the difference between the average autocorrelogram for both stimulus polarities and the cross-correlogram for opposite stimulus polarities (Sec 3.8.4). These functions have been preferred over PSTH-based analyses for estimating correlation sequences and response spectra (Cedolin and Delgutte, 2005; Joris et al., 2006; Rallapalli and Heinz, 2016). Shuffled autocorrelograms have also been used to derive temporal metrics, such as the correlogram peak-height and half-width, to quantify the strength and precision of temporal coding in the response, respectively (Louage et al., 2004, including for nonstationary stimuli (Paraouty et al., 2018; Sayles et al., 2015; Sayles and Winter, 2008). In addition, cross-correlograms have been used to develop metrics to quantify ENV/TFS similarity between responses to different stimuli recorded from the same neuron (e.g., speech stimuli (Heinz and Swaminathan, 2009; Rallapalli and Heinz, 2016; Swaminathan and Heinz, 2012), or between responses from different neurons (Heinz et al., 2010; Joris et al., 2006; Swaminathan and Heinz, 2011).

Although correlogram-based analyses provide a rich set of temporal metrics, they suffer from three major limitations. First, correlograms discard phase information in the response. Response phase can convey important information, especially for complex stimuli, like speech (Delgutte et al., 1998; Greenberg and Arai, 2001; Paliwal and Alsteris, 2003). Second, metrics derived from the shuffled autocorrelogram and the *sumcor* are corrupted by rectifier distortions (e.g., Fig 3.2H). Third, spectral estimates based on correlograms are appropriate for second-order stationary signals. To accommodate for nonstationary signals, usually a sliding-window-based approach is employed where in each temporal window the spectrum and/or correlogram is computed (Sayles and Winter, 2008). This windowing-based approach faces the familiar problem of the time-frequency resolution trade-off. In addition, the smoothing benefit provided by the correlogram comes at large computational cost as its computation requires all-order spike-time differences across all trials. This computation cost scales quadratically (N^2) with the number of spikes (N) and can be cumbersome for large N.

3.4 A unified framework for quantifying temporal coding based on alternating-polarity PSTHs (*apPSTHs*)

In this section, we first show that *apPSTHs* can be used to unify classic metrics, e.g., *VS* and correlograms, in a computationally efficient manner. Then, we show that *apPSTHs* offer more precise spectral estimates compared to correlograms, and allow for perceptually relevant analyses that are not possible with classic metrics.

3.4.1 Alternating-polarity PSTHs (*apPSTHs*)

Let us denote the PSTHs in response to the positive and negative polarities of a stimulus as $p(t)$ and $n(t)$, respectively. Then, the *sum PSTH*, $s(t)$, which represents the polarity-tolerant component in the response, is estimated as

$$s(t) = \frac{p(t) + n(t)}{2}. \tag{3.6}$$

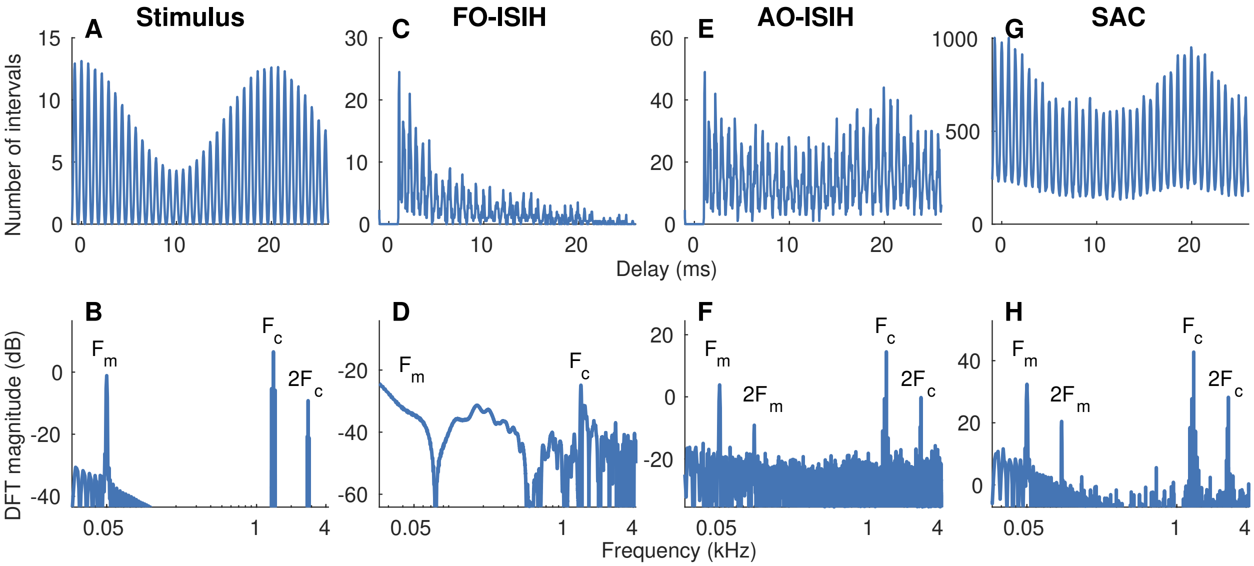

Figure 3.2. The shuffled autocorrelogram is better than the first-order and all-order ISI histogram, both in the time and frequency domains. Example correlograms (top) and associated spectra (bottom) constructed using spike trains recorded from an AN fiber (CF = 1.4 kHz, medium SR) in response to a SAM-tone at F_c = CF (50-Hz modulation frequency or F_m, 0-dB modulation depth, 700-ms duration, 27 repetitions, 50 dB SPL). (A) Autocorrelation function of the half-wave rectified stimulus. (B) The discrete Fourier transform (DFT) of A. (C) The first-order (FO) ISI histogram. (D) DFT of C. The first-order ISI histogram poorly captures the carrier (TFS) and fails to capture the modulator (ENV). (E) The all-order (AO) ISI histogram. (F) DFT of E. The all-order ISI histogram captures both the carrier and modulator despite being noisy. Both the first-order (C) and the all-order (E) ISI histograms show dips for intervals less than the refractory period (∼0.6 ms), with the corresponding spectra corrupted by these refractory effects. (G) The shuffled autocorrelogram. (H) DFT of G. The shuffled autocorrelogram is smoother compared to the other correlograms, which also leads to improved SNR in the spectrum at both the carrier and modulator frequencies. All these ISI histograms are corrupted by rectifier distortion at twice the carrier frequency ($2F_c$). Bin width = 50 μs for histograms in C, E, and G.

The *difference PSTH*, $d(t)$, which represents the polarity-sensitive component in the response, is estimated as

$$d(t) = \frac{p(t) - n(t)}{2}. \tag{3.7}$$

The difference PSTH has been previously described as the compound PSTH (Goblick and Pfeiffer, 1969). Here we use the terms *sum* and *difference* for $s(t)$ and $d(t)$, respectively, for simplicity. Compared to the spectra of the single-polarity PSTHs [i.e., of $p(t)$ or $n(t)$], the spectrum of the difference PSTH, $D(f)$, is substantially less affected by rectifier distortion artifacts (Sinex and Geisler, 1983). This improvement occurs because even-order distortions, which strongly contribute to these artifacts, are effectively canceled out by subtracting PSTHs for opposite polarities. The Fourier magnitude spectrum of the difference PSTH has been referred to as the synchronized rate. We show that the synchronized rate relates to VS by

$$VS(f) = \frac{|D(f)|}{N}, \tag{3.8}$$

where f is frequency in Hz, and N is the total number of spikes (Sec 3.8.2).

It can also be shown that the autocorrelogram and the autocorrelation function of the PSTH are related (Sec 3.8.3). In particular the SAC for a set of M spike trains $X = \{\underline{x_1}, \underline{x_2}, ..., \underline{x_M}\}$ can be estimated as

$$SAC(X) = \mathcal{R}_{\mathcal{X}}(PSTH_X) - \sum_{\mathrm{i}=1}^{M} \mathcal{R}_{\mathcal{X}}(\underline{x_{\mathrm{i}}}), \tag{3.9}$$

where $\mathcal{R}_{\mathcal{X}}$ is the autocorrelation operator, and $PSTH_X$ is the PSTH constructed using X. Similarly, the SCC for two sets of spike trains $X = \{\underline{x_1}, \underline{x_2}, ..., \underline{x_L}\}$ and $Y = \{\underline{y_1}, \underline{y_2}, ..., \underline{y_M}\}$ can be estimated as

$$SCC(X, Y) = \mathcal{R}_{\mathcal{XY}}(PSTH_X, PSTH_Y), \tag{3.10}$$

where $PSTH_X$ and $PSTH_Y$ are PSTHs constructed using X and Y, respectively, and $\mathcal{R}_{\mathcal{XY}}$ is the cross-correlation operator. Since SACs and SCCs can be computed using *apPSTHs*, it follows that *sumcor* and *difcor* can also be computed using *apPSTHs* (S4 Appendix). As *apPSTHs* can be used to compute correlograms, *apPSTHs* offer the same degree of smoothing as correlograms.

The use of *apPSTHs* to compute correlograms is computationally more efficient compared to the existing correlogram-estimation method, i.e., by tallying all interspike intervals. For a fixed stimulus duration and PSTH resolution, estimating the autocorrelation function of the PSTH requires constant time complexity [$\mathcal{O}(1)$]. Thus, for N spikes, the SAC and SCC can be computed with $\mathcal{O}(N)$ complexity that is needed for constructing the PSTH using Eqs 3.9 and 3.10. This is substantially better than the $\mathcal{O}(N^2)$ complexity needed to compute the correlograms by tallying shuffled all-order interspike intervals. For example, consider a spike-train dataset that consists of 50 repetitions of a stimulus with 100 spikes per repetition. To compute the SAC using (all-order) ISIs, each spike time (5000 unique spikes) has to be compared with spike times from all other repetitions (4900 spike times). This tallying method requires 24.5×10^6 (i.e., 5000×4900) operations to compute the SAC, where one operation consists of comparing two spike times and incrementing the corresponding SAC-bin by 1. In contrast, only 5000 operations are needed to construct the PSTH for 5000 (50×100) total spikes. The PSTH can then be used to estimate the SAC with constant time complexity. In addition to their computational efficiency, *apPSTHs* offer additional benefits for relating single-unit responses to far-field responses, for spectral estimation, and for speech-intelligibility modeling as discussed below.

3.4.2 *apPSTHs* unify single-unit and far-field analyses

The PSTH is particularly attractive because the PSTH from single neurons or a population of neurons, by virtue of being a continuous signal, can be directly compared to evoked potentials in response to the same stimulus (e.g., Fig 3.3). In this example, the speech sentence s_3 was used as the stimulus to record the frequency following response (FFR) from one animal. The same stimulus was also used to record spike trains from AN fibers (N=246) from 13 animals. The mean $d(t)$ and mean $s(t)$ were computed by pooling PSTHs across all neurons. The difference and sum FFRs were estimated by subtracting and averaging the FFRs to alternating polarities, respectively. This approach of estimating polarity-tolerant and polarity-sensitive components from FFRs is well established (Aiken and Picton, 2008; Ananthakrishnan et al., 2016; Shinn-Cunningham et al., 2013). Qualitatively, the periodicity information in the mean $d(t)$ and the difference FFR were similar (Fig 3.3A); this is expected because the difference FFR receives significant contributions from the auditory nerve (King et al., 2016). To compare the spectra for the two responses, a 100-ms segment was considered. The first formant (F_1) and the first few harmonics of the fundamental frequency (F_0) were well captured in both spectra. F_2 was also well captured in the difference FFR, and to a lesser extent, in the mean $d(t)$.

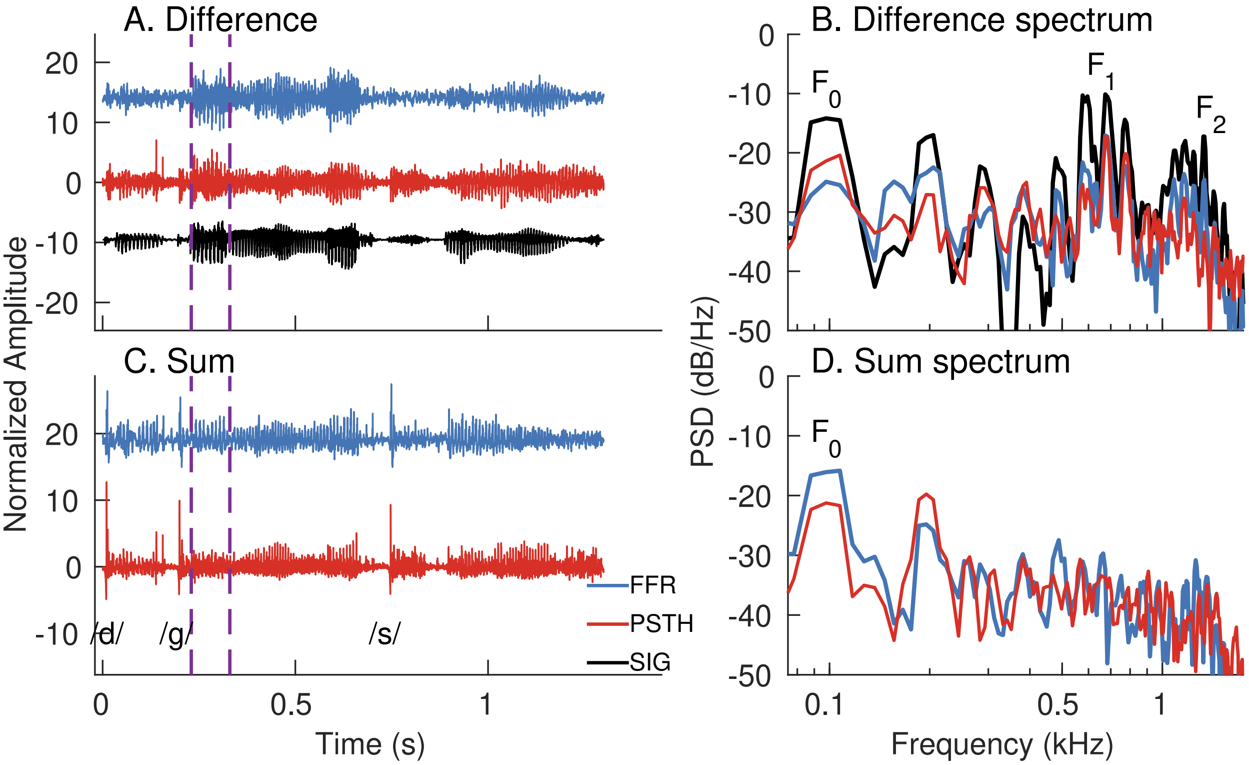

Figure 3.3. *apPSTHs* can be directly compared to evoked potentials in response to the same stimulus. (A) Time-domain waveforms for the difference FFR (blue) and the mean difference PSTH [$d(t)$, red] in response to the Danish speech stimulus, s_3 (black). The mean $d(t)$ was computed by taking the grand average of the $d(t)s$ from 246 AN fibers from the 13 animals (CFs from 0.2 to 11 kHz). The difference FFR was estimated by subtracting FFRs to alternating polarities of the stimulus. (B) Spectra for the signals in A for a 100-ms segment (delimited by purple dashed lines in A). (C) Time-domain waveforms for the sum FFR (blue) and the mean sum PSTH [$s(t)$, red] for the same stimulus, s_3. Both responses show sharp onsets for plosive (/d/ and /g/) and fricative (/s/) consonants. (D) Spectra for the responses in C for the same segment considered in B. The mean $s(t)$ was estimated as the grand average of the $s(t)s$ from the 246 neurons. Sum FFR was estimated by halving the sum of the FFRs to both polarities. Stimulus intensity = 65 dB SPL.

The mean $s(t)$ and the sum FFR also show comparable temporal features in these nonstationary responses (Fig 3.3C). For example, both responses show sharp onsets for plosive and fricative consonants. The segment considered in Fig 3.3B was used to compare the spectra for the two sum responses. Both spectra show similar spectral peaks near the first two harmonics of F_0 (Fig 3.3D), which indicates that pitch-related periodicity is well captured in both the sum FFR and the mean $s(t)$. However, there are some discrepancies between the relative heights of the first two F_0-harmonics. These discrepancies could arise because the average FFR primarily reflects activity of high-frequency neurons from rostral generators (e.g., the inferior colliculus, King et al., 2016), which show stronger polarity-tolerant responses compared to the auditory nerve (Joris, 2003). In contrast, the mean $s(t)$ is based on responses of AN fibers, which show strong polarity-sensitive responses to F_0 due to tuning-curve tail responses at high sound levels like that used here. These tail responses contribute to power at $2F_0$ as rectifier distortion. Overall, using *apPSTHs* for invasive spike-train recordings allows direct comparison of invasive single-unit data with noninvasive continuous-valued evoked potentials.

3.4.3 Variance of *apPSTH*-based spectral estimates can be reduced relative to the correlogram-based spectral estimates

Temporal information in a signal can be studied not only in the time domain (e.g., using correlograms) but also in the frequency domain (e.g., using the power spectral density, PSD). The frequency-domain representation often provides a compact alternative compared to the time-domain counterpart. In the framework of spectral estimation, the source ("true") spectrum, which is unknown, is regarded as a parameter of a random process that is to be estimated from the available data (i.e., from examples of the random process). Spectral estimation is complicated by two factors: (1) finite response length, and (2) stochasticity of the system. The former introduces bias to the estimate, i.e., the PSD at a given frequency can differ from the true value. This bias reflects the leakage due to power at nearby (narrowband bias) and far-away (broadband bias) frequencies (due to the inherent temporal windowing from the finite-duration response). Stochasticity of the system adds randomness to the sampled data, which creates variance in the estimate. Desirable properties of PSD estimators are minimized bias and variance. Bias can be reduced by multiplying the data (prior to spectral estimation) with a taper that has a strong energy concentration near 0 Hz. Variance can be reduced by using a greater number of tapers to estimate multiple (independent) PSD estimates, which can be averaged to compute the final estimate. The multitaper approach optimally reduces the bias and variance of the PSD estimate (Babadi and Brown, 2014; Thomson, 1982). In this approach, for a given data length, a frequency resolution is chosen, based on which a set of orthogonal tapers are computed. These tapers include both even and odd tapers, which can be used to obtain the independent PSD estimates to be averaged. In contrast, only even tapers can be used with correlograms as they are even sequences (Oppenheim, 1999; Rangayyan, 2015). Therefore, variance in the PSD estimate can be reduced by a factor of up to 2 by using *apPSTHs* instead of correlograms.

For example, the benefit (in terms of spectral-estimation variance) of using the multitaper spectrum of $d(t)$, as opposed to the discrete Fourier transform (DFT) of the *difcor*, can be quantified by comparing the two spectra at a single frequency (Fig 3.4). In this example, a 100-ms segment of the s_3 speech stimulus was used as the analysis window. The segment had an F_0 of 98 Hz and F_1 of 630 Hz (Fig 3.4A). Fig 3.4B shows example spectra estimated using spike trains recorded from a low-frequency AN fiber [CF = 900 Hz, SR = 81 spikes/s]. To compare the variances in the two estimated spectra, fractional power at the 6th harmonic was considered, as this harmonic was the closest to F_1. This analysis was restricted to neurons (N=10) for which data was available for at least 75 repetitions per polarity of the stimulus and that had a CF between 0.3 and 2 kHz. For each neuron, 25 spike trains per polarity were chosen randomly 12 times to estimate fractional power at the 6th harmonic. The same set of spike trains were used to estimate distributions for both the *difcor*-spectrum and *D(f)*. The ratio of *difcor*-based fractional power variance to the *apPSTH*-based fractional power variance at $6F_0$ was >1 for all 10 neurons considered (Fig 3.4D), demonstrating the benefit of being able to compute a multitaper spectrum from $d(t)$ compared to the *difcor*-spectrum in reducing variance. Overall, these results indicate that less data are required to achieve the same level of precision in a spectral metric based on the multitaper spectrum of an *apPSTH* compared to the same metric derived from the DFT of the correlogram.

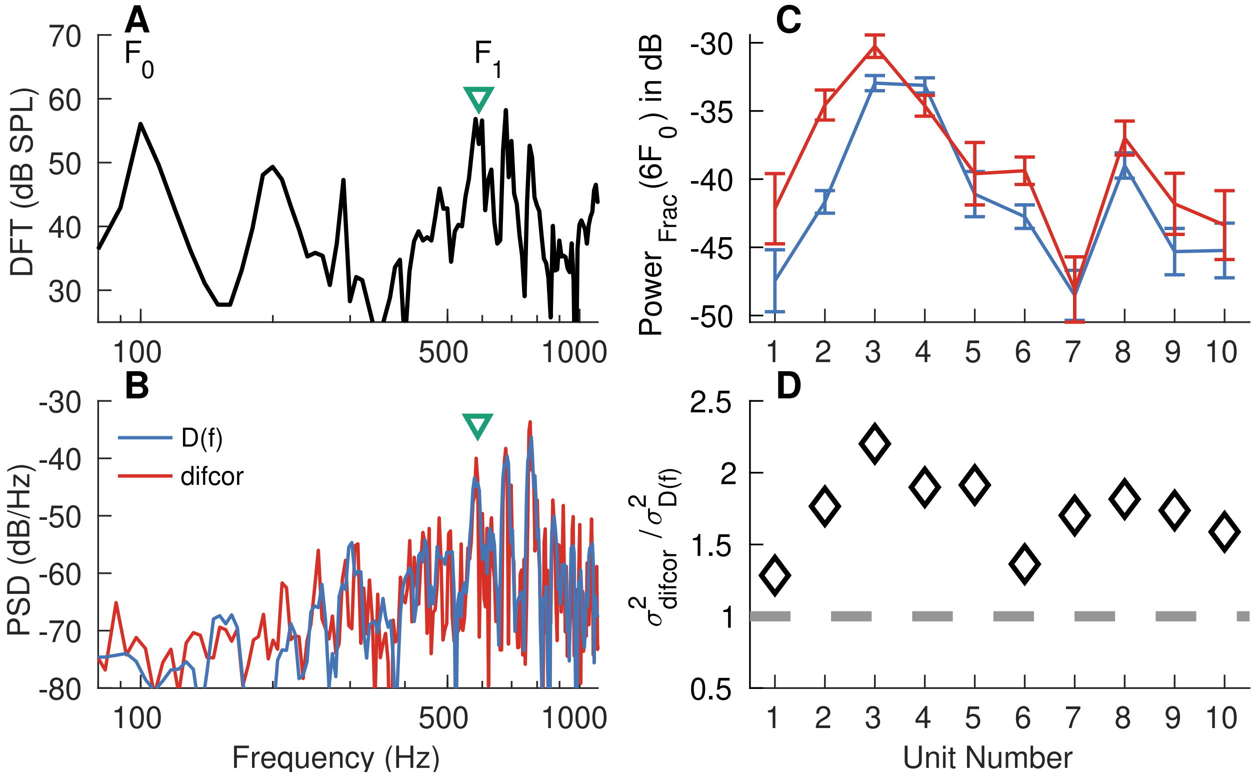

Figure 3.4. Lower spectral-estimation variance can be achieved using *apPSTHs* (with multiple tapers) compared with *difcor* correlograms. (A) Spectrum for the 100-ms segment in the speech sentence *s3* ($F_0 \sim 98$ Hz, $F_1 \sim 630$ Hz) used for analysis. (B) Example spectra for an AN fiber (CF=900 Hz, high SR) with spikes from 25 randomly chosen repetitions per polarity. The first two discrete-prolate spheroidal sequences were used as tapers corresponding to a time-bandwidth product of 3 to estimate $D(f)$, the spectrum of $d(t)$. No taper (i.e., rectangular window) was used to estimate the *difcor* spectrum. The AN fiber responded to the 6th, 7th and 8th harmonic of the fundamental frequency. (C) Error-bar plots for fractional power ($Power_{Frac}$) at the frequency (green triangle) closest to the 6th harmonic. Error bars were computed for 12 randomly and independently drawn sets of 25 repetitions per polarity. The same spikes were used to compute the spectra for $d(t)$ (blue) and *difcor* (red). (D) Diamonds denote the ratio of variances for the *difcor*-based estimate to the *d(t)*-based estimate. This ratio was greater than 1 (i.e., above the dashed gray line) for all units considered, which demonstrates that the variance for the multitaper-*d(t)* spectrum was lower than the *difcor*-spectrum variance. AN fibers with CFs between 0.3 and 2 kHz and with at least 75 repetitions per polarity of the stimulus were considered. Bin width = 0.1 ms for PSTHs. Sampling frequency = 10 kHz for FFRs. Stimulus intensity = 65 dB SPL.

3.4.4 Benefits of *apPSTHs* for speech-intelligibility modeling

Speech-intelligibility (SI) models aim to predict the effects of acoustic manipulations of speech on perception. Thus, SI models allow for quantitative evaluation of the perceptually relevant features in speech. More importantly, SI models can guide the development of optimal hearing-aid strategies for hearing-impaired listeners. However, state-of-the-art SI models are largely based on the acoustic signal, where there is no physiological basis to capture the various effects of sensorineural hearing loss (SNHL, Cooke, 2006; Houtgast and Steeneken, 1973; Kryter, 1962; Relaño-Iborra et al., 2016; Taal et al., 2011). In contrast, neurophysiological SI models (i.e., SI models based on neural data) are particularly important in this regard since spike-train data from preclinical animal models of various forms of SNHL provide a direct way to evaluate the effects of SNHL on speech-intelligibility modeling outcomes (Heinz, 2015; Rallapalli and Heinz, 2016).

A major advantage of PSTH-based approaches over correlogram-based approaches is that they can be used to extend a wider variety of acoustic SI models to include neurophysiological data. In particular, correlograms can be used to extend power-spectrum-based SI models (Cooke, 2006; Houtgast and Steeneken, 1973; Jørgensen and Dau, 2011; Kryter, 1962; Taal et al., 2011) but not for the more recent SI models that require phase information of the response (Relaño-Iborra et al., 2016; Scheidiger et al., 2018). For example, the speech envelope-power-spectrum model (sEPSM) has been evaluated using simulated spike trains since sEPSM only requires power in the response envelope, which can be estimated from the *sumcor* spectrum (Rallapalli and Heinz, 2016). However, *sumcor* cannot be used to evaluate envelope-phase-based SI models since it discards phase information. Studies have shown that the response phase can be important for speech intelligibility (Delgutte et al., 1998; Paliwal and Alsteris, 2003). In contrast to the *sumcor*, the time-varying PSTH contains both phase and magnitude information, and thus, can be used to evaluate both power-spectrum- and phase-spectrum-based SI models. For example, because the PSTH $p(t)$ [or $n(t)$] is already rectified, it can be filtered through a modulation filter bank to estimate “internal representations” in the modulation domain (Fig 3.5). These spike-train-derived “internal representations” are analogous to those used in phase-spectrum-based SI models (Relaño-Iborra et al., 2016;

Scheidiger et al., 2018) and can be further processed by existing SI back-ends to estimate SI values. In summary, *apPSTHs* can be used to estimate complete (amplitude and phase) spectrally specific modulation-domain representations using modulation filter banks. These analyses allow for the evaluation of a wider variety of acoustic-based SI models in the neural domain, where translationally relevant data can be obtained from preclinical animal models of various forms of SNHL.

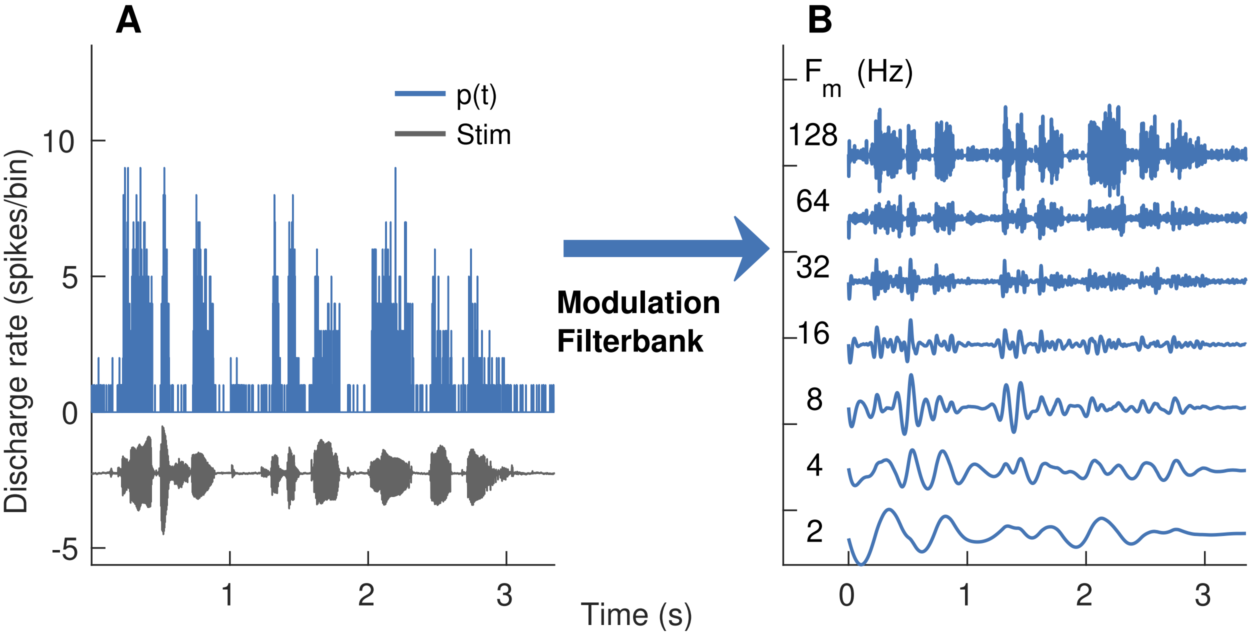

Figure 3.5. Modulation-domain internal representations for speech coding can be obtained from PSTH-based envelopes. PSTH response $[p(t)]$ from one AN fiber (CF=290 Hz, SR= 12 spikes/s) is shown. (A) Time-domain waveforms for the stimulus (gray) and $p(t)$ (blue). (B) Output of a modulation filter bank after the processing of $p(t)$. Modulation filters were zero-phase, fourth-order, and octave-wide IIR filters. Center frequencies (F_m) for these filters ranged from 2 to 128 Hz (octave spacing), similar to those used in recent psychophysically based SI models (e.g., (Relaño-Iborra et al., 2016). PSTH bin width = 0.5 ms. 15 stimulus repetitions. Stimulus intensity = 60 dB SPL.

3.5 Quantifying ENV and TFS using *apPSTHs* for stationary signals

In this section, we first describe existing and novel ENV and TFS components that can be derived from *apPSTHs*. Next, we compare the relative merits of the novel components over existing ENV and TFS components using simulated AN fiber data. Finally, we apply these *apPSTHs* to analyze spike-train data recorded in response to speech and speech-like stimuli.

3.5.1 Several ENV and TFS components can be derived from *apPSTHs*

The neural response envelope can be obtained from *apPSTHs* in two orthogonal ways: (1) the low-frequency signal, $s(t)$, and (2) the Hilbert envelope of the high-frequency carrier-related energy in $d(t)$. $s(t)$ is thought to represent the polarity-tolerant response component, which has been defined as the envelope response (Joris, 2003; Louage et al., 2004). For a stimulus with harmonic spectrum, $s(t)$ captures the envelope related to the beating between harmonics. In addition, onset and offset responses (e.g., in response to high-frequency fricatives, Fig 3.3C) are also well captured in $s(t)$. Although *sumcor* and $s(t)$ are related, dynamic features like onset and offset responses are captured in $s(t)$, but not in the *sumcor* since the *sumcor* discards phase information by essentially averaging ENV coding across the whole stimulus duration. The use of sum envelope is popular in far-field responses (Aiken and Picton, 2008; Ananthakrishnan et al., 2016; Shinn-Cunningham et al., 2013) but not directly in auditory neurophysiology studies. A major disadvantage of $s(t)$ is that it is affected by rectifier distortions if a neuron phase locks to low-frequency energy in the stimulus (e.g., Fig 3.6A). We discuss this issue of rectifier distortion in more detail later in the following section.

A second way envelope information in the neural response can be quantified is by computing the envelope of the difference PSTH, $d(t)$. This envelope, $\mathrm{e}(t)$, can be estimated as the magnitude of the analytic signal, $a(t)$, of the difference PSTH

$$\mathrm{e}(t) = \frac{|a(t)|}{\sqrt{2}}, \tag{3.11}$$

where $a(t) = d(t) + \jmath\mathcal{H}\{d(t)\}$, and $\mathcal{H}\{\cdot\}$ is the Hilbert transform operator. The factor $\sqrt{2}$ normalizes for the power difference after applying the Hilbert transform. $d(t)$ is substantially less affected by rectifier distortion (Sinex and Geisler, 1983), and thus, so is $\mathrm{e}(t)$. The use of $\mathrm{e}(t)$ parallels the procedure followed by many computational models that extract envelopes from the output of cochlear filter banks (Dubbelboer and Houtgast, 2008; Jørgensen and Dau, 2011; Sadjadi and Hansen, 2011). The relative merits of e(t) and $s(t)$ to represent the response envelope is discussed in the following section.

The TFS component can also be estimated in two ways: (1) $d(t)$, and (2) the cosine of the Hilbert phase of $d(t)$. The difference PSTH has been traditionally referred to as the TFS response because it is the polarity-sensitive component. *difcor* and metrics derived from it relate to $d(t)$ as the *difcor* is related to the autocorrelation function of $d(t)$ (see Section 3.8.4). However, $d(t)$ does not represent the response to only the carrier (phase) since it also contains envelope information in e(t). We propose a novel representation of the TFS component in the response, $\phi(t)$, estimated as the cosine phase of the analytic signal

$$\phi(t) = \sqrt{2} \times rms[d(t)] \times cos[\angle a(t)], \tag{3.12}$$

where normalization by $\sqrt{2} \times rms[d(t)]$ is used to match the power in $\phi(t)$ with the power in $d(t)$ since $cos[\angle a(t)]$ is a constant-rms ($rms = 1/\sqrt{2}$) signal.

3.5.2 Relative merits of sum and Hilbert-envelope PSTHs in representing spike-train envelope responses

The relative merits of the two envelope PSTHs, $s(t)$ and e(t), were evaluated based on simulated spike-train data generated using a computational model of AN responses (Bruce et al., 2018). The model includes both cochlear-tuning and hair-cell transduction nonlinearities in the auditory system. Responses were generated for 24 AN fibers whose CFs were logarithmically spaced between 250 Hz and 8 kHz. Model parameters are listed in S1 Table. For each simulated fiber, a SAM tone was used as the stimulus. The SAM tone carrier (F_c) was placed at CF for each fiber and was modulated by a 20-Hz modulator (F_m) at 100% (i.e., 0-dB) modulation depth. The intensity was ∼65 dB SPL for all CFs, with slight adjustments to maintain a consistent discharge rate of 130 spikes/s (e.g., to account for the middle-ear transfer function that is included in the model). The number of repetitions per stimulus polarity was set to 25, which is typical of auditory-neurophysiology experiments.

Modulation spectra were estimated for $s(t)$ and e(t) [denoted by $S(f)$ and $E(f)$, respectively] for individual-fiber responses (Figs 3.6A-3.6C). $d(t)$ was band-pass filtered near CF (200-Hz bandwidth, 2nd order filter) before applying the Hilbert transform. This filtering minimized the spectral energy in $d(t)$ that was not stimulus related. The two envelopes were evaluated based on their representations of the modulator and rectifier distortion. Rectifier distortions are expected to occur at even multiples of the carrier frequency and nearby sidebands (i.e., $2nF_c$, $2nF_c - F_m$, and $2nF_c + F_m$ for integers n, Fig 3.6A). It is desirable for an envelope metric to consistently represent envelope coding across CFs and be less affected by rectifier-distortion artifacts. Modulation coding for the simulated responses was quantified as the power in 10-Hz bands centered at the first three harmonics of F_m (i.e., 15 to 25 Hz, 35 to 45 Hz, and 55 to 65 Hz) for both $s(t)$ and e(t) (Fig 3.6D). The need to include multiple harmonics of F_m arises because the response during a stimulus cycle departs from sinusoidal shape due to the saturating nonlinearity associated with the inner-hair-cell transduction process (S1 Fig). While F_m-related power was nearly constant across CF for $s(t)$, it was nearly constant for e(t) only up to 1.2 kHz, after which it rolled off. This roll-off for e(t) is not surprising since e(t) relies on phase-locking near the carrier and the sidebands, as confirmed by the strong correspondence between tonal phase-locking at the carrier frequency and F_m-related power in e(t) (Fig 3.6D).

The analysis of rectifier distortion was limited to only the distortion components near the second harmonic of the carrier (i.e., $2F_c$, $2F_c - F_m$, and $2F_c + F_m$) since this harmonic is substantially stronger than higher harmonics (e.g., Fig 3.6A). Rectifier distortion was quantified as the sum of power in 10-Hz bands centered at the three distortion frequency components. Because e(t) was estimated from spectrally specific $d(t)$, which was band-limited to 200 Hz near the carrier frequency, e(t) was virtually free from rectifier distortion. In contrast, $s(t)$ was substantially affected by rectifier distortion for simulated fibers with CFs below $\sim$2 kHz (Fig 3.6E). Rectifier distortion in $S(f)$ dropped for fibers with CF above $\sim$0.8 kHz because phase locking at distortion frequencies (i.e., twice the carrier frequencies) was attenuated by the roll-off in tonal phase locking. For example, the simulated AN fiber in Fig 3.6B (CF = 1.7 kHz) maintained comparable F_m-related power for both envelopes, but rectifier distortion for $s(t)$ was substantially diminished because the distortion frequency (3.4

kHz) is well above the phase-locking roll-off. These results indicate that $s(t)$ is substantially corrupted by rectifier distortion (at twice the stimulus frequency) when the neuron responds to stimulus energy that is below half the phase-locking cutoff.

Next, these spectral power metrics were compared with the correlogram-based metric, *sumcor* peak-height (Figs 3.6F-3.6H). The *sumcor* peak-height metric is defined as the maximum value of the normalized time-domain *sumcor* function (Louage et al., 2004). Prior to estimating the peak-height, the *sumcor* is sometimes adjusted by adding an inverted triangular window to compensate for its triangular shape (Heinz and Swaminathan, 2009). Here, *sumcors* were compensated by subtracting a triangular window from it so that the baseline *sumcor* is a flat function with a value of 0 (instead of 1) in the absence of ENV coding. This baseline value of 0 for the *sumcor* is the same as the baseline value for the *difcor* in the absence of TFS coding. In Sec 3.8.5, we show that the *sumcor* peak-height is a

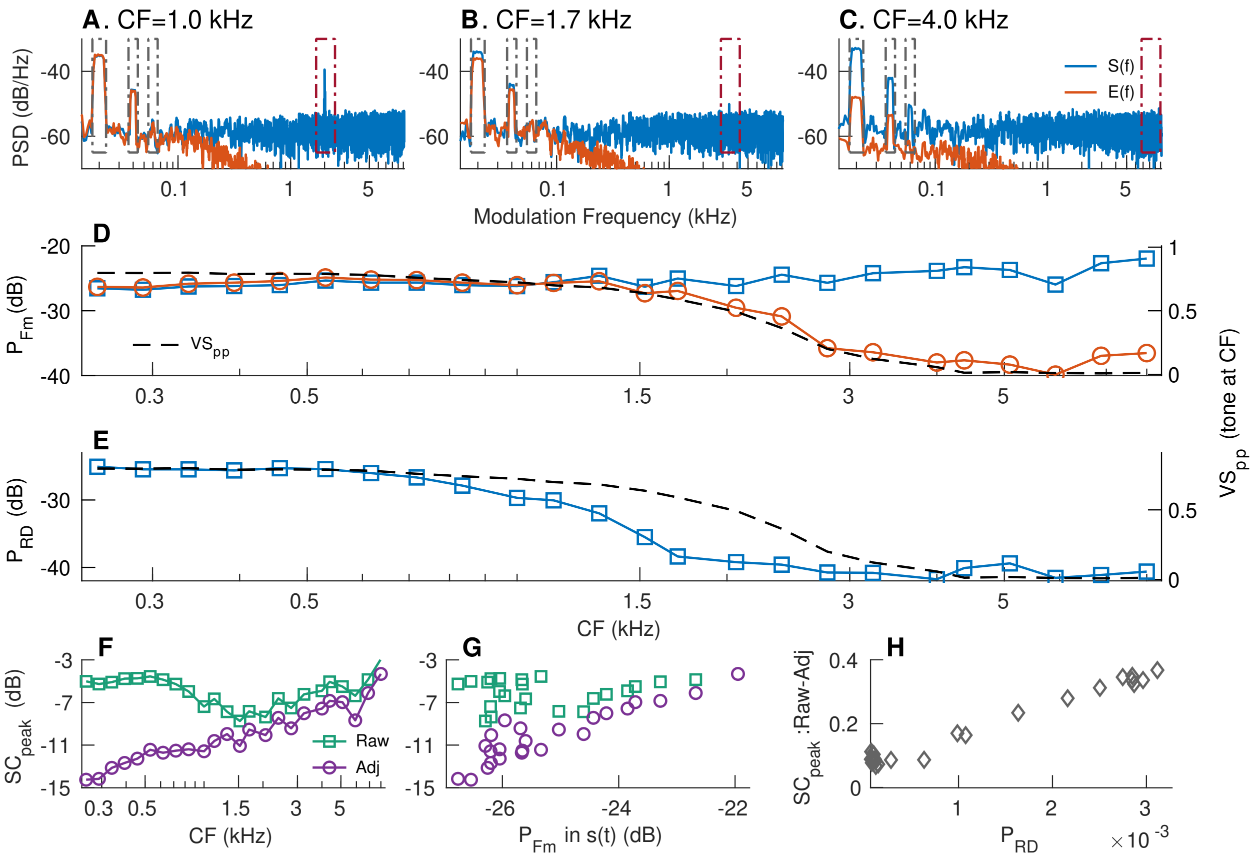

Figure 3.6. *(see next page)*

Figure 3.6 *(continued from previous page.)* **Envelope-coding metrics should be spectrally specific to avoid artifacts due to rectifier distortion and neural stochasticity.** Simulated responses for 24 AN fibers (log-spaced between 250 Hz and 8 kHz) were obtained using a computational model (see text) using SAM tones at CF (modulation frequency, F_m=20 Hz; 0-dB modulation depth) as stimuli. Stimulus intensity $\sim$ 65 dB SPL. $S(f)$ (blue) and $E(f)$ (red) for three example model fibers with CFs = 1.0, 1.7, and 4 kHz (panels A-C) illustrate the relative merits of $s(t)$ and e(t), and the potential for rectifier distortion to corrupt envelope coding metrics. $d(t)$ was band-limited to a 200-Hz band near F_c for each AN fiber prior to estimating e(t) from the Hilbert transform of $d(t)$. (A) For the 1-kHz fiber, $S(f)$ and $E(f)$ are nearly identical in the F_m band. $S(f)$ is substantially affected by rectifier distortion at 2×CF, which can be ignored using spectrally specific analyses. (B) The two envelope spectra are largely similar near the F_m bands since phase-locking near the carrier (1.7 kHz) is still strong (panel D). Rectifier distortion in $S(f)$ is greatly reduced since phase-locking at twice the carrier frequency (3.4 kHz = 2×1.7 kHz) is weak. (C) F_m-related power in $E(f)$ and rectifier distortion in $S(f)$ are greatly reduced as the frequencies for the carrier and twice the carrier are both above the phase-locking roll-off. (D) The strength of modulation coding was evaluated as the sum of the power near the first three harmonics of F_m (gray boxes in panels A-C) for $S(f)$ (blue squares) and $E(f)$ (red circles). VS_{pp} was also quantified to CF-tones for each AN fiber (black dashed line, right Y axis). (E) Rectifier distortion (RD) analysis was limited to the second harmonic of the carrier (brown boxes in panels A-C). RD was quantified as the sum of power in 10-Hz bands around twice the carrier frequency ($2 \times CF$) and the adjacent sidebands ($2 \times CF \pm F_m$). RD for $E(f)$ is not shown because $E(f)$ was virtually free from RD. (F) Raw and adjusted *sumcor* peak-heights across CFs. *sumcors* were adjusted by band-pass filtering them in the three F_m-related bands. Large differences between the two metrics at low frequencies indicate that the raw *sumcor* peak-heights are corrupted by rectifier distortion at these frequencies. (G) Relation between raw and adjusted *sumcor* peak-heights with F_m-related power (from panel D) in $S(f)$. Good correspondence between F_m-related power in $S(f)$ and adjusted *sumcor* peak-height supports the use of spectrally specific envelope analyses. (H) The difference between raw and adjusted *sumcor* peak-heights was largely accounted for by RD power. However, this difference was always greater than zero, suggesting broadband metrics can also be biased because of noise related to neural stochasticity.

broadband metric and it is related to the total power in $s(t)$, including rectifier distortions. When the *sumcor* is used to analyze responses of low-frequency AN fibers to broadband noise stimuli, the *sumcor*-spectrum, and thus, the *sumcor* peak-height, are corrupted by rectifier distortions (Heinz and Swaminathan, 2009). Similar to $S(f)$ for low-frequency SAM tones (Fig 3.6A), these distortions show up at 2×CF in the *sumcor*-spectrum, whereas the *difcor*-spectrum has energy only near CF (Heinz and Swaminathan, 2009). Heinz and colleagues accounted for these distortions by low-pass filtering the *sumcor* below CF to remove the effects of rectifier distortion at 2×CF. Here, we generalize this issue by comparing the *sumcor* and spectrally specific ENV metrics for narrowband SAM-tone stimuli to demonstrate the limitations of any broadband ENV metric. *sumcors* were adjusted by band-limiting them to 10-Hz bands near the first three harmonics of F_m. As expected, the difference between the raw and adjusted *sumcor* peak-heights was large at low CFs (Fig 3.6F), where rectifier distortion corrupts the broadband *sumcor* peak-height metric. At high CFs (above 1.5 kHz), the difference between raw and adjusted *sumcor* peak-heights was small but nonzero. These differences correspond to power in $S(f)$ at frequencies other than the modulation-related bands and reflect the artifacts of neural stochasticity due to finite number of stimulus trials. As power is always nonnegative, including power at frequencies unrelated to the target frequencies adds bias and variance to any broadband metric. The adjusted *sumcor* peak-height, unlike the raw *sumcor* peak-height, showed good agreement with spectrally specific F_m-related power in $S(f)$ (Fig 3.6G).

Overall, these results support the use of spectrally specific analyses to quantify ENV coding in order to minimize artifacts due to rectifier distortion as well as the effects of neural stochasticity. Of the two candidate *apPSTHs* to quantify response envelope, $\mathrm{e}(t)$ had the benefit of minimizing rectifier distortion. However, $\mathrm{e}(t)$'s reliance on carrier-related phase locking limits the use of $\mathrm{e}(t)$ as a unifying ENV metric across the whole range of CFs. Instead, spectrally specific $s(t)$ is more attractive because of its robustness in representing the response envelope across CFs (Fig 3.6D).

3.5.3 Relative merits of difference and Hilbert-phase PSTHs in representing spike-train TFS responses

In order to evaluate the relative merits of $d(t)$ and $\phi(t)$ in representing the neural response TFS, the same set of simulated AN spike-train responses were used as in Fig 3.6. The stimulus, a CF-centered SAM-tone, has energy at the carrier frequency (F_c) and the sidebands ($F_c \pm F_m$). The sidebands are 6 dB lower in amplitude compared to the carrier amplitude for the 100% modulated stimulus. Although the stimulus has power at the carrier and the sidebands, only the carrier representation should be considered towards quantifying the response TFS because the energy at the sidebands arises due to the modulation of the carrier by the modulator (ENV). As the carrier has energy at a single frequency (F_c) for a SAM tone, the desirable TFS response should have maximum energy concentrated at the carrier frequency and not at the sidebands. Therefore, the merits of $d(t)$ and $\phi(t)$ were evaluated based on how well they capture the carrier and suppress the sidebands (Fig 3.7).

As mentioned previously, $d(t)$ was band-limited to a 200-Hz bandwidth near the carrier frequency before estimating $\phi(t)$. $D(f)$ at low CFs contained substantial energy at both the carrier and the sidebands (Figs 3.7A and 3.7B). This indicates that $d(t)$ represents the complete neural coding of the SAM tone (both the envelope and the carrier) and not just the carrier. Furthermore, $D(f)$ has additional sidebands ($F_c \pm 2F_m$) around the carrier frequency. These sidebands arise as a result of the saturating nonlinearity associated with inner-hair-cell transduction (S1 Fig), and thus, should not be considered towards TFS response. In contrast, $\Phi(f)$, the spectrum of $\phi(t)$ had most of its power concentrated at the carrier frequency, with substantially less power in the sidebands (Figs 3.7A and 3.7B). These results were consistent across a wide range of CFs and for both sidebands (Figs 3.7D and 3.7E). Overall, these results show that $\phi(t)$ is a better PSTH compared to $d(t)$ in quantifying the response TFS since $\phi(t)$ emphasizes power at the carrier frequency and not at the sidebands.

In the following, we apply *apPSTH*-based analyses on spike-train data recorded from chinchilla AN fibers in response to speech and speech-like stimuli. In these examples, we particularly focus on certain ENV features, such as pitch coding for vowels and response onset for consonants, and TFS features, such as formant coding for vowels.

3.5.4 Neural characterization of ENV and TFS using ***apPSTHs*** for a synthesized stationary vowel

Fig 3.8 shows the response spectra obtained using various *apPSTHs* [$p(t)$, $s(t)$, $d(t)$, and $\phi(t)$] for a low-frequency AN fiber. The stationary vowel, s_1, was used as the stimulus. The neuron's CF was close to the first stimulus formant (Fig 3.8B). $P(f)$ shows a strong response

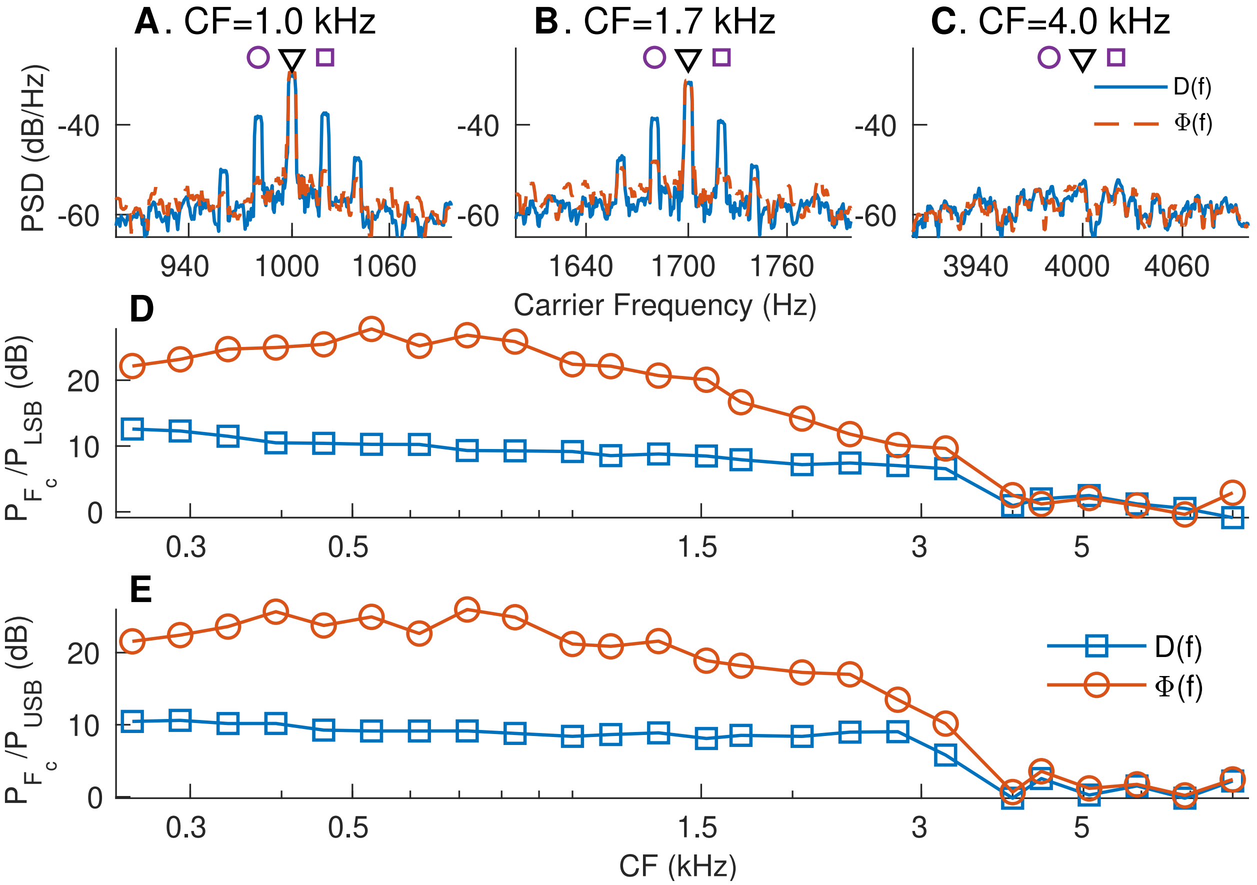

Figure 3.7. Compared to the $d(t)$, the *apPSTH* $\phi(t)$ provides a better TFS representation. (A-C) Spectra of $d(t)$ and $\phi(t)$ for the same three simulated AN fiber responses for which ENV spectra were shown in Fig 3.6. $D(f)$ has substantial power at CF (black triangle), as well as at lower (purple circle) and upper (purple square) sidebands. $\Phi(f)$, the spectrum of $\phi(t)$, shows maximum power concentration at CF (carrier frequency), with greatly reduced sidebands. (D) Ratio of power at CF (carrier, black triangle in panels A-C) to power at lower sideband (LSB, $F_c - F_m$, purple circles in panels A-C). (E) Ratio of power at CF (carrier) to power at upper sideband (USB, $F_c + F_m$, purple squares in panels A-C). $\phi(t)$ highlights the carrier and not the sidebands, and thus, compared to $d(t)$, $\phi(t)$ is a better representation of the true TFS response.

to the 6th harmonic (first formant). In addition, there is substantial energy near 1200 Hz, the frequency corresponding to the second formant. However, this peak near 1200 Hz results from rectifier distortion from the first formant and not in response to the second formant itself; this is confirmed by $D(f)$ (Fig 3.8D), which shows a clear peak at the 6th harmonic and little energy near the 12th harmonic. Similar to $P(f)$, $S(f)$ shows substantial energy

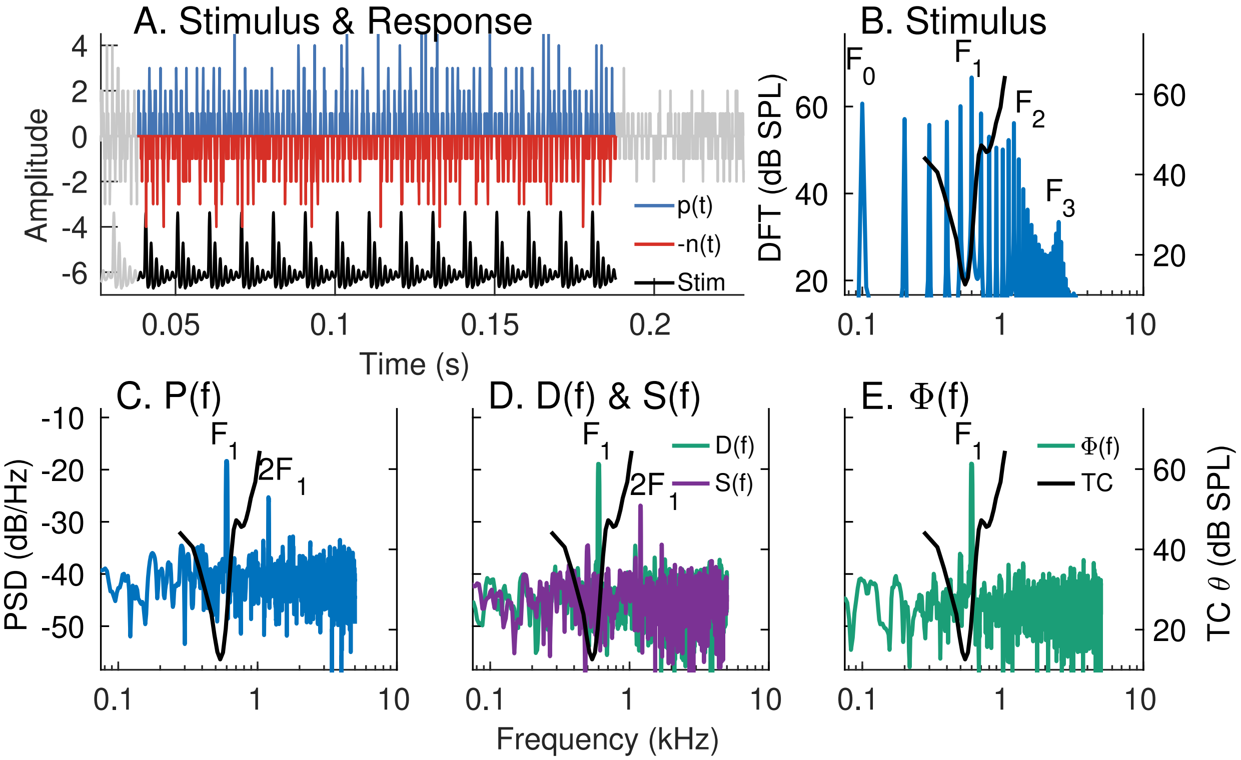

Figure 3.8. Spectral-domain application of various *apPSTHs* to spike trains recorded in response to a stationary vowel. Example of spectral analyses of spike trains recorded from an AN fiber (CF= 530 Hz, SR=90 spikes/s) in response to a synthesized stationary vowel (s_1 described in *Materials and Methods*, fundamental frequency: $F_0 = 100$ Hz, first formant: $F_1 = 600$ Hz). (A) Time-domain representation of $p(t)$, $n(t)$, and the stimulus ($Stim$). $n(t)$ is flipped along the y-axis for display. Signals outside the analysis window are shown in gray. PSTH bin width = 0.1 ms. Number of stimulus repetitions per polarity = 30. Stimulus intensity = 65 dB SPL. (B) Stimulus spectrum (blue, left yaxis). (C) $P(f)$. (D) Spectra for difference [$D(f)$, green] and sum [$S(f)$, purple] PSTHs. (E) Spectra of Hilbert-based TFS PSTH [$\Phi(f)$, green]. In panels B-E, the frequency-threshold tuning curve (TC θ, black) of the neuron is plotted on the right y-axis. $P(f)$ and $S(f)$ are corrupted by rectifier distortion at $2F_1$ frequency. The response primarily reflects TFS-based F_1 coding (E) and little envelope coding (D), which is consistent with the "synchrony capture" phenomenon for stationary vowel coding.

at $2F_1$ due to rectifier distortion at twice the TFS (F_1) frequency. Except for this rectifier distortion, there is little energy at other frequencies, including at the fundamental frequency, in $S(f)$ demonstrating weak envelope coding in the response. The response of this neuron primarily reflects TFS coding of F_1 [$\Phi(f)$ in Fig 3.8E] where the spectrum shows substantial energy only at the 6th harmonic of the fundamental frequency. These results are consistent with the previously reported phenomenon of “synchrony-capture” in the neural coding of stationary vowels (Delgutte and Kiang, 1984a; Young and Sachs, 1979). In “synchrony-capture”, the response of a neuron with CF near a formant is dominated by the harmonic of the fundamental frequency closest to the formant. As the response primarily follows a single sinusoid, the Hilbert-envelope PSTH, e(t), is essentially flat across the vowel duration and has little energy other than at 0 Hz [not shown for ease of visualization of $\Phi(f)$]. As a result, there is little difference between $\phi(t)$ and $d(t)$ for this stationary vowel.

3.5.5 Neural characterization of ENV and TFS using *apPSTHs* for a natural speech segment

Most previous studies have used the period histogram to study speech coding in the spectral domain (Delgutte and Kiang, 1984a; Young and Sachs, 1979). The period histogram is limited to stationary periodic stimuli, which were employed in those studies. In contrast, the use of *apPSTHs* facilitates the spectral analysis of neural responses to natural speech stimuli, which need not be stationary (Fig 3.9). In this example, the response of a low-frequency AN fiber to a 100-ms vowel segment of the s_3 natural speech sentence was considered. The CF (1.1 kHz) of this neuron is close to the second formant (F_2) of this segment (Fig 3.9B). $P(f)$ shows peaks corresponding to F_2 (∼1.2 kHz) and F_0 (∼130 Hz, Fig 3.9C). Similar to Fig 3.8, both $D(f)$ and $\Phi(f)$ show substantial energy near the formant closest to the neuron’s CF. In contrast to Fig 3.8, $S(f)$ [and $E(f)$] shows substantial energy near the fundamental frequency (inconsistent with synchrony capture). A detailed discussion of this discrepancy is beyond the scope of the present report, except to say that this lack of synchrony capture for natural vowels is a consistent finding that will be reported in a future study. The presence of substantial energy near F_0 in $E(f)$ indicates that $d(t)$ is corrupted by pitch-related modulation in e(t). This is because, mathematically, $D(f)$ is the convolution of the true TFS spectrum [$\Phi(f)$] and the Hilbert-envelope spectrum [$E(f)$]. Overall, these results demonstrate the application of various *apPSTHs* to study the neural representation of natural nonstationary speech stimuli in the spectral domain.

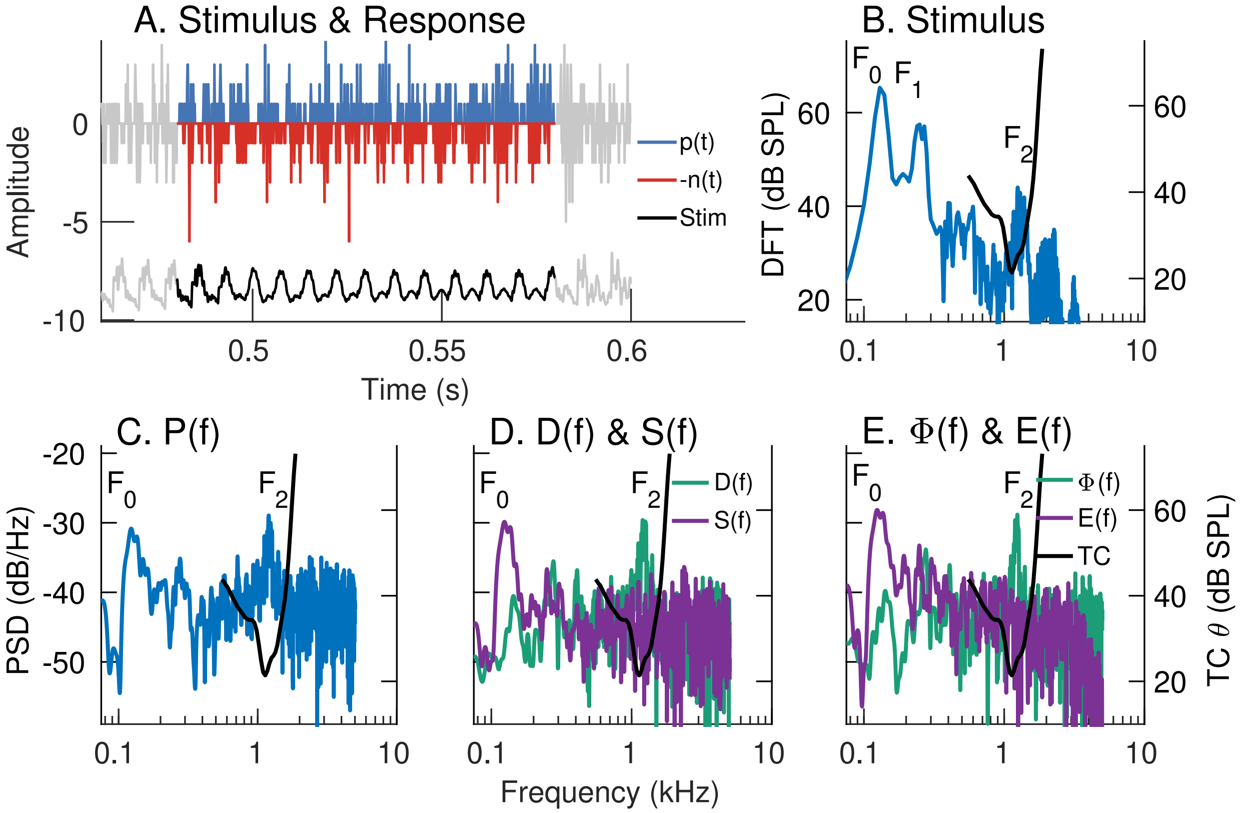

Figure 3.9. Spectral-domain application of various *apPSTHs* to spike trains recorded in response to natural speech. Example of spectral analyses of spike trains recorded from an AN fiber (CF= 1.1 kHz, SR=64 spikes/s) in response to a vowel snippet of a speech stimulus (s_3). Same format as Fig 3.8. PSTH bin width = 0.1 ms. Number of stimulus repetitions per polarity = 50. Stimulus intensity = 65 dB SPL. $P(f)$ shows comparable energy at F_0 (130 Hz) and F_2 (1.2 kHz). Both $S(f)$ and $E(f)$ show peaks near F_0. Similarly, both $D(f)$ and $\Phi(f)$ show good F_2 representations, although $D(f)$ is corrupted by the strong F_0-related modulation in e(t) as $d(t) = \mathrm{e}(t) \times \phi(t)$. The significant representation of F_0 in this near-F_2 AN fiber response to a natural vowel is inconsistent with the synchrony-capture phenomenon for synthetic stationary vowels.

3.5.6 Onset envelope is well represented in the sum PSTH but not in the Hilbert-envelope PSTH

In addition to analyzing spectral features, *apPSTHs* can also be used to analyze temporal features in the neural response. An example temporal feature is the onset envelope, which has been shown to be important for the neural coding of consonants (Delgutte, 1980; Heil, 2003), in particular fricatives (Delgutte and Kiang, 1984b). A diminished onset envelope in the peripheral representation of consonants has been hypothesized to be a contributing factor for perceptual deficits experienced by hearing-impaired listeners (Allen and Li, 2009), and thus is an important feature to quantify. Fig 3.10 shows example onset responses for a high-frequency AN fiber (CF= 5.8 kHz, SR= 70 spikes/s) for a fricative (/s/) portion of the speech stimulus s_3. The onset is well captured in single-polarity PSTHs [$p(t)$ and $n(t)$, Fig 3.10A] and in the sum envelope [$s(t)$, Fig 3.10B]. Since the onset is a polarity-tolerant feature, it is greatly reduced by subtracting the PSTHs to opposite polarities. As a result, the response onset is poorly captured in $d(t)$ (Fig 3.10C) and its Hilbert envelope, e(t) (Fig 3.10D).

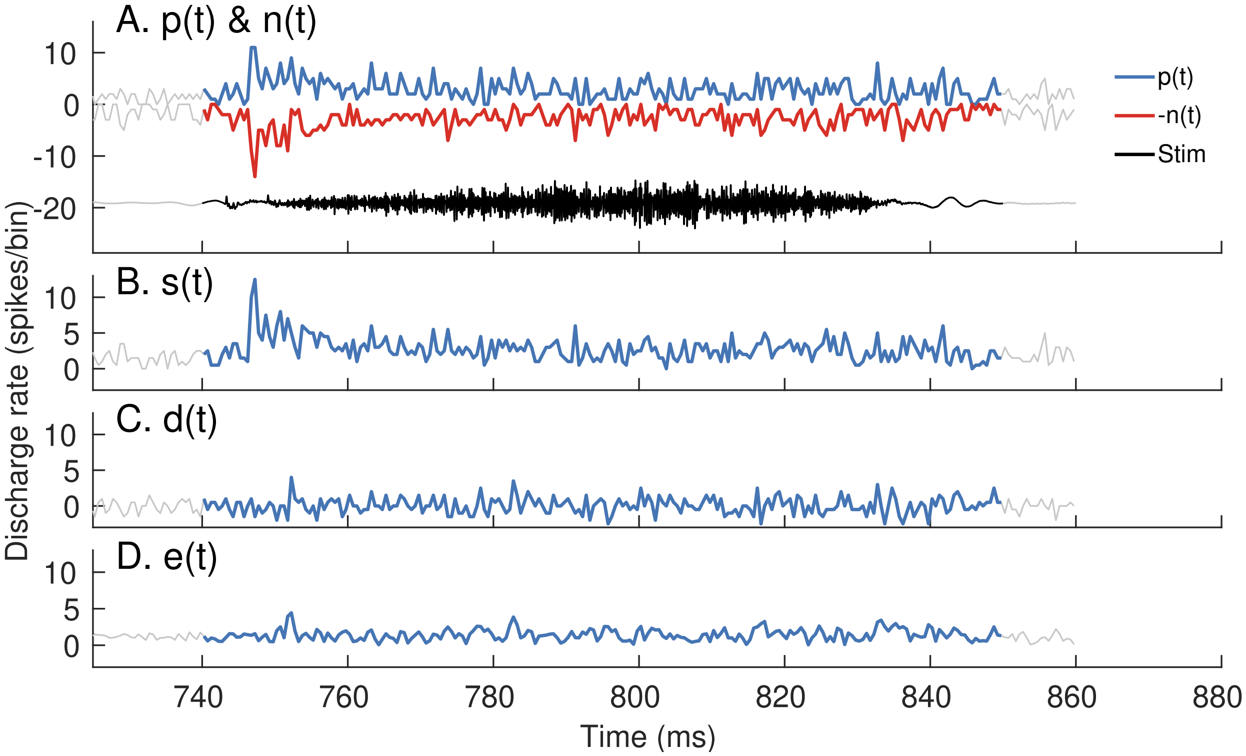

Figure 3.10. $p(t)$, $n(t)$, and $s(t)$ have robust representations of the onset response, whereas e**(t) and $d(t)$ do not.** Response of a high-frequency fiber (CF= 5.8 kHz, SR= 70 spikes/s) to a fricative portion (/s/) of the speech stimulus, s_3. Stimulus intensity = 65 dB SPL. *(A)* Stimulus (black, labeled *Stim*), $p(t)$ (blue) and $n(t)$ (red, flipped along the y-axis for display). PSTH bin width = 0.5 ms. Number of stimulus repetitions per polarity = 50. (B) The sum envelope, $s(t)$ (C) The difference PSTH, $d(t)$, and (D) the Hilbert-envelope PSTH, e(t). Since the onset envelope is a polarity-tolerant response, all PSTHs capture the response onset except for $d(t)$ and e(t).

Overall, these examples show that *apPSTHs* can be used to study various spectral and temporal features in the neural response for natural and synthesized stimuli in the ENV/TFS dichotomy. These *apPSTHs* are summarized in Table 3.2.

Table 3.2. *apPSTH*-taxonomy for ENV & TFS. We define *apPSTHs* as the collection of PSTHs derived using both polarities of the stimulus. The pair of PSTHs, $p(t)$ and $n(t)$, is a sufficient statistic for *apPSTHs* since all other PSTHs in the group can be derived from the two. Alternatively, the pair, $d(t)$ and $s(t)$, is also a sufficient statistic for *apPSTHs*. Each PSTH (e.g., the positive polarity PSTH) can be expressed in the time domain $[p(t)]$ or in the frequency domain $[P(f)]$. RD = Rectifier distortion. (T,F) = (Time, Frequency).

PSTH name	Notation: (T,F)	Definition	ENV and/or TFS representation	RD	Comments
Positive	$p(t), P(f)$	+ve polarity	TFS & ENV	Large	
Negative	$n(t), N(f)$	-ve polarity	TFS & ENV	Large	
Difference	$d(t), D(f)$	$\frac{p(t)-n(t)}{2}$	TFS & ENV	Small	Includes both the carrier and sideband components (thus not a clean representation of TFS)
Sum	$s(t), S(f)$	$\frac{p(t)+n(t)}{2}$	ENV	Large	Consistent representation of spectrally specific modulation strength, but corrupted by rectifier distortion at 2×CF
Analytic	$a(t), A(f)$	$d(t) + \jmath\mathcal{H}\{d(t)\}$	TFS & ENV	Small	$\mathcal{H}\{\cdot\}$ is the Hilbert transform operator
Hilbert envelope	$\mathrm{e}(t), E(f)$	$\|a(t)\|/\sqrt{2}$	ENV	Small	Polarity-sensitive ENV (subject to TFS phase locking)
Hilbert phase	$\phi(t), \Phi(f)$	$\sqrt{2} \times rms[d(t)] \times cos[\angle a(t)]$	TFS	Small	Carrier TFS (subject to TFS phase locking)

3.6 Quantifying ENV and TFS using *apPSTHs* for nonstationary signals

In the discussion so far, we have argued for using spectrally specific metrics to analyze neural responses to stationary stimuli. Another example where spectral specificity is needed is in evaluating the neural coding of nonstationary speech features (e.g., formant transitions). Speech is a nonstationary signal and conveys substantial information in its dynamic spectral trajectories (e.g., Fig 3.1A). A number of studies have investigated the robustness of the neural representation of dynamic spectral trajectories using frequency glides and frequency-modulated tones as the stimulus (Billings et al., 2019; Clinard and Cotter, 2015; Krishnan and Parkinson, 2000; Skoe and Kraus, 2010). These studies have usually employed a spectrogram analysis. While a spectrogram is effective for analyzing responses to nonstationary signals with unknown parameters, it does not explicitly incorporate information about the stimulus, which is often designed by the experimenter. Since the spectrogram relies on a narrow moving temporal window, it offers poor spectral resolution due to the time-frequency uncertainty principle. Instead of using conventional spectrogram analyses, frequency demodulation and filtering can be used together to estimate power along a spectrotemporal trajectory more accurately as we describe below. These spectrally specific analyses will facilitate more sensitive metrics to investigate the coding differences between nonstationary features in natural speech and extensively studied stationary features in synthetic stationary speech.

3.6.1 Frequency-demodulation-based spectrotemporal filtering

First, we describe the spectrotemporal filtering technique using an example stimulus with dynamic spectral components (Fig 3.11). The 2-second-long stimulus consists of three spectrotemporal trajectories: (1) a stationary tone at 1.4 kHz, (2) a stationary tone at 2 kHz, and (3) a dynamic linear chirp that moves from 400 to 800 Hz over the stimulus duration. We are interested in estimating the power of the nonstationary component, the linear chirp. In order to estimate the power of this chirp, conventional spectrograms will employ one of the following two approaches. First, one can use a long window (e.g., 2 seconds) and compute power over the 400-Hz bandwidth from 400 to 800 Hz. In the second approach,

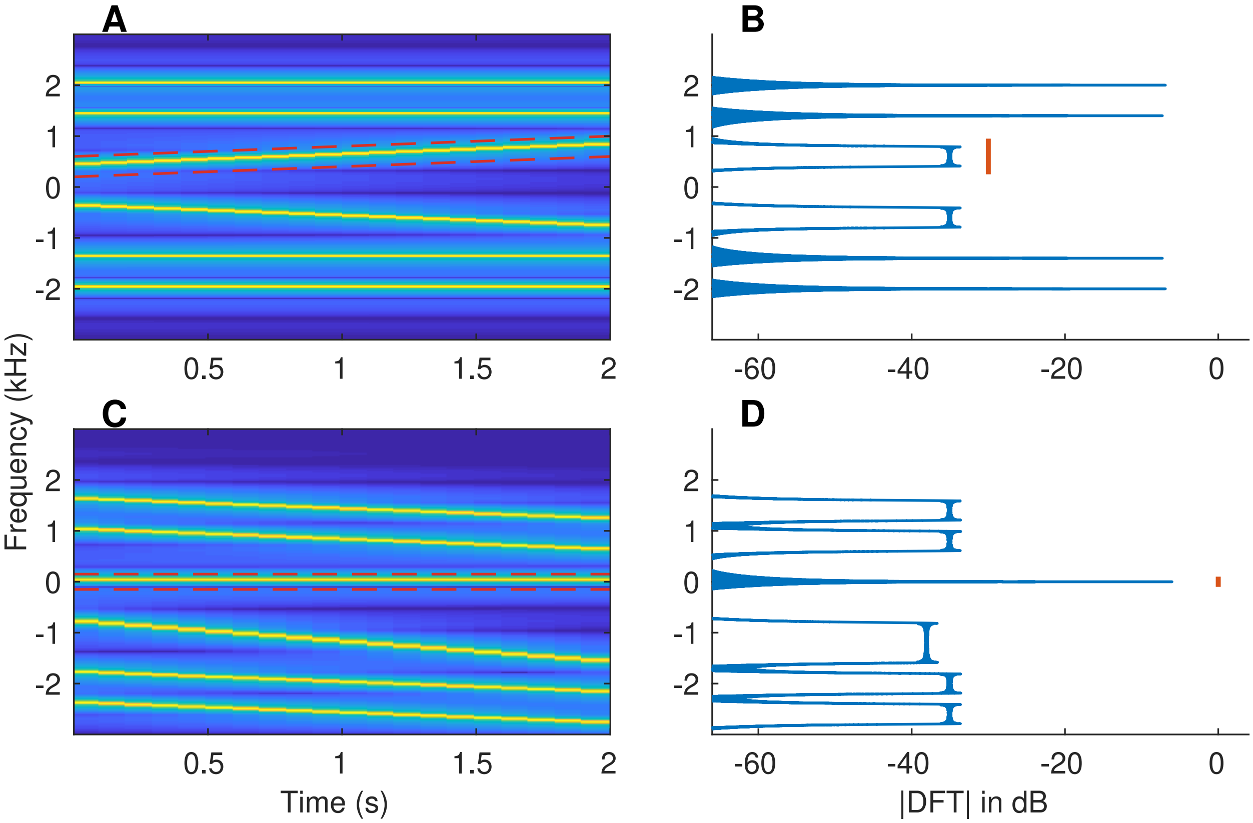

Figure 3.11. More accurate estimates of power along a spectrotemporal trajectory can be obtained using frequency demodulation. (A) The spectrogram of a synthesized example signal that mimics a single speech-formant transition. The 2-s long signal consists of two stationary tones (1.4 and 2 kHz) and a linear frequency sweep (400 to 800 Hz). Red dashed lines outline the spectrotemporal trajectory along which we want to compute the power. Both positive and negative frequencies are shown for completeness. (B) The Fourier-magnitude spectrum of the original signal. The energy related to the target spectrotemporal trajectory is spread over a wide frequency range (400 to 800 Hz, red line). (C) The spectrogram of the frequency-demodulated signal, where the target trajectory was used for demodulation (i.e., shifted down to 0 Hz). (D) The magnitude-DFT of the frequency-demodulated signal. The desired trajectory is now centered at 0 Hz, with its (spectral) energy spread limited only by the signal duration (i.e., equal to the inverse of the signal duration), and hence, is much narrower.

one can use moving windows that are shorter in duration (e.g., 50 ms) and compute power with a resolution of 30 Hz (20-Hz imposed by inverse of the window duration and 10-Hz imposed by change in chirp frequency over 50 ms). As an alternative to these conventional approaches, one can demodulate the spectral trajectory of the linear chirp so that the chirp is demodulated to near 0 Hz (Fig 3.11C and 3.11D, see *Materials and Methods*). Then, a low-pass filter with 0.5-Hz bandwidth (as determined by the reciprocal of the 2-s stimulus duration) can be employed to estimate the time-varying power along the chirp trajectory. This time-varying power is estimated at the stimulus sampling rate, similar to the temporal sampling of the output of a band-pass filter applied on stationary signals. While the same temporal sampling can be achieved using the spectrogram by sliding the window by one sample and estimating the chirp-related power for each window, it will be computationally much more expensive compared to the frequency-demodulation-based approach. Furthermore, the spectral resolution of 0.5 Hz is the same as that for a stationary signal, which demonstrates a 60-fold improvement compared to the 50-ms window-based spectrogram approach.

3.6.2 The *harmonicgram* for synthesized nonstationary speech

As shown in Fig 3.11, the combined use of frequency demodulation and low-pass filtering can provide an alternative to the spectrogram for analyzing signals that have time-varying frequency components. Such an approach can also be used to study the coding of dynamic stimuli that have harmonic spectrum with time-varying F_0, such as music and voiced speech. At any given time, a stimulus with a harmonic spectrum has substantial energy only at multiples of the fundamental frequency, F_0, which itself can vary with time [i.e., $F_0(t)$]. We take advantage of this spectral sparsity of the signal to introduce a new compact representation, the *harmonicgram*. Consider the k-th harmonic of the time-varying $F_0(t)$; power along this trajectory $[kF_0(t)]$ can be estimated using the frequency-demodulation-based spectrotemporal filtering technique. One could estimate the time-varying power along all integer multiples (k) of $F_0(t)$. This combined representation of the time-varying power across all harmonics of F_0 is the *harmonicgram* (see *Materials and Methods*). This name derives from the fact that this representation uses harmonic number instead of frequency (or spectrum) as in the conventional spectrogram.

Fig 3.12 shows harmonicgrams derived from *apPSTHs* in response to the nonstationary synthesized vowel, s_2. The first two formants are represented by their harmonic numbers, $F_1(t)/F_0(t)$ and $F_2(t)/F_0(t)$, which are known a priori in this case. Two harmonicgrams were constructed using responses from two AN fiber pools: (1) AN fibers that had a low CF (CF

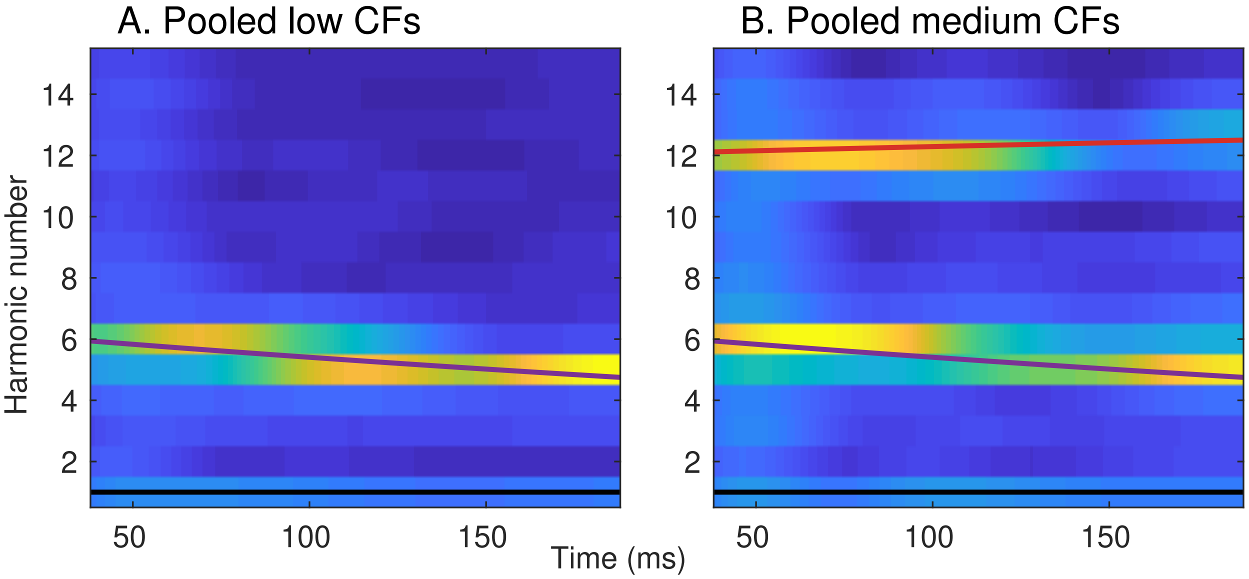

Figure 3.12. The harmonicgram can be used to visualize formant tracking in synthesized nonstationary speech. Neural harmonicgrams for fibers with a CF below 1 kHz (A, N=16) and for fibers with a CF between 1 and 2.5 kHz (B, N=29) in response to the dynamic vowel, s_2. Stimulus intensity = 65 dB SPL. The formant frequencies mimic formant trajectories of a natural vowel (Hillenbrand and Nearey, 1999). A 20-Hz bandwidth was employed to low-pass filter the demodulated signal for each harmonic. The harmonicgram for each AN-fiber pool was constructed by averaging the Hilbert-phase PSTHs of all AN fibers within the pool. PSTH bin width = 50 μs. Data are from one chinchilla. The black, purple and, red lines represent the fundamental frequency (F_0/F_0), the first formant (F_1/F_0) and the second formant (F_2/F_0) contours, respectively. The time-varying formant frequencies were normalized by the time-varying F_0 to convert the spectrotemporal representation into a harmonicgram.

< 1 kHz), and (2) AN fibers that had a medium CF (1 kHz < CF < 2.5 kHz). Previous neurophysiological studies have shown that AN fibers with CF near and slightly above a formant strongly synchronize to that formant, especially at moderate to high intensities (Delgutte and Kiang, 1984a; Young and Sachs, 1979). Therefore, the low-CF pool was expected to capture F_1, which changed from 630 Hz to 570 Hz. Similarly, the medium-CF pool was expected to capture F_2, which changed from 1200 Hz to 1500 Hz. The harmonicgram for each pool was constructed by using the average Hilbert-phase PSTH, $\phi(t)$, of all AN fibers in the pool. The harmonicgram is shown from 38 ms to 188 ms to optimize the dynamic range to visually highlight the formant transitions by ignoring the onset response. The dominant component

in the neural response for F_1 was expected at the harmonic number closest to F_1/F_0. For this stimulus, F_1/F_0 started at a value of 6.3 (630/100) and reached 4.75 (570/120) at 188 ms crossing 5.5 at 88.5 ms (Fig 3.12A). This transition of F_1/F_0 was faithfully represented in the harmonicgram where the dominant response switched from the 6th to the 5th harmonic near 90 ms. Similarly, F_2/F_0 started at 12, consistent with the dominant response at the 12th harmonic before 100 ms (Fig 3.12B). Towards the end of the stimulus, F_2/F_0 reached 12.5, which is consistent with the near-equal power in the 12th and the 13th harmonic in the harmonicgram. In contrast to findings from previous studies, the harmonicgram for the medium-CF pool indicates that these fibers respond to both the first and second formants. Such a complex response with components corresponding to multiple formants is likely due to the steep slope of the vowel spectrum (S2 Fig).

3.6.3 The harmonicgram for natural speech

The harmonicgram analysis is not limited to synthesized vowels, but it can also be applied to natural speech (Fig 3.13). These harmonicgrams were constructed for the natural speech stimulus, s_3, using average $\phi(t)$ for the same low-CF and medium-CF AN fiber pools that were used in Fig 3.12. Here, we consider a 500-ms segment of the stimulus, which contains multiple phonemes. Qualitatively, similar to Fig 3.12, these harmonicgrams capture the formant contours across phonemes. The harmonicgram for the low-CF pool emphasizes the F_1 contour, whereas the harmonicgram for the medium-CF pool primarily emphasizes the F_2 contour, and to a lesser extent, the F_1 contour. Compared to the spectrogram, the harmonicgram representation for these responses are more compact and spectrally specific. Furthermore, from a neural-coding perspective, quantifying how individual harmonics of F_0 are represented in the response is more appealing than the spectrogram since response energy is concentrated only at these F_0 harmonics.

The harmonicgram not only provides a compact representation for nonstationary signals with harmonic spectra, it can also be used to quantify the coding strength of time-varying features, such as formants for speech (Figs 3.13E and 3.13F). In these examples, the strength of formant coding at each time point, t, was quantified as the sum of power in the three harmonics closest to the F_0-normalized formant frequency at that time [e.g., $F_1(t)/F_0(t)$ for the first formant]. As expected, power for the harmonics near the first formant was substantially greater than that for the second formant for the low-CF pool (Fig 3.13E). For the medium-CF pool, F_2 representation was robust over the whole stimulus duration, although F_1 representation was largely comparable (Fig 3.13F). These examples demonstrate novel analyses using the *apPSTH*-based harmonicgram to quantify time-varying stimulus features in single-unit neural responses at high spectrotemporal resolution that are not possible with conventional windowing-based approaches.

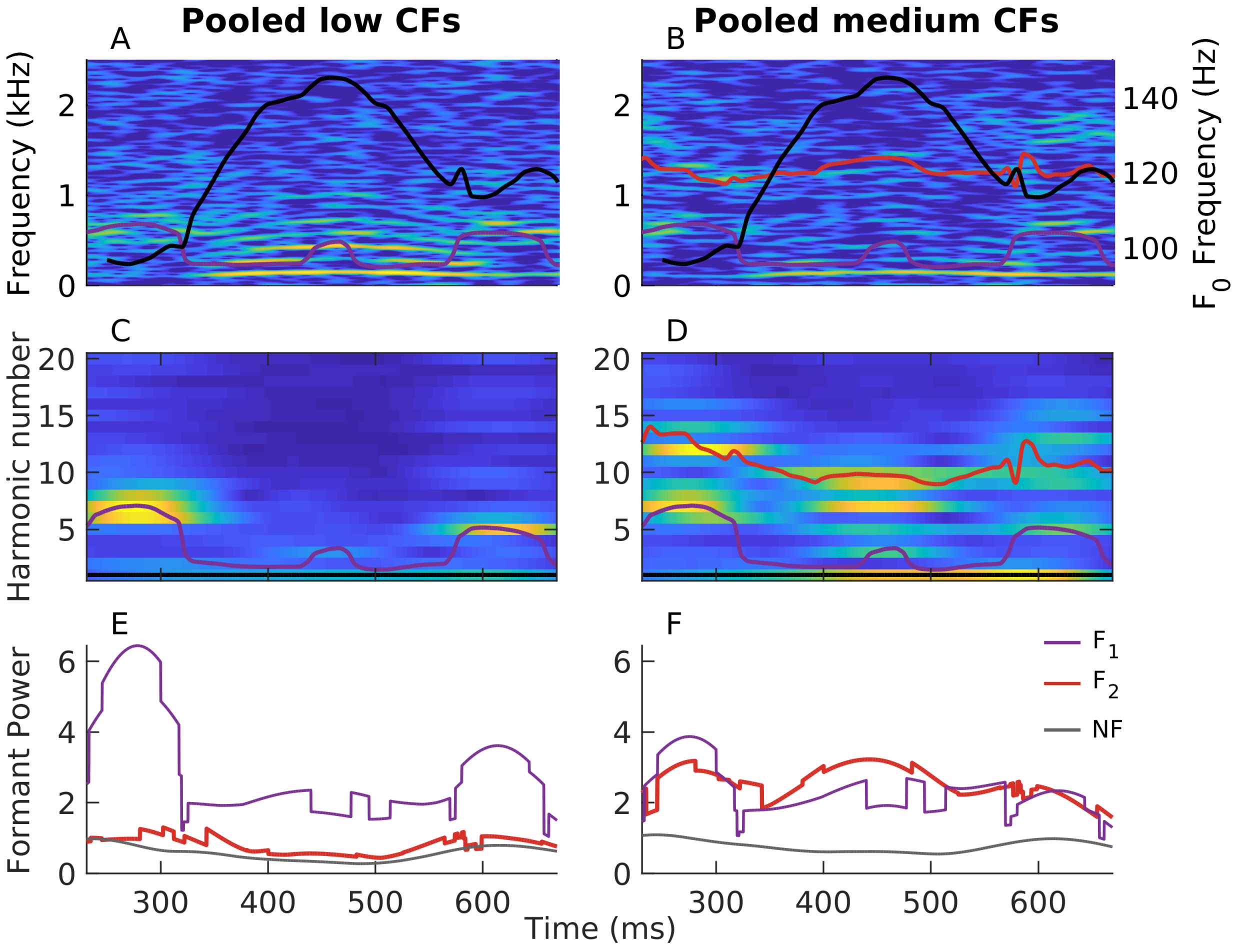

Figure 3.13. *See next page*

Figure 3.13 *(continued from previous page.)* **The harmonicgram can be used to quantify the coding of time-varying stimulus features at superior spectrotemporal resolution compared to the spectrogram.** Harmonicgrams were constructed using $\phi(t)$ for the same two AN-fiber pools described in Fig 3.12. PSTH bin width = 50 μs. A 9-Hz bandwidth was employed to low-pass filter the demodulated signal for each harmonic. The data were collected from one chinchilla in response to the speech stimulus, s_3. Stimulus intensity = 65 dB SPL. A 500-ms segment corresponding to the voiced phrase "amle" was considered. (A, B) Spectrograms constructed from the average $\phi(t)$ for the low-CF pool (A) and from the medium-CF pool (B). (C, D) Average harmonicgrams for the same set of fibers as in A and B, respectively. Warm (cool) colors represent regions of high (low) power. The first-formant contour (F_1 in A and B, F_1/F_0 in C and D) is highlighted in purple. The second-formant contour (F_2 in A and B, F_2/F_0 in C and D) is highlighted in red. Trajectories of the fundamental frequency (black in A and B, right Y axis) and the formants were obtained using Praat (Boersma, 2001). (E, F) Harmonicgram power near the first formant (purple) and the second formant (red) for the low-CF pool (E) and the medium-CF pool (F). Harmonicgram power for each formant at any given time (t) was computed by summing the power in the three closest F_0 harmonics adjacent to the normalized formant contour [e.g., $F_1(t)/F_0(t)$] at that time. The noise floor (NF) for power was estimated as the sum of power for the 29th, 30th, and 31st harmonics of F_0 because the frequencies corresponding to these harmonics were well above the CFs of both fiber pools. These time-varying harmonicgram power metrics are spectrally specific to F_0 harmonics and are computed with high temporal sampling rate (same as the original signal). This spectrotemporal resolution is much better than the spectrotemporal resolution that can be obtained using spectrograms.

3.6.4 The harmonicgram can also be used to analyze FFRs in response to natural speech

As mentioned earlier, a major benefit of using *apPSTHs* to analyze spike trains is that the same analyses can also be applied to evoked far-field potentials. In Fig 3.14, the harmonicgram analysis was applied to the difference FFR recorded in response to the same speech sentence (s_3) that was used in Fig 3.13. In fact, these FFR data and spike-train data used in Fig 3.13 were collected from the same chinchilla. The difference FFR was computed as the difference between FFRs to opposite polarities of the stimulus. The spectrogram and harmonicgram can also be constructed using the Hilbert-phase FFR to highlight the TFS component of the response (S3 Fig). Unlike the *apPSTHs* for AN fibers, the FFR cannot be used to construct two sets of harmonicgrams corresponding to different populations of neurons because the FFR lacks tonotopic specificity. Nevertheless, this FFR-harmonicgram is strikingly similar to the medium-CF pool harmonicgram in Fig 3.13D. The dynamic representations of the first two formants are robust in both the representations. In fact, the FFR representations seem more robust in formant tracking compared to PSTH-derived representations, qualitatively, especially for the harmonicgram. A more uniform sample of neurons contribute to evoked responses compared to the AN fiber sample corresponding to Fig 3.13, which could be a factor for the robustness of the FFR representations. Overall, these results reinforce the idea that using *apPSTHs* to analyze spike trains offers the same spectrally specific analyses that can be applied to evoked far-field potentials, e.g., the FFR, thus allowing a unifying framework to study temporal coding for both stationary and nonstationary signals in the auditory system.

3.7 Discussion

3.7.1 Use of *apPSTHs* underlies a unifying framework to study temporal coding in the auditory system

A better understanding of the neural correlates of perception requires the integration of electrophysiological, psychophysical, and neurophysiological analyses in the same framework. Although extensive literature exists in both electrophysiology and neurophysiology on the neural correlates of perception, the analyses employed in these studies have diverged. This disconnect is largely because the forms of the neural data are different (i.e., continuous-valued waveforms versus point-process spike trains). The present report provides a unifying framework for analyzing spike trains using *apPSTHs*, which offers numerous benefits over previous neurophysiological analyses. Specifically, the use of *apPSTHs* incorporates many of the previous ad-hoc approaches, such as VS and correlograms (Eqs 3.8 to 3.10). In fact, correlograms and metrics derived from them can be estimated using *apPSTHs* in a computationally efficient way. The *apPSTHs* essentially convert the naturally rectified neurophysiological point-process data into a continuous-valued signal, which allows advanced signal processing tools designed for continuous-valued signals to be applied to spike-train

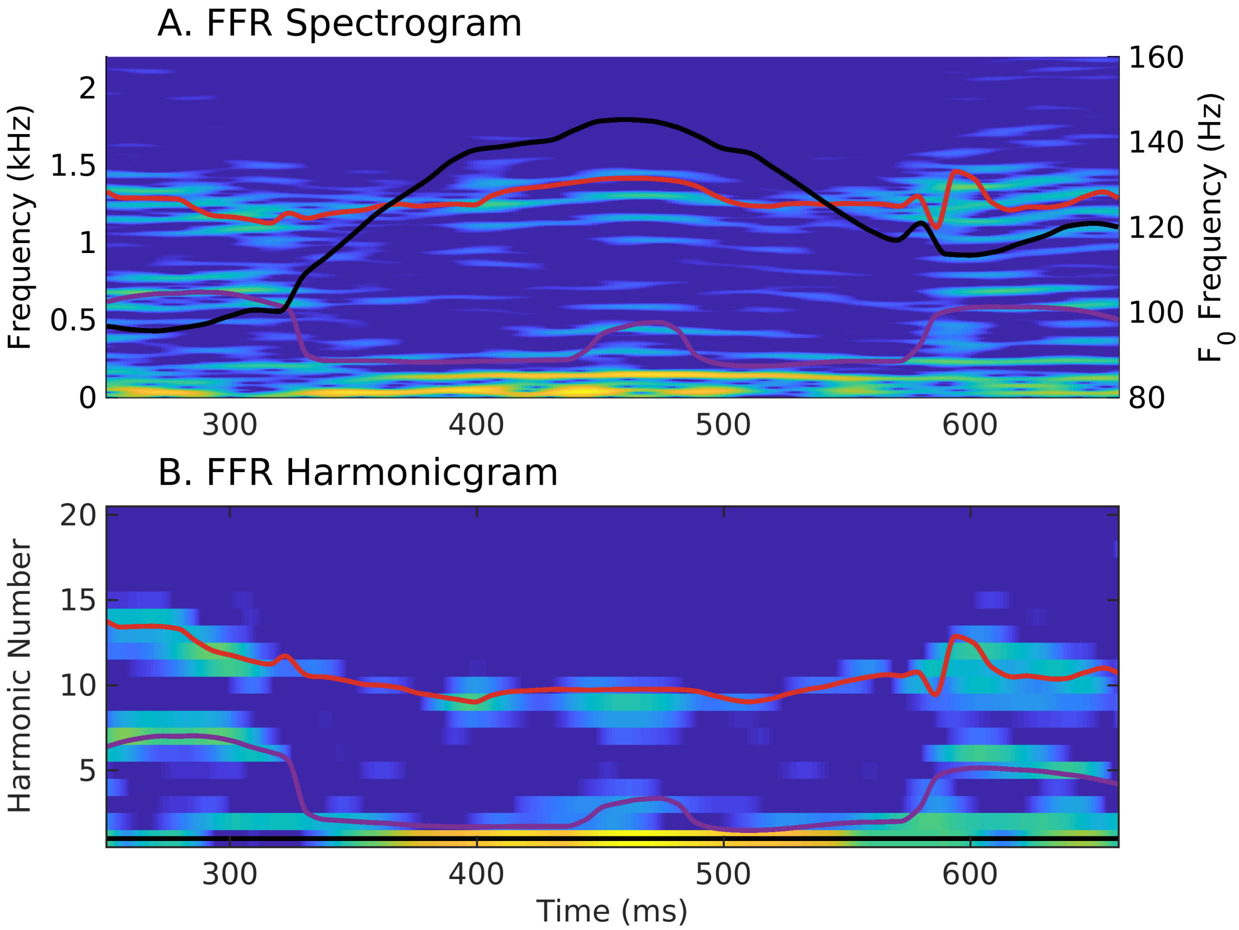

Figure 3.14. The harmonicgram of the FFR to natural speech shows robust dynamic tracking of formant trajectories, similar to the AN-fiber harmonicgram. Comparison of the spectrogram (A) and the harmonicgram (B) for the FFR recorded in response to the same stimulus, s_3 that was used to analyze *apPSTHs* in Fig 3.13. Stimulus intensity = 65 dB SPL. Lines and colormap are the same as in Fig 3.13. These plots were constructed using the difference FFR, which reflects the neural coding of both stimulus TFS and ENV. To highlight the coding of stimulus TFS, Hilbert-phase [$\phi(t)$] FFR can be used instead of the difference FFR (S3 Fig). The FFR harmonicgram (A) is strikingly similar to the AN-fiber harmonicgrams in Figs 3.13C and 3.13D in that the representations of the first two formants are robust. The FFR data here and spike-train data used in Fig 3.13 were obtained from the same animal.

data. For example, *apPSTHs* can be used to derive spectrally specific TFS components [e.g., $\phi(t)$, Fig 3.7], multitaper spectra (Fig 3.4), modulation-domain representations (Fig 3.5), and harmonicgrams (Figs 3.12 and 3.13). *apPSTHs* can also be directly compared to evoked far-field responses for both stationary and nonstationary stimuli (e.g., Figs 3.13 and 3.14).

3.7.2 Temporal coding metrics should be spectrally specific

The various analyses explored in this report advocate for spectral specificity of temporal coding metrics in a variety of ways. The need for spectrally specific analyses arises for two reasons: (1) neural data is finite and inherently stochastic, and (2) spike-train data are rectified. Neural stochasticity exacerbates the spectral-estimate variance at all frequencies; therefore, time-domain (equivalently broadband) metrics will be noisier compared to narrowband metrics. Similarly, the rectified nature of spike-train data can introduce harmonic distortions in the response spectrum, which can corrupt broadband metrics (e.g., TFS distortion at two times the carrier frequency corrupting estimates of ENV coding, Figs 3.6A and 3.6B).

These issues requiring spectral specificity are not unique to the *apPSTH* analyses but also apply to classic metrics, e.g., correlograms. For example, the broadband correlation index (CI) metric is appropriate to analyze responses of neurons with high CFs, but the CI metric is corrupted by rectifier distortions for neurons with low CFs (Heinz and Swaminathan, 2009; Joris et al., 2006). Studies have previously tried to avoid these distortions in the *sumcor* by restricting the response bandwidth to below the CF because, for a given filter, the envelope bandwidth cannot be greater than the filter bandwidth (Heinz and Swaminathan, 2009; Kale and Heinz, 2010).

Here, we have extended and generalized the analysis of these issues using narrowband stimuli. In particular, when a neuron responds to low-frequency stimulus energy that is below half the phase-locking cutoff, responses that contain any polarity-tolerant component [e.g., $p(t)$, $n(t)$, $s(t)$, *SAC*, and *sumcor*] will be corrupted by rectifier distortion of the polarity-sensitive component (Fig 3.6E). Any broadband metric of temporal coding should exclude these distortions at twice the carrier frequency. Beyond avoiding rectifier distortion, limiting the bandwidth of a metric to only the desired bands will lead to more precise analyses by minimizing the effects of neural stochasticity (Fig 3.6H). For example, envelope coding metrics for SAM-tone stimuli should consider the spectrum power only at F_m and its harmonics (Vasilkov and Verhulst, 2019), rather than the simple approach of always low-pass filtering at CF (Heinz and Swaminathan, 2009).

Similar to envelope-based metrics, metrics that quantify TFS coding should also be spectrally specific to the carrier frequency. Previous metrics of TFS coding, such as $d(t)$ and *difcor*, are not specific to the carrier frequency but rather include modulation sidebands as well as additional sidebands due to transduction nonlinearities (Fig 3.7). In contrast, $\phi(t)$ introduced here emphasizes the carrier and suppresses the sidebands (Fig 3.7). Thus, the spectrally specific $\phi(t)$ is a better TFS response, which relates to the zero-crossing signal used in the signal processing literature (Logan, 1977; Voelcker, 1966b; Wiley, 1981).

3.7.3 Spectral-estimation benefits of using *apPSTHs*

Neurophysiological studies have usually favored the DFT to estimate the response spectrum. For example, the DFT has been applied to the period histogram (Delgutte and Kiang, 1984a; Young and Sachs, 1979), the single-polarity PSTH (Carney and Geisler, 1986; Miller and Sachs, 1983), the difference PSTH (Sinex and Geisler, 1983), and correlograms (Louage et al., 2004). Since spike-train data are stochastic and usually sparse and finite, there is great scope for spectral estimates, including the DFT spectrum, to suffer from bias and variance issues. The multitaper approach optimally uses the available data to minimize the bias and variance of the spectral estimate (Babadi and Brown, 2014; Percival and Walden, 1993; Thomson, 1982). The multitaper approach can be used with both *apPSTHs* and correlograms, but using *apPSTHs* offers additional variance improvement up to a factor of 2 (Fig 3.4). This improvement is because twice as many tapers (both odd and even) can be used with an *apPSTH* compared to a correlogram, which is an even sequence and limits analyses to only using even tapers.

3.7.4 Benefits of spectrotemporal filtering

Analysis of neural responses to nonstationary signals has been traditionally carried out using windowing-based approaches, such as the spectrogram. Shorter windows help with tracking rapid temporal structures, but they offer poorer spectral resolution. On the other hand, larger windows allow better spectral resolution at the cost of smearing rapid dynamic features. As an alternative to windowing-based approaches, spectrotemporal filtering can improve the spectral resolution of analyses by taking advantage of stimulus parameters that are known a priori (Fig 3.11). This approach is particularly efficient to analyze spectrally sparse signals (i.e., signals with instantaneous line spectra, such as voiced speech). In particular, the spectral resolution is substantially improved compared to the spectrogram. In addition, while the same temporal sampling can be obtained using the spectrogram, it will be much more computationally expensive compared to the spectrotemporal filtering approach as discussed in the following example.

Consider the signal in Fig 3.11). The neural response for this signal will have noise over the whole response bandwidth due to neural stochasticity, in addition to the representation of the signal components. Let us assume that we are interested in comparing the coding of the 1400-Hz stationary tone and the linear chirp (400 to 800 Hz sweep over 2 seconds). The coding of these signals can be quantified by estimating the power of these components in the response. For the stationary tone, power can be estimated over a 0.5-Hz band around 1400 Hz. For the chirp, one of the following approaches can be employed. Power can be estimated in a long temporal window, which improves the spectral resolution of the analysis. However, estimating power for the chirp will require a 400 Hz spectral window, which is determined by the chirp bandwidth. Since at any given time the signal has power only at three frequencies, use of such a large spectral window will allow the response noise to introduce significant bias to the estimated chirp-related power in the response since power is a nonnegative random variable. On the other extreme, one could use a shorter window, say 2 ms, such that the chirp frequency is nearly constant over this window. However, a 500-Hz bandwidth limitation is posed due to time-frequency uncertainty. In contrast to the spectrogram, the spectrotemporal filtering approach demodulates the chirp trajectory onto a single frequency, and thus, has a spectral resolution of 0.5 Hz (inverse of signal duration). This spectral resolution is the same as the one used for the 1400-Hz stationary tone, permitting an equivalent comparison. Moreover, the filtered signal has the same temporal sampling rate as the original response. Achieving the same temporal sampling using spectrograms will be computationally extremely expensive because one has to slide the window by one sample and compute the PSD for each window. Thus, by combining *apPSTHs* with advanced signal processing approaches, the response to the 1400-Hz tone can be compared with the response to the chirp, at the narrowest spectral resolution possible (inverse of the signal duration) and at the original temporal sampling rate.

The benefits of spectrotemporal filtering extend to other spectrally sparse signals, like harmonic complexes. A priori knowledge of the fundamental frequency can be used to construct the harmonicgram, which takes advantage of power concentration at harmonics of F_0. Such an approach contrasts with the spectrogram, which computes power at all frequencies uniformly. The harmonicgram can be used to analyze both kinematic synthesized vowels (Fig 3.12) as well as natural speech (Fig 3.13). The harmonicgram is particularly useful in quantifying dominant harmonics in the response at high temporal sampling, and is thus applicable to nonstationary signals. The harmonicgram can also be applied to evoked far-field potentials (e.g., the FFR in Fig 3.14). While alternatives exist to analyze spike-train data in response to time-varying stimuli (Brown et al., 2002), the present spectrotemporal technique is simpler and can be directly applied to both spike-train data and far-field responses. Overall, these results support the idea that using *apPSTHs* to analyze spike trains provides a unifying framework to study temporal coding in the auditory system across modalities. Furthermore, this framework facilitates the study of dynamic-stimulus coding by the nonlinear and time-varying auditory system.

3.7.5 *apPSTHs* allow animal models of sensorineural hearing loss to be linked to psychophysical speech-intelligibility models

Speech-intelligibility models not only improve our understanding of perceptually relevant speech features, they can also be used to optimize hearing-aid and cochlear-implant strategies. However, existing SI models work well for normal-hearing listeners but have not been widely extended for hearing-impaired listeners. This gap is largely because of the fact that most SI models are based on signal-processing algorithms in the acoustic domain, where individual differences in the physiological effects of various forms of sensorineural hearing loss on speech coding are difficult to evaluate. This gap can be addressed by extending acoustic SI models to the neural spike-train domain. In particular, spike-train data obtained from preclinical animal models of sensorineural hearing loss can be used to explore the neural correlates of perceptual deficits faced by hearing-impaired listeners (Trevino et al., 2019). These insights will be crucial for developing accurate SI models for hearing-impaired listeners.

apPSTHs offer a straightforward means to study various speech features in the neural spike-train domain. As *apPSTHs* are in the same discrete-time continuous-valued form as acoustic signals, acoustic SI models can be directly translated to the neural domain. Many successful SI models are based on the representation of the temporal envelope (Jørgensen and Dau, 2011; Relaño-Iborra et al., 2016), although the role of TFS remains a matter of controversy (Lorenzi et al., 2006). In fact, recent studies have suggested that the peripheral representation of TFS can shape the central envelope representation, and thereby alter speech perception outcomes (Ding et al., 2014; Viswanathan et al., 2019). *apPSTHs* can be used to derive modulation-domain representations so that envelope-based SI models can be evaluated in the neural domain (Fig 3.5). Similarly, the Hilbert-phase PSTH, $\phi(t)$, can be used to evaluate the neural representation of TFS features. These TFS results will be particularly insightful for cochlear-implant stimulation strategies that rely on the zero-crossing component of the stimulus, which closely relates to $\phi(t)$ (Chen and Zhang, 2011; Grayden et al., 2004).

Translational benefits of animal models

A key motivation of this paper was to develop a framework so that insights and findings from animal models can ultimately improve our understanding of how the human auditory system processes real-life sounds, like speech. Experiments involving human subjects are typically limited to far-field responses, such as compound action potential, frequency-following responses, and auditory brainstem responses. However, these evoked responses include contributions from multiple sources such as the cochlear microphonic, electrical interferences, and responses from several neural substrates (King et al., 2016; Verschooten and Joris, 2014); these contributions are not clearly understood. The *apPSTH*-based framework offers a straightforward way to study these contributions by comparing anatomically specific spike-train responses with clinically viable noninvasive responses.

This framework is also beneficial to develop and validate noninvasive metrics using animal models and apply these metrics to humans. For example, we demonstrated the applicability of the new spectrally compact harmonicgram approach on both spike-train data and FFR data recorded from chinchillas to evaluate speech coding. This harmonicgram analysis can also be applied to FFR data recorded from humans to study natural speech coding in both normal and impaired auditory systems. Similarly, the representation of other important response features such as the onset and adaptation can also be linked between invasive and noninvasive data using animal models of different SNHL. Overall, these insights will be informative regarding the anatomical and physiological state of humans using noninvasive measures.

3.7.6 Limitations

Biological feasibility

The analyses proposed here aim to rigorously quantify the dichotomous ENV/TFS information in the neural response and bridge the definitions between the audio and neural spike-train domains. Methods discussed here may not all be biologically feasible. For example, the brain does not have access to both polarities of the stimulus. Thus, the PSTHs that require two polarities to be estimated, e.g., $s(t)$, $d(t)$, and $\phi(t)$, may not have an "internal representation" in the brain. This limitations also applies to correlogram metrics based on *sumcor* and *difcor*, which require two polarities of the stimulus. Thus, the use of the single-polarity PSTH [$p(t)$] to derive the central "internal representations" is more appropriate from a biological feasibility perspective (e.g., Fig 3.5). However, these various ENV/TFS components allow a thorough characterization of the processing of spectrotemporally complex signals by the nonlinear auditory system and can guide the development of more accurate speech-intelligibility models and help improve signal processing strategies for hearing-impaired listeners.

Alternating-polarity stimuli

Use of two polarities may not be sufficient to separate out all the components underlying the neural response when more than two components contribute to the neural response at a given frequency. In particular, it may be intractable to separate out rectifier distortion when the bandwidths of ENV and TFS in the response overlap. For example, consider the response of a broadly tuned AN fiber in response to a vowel, which has a fundamental frequency of F_0. The energy at $2F_0$ in $S(f)$ may reflect one or more of the following sources: (1) rectifier distortion to carrier energy at F_0, (2) beating between (carrier) harmonics that are separated by $2F_0$, and (3) effects of transduction nonlinearities on the beating between (carrier) harmonics that are separated by F_0. In these special cases, additional stimulus phase variations can be used to separate out these components (Billings and Zhang, 1994; Lucchetti et al., 2018).

The harmonicgram

A key drawback of applying the harmonicgram to natural speech is the requirement of knowing the F_0 trajectory. F_0 estimation is a difficult problem, especially in degraded speech. Thus, the harmonicgram could be inaccurate unless the F_0 trajectory is known, or at least the original stimulus is known so that F_0 can be estimated. A second confound is the unknown stimulus to response latency for different systems. Latencies for different neurons vary with their characteristic frequency, stimulus frequency, as well as stimulus intensity. Thus, even if the acoustic spectrotemporal trajectory is precisely known, errors may accumulate if latencies are not properly accounted for. This issue will likely be minor for spectrotemporal trajectories with slow dynamics. For stimuli with faster dynamics, latency confounds can be easily minimized by estimating stimulus-to-response latency by cross-correlation and using a larger cutoff frequency for low-pass filtering.

3.8 Appendix

3.8.1 Vector strength metric definitions

Vector Strength. The *vector strength* (*VS*) metric is used to quantify how well spikes in a spike train are synchronized to a frequency, f (Goldberg and Brown, 1969; Johnson, 1980). Let us denote a spike train with N spikes as $\underline{\zeta}$ such that $\underline{\zeta} = \{t_1, t_2, ..., t_N\}$ and the $\{t_\mathrm{i}\}$s are individual spike times. To compute the vector strength, these spike times are first transformed onto the unit circle such that t_i maps to z_i as

$$z_\mathrm{i} = \mathrm{e}^{\jmath 2\pi f t_\mathrm{i}}.$$

The mean of the set of complex vectors corresponding to all N spikes is

$$\rho(f) = \frac{1}{N}\sum_{\mathrm{i}=1}^{N} z_\mathrm{i} = \frac{1}{N}\sum_{\mathrm{i}=1}^{N} \mathrm{e}^{\jmath 2\pi f t_\mathrm{i}}. \tag{3.13}$$

Then, *VS* at frequency f is defined as the magnitude of $\rho(f)$.

$$\begin{aligned} VS(f) &= |\rho(f)| \\ &= \left|\frac{1}{N}\sum_{\mathrm{i}=1}^{N} z_\mathrm{i}\right| \\ &= \left|\frac{1}{N}\sum_{\mathrm{i}=1}^{N} [cos(2\pi f t_\mathrm{i}) + \jmath\ sin(2\pi f t_\mathrm{i})]\right| \\ &= \left\{\left[\frac{1}{N}\sum_{\mathrm{i}=1}^{N} cos(2\pi f t_\mathrm{i})\right]^2 + \left[\frac{1}{N}\sum_{\mathrm{i}=1}^{N} sin(2\pi f t_\mathrm{i})\right]^2\right\}^{\frac{1}{2}} \end{aligned} \tag{3.14}$$

Phase-projected Vector Strength. The *phase-projected vector strength* (VS_{pp}) is identical to the *VS* for a single spike train (i.e., for a single stimulus repetition), but these metrics differ when multiple (R) stimulus repetitions are used. VS_{pp} is advantageous relative to VS when there are relatively fewer spikes per repetition (Yin et al., 2010). To estimate VS_{pp} at frequency f, the magnitude (i.e., VS) and phase $[\phi_r(f)]$ of the mean complex vector are first calculated for individual repetitions using Eqs 3.13 and 3.14 (instead of pooling spike times across all R repetitions). The per-repetition VS estimates, called $VS^r(f)$, are weighted by the cosine of the phase difference between $\phi^r(f)$ of the repetition and the mean phase based on all spikes from all repetitions, $\phi^{ref}(f)$, to estimate the *phase-projected vector strength*, $VS^r_{pp}(f)$, for the repetition.

$$VS^r_{pp}(f) = VS^r(f)\ cos\left[\phi^r(f) - \phi^{ref}(f)\right],$$

where $\phi^r(f)$ for repetition r with N_r spikes $\{t^r_1, t^r_2, ..., t^r_{N_r}\}$ is computed as

$$\phi^r(f) = \tan^{-1}\frac{\sum_{\mathrm{i}=1}^{N_r} sin(2\pi f t^r_{\mathrm{i}})}{\sum_{\mathrm{i}=1}^{N_r} cos(2\pi f t^r_{\mathrm{i}})},$$

and $\phi^{ref}(f)$ is computed using all spikes across all R repetitions as

$$\phi^{ref}(f) = \tan^{-1}\frac{\sum_{r=1}^{R}\sum_{\mathrm{i}=1}^{N_r} sin(2\pi f t^r_{\mathrm{i}})}{\sum_{r=1}^{R}\sum_{\mathrm{i}=1}^{N_r} cos(2\pi f t^r_{\mathrm{i}})}.$$

$VS_{pp}(f)$ for R repetitions is computed as the mean $VS^r_{pp}(f)$ across all repetitions,

$$VS_{pp}(f) = \frac{1}{R}\sum_{\mathrm{i}=1}^{R} VS^r_{pp}(f).$$

3.8.2 Relation between the *vector strength* metric and the *difference PSTH*

Let us assume that we have R sets of spike trains $\{\underline{\zeta}_{\mathrm{i}}\} : \mathrm{i} \in [1, .., R]$ for a tone stimulus with duration D and frequency f_0. Let the corresponding PSTH be $p(t)$, and the total number of spikes be N.

In Eq 3.13, $\sum_{\mathrm{i}=1}^{N} \mathrm{e}^{\jmath 2\pi f t_{\mathrm{i}}}$ can be written as (van Hemmen, 2013)

$$\sum_{\mathrm{i}=1}^{N} \mathrm{e}^{\jmath 2\pi f t_{\mathrm{i}}} = \int_{t=0}^{D} p(t)\mathrm{e}^{\jmath 2\pi f t} dt. \tag{3.15}$$

Using Eq 3.15 in Eq 3.13, we get

$$\begin{aligned} \rho(f) &= \frac{1}{N}\int_{t=0}^{D} p(t)\mathrm{e}^{\jmath 2\pi f t} dt \\ \Longrightarrow \rho(f_0) &= \frac{1}{N}\int_{t=0}^{D} p(t)\mathrm{e}^{\jmath 2\pi f_0 t} dt. \end{aligned} \tag{3.16}$$

If we assume response phase locking to positive and negative polarity of a sinusoid (f_0) differ by a phase of π [i.e., a time difference of $T_0/2(= 1/2f_0)$ such that $p(t) \simeq n(t)\mathrm{e}^{\jmath 2\pi f T_0/2}$], we can write

$$\begin{aligned}
\rho(f) &= \frac{1}{N}\int_{t=0}^{D} p(t)\mathrm{e}^{\jmath 2\pi f t}dt \\
&= \frac{1}{N}\int_{t=0}^{D} n(t)\mathrm{e}^{\jmath 2\pi f T_0/2}\mathrm{e}^{\jmath 2\pi f t}dt.
\end{aligned} \tag{3.17}$$

For $f \neq f_0$, the integral in Eq 3.17 will be zero. For $f = f_0$,

$$\begin{aligned}
\rho(f_0) &= \frac{1}{N}\int_{t=0}^{D} n(t)\mathrm{e}^{\jmath 2\pi f_0 \frac{1}{f_0}\frac{1}{2}}\mathrm{e}^{\jmath 2\pi f_0 t}dt \\
&= \frac{1}{N}\int_{t=0}^{D} n(t)\mathrm{e}^{\jmath \pi}\mathrm{e}^{\jmath 2\pi f_0 t}dt \\
\Longrightarrow \rho(f_0) &= \frac{1}{N}\int_{t=0}^{D} -n(t)\mathrm{e}^{\jmath 2\pi f_0 t}dt.
\end{aligned} \tag{3.18}$$

Adding Eqs. 3.16 and 3.18, we get

$$\begin{aligned}
2\rho(f_0) &= \frac{1}{N}\int_{t=0}^{D} \left[p(t) - n(t)\right]\mathrm{e}^{\jmath 2\pi f_0 t}dt \\
&= \frac{1}{N}\int_{t=0}^{D} 2d(t)\mathrm{e}^{\jmath 2\pi f_0 t}dt \\
\Longrightarrow \rho(f_0) &= \frac{1}{N}\int_{t=0}^{D} d(t)\mathrm{e}^{\jmath 2\pi f_0 t}dt \\
&= \frac{D(-f_0)}{N},
\end{aligned}$$

where $D(f) = \int_{t=0}^{D} d(t)\mathrm{e}^{-\jmath 2\pi f t}dt$ is the Fourier transform of $d(t)$. Since $d(t)$ is a real signal, $|D(f)| = |D(-f)|$. Thus, the relation between VS and the difference PSTH becomes,

$$VS(f) = |\rho(f)| = \frac{|D(f)|}{N}. \tag{3.19}$$

3.8.3 Relation between *shuffled correlograms* and *apPSTHs*

Consider $\mathbb{X}$: a set of T_X spike trains $\{\underline{\zeta_1}, \underline{\zeta_2}, ..., \underline{\zeta_{T_X}}\}$ in response to a stimulus of duration D. For each spike train $\underline{\zeta_{\mathrm{i}}}$, we can construct a PSTH, $\underline{x_{\mathrm{i}}}$, with PSTH bin width Δ so that the length of the single-trial PSTH $\underline{x_{\mathrm{i}}}$ is $M = D/\Delta$. The single-trial PSTH is a binary-valued vector because each element in the vector is either 0 or 1. Let us denote the PSTH for $\mathbb{X}$ by $PSTH_X$ such that $PSTH_X = \sum_{\mathrm{i}=1}^{T_X} \underline{x_{\mathrm{i}}}$. Consider $\mathbb{Y}$: another set of T_Y spike trains, with $\underline{y_{\mathrm{i}}}$ and $PSTH_Y$ defined similarly to $\underline{x_{\mathrm{i}}}$ and $PSTH_X$, respectively. Let us assume that the stimulus duration and bin width for $\underline{y_{\mathrm{i}}}$ are the same as that for $\underline{x_{\mathrm{i}}}$. Let the average discharge rates (in spikes/s) for $\mathbb{X}$ and $\mathbb{Y}$ be r_X and r_Y, respectively. The shuffled cross-correlogram (SCC) for two spike trains $\underline{\zeta_{\mathrm{i}}}$ and $\underline{\zeta_{\mathrm{j}}}$ computed using tallying (Louage et al., 2004) is identical to the cross-correlation function (denoted by $\mathcal{R}_{\mathcal{X}\mathcal{Y}}$) between their respective PSTHs, ($\underline{x_{\mathrm{i}}}$ and $\underline{x_{\mathrm{j}}}$). Thus, the raw (not normalized) shuffled cross-correlogram (SCC^{raw}) at τ delay can be computed as

$$\begin{aligned}
SCC^{raw}_{\mathbb{X},\mathbb{Y}}(\tau) &= \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_1}, \{\underline{y_1}, \underline{y_2}, ..., \underline{y_{T_Y}}\}) + ... + \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_{T_X}}, \{\underline{y_1}, \underline{y_2}, ..., \underline{y_{T_Y}}\}) \\
&= \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_1}, [\underline{y_1} + \underline{y_2} + ... + \underline{y_{T_Y}}]) + ... + \qquad (3.20) \\
&\quad \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_{T_X}}, [\underline{y_1} + \underline{y_2} + ... + \underline{y_{T_Y}}]) \\
&= \sum_{\mathrm{i}=1}^{T_X} \sum_{\mathrm{j}=1}^{T_Y} \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_{\mathrm{i}}}, \underline{y_{\mathrm{j}}}) \\
&= \mathcal{R}_{\mathcal{X}\mathcal{Y}}(PSTH_X, PSTH_Y) \\
\Longrightarrow SCC^{norm}_{\mathbb{X},\mathbb{Y}}(\tau) &= \frac{\mathcal{R}_{\mathcal{X}\mathcal{Y}}(PSTH_X, PSTH_Y)}{T_X T_Y r_X r_Y D\Delta}, \qquad (3.21)
\end{aligned}$$

where SCC^{norm} is the normalized SCC (Heinz and Swaminathan, 2009; Louage et al., 2004).

Similarly, the raw shuffled autocorrelogram (SAC^{raw}) at τ delay can be computed as,

$$\begin{aligned}
SAC^{raw}_{\mathbb{X}}(\tau) &= \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_1}, \{\underline{x_2}, \underline{x_3}, ..., \underline{x_{T_X}}\}) + \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_2}, \{\underline{x_1}, \underline{x_3}, ..., \underline{x_{T_X}}\}) + ... \\
&\quad + \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_{T_X}}, \{\underline{x_1}, \underline{x_2}, ... + \underline{x_{T_X-1}}\}) \\
&= \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_1}, [\underline{x_2} + \underline{x_3} + ... + \underline{x_{T_X}}]) + \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_2}, [\underline{x_1} + \underline{x_3} + ... + \underline{x_{T_X}}]) \\
&\quad + ... + \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_{T_X}}, [\underline{x_1} + \underline{x_2} + ... + \underline{x_{T_X-1}}]) \\
&= \sum_{\mathrm{i}=1}^{T_X} \sum_{\mathrm{j}=1,\mathrm{j}\neq\mathrm{i}}^{T_X} \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_{\mathrm{i}}}, \underline{x_{\mathrm{j}}}) \\
&= \sum_{\mathrm{i}=1}^{T_X} \sum_{\mathrm{j}=1}^{T_X} \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_{\mathrm{i}}}, \underline{x_{\mathrm{j}}}) - \sum_{\mathrm{i}=1}^{T_X} \mathcal{R}_{\mathcal{X}\mathcal{Y}}(\underline{x_{\mathrm{i}}}, \underline{x_{\mathrm{i}}}) \\
&= \mathcal{R}_{\mathcal{X}}(PSTH_X) - \sum_{\mathrm{i}=1}^{T_X} \mathcal{R}_{\mathcal{X}}(\underline{x_{\mathrm{i}}}) \\
\Longrightarrow SAC^{norm}_{\mathbb{X}}(\tau) &= \frac{\mathcal{R}_{\mathcal{X}}(PSTH_X) - \sum_{\mathrm{i}=1}^{T_X} \mathcal{R}_{\mathcal{X}}(\underline{x_{\mathrm{i}}})}{T_X(T_X-1)r_X^2 D\Delta}, \qquad (3.22)
\end{aligned}$$

where $\mathcal{R}_{\mathcal{X}}$ denotes the autocorrelation function. Similar to autocorrelation functions, the SAC^{norm} has its maximum at zero delay.

In the numerator of Eq 3.22, the term $\sum_{\mathrm{i}=1}^{T_X} \mathcal{R}_{\mathcal{X}}(\underline{x_{\mathrm{i}}})$ is negligible compared to $\mathcal{R}_{\mathcal{X}}(PSTH_X)$ for $\tau \neq 0$. For $\tau = 0$, $\sum_{\mathrm{i}=1}^{T_X} \mathcal{R}_{\mathcal{X}}(\underline{x_{\mathrm{i}}})$ is equal to the total number of spikes (N) in $\mathbb{X}$. Thus, Eq 3.22 can be further approximated by,

$$\begin{aligned}
SAC^{norm}_{\mathbb{X}}(\tau) &\simeq \frac{\mathcal{R}_{\mathcal{X}}(PSTH_X) - N\delta(\tau)}{T_X(T_X-1)r_X^2 D\Delta} \qquad (3.23) \\
&= \frac{\mathcal{R}_{\mathcal{X}}(PSTH_X)}{T_X(T_X-1)r_X^2 D\Delta} - \frac{\delta(\tau)}{(T_X-1)r_X\Delta} \\
&\simeq \frac{\mathcal{R}_{\mathcal{X}}(PSTH_X)}{T_X^2 r_X^2 D\Delta} - \frac{\delta(\tau)}{T_X r_X \Delta}, \qquad (3.24)
\end{aligned}$$

where $N = r_X D T_X$, and δ is the Dirac delta function. The simplifying approximation in Eq 3.24 is valid for typically used T_X values in neurophysiological experiments, and equates the normalization factors between $SACs$ and $SCCs$ when working with *difcor* and *sumcor* (e.g., Sec 3.8.4). Eqs 3.21 to 3.24 indicate that correlograms can be computed much more efficiently using *apPSTHs* instead of by tallying spike times [$\mathcal{O}(N)$ instead of $\mathcal{O}(N^2)$, see main text].

3.8.4 Relation between *difcor/sumcor* and *difference/sum PSTHs*

Consider $\mathbb{X}_+$: spike trains in response to the positive polarity of a stimulus, and $\mathbb{X}_-$: spike trains in response to the negative polarity of the stimulus. Then, the *difcor* at τ delay can be computed as

$$\begin{aligned} difcor_{\mathbb{X}}(\tau) &= \frac{1}{2}\left[\frac{SAC^{norm}_{\mathbb{X}_+} + SAC^{norm}_{\mathbb{X}_-}}{2} - \frac{SCC^{norm}_{\mathbb{X}_+,\mathbb{X}_-} + SCC^{norm}_{\mathbb{X}_-,\mathbb{X}_+}}{2}\right] \\ &= \frac{1}{4}\left[SAC^{norm}_{\mathbb{X}_+} + SAC^{norm}_{\mathbb{X}_-} - SCC^{norm}_{\mathbb{X}_+,\mathbb{X}_-} - SCC^{norm}_{\mathbb{X}_-,\mathbb{X}_+}\right] \end{aligned}$$

For analytic simplicity, we use Eq 3.24 for SAC^{norm} instead of Eq 3.22. Let us assume that the number of repetitions and average rates for both polarities are the same. Thus,

$$\begin{aligned} difcor_{\mathbb{X}}(\tau) = \frac{1}{4\mathcal{K}}[&\mathcal{R}_{\mathcal{X}}(PSTH_{\mathbb{X}_+}) - N\delta(\tau) + \mathcal{R}_{\mathcal{X}}(PSTH_{\mathbb{X}_-}) - N\delta(\tau) \\ &- \mathcal{R}_{\mathcal{X}\mathcal{Y}}(PSTH_{\mathbb{X}_+}, PSTH_{\mathbb{X}_-}) - \mathcal{R}_{\mathcal{X}\mathcal{Y}}(PSTH_{\mathbb{X}_-}, PSTH_{\mathbb{X}_+})], \end{aligned}$$

where $\mathcal{K} = T_X^2 r_X^2 D\Delta$ is a constant. Now, $PSTH_{\mathbb{X}_+} = p(t)$, $PSTH_{\mathbb{X}_-} = n(t)$, and the difference PSTH $d(t) = [p(t) - n(t)]/2$. Then, the *difcor* for $\mathbb{X}$ at delay τ is

$$\begin{aligned} difcor_{\mathbb{X}}(\tau) = &\frac{1}{4\mathcal{K}}\left\{\mathcal{R}_{\mathcal{X}}\left[p(t)\right] + \mathcal{R}_{\mathcal{X}}\left[n(t)\right] - \mathcal{R}_{\mathcal{X}\mathcal{Y}}\left[p(t), n(t)\right] - \mathcal{R}_{\mathcal{X}\mathcal{Y}}\left[n(t), p(t)\right]\right\} \\ &- \frac{N\delta(\tau)}{2\mathcal{K}} \end{aligned}$$

Now,

$$\begin{aligned} &\mathcal{R}_{\mathcal{X}}\left[p(t)\right] + \mathcal{R}_{\mathcal{X}}\left[n(t)\right] - \mathcal{R}_{\mathcal{X}\mathcal{Y}}\left[p(t), n(t)\right] - \mathcal{R}_{\mathcal{X}\mathcal{Y}}\left[n(t), p(t)\right] \\ &= \int_{t=0}^{D} p(t)p(t-\tau)dt + \int_{t=0}^{D} n(t)n(t-\tau)dt - \int_{t=0}^{D} p(t)n(t-\tau)dt - \int_{t=0}^{D} n(t)p(t-\tau)dt \\ &= \int_{t=0}^{D} p(t)\left[p(t-\tau) - n(t-\tau)\right]dt - \int_{t=0}^{D} n(t)\left[p(t-\tau) - n(t-\tau)\right]dt \\ &= \int_{t=0}^{D} 2p(t)d(t-\tau)dt - \int_{t=0}^{D} 2n(t)d(t-\tau)dt \end{aligned}$$

$$
\begin{aligned}
&= \int_{t=0}^{D} 2\left[p(t) - n(t)\right] d(t-\tau)dt \\
&= \int_{t=0}^{D} 4d(t)d(t-\tau)dt \\
&= 4\mathcal{R}_{\mathcal{X}}[d(t)]
\end{aligned}
$$

Thus,

$$
\begin{aligned}
difcor_{\mathbb{X}}(\tau) &= \frac{1}{4\mathcal{K}}\left\{\mathcal{R}_{\mathcal{X}}\left[p(t)\right] + \mathcal{R}_{\mathcal{X}}\left[n(t)\right] - \mathcal{R}_{\mathcal{X}\mathcal{Y}}\left[p(t), n(t)\right] - \mathcal{R}_{\mathcal{X}\mathcal{Y}}\left[n(t), p(t)\right]\right\} \\
&\quad - \frac{N\delta(\tau)}{2\mathcal{K}} \\
&= \frac{1}{4\mathcal{K}} \times 4\mathcal{R}_{\mathcal{X}}[d(t)] - \frac{N\delta(\tau)}{2\mathcal{K}} \\
&= \frac{\mathcal{R}_{\mathcal{X}}[d(t)]}{\mathcal{K}} - \frac{N\delta(\tau)}{2\mathcal{K}} \\
&= \frac{\mathcal{R}_{\mathcal{X}}[d(t)]}{T_X^2 r_X^2 D\Delta} - \frac{r_X D T_X \delta(\tau)}{2T_X^2 r_X^2 D\Delta} \\
\Longrightarrow difcor_{\mathbb{X}}(\tau) &= \frac{\mathcal{R}_{\mathcal{X}}[d(t)]}{T_X^2 r_X^2 D\Delta} - \frac{\delta(\tau)}{2T_X r_X \Delta} \qquad (3.25)
\end{aligned}
$$

Similarly, it can be shown that

$$
\begin{aligned}
sumcor_{\mathbb{X}}(\tau) &= \frac{1}{2}\left[SAC_{\mathbb{X}}^{norm} + SCC_{\mathbb{X}+,\mathbb{X}-}^{norm}\right] \\
&= \frac{1}{2}\left[\frac{SAC_{\mathbb{X}_+}^{norm} + SAC_{\mathbb{X}_-}^{norm}}{2} + \frac{SCC_{\mathbb{X}+,\mathbb{X}-}^{norm} + SCC_{\mathbb{X}-,\mathbb{X}+}^{norm}}{2}\right] \\
&= \frac{1}{4\mathcal{K}} \times 4\mathcal{R}_{\mathcal{X}}[s(t)] - \frac{N\delta(\tau)}{2\mathcal{K}} \\
&= \frac{\mathcal{R}_{\mathcal{X}}[s(t)]}{\mathcal{K}} - \frac{N\delta(\tau)}{2\mathcal{K}} \\
\Longrightarrow sumcor_{\mathbb{X}}(\tau) &= \frac{\mathcal{R}_{\mathcal{X}}[s(t)]}{T_X^2 r_X^2 D\Delta} - \frac{\delta(\tau)}{2T_X r_X \Delta} \qquad (3.26)
\end{aligned}
$$

where $s(t)$ is the sum PSTH, i.e., $s(t) = [p(t) + n(t)]/2$.

Eqs 3.25 and 3.26 indicate that *sumcor* and *difcor* are related to the autocorrelation function of the *sum* and *difference* PSTHs, respectively, and thus can be computed much more efficiently [$\mathcal{O}(N)$ rather than $\mathcal{O}(N^2)$].

3.8.5 Relation between *shuffled-correlogram* peak-height and *apPSTHs*

Consider a difference PSTH, $d(t)$, based on a set of spike trains $\mathbb{X}$ in response to a stimulus of duration D. Let us denote the Fourier transform of $d(t)$ by $D(f)$. Then, from Eq 3.25, the *difcor* peak-height, i.e., *difcor* value at zero delay (τ), can be computed as

$$\begin{aligned} dif cor_X(\tau = 0) &= \left.\frac{\mathcal{R}_{\mathcal{X}}\{d(t)\}}{\mathcal{K}}\right|_{\tau=0} - \left.\frac{N\delta(\tau)}{2\mathcal{K}}\right|_{\tau=0} \\ &= \frac{1}{\mathcal{K}} \int_{t=0}^{D} d^2(t)dt - \frac{N}{2\mathcal{K}} \\ &= \frac{1}{\mathcal{K}} \int_{f=-\infty}^{\infty} |D(f)|^2 \, df - \frac{N}{2\mathcal{K}}, \quad \text{(by Parseval's theorem)} \end{aligned} \tag{3.27}$$

Following similar steps from Eq 3.26, it can also be shown that the *sumcor* peak-height can be computed as

$$sumcor_X(\tau = 0) = \frac{1}{\mathcal{K}} \int_{f=-\infty}^{\infty} |S(f)|^2 \, df - \frac{N}{2\mathcal{K}} \tag{3.28}$$

where $S(f)$ is the Fourier transform of the sum PSTH, $s(t)$.

Comparing Eq 3.19 with Eqs. 3.27 and 3.28, we see that vector strength is a frequency-specific metric, whereas correlogram peak-heights are broadband measures, which are thus susceptible to rectifier distortion (see Fig 3.6).

3.8.6 Parameters for the AN model

Table 3.3. AN model parameters. List of parameters used in the AN model to generate simulated spike-train data.

Parameter	Value
Sampling Frequency	100 kHz
Number of Repetitions (per polarity)	25
Spontaneous firing rate (SR)	70 spikes/s
Absolute refractory period	0.6 ms
Baseline mean relative refractory period	0.6 ms
OHC health value	1.0 (normal)
IHC health value	1.0 (normal)
Species	1 (cat)
Fractional Gaussian noise type	0 (fixed)
Implementation type of the power-law functions in the Synapse	0 (approximate)
Spike time resolution	10 μs

3.8.7 Nonlinear inner-hair-cell transduction function introduces additional sidebands in the spectrum for a SAM tone.

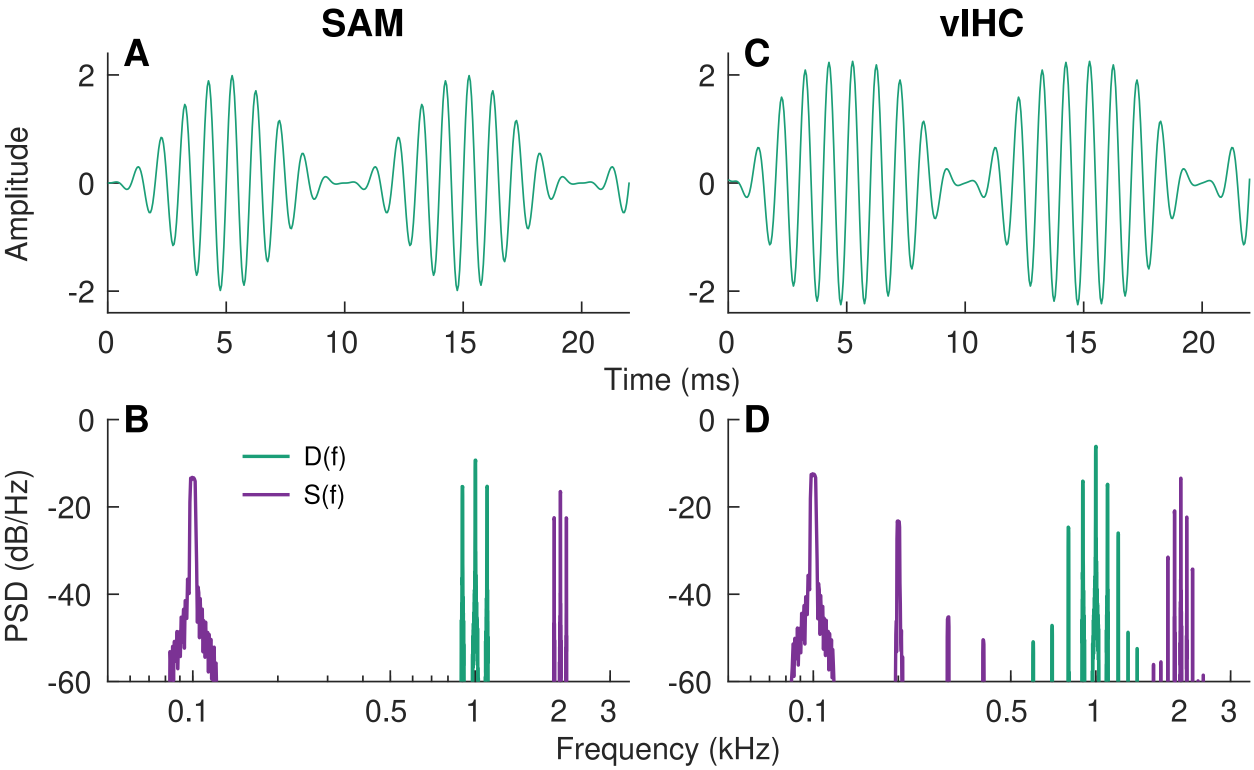

Figure 3.15. Nonlinear inner-hair-cell transduction function introduces additional sidebands in the spectrum for a SAM tone. (A) Waveform for a SAM tone (F_c=1 kHz, F_m=100 Hz, 0-dB modulation depth). (B) $D(f)$ and $S(f)$ for the SAM tone in A. (C) Waveform of the output after processing the SAM tone through a sigmoid function. The sigmoid function was used as a simple proxy for the inner-hair-cell transduction function. This output (vIHC) was further low-pass filtered at 2 kHz to mimic the membrane properties of inner hair cells. (D) $D(f)$ and $S(f)$ for the signal in C. In addition to having power at F_c and $F_c \pm F_m$, $D(f)$ for vIHC has substantial energy at $F_c \pm 2F_m$ (plus reduced energy at higher multiple F_m-offsets from F_c). Similarly, $S(f)$ for vIHC has substantial energy at F_m as well as at the first few harmonics of F_m. $S(f)$ is also corrupted by rectifier distortion at $2F_c$ (and multiple F_m-offsets from $2F_c$) as expected.

3.8.8 DFT-magnitude for the nonstationary vowel, s_2

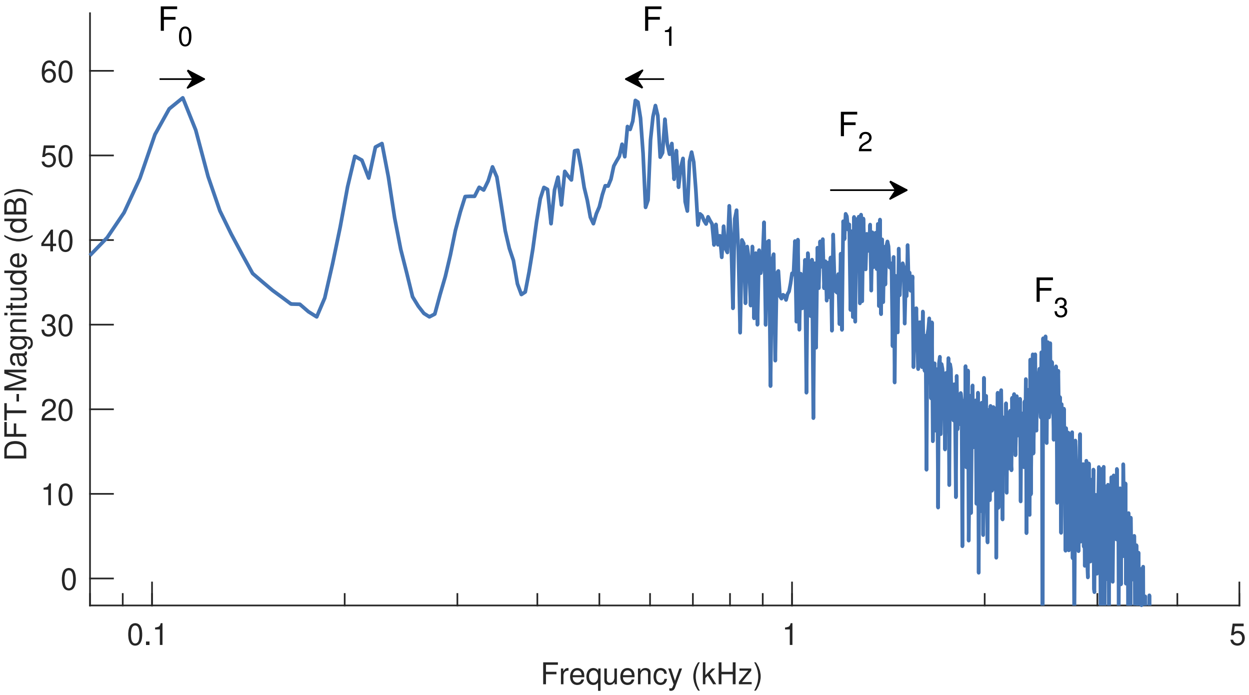

Figure 3.16. DFT-magnitude for the nonstationary vowel, s_2. The stimulus duration was 188 ms. The movements of F_0 (100 to 120 Hz), F_1 (630 to 570 Hz), and F_2 (1200 to 1500 Hz) are indicated by arrows. F_3 was fixed at 2500 Hz.

3.8.9 FFR harmonicgram constructed using the Hilbert-phase FFR.

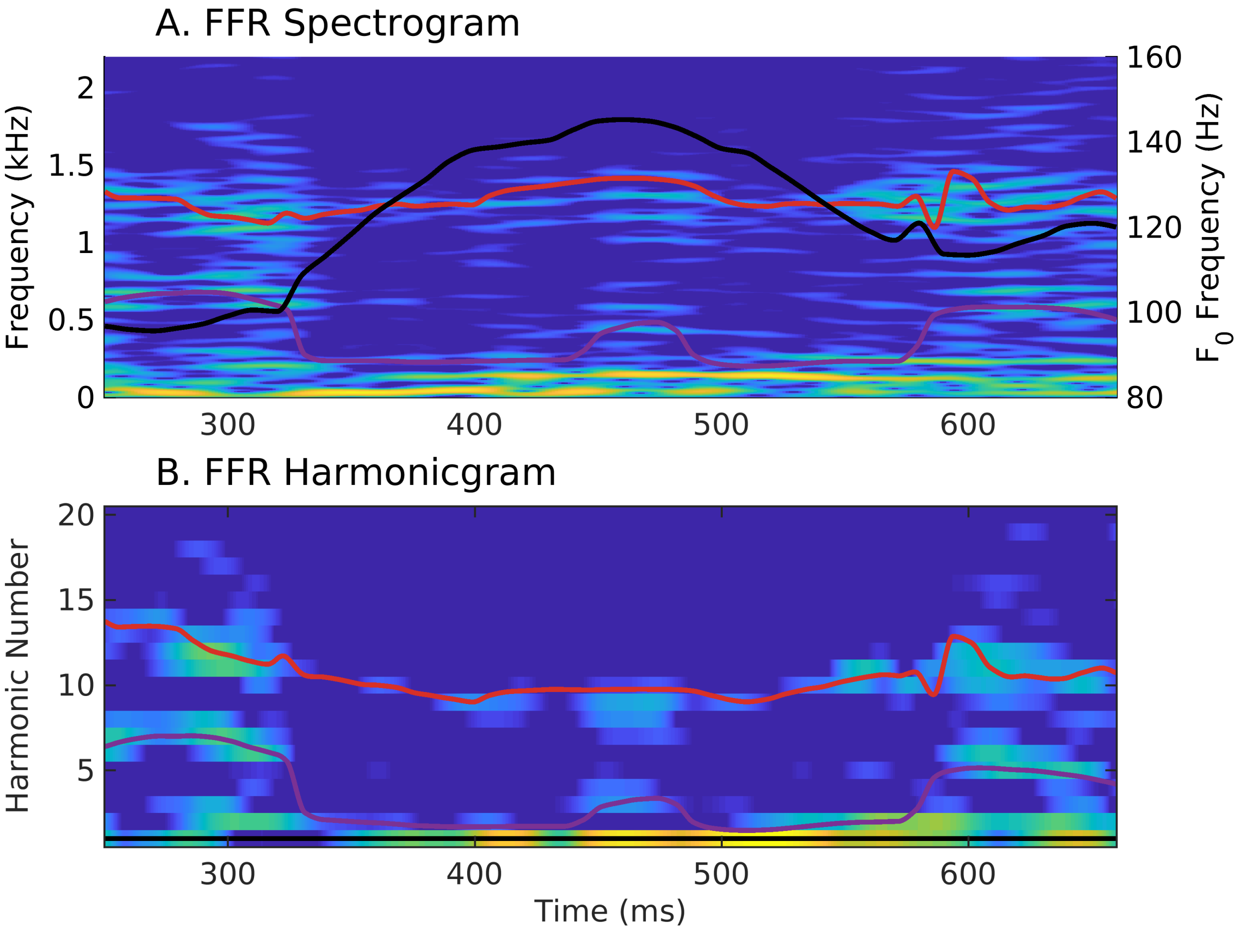

Figure 3.17. FFR harmonicgram can be constructed using the Hilbert-phase response. Same format as Fig 3.14. The spectrogram (A) and the harmonicgram (B) were constructed using $\phi(t)$.

4. DISTORTED TONOTOPY SEVERELY DEGRADES NEURAL REPRESENTATIONS OF NATURAL SPEECH IN NOISE FOLLOWING ACOUSTIC TRAUMA

SUMMARY[1]

Listeners with noise-induced hearing loss demonstrate speech-perception deficits, especially in noisy environments, which still remain despite audibility compensation. These suprathreshold deficits are thought to arise from increased auditory bandwidth and reduced temporal-coding precision. To test these hypotheses, we measured single auditory-nerve-fiber and evoked frequency-following responses to natural speech in noise from anesthetized chinchillas with normal hearing and mild-to-moderate hearing loss. Our results show several coding deficits for vowels and consonants, but temporal coding precision is unaffected. Although increased bandwidth contributes to these deficits, it is not the primary effect. Rather, two key factors explain these deficits: (1) distorted tonotopy, which reflects oversensitivity of neurons to low-frequency energy, disrupts the coding of informative high-frequency speech features and increases susceptibility to masking noise, and (2) audiometric threshold, which relies on the most sensitive group of neurons, underestimates average audibility deficits at the population level. These findings help explain the neural origins of common perceptual difficulties that hearing-impaired listeners experience; furthermore, they offer insight into individualized strategies for hearing rehabilitation.

4.1 Introduction

Individuals with sensorineural hearing loss (SNHL) demonstrate speech perception deficits that are not resolved despite compensating for audibility, even in quiet environments (Dubno et al., 1982). More importantly, these patients particularly struggle in noisy environments despite state-of-the-art hearing aid strategies and noise-reduction algorithms (Lesica, 2018; McCormack and Fortnum, 2013). In fact, difficulty understanding speech in noise is the number one complaint in the audiology clinic (Chung, 2004; Souza, 2016). Clinical outcomes of

[1] Manuscript *in prep*

hearing aids can be improved by studying the neural representation of speech-in-noise in the peripheral nervous system (i.e., the auditory nerve or AN) because the AN serves as the bottleneck of auditory information that is accessible to the brain. Furthermore, key physiological and anatomical effects of SNHL that contribute to perceptual deficits of hearing-impaired listeners are evident at this peripheral level (Trevino et al., 2019). Understanding the neural representation of natural speech-in-noise in normal and impaired auditory systems can provide invaluable insight to enhance perceptually informative cues in hearing aids as well as to improve algorithms used by machine listening devices, particularly in adverse situations. To achieve these feats, several gaps in our understanding of impaired auditory processing need to be addressed. First, the study of speech coding in the AN, in quiet and in noise, has been predominantly carried out using short-duration synthesized stationary stimuli, primarily with normal-hearing animals (Delgutte and Kiang, 1984a; Sinex and Geisler, 1983; Young and Sachs, 1979) and rarely with hearing-impaired animals (Miller et al., 1997; Schilling et al., 1998). Only a handful of studies have employed longer natural phrases or sentences (Delgutte et al., 1998; Kiang and Moxon, 1974; Young, 2008), but none of these examined hearing loss. Real-world speech differs from stationary speech in that it is highly dynamic in nature and most of the information is present in its time varying properties like formant transitions and spectrotemporal modulations (Elliott and Theunissen, 2009; Greenberg et al., 2003; Jakobson and Fant, 1963; Stilp et al., 2010; Zeng et al., 2005). Therefore, understanding the neural representation of natural speech is important. Second, the handful of studies that have investigated speech-in-noise coding have been limited to normal-hearing animals and have used positive-SNR conditions. However, studies have shown that negative SNRs are particularly important because these are where perceptual results differ substantially between normal-hearing and hearing-impaired listeners (Festen and Plomp, 1990).

In this work, we used a natural speech sentence as a stimulus and collected spike trains from AN fibers of anesthetized chinchillas that either had normal hearing or noise-induced hearing loss (NIHL). The stimulus was also presented after mixing with speech shaped noise at several perceptually relevant SNRs. Previous psychoacoustic (Fant, 1973; Klatt, 1982) and neurophysiological (May, 2003; Young, 2008) studies have shown that the first three formants are important for perception and neural coding of voiced segments, e.g., vowels.

Therefore, the coding of voiced segments was evaluated based on the robustness of the first three formants (F1 to F3) and their transitions. The sentence also contained a high-frequency fricative (/s/) and two plosive consonants (/d/ and /g/), whose representations were compared between the normal and the hearing-impaired neurons. /s/ is one of the most frequent phonemes in English (Denes, 1963; Tobias, 1959), and is commonly misperceived by listeners with hearing impairment (Bilger and Wang, 1976; Owens et al., 1972; Turner and Robb, 1987). Difficulties in /d/ and /g/ perception by hearing-impaired listeners are also well documented (Bilger and Wang, 1976; Owens et al., 1972; Turner and Robb, 1987). While consonants are softer (less intense) than vowels, their perception is robust due to a number of features like formant transition, spectral content, and better neural rate representations that occur at lower intensities (Blumstein and Stevens, 1979; Diehl, 2008; Pickett, 1999; Smits et al., 1996; Young, 2008). Studies have shown that onset responses of AN fibers carry sufficient information for fricative discrimination (Delgutte, 1980; Delgutte and Kiang, 1984b). Auditory nerve fibers can also track spectral changes like consonant-vowel syllable coarticulation in addition to encoding consonant-related information in their sustained response (Bandyopadhyay and Young, 2004; Carney and Geisler, 1986; Deng and Geisler, 1987; Miller and Sachs, 1983; Shamma, 1985; Sinex and Geisler, 1983). Therefore, consonant coding was directly evaluated by quantifying onset and sustained rate profiles, and indirectly evaluated by quantifying the representation of dynamic formants.

Our results demonstrate a number of significant degradations in natural speech coding, which provide new insights into the challenges faced by listeners with SNHL. A disruption in tonotopic speech coding at the AN level following acoustic trauma resulted in a degradation of highly informative features such as second and third formants at the expense of an overrepresented first formant. These hearing-loss effects were exacerbated in noise. Driven rate was restored for voiced segments but not for unvoiced consonants, demonstrating that amplifying to restore audibility does not restore suprathreshold coding. The distorted tonotopic mapping substantially diminished the noise-resistant coding ability of high-threshold low and medium spontaneous rate fibers, which are usually less affected by noise compared to high spontaneous rate fibers (Geisler and Silkes, 1991; Silkes and Geisler, 1991). Finally, temporal coding precision in AN fibers was not diminished following acoustic trauma, inconsistent with suggestions from perceptual studies (e.g., Halliday et al., 2019; Lorenzi et al., 2006). Overall, these results establish neural correlates of several suprathreshold deficits in speech-in-noise perception with SNHL.

4.2 Methods

All procedures described below followed PHS-issued guidelines and were approved by Purdue University Animal Care and Use Committee (Protocol No: 1111000123).

4.2.1 Noise exposure and electrophysiological recordings

For detailed descriptions of noise exposure and electrophysiological recordings, see JARO paper. Briefly, male chinchillas (<1-year old, weight between 400 and 700 gm) were used in all experiments. ABRs and FFRs were recorded using needle electrodes in a vertical montage (vertex to mastoid, differential mode, common ground near animals' nose). A single discrete noise exposure (116 dB SPL, 2-hour duration, octave-band noise centered at 500 Hz) was used to induce NIHL. Animals were anesthetized using xylazine (2 to 3 mg/kg, subcutaneous) and ketamine (30 to 40 mg/kg, intraperitoneal) during data recordings and noise exposure. DPOAE was measure using an in-ear mic.

4.2.2 Surgical preparation and neurophysiological recordings

For detailed surgical preparation and neurophysiological recordings, see (Henry et al., 2016). Briefly, spike trains were recorded from single AN fibers of anesthetized chinchillas using glass micropipettes (impedance > 10 MΩ). Anesthesia was induced with previously mentioned dose of xylazine/ketamine and maintained with sodium pentobarbital. A posterior fossa approach was employed for craniotomy. Identified spikes were stored digitally with 10-µs resolution.

4.2.3 Stimuli

Screening experiments

For ABRs, tone pips (5-ms duration, 1-ms on and off ramp) ranging from 0.5 to 8 kHz (octave spaced) were played at 0 dB SPL to 80 dB SPL in 10-dB steps. 500 repetitions of both positive and negative polarities were played for each intensity condition. ABR threshold was calculated based on a cross-correlation analysis (Henry et al., 2011). Another intermediate (odd multiple of 5 dB) step was used near preliminary ABR threshold estimate to fine-tune the final estimate. DPOAEs were measured for pairs of tones (f1, f2) presented simultaneously with f2/f1 =1.2 at 75 (f1) and 65 (f2) dB SPL.

FFR experiments

A naturally spoken speech sentence (list #3, sentence #1) from the Danish speech intelligibility test (CLUE, Nielsen and Dau, 2009) was used for FFR experiments. Intensity was set to 70 dB SPL for both groups. Both positive and negative polarities (500 repetitions/polarity) of the stimulus were used to allow estimation of envelope and temporal fine structure components from the FFR. Both envelope (sum of FFRs to opposite polarities) and temporal fine structure (difference between FFRs to opposite polarities) components in response to the voiced portions of speech had near-normal amplitude for the hearing-impaired chinchillas at this intensity (JARO). This restoration in amplitude likely reflects convergence in audibility at moderate intensities for mild-moderate hearing loss as well as central-gain related changes in the midbrain, which substantially contributes to FFR responses.

4.2.4 AN experiment

For AN experiments, the same speech sentence was used. Intensity was set to 65 dB SPL for normal-hearing chinchillas and 80 dB SPL for hearing-impaired chinchillas. The spectrally flat gain of 15 dB was roughly based on half-gain rule (Lybarger, 1978), which has been used in AN studies of NIHL (Schilling et al., 1998). Speech was also presented after mixing with frozen steady-state noise at three different (-10, -5, and 0 dB) SNRs. Noise was spectrally matched to 10 sentences spoken by the same speaker using autoregressive modelling. Both polarities of stimuli were presented (25 trials per polarity for most AN fibers). Prior to collecting spike-train data in response to speech or noise, FTC (Chintanpalli and Heinz, 2007) and SR (over 30-s duration) were estimated for individual AN fibers.

4.2.5 Analysis

4.2.6 Quantifying formant coding strength using the harmonicgram

The harmonicgram ($\mathcal{HG}$) provides superior spectrotemporal resolution for analyzing non-stationary signals, like speech, compared to the spectrogram (Parida et al., 2020). Briefly, the harmonicgram was constructed as follows. Difference PSTH ($d[n]$) was constructed by halving the difference between PSTHs to opposite stimulus polarities. The fundamental frequency contour ($F_0[n]$) was estimated using Praat (Boersma, 2001). Response components ($\mathcal{HG}[k,n]$) along kth-harmonic of were estimated using frequency demodulation and low-pass filtering (LPF). Low-pass filter was 6th order zero-phase IIR filter with 10-Hz cut-off frequency.

$$\phi^k_{traj}[n] = \frac{1}{f_s}\sum_{m=1}^{n} kF_0[n] \tag{4.1}$$

$$d_{demod}[n] = d[n]\mathrm{e}^{-\jmath 2\pi\phi^k_{traj}[n]} \tag{4.2}$$

$$\mathcal{HG}[k,n] = LPF(d_{demod}[n]) \tag{4.3}$$

To evaluate the coding strength of a formant, F_X, in the response $d[n]$, the fractional power ($FracPower$) of three harmonics near F_0-normalizedF_X was calculated as:

$$FracPower(F_X) = \frac{\sum_{n=1}^{N}\sum_{k=\alpha[n]-1}^{\alpha[n]+1}(\mathcal{HG}[k,n])^2}{\sum_{n=1}^{N}\sum_{k=1}^{K}(\mathcal{HG}[k,n])^2}, \tag{4.4}$$

where $\alpha[n] = round(F_X[n]/F_0[n])$, N is the length of $d[n]$, $round$ is the nearest-integer operator, and K is the number of F_0 harmonics considered. Because the lowest value for

$F_0[n]$ was ∼100 Hz and phase-locking for chinchillas is substantially degraded beyond 3500 Hz, K was set to 35.

For a given SNR conditions, fractional power metrics were computed for responses to noisy-speech [$FracPower_{SN}(F_X)$] and noise-alone [$FracPower_N(F_X)$] for individual AN fibers. The difference in these two power metrics [i.e., $FracPower_{SN}(F_X) - FracPower_{SN}(F_X)$] was used as the strength of F_X coding for the fibers for that SNR condition. For the quiet condition, $FracPower_N(F_X)$ was set to 0.

4.2.7 Correlation analysis for fricative coding in noise

Response envelopes were obtained from single-polarity PSTHs using a low-pass filter (fourth-order, cut-off = 32 Hz). Let the response envelope to speech-alone, noisy-speech, and noise-alone be denoted by R_S, R_SN, and R_N, respectively. Then, the corrected correlation between R_S and R_SN was quantified as:

$$corr_{norm}(R_S, R_{SN}) = corr(R_S, R_{SN}) - corr(R_S, R_N), \tag{4.5}$$

where $corr(X, Y) = \mathcal{E}(XY)$ and $\mathcal{E}$ is the expectation operator.

4.2.8 Spike-train precision analysis

Precision of spike trains was measured using the Victor-Purpura (VP) distance metric (Victor and Purpura, 1996), which quantifies the dissimilarity between two sets of spike trains. If a neuron responds precisely and consistently across different trials of a stimulus, the VP distance between the two spike trains will be small. VP distance between two spike trains, X and Y, is defined as the total cost involved in transforming X so that it matches Y. This transformation allows three operations: (1) addition of spikes, (2) deletion of spikes, and (3) shifting of spikes. Cost of adding or deleting a spike is set to 1. Cost of shifting a spike by Δt is proportional to a free shifting parameter (c_v), which controls the temporal resolution of analysis. For a given , the most optimal (involving smallest total cost) set of operations to transform X to Y are determined based on dynamic programming; VP distance between X and Y is computed as the total cost of all the operations involved in this optimal transformation.

For a given neuron, first a window length (ω) and a shifting cost (c_v) were chosen. Spike trains were divided into windows of ω (without overlap). For each window, average VP distance was computed across all combinations of spike trains (i.e., corresponding to different trials) using the shifting cost c_v. The average VP distance across all windows was the final VP distance estimate for the neuron considered.

4.2.9 Statistical analysis

FFR onset responses were compared between normal-hearing and hearing-impaired groups unpaired t-test. To test whether spike-train precision was different between the two groups, a linear regression model was used with VP distance as the dependent variable. Driven rate, characteristic frequency, and hearing status (categorical) were independent variables.

4.3 Results

4.3.1 Acoustic exposure yielded a mild-moderate hearing loss

Mild-to-moderate hearing loss is the most prevalent degree among patients with hearing loss (Goman and Lin, 2016). To investigate the neural coding deficits that these patients experience, we used a noise-exposure protocol (two-hour-long 116 dB SPL octave-band noise centered at 500 Hz) that has been previously used in our laboratory to induce mild-to-moderate hearing loss (Kale and Heinz, 2010). Thresholds for auditory brainstem responses (ABRs) to tone bursts increased by 20 dB following noise exposure (Fig 4.1A). Similarly, the level of distortion product otoacoustic emissions decreased by ∼15 dB SPL following noise exposure (Fig 4.1B), indicating the presence of substantial outer hair cell damage. These electrophysiological changes are consistent with a mild-to-moderate, permanent hearing loss model as per ASHA guidelines (Clark, 1981).

Thresholds were also elevated at a single neuron level in the AN by ∼35 dB SPL on average (Fig 4.1C). This threshold shift was accompanied by a substantial loss of tuning as quantified by a reduction in Q10 value for frequency tuning curves (FTCs) of individual

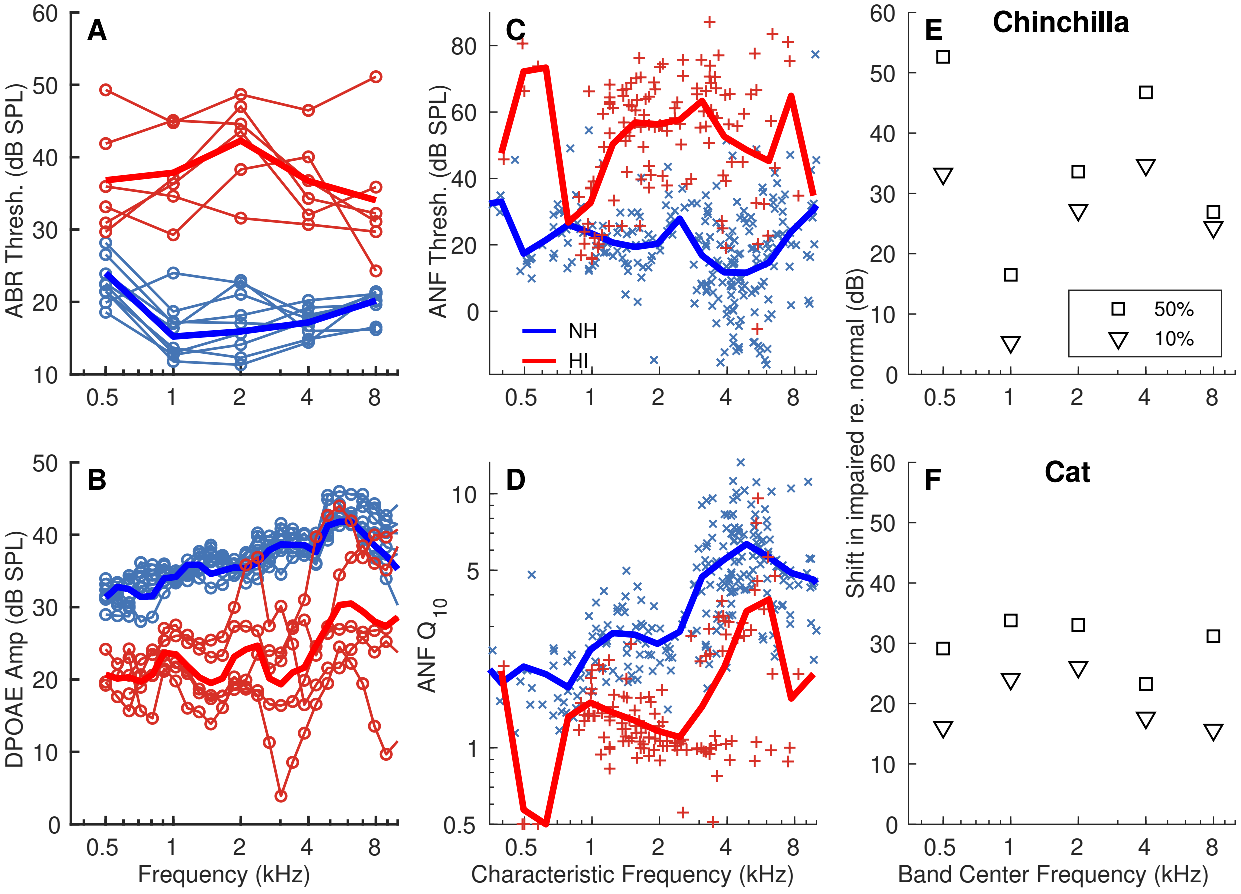

Figure 4.1. Acoustic overexposure yielded a flat mild-moderate hearing loss model with substantial outer hair cell damage. (A) ABR thresholds for the hearing-impaired (HI) chinchillas were elevated by 20 dB on average. (B) DPOAE amplitude was also reduced by 15 dB on average. In A and B, thin lines are for individual animals; thick lines represent group averages. (C) Compared to the normal-hearing (NH) AN fiber population, average threshold for impaired AN fibers was elevated by 40 dB, twice the threshold shift based on ABRs. (D) Q10 values were consistently reduced for impaired AN fibers. In C and D, markers represent data from individual AN fibers. Thick lines in C and D represent two-third octave averages along CF. (E) Shift in 10 (triangles) and 50 (squares) percentile point of band-specific ANF threshold distributions data following NIHL. Band-specific threshold distributions for each group were estimated as follows. First, threshold data were grouped into five octave-wide CF bands centered at 0.5, 1, 2, 4, and 8 kHz. Next, threshold distributions were estimated for each band, which were used to estimate 10 and 50 percentile points. (F) Same format as E but for extensive cat threshold data from (Heinz et al., 2005). For both chinchilla and cat data, shift in the 10-percentile point (which likely drives ABR/audiogram threshold) was smaller than shift in the 50-percentile point.

AN fibers (Fig 4.1D). These results were consistent across a wide range of characteristic frequencies (CFs). These physiological properties were similar to results from previous studies from our laboratory (Henry et al., 2016; Sayles et al., 2016). Single-unit data are from 13 normal-hearing (250 AN fibers) and 6 hearing-impaired (124 AN fibers) chinchillas.

4.3.2 ABR threshold may underestimate average AN population audiometric deficits

ABR thresholds at a frequency correlate with the most sensitive AN fibers near that frequency (Ngan and May, 2001). Thus, ABR thresholds reflect the activity of the most sensitive fibers. However, speech perception at suprathreshold levels, especially in complex environments, likely requires response averaging across many neurons and not just the most sensitive AN fibers (Bharadwaj et al., 2014). Therefore, linear gain aimed at restoring ABR threshold may not be sufficient to restore suprathreshold audibility if a large portion of the AN-fiber population has threshold shift that is greater than ABR threshold shift.

Table 4.1. Standard deviation (*std*) of ANF threshold distributions was greater for the impaired population compared to the normal population in each frequency band. ANF threshold data were divided into five groups based on octave-wide bands centered at .5 to 8 kHz. For each band, ANF threshold distributions were calculate for each group.

Band center frequency (kHz)	*std* (Chinchilla data)			*std* (Cat data)		
	NH	HI	HI-NH	NH	HI	HI-NH
0.5	10	15.1	5.1	9.2	13.1	3.9
1	10	18.5	8.4	8.4	16	7.6
2	10.9	12.3	1.4	7.5	17.2	9.6
4	14	19.4	5.5	10.7	16.1	5.3
8	16.9	23.9	7	9.9	17.4	7.5

In order to evaluate this hypothesis, we estimated AN-fiber threshold distribution for NH and HI groups in five octave-wide CF bands centered at .5 to 8 kHz (octave spaced). For each CF band, 10- and 50-percentile points were estimated. Our results showed greater shift for 50-percentile point compared to 10-percentile point following NIHL for all five bands.

We also reanalyzed previously published extensive cat data (N=738) from (Heinz et al., 2005). Similar to chinchilla data, cat data also showed a larger shift in 50-percentile points compared to the shift in 10-percentile points for all five bands (Fig 4.1F). These results support the hypothesis that any audiometric indicator that relies on a few sensitive AN fibers may underestimate audibility deficits at the population level.

4.3.3 Trial-to-trial precision was unaffected following NIHL

While one study has found degradation in spike-train precision following SNHL (Woolf et al., 1981), other studies have found no such degradation (Harrison and Evans, 1979; Kale and Heinz, 2010). To test whether there was any degradation in the ability of AN fibers to precisely encode temporal information in response to speech, trial-to-trial precision was quantified using Victor-Purpura (VP, Victor and Purpura, 1996) distance metric (Fig 4.2). VP distance inversely relates to precision. A linear regression model was used to test the effect of hearing status on VP distance(see Experimental Procedures).There was no effect of hearing loss on VP distance estimates for a range of conditions (Fig 4.2, results are shown for two conditions). Thus, there was no degradation in the ability of a neuron to precisely phase-lock to temporal features in the stimulus following NIHL.

4.3.4 Distorted tonotopy diminished near-CF cues at the expense of enhanced low-frequency cues following NIHL

Normal AN fibers are characterized by their high sensitivity (low threshold) and spectral specificity (sharp tuning), which allow these fibers to selectively respond to stimulus energy near their CF (e.g., Fig 4.3). To minimize rectifier-distortion effects on response spectral estimates, difference PSTH ($d[n]$) was used (Parida et al., 2020). $d[n]$ was computed as half the difference between PSTHs for opposite stimulus polarities. In this example, response of the normal AN fiber was dominated by the first formant or F_1, the spectral feature closest to the fiber's CF for the segment considered (Fig 4.3B). This effect can be seen in the fast-spiking activity in the peristimulus time histogram (PSTH, Fig 4.3A) and the corresponding $d[n]$ spectrum (Fig 4.3C), which shows a clear peak near F_1 (blue arrow). In contrast, following NIHL, AN fibers often show elevated thresholds, hypersensitive tails,

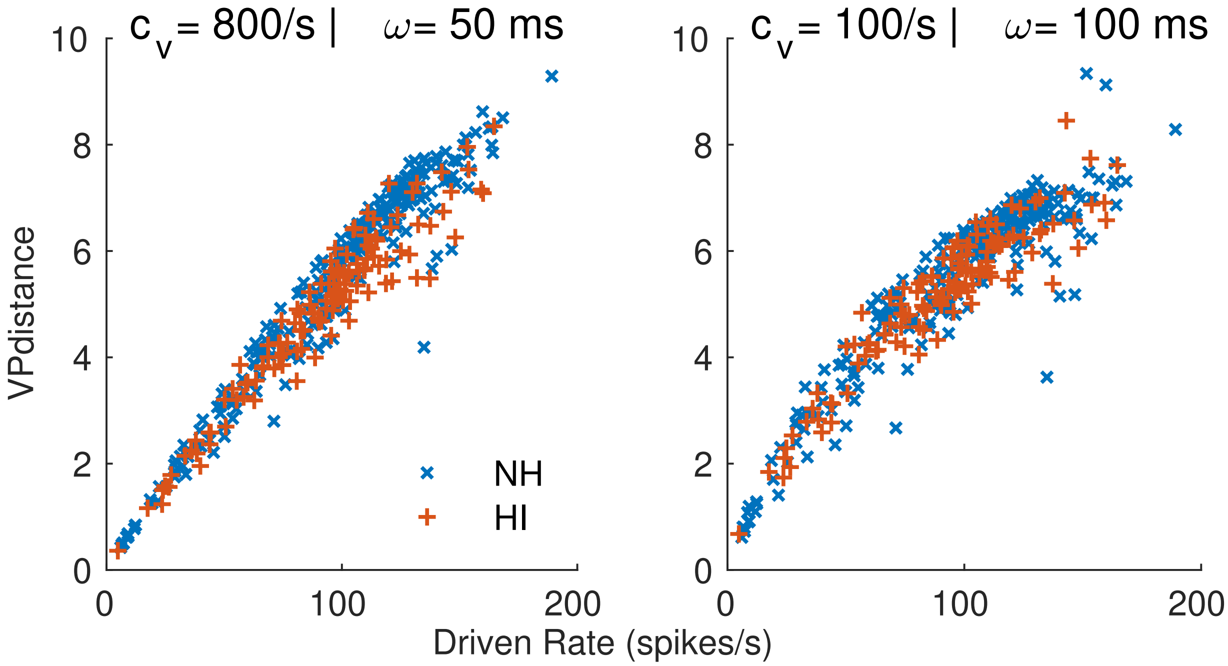

Figure 4.2. Precision of temporal coding was not affected following acoustic trauma. Average across-trial Victor-Purpura (VP) distance for normal (blue) and impaired (red) neurons for two window (ω) and cost (c_v) combinations. For each combination (ω/c_v), average across-trial VP distance for a given AN fiber was computed by first computing the average across-trial VP distance in windows of duration (without overlap), and then, taking the average distance for all windows. VP distance estimates, which inversely relate to precision, were similar between the two groups.

and broader tuning putatively associated with downward shifts in functional characteristic frequency (i.e., best frequency, Liberman, 1984; Liberman and Dodds, 1984b). As a result, responses of impaired neurons were often dominated by low-frequency stimulus energy that is far-below the neuron's CF, even though the neuron's CF is near a resonance (i.e., a formant) in the stimulus spectrum (Fig 4.3B). This is evident from the PSTH of the impaired neuron, which shows strong phase-locked activity to the fundamental frequency or F0 (Figs 4.3A and 4.3C).

To confirm that audibility was not a primary contributing factor to these results, driven rates during voiced portions of the stimulus were compared between the two groups. Despite the expected reduction in spontaneous rate (Liberman and Dodds, 1984a), driven rates were largely compensated by the "half-gain amplification strategy" employed in this study

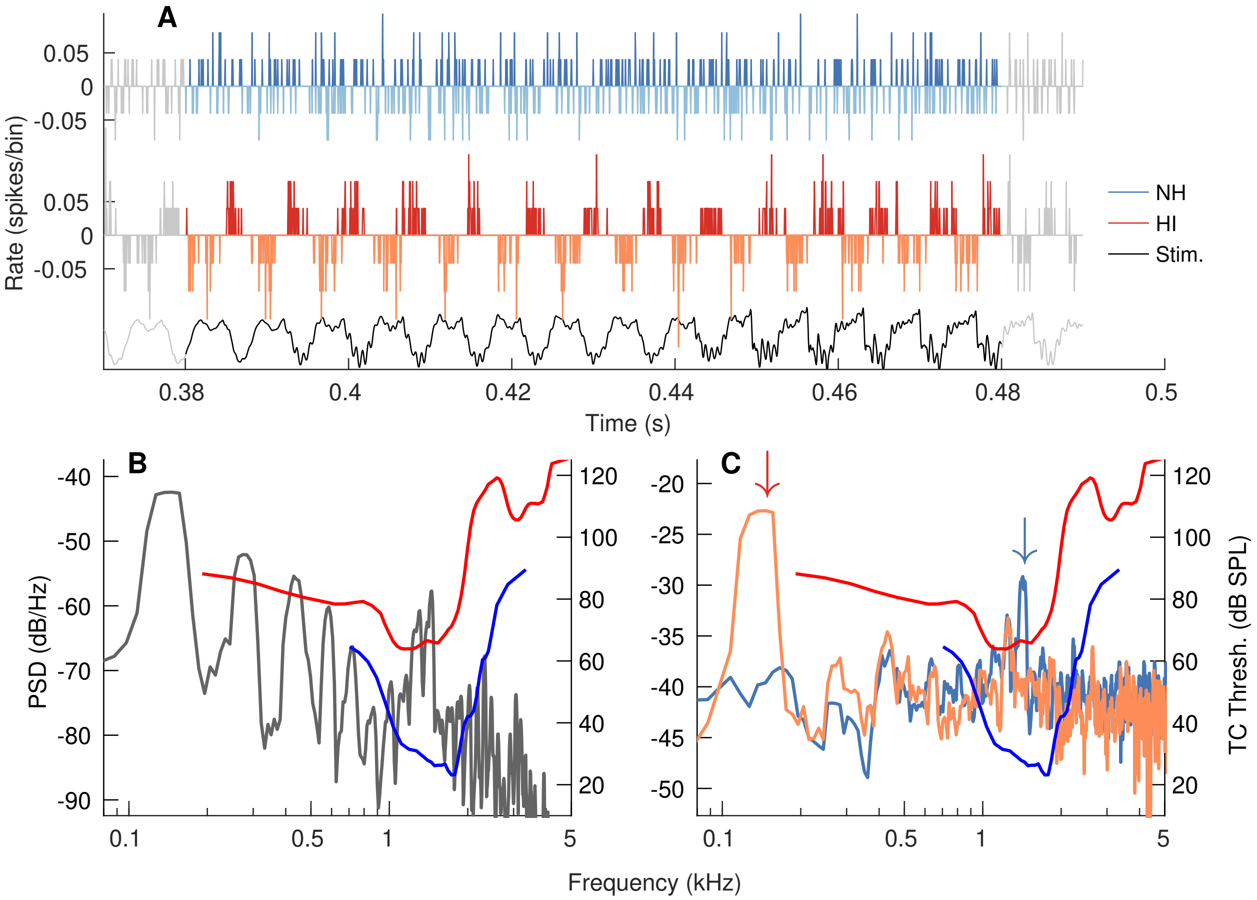

Figure 4.3. Example data demonstrate that tonotopic distortion leads to severely degraded near-CF representation and enhanced low-frequency representation in the response of an impaired AN fiber. (A) Alternating-polarity PSTHs in response to a "quasi-stationary" stimulus segment (dark gray) from a normal (blue) and impaired (red) AN fiber. Darker (lighter) shade represents PSTH in response to positive (negative) stimulus polarity. Negative polarity PSTH is flipped along the Y-axis for display. Stimulus is scaled arbitrarily. PSTH bin width = 200 μs. (B) Spectrum of the stimulus segment (dark-gray, left Y-axis) and tuning curves (right Y-axis) of the two AN fibers considered in A. (C) Spectrum of difference PSTHs (d[n], n=bin index) based on PSTHs in A. d[n] was computed by subtracting PSTHs to opposite polarities. As expected, d[n]-spectrum for the normal AN fiber peaks near the first formant (blue arrow), which is near its CF. In contrast, d[n]-spectrum for the impaired fiber peaks at the fundamental frequency (red arrow), as a direct result of distorted tonotopy.

(Fig 4.4A; p>.05, unpaired *t-test*). Next, to quantify the effects of distorted tonotopy at a population level, multitaper spectrum of the difference PSTH ($d[n]$) was considered for individual neurons (Parida et al., 2020). As a metric for tonotopic coding, fractional power was quantified in a band-pass spectral window near CF (Fig 4.4B). The spectral window was fourth order, was centered at CF for each fiber, and its 3-dB bandwidth was set to 50 percentile fit for AN-FTC bandwidth (Q3-dB) for normal-hearing chinchillas (Kale and Heinz, 2010). Use of fractional power with $d[n]$ was to minimize the effects of overall power and rate differences, if any. Near-CF power estimates were significantly lower for impaired neurons for frequencies below 3 kHz (p<.0001; unpaired *t-test*). Above 3 kHz, the difference vanished (p>.05; unpaired *t-test*) likely because phase-locking is significantly reduced for chinchilla AN fibers beyond this frequency (Temchin et al., 2005).

To quantify the susceptibility of neurons with higher CFs (>0.6 kHz) to very low-frequency (<400 Hz) stimulus, power in $d[n]$-spectrum was computed in a low-pass spectral window (3-dB cutoff = 400 Hz, fourth order). As expected, these low-frequency power estimates were significantly (p<0.01; unpaired *t-test*) higher for impaired neurons than for normal neurons (Fig 4.4C). The ratio of fractional power near CF to fractional power at low frequencies, which indicates the strength of tonotopic coding (relative to low-frequency coding), was 0 dB on average for normal neurons with CF below 2.5 kHz. In contrast, this relative power metric was 0 dB on average neurons in the same CF region. Overall, these results demonstrate a disruption of near-CF stimulus energy coding at the expense of very low-frequency stimulus energy coding at a single AN fiber level.

4.3.5 Temporal representation of dynamic F_1 contour was enhanced, whereas representation of F_2 and F_3 contours were degraded following NIHL

The effects of distorted tonotopy on formant coding were evaluated for speech in quiet or in noise. Formant coding is traditionally quantified using the Fourier spectrum of the period histogram or the difference PSTH (Sinex and Geisler, 1983; Young and Sachs, 1979); these analyses provide sufficient spectrotemporal resolution for analyzing responses to stationary speech tokens like those used in the mentioned studies. To quantify power along the formant

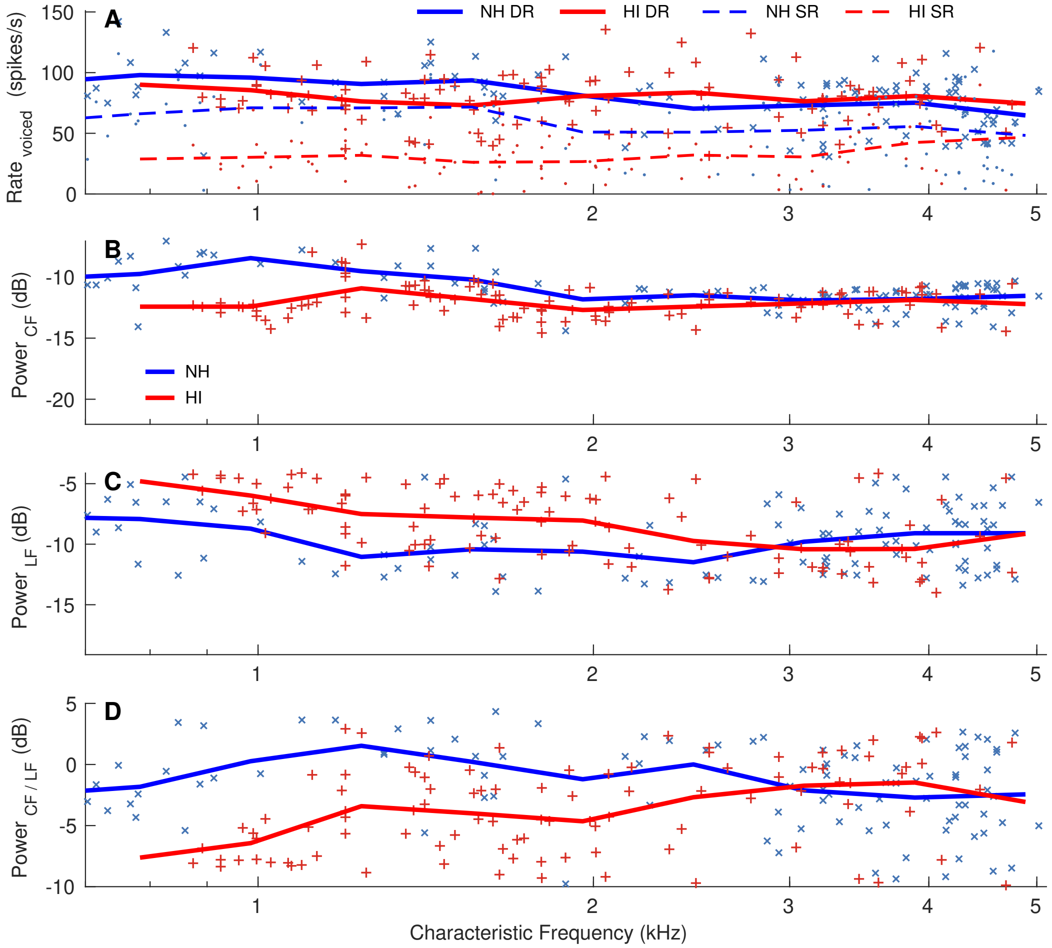

Figure 4.4. For voiced segments, effects of distorted tonotopy primarily manifest as spectral changes in the response and not as a change in driven rate. (A) Driven rates for normal and impaired AN fibers were comparable during voiced portions of the speech stimulus. Cross and plus markers represent driven rates for normal (blue) and impaired (red) AN fibers. Dots represent spontaneous rate. Solid and dashed lines are two-third octave averages for driven rate and spontaneous rate, respectively. (B) Fractional power near CF (based on d[n]) for individual AN fibers. For a given fiber (CF), fractional power was quantified as the power in a rectangular window (centered at CF, bandwidth based on 50th percentile Q10 fit for normal-hearing chinchillas, (Kale and Heinz, 2010) normalized by the total power of d[n]-spectrum. (C) Fractional power in a low-frequency band (< 400 Hz) for AN fibers. (D) Ratio of fractional power near CF to fractional power in the low-frequency band for AN fibers. While this fractional power ratio for normal AN fibers with CF below 2.5 kHz was 0 dB in general, it was substantially reduced (<-4 dB) for the impaired population in the same CF range.

trajectories in a nonstationary speech stimulus, the harmonicgram can be used as it offers superior spectrotemporal resolution compared to the spectrogram (Parida et al., 2020).

Formant coding strength was quantified as the power estimate along the trajectories of the first three formants for different SNR conditions (Fig 4.5). Results for -10 dB SNR are not shown as response to speech was virtually absent in most fibers. For speech in quiet, fractional power for individual formant trajectories peaked at CFs near or slightly above their mean frequencies (Figs 4.4A-4.4C), a consistent result that has previously been reported (Delgutte and Kiang, 1984a; Young and Sachs, 1979). Lower limit of CF has been set to 500 Hz because of a scarcity of AN fibers below this frequency range in our fiber pool. As expected, representation of F_1 was enhanced for the impaired group, and was extended to CFs substantially above the F_1 trajectory range, (i.e., CFs in the F_2 and F_3 ranges). Coding of F_2 and F_3 was diminished for the impaired group, and the corresponding peaks were shifted to a higher frequency than the expected CF ranges based on mean formant trajectories. This is likely a result of the best frequency lowering, a classic physiological change that accompanies distorted tonotopy (Henry et al., 2016; Liberman, 1984).

Noise had a much more detrimental impact on already degraded representations of F_2 and F_3 for the impaired population (Figs 4.4D-4.4H). At 0 dB SNR, peaks for all formants were still discernible for the normal-hearing pool (Figs 4.4D-4.4F). For impaired neurons, while F_2 was somewhat represented (at a CF location much higher than the expected mean F_2 frequency), F_3 coding was almost nonexistent. Similarly, at -5 dB SNR, the normal-hearing pool showed a discernible peak near F_2, whereas the impaired pool did not show any F_2 coding. Although F_1 was overrepresented in the temporal representation by impaired neurons at all these SNRs, a temporal-place code scheme is likely to show a deficit based on the tonotopically nonspecific F_1 representation in the same pool.

Overall, these results show a degradation of higher-formant (F_2 and F_3) coding, which are known to be important for perception of vowels as well as consonants, in responses of impaired neurons following acoustic insult. Moreover, these degradations are more severe in the presence of noise, which potentially contributes to the perceptual deficits that hearing-impaired listeners experience in noisy environments.

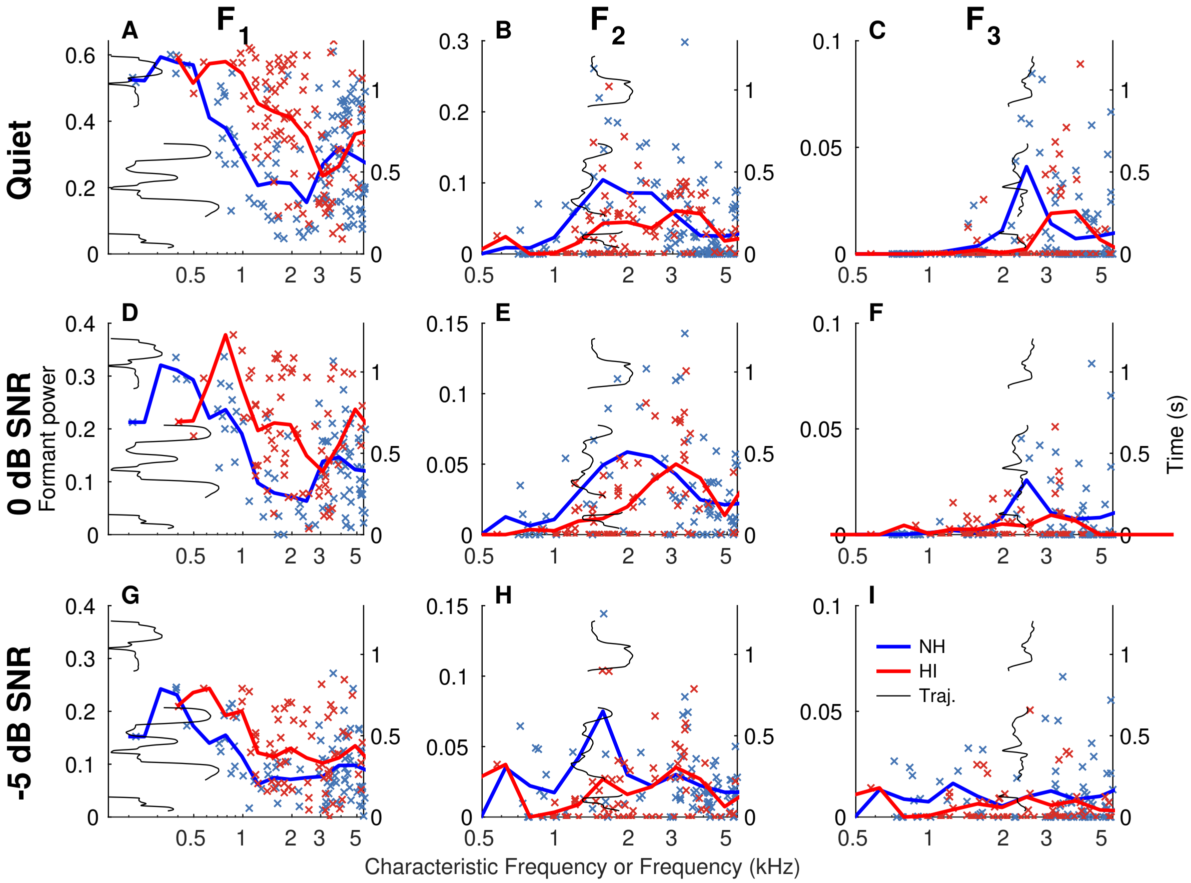

Figure 4.5. Noise affected the temporal-place formant representation for the hearing-impaired population much more than that for the normal-hearing population. The nine panels represent the strength of the first three formants (F1, F2, and F3) for several SNR (quiet, 0 dB, and -5 dB) conditions for normal (blue) and impaired (red) AN fibers. In each panel, markers (left Y-axis) represent the fractional power estimate along the corresponding formant trajectory for noisy-speech response relative to the fractional power for the noise-alone response. d[n] was considered to quantify fractional power. Thick lines indicate two-third octave averages per group. Formant trajectories (thin black) are denoted on the right Y-axis. X-axis is limited to CFs above 0.5 kHz due to sparsity of AN fibers with CF $<$ 0.5 kHz in either group. For the normal-hearing population, average lines peak near formants, as expected, even in noise (e.g., E, F and H). In contrast, average lines for the quiet condition for the impaired population peaked at frequencies well above formants' trajectory. This already degraded formant representation was further diminished in the presence of noise, much more severely compared to the normal population (e.g., E, F, and H).

4.3.6 Unlike voiced segments, driven rate for a fricative consonant was not restored despite compensating for audibility loss

Fricatives constitute a substantially large portion of phoneme confusions among hearing-impaired listeners (Bilger and Wang, 1976; Dubno et al., 1982; Van de Grift Turek et al., 1980). To explore the potential neural bases underlying these deficits, neural responses to a fricative (/s/) and two stop consonants (/d/ and /g/) were analyzed.

Previous studies have reported robust fricative coding by normal AN fibers in terms of onset and sustained rate responses (Delgutte and Kiang, 1984b). Normal fibers with CF near "high" frequencies (i.e., near frequencies where fricative has strong energy, e.g., Fig 4.6A) show a sharp onset, followed by a sustained discharge rate that is well above spontaneous activity (Fig 4.6B). In contrast, impaired neurons showed a substantial reduction in onset response (e.g., Fig 4.6B), although sustained rate remained above spontaneous activity.

Onset and sustained rates were estimated for individual AN fibers by windowing their responses at the onset and steady-state part of the fricative (Fig 4.6B). These windows were based on a previous neurophysiological study on fricatives (Delgutte and Kiang, 1984b). In 3 to 8 kHz CF range, i.e., where the fricative had substantial energy, onset rates were significantly lower for the impaired population (Fig 4.6C; p<.0001, unpaired *t-test*). These results starkly contrast with results for voiced portion of the stimulus, where the driven rate was comparable between the two groups (Fig 4.4A). Sustained rate was also reduced for the impaired population compared to the normal population, albeit to a lesser extent (Fig 4.6D); p<.05, unpaired *t-test*.

To evaluate the noninvasive correlates of these single-neuron level changes in consonant coding, frequency following responses (FFRs) were recorded in response to the speech stimulus from the same animal pool. These FFR data have been described previously in Chapter 7. Only the onset response was considered to evaluate consonant coding because evoked responses like the FFR require synchronous activity (e.g., the onset) across populations of neurons. As sustained responses to fricatives lack a clear temporal pattern, they are rather weakly represented in the FFR, if at all (Skoe and Kraus, 2010).

Two representative FFR responses, one from each group, demonstrate the effects of NIHL on consonant coding (Fig 4.6E). Responses during voiced speech (e.g., before 650 ms) were

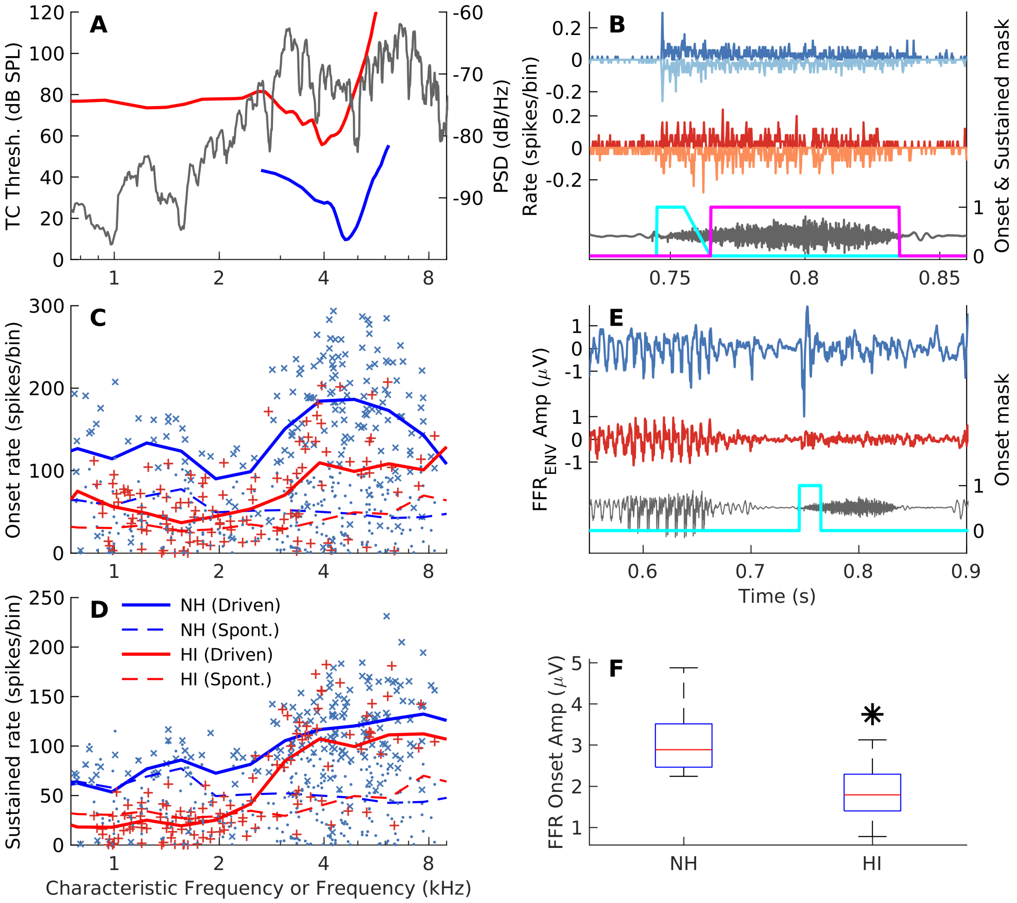

Figure 4.6. The flat amplification strategy did not restore the driven rate in response to fricative /s/ for the impaired group. (A) Spectrum of /s/ (gray, right Y-axis) and tuning curves of example normal (blue) and impaired (red) AN fibers. (B) Time-domain waveforms of /s/ (gray), and alternating-polarity PSTHs of normal (blue) and impaired (red) AN fibers considered in A. PSTH bin width = 200 µs. Same format as Fig 2A. While both example AN fibers had comparable sustained response, onset response for the impaired neuron was substantially degraded compared to the normal AN fiber. Cyan and magenta windows (right Y-axis) denote the masks used to compute onset and sustained rates, respectively, for individual AN fibers. (C-D) Same format as Fig 4A. Onset rate for the impaired population was substantially reduced compared to the normal population for all CFs (C). Sustained rate was slightly reduced for impaired AN fibers in the high CF (> 3 kHz) region where /s/ had substantial energy (D). (E) Example FFR_{ENV} data from a normal-hearing (blue) and a hearing-impaired (red) chinchilla demonstrate reduction in onset response for /s/. FFR_{ENV} was estimated by averaging the responses to opposite polarities of the stimulus. Cyan window (right Y-axis) indicates the onset mask. (F) Distributions of peak-to-peak amplitude data (in the onset window in E) show a significant reduction in FFR onset in response to /s/ for hearing-impaired chinchillas.

comparable between the two FFRs (also see Chapter 7). In contrast, the two responses strongly diverged during the onset window. While FFR from the normal-hearing animal had a sharp onset, the FFR from the hearing-impaired animal lacked any clear onset.

To quantify the strength of onset in the FFR of the normal and impaired populations, peak-to-peak amplitude in the onset window was quantified. Results showed a significant reduction in the onset amplitude for the impaired population (p<.05; unpaired *t-test*). Overall, these results indicate that audibility profiles for low-intensity fricatives are fundamentally different from those of high-intensity voiced speech.

4.3.7 Driven rates for stop consonants were also not restored following NIHL despite compensating for audibility loss

Similar to fricatives, stop consonants are also among the most confused phonemes for hearing-impaired listeners (Bilger and Wang, 1976; Van de Grift Turek et al., 1980). The neural representation of the two stop consonants (/d/ and /g/) that were present in the speech stimulus used here were evaluated based on onset and sustained rates. While /g/ is characterized by more low-frequency (< 2 kHz) energy, /d/ is characterized by a flatter spectrum with more energy at slightly higher frequencies (up to 4 kHz, e.g., Fig 4.7A). Responses of the two example neurons in Fig 4.7A are representative of the two groups. In response to /d/ and /g/, the normal AN fiber showed a strong onset response (Figs 4.7B and 4.7C), followed by sustained activity that was well-above spontaneous activity. In contrast, for the impaired AN fiber, onset was nonexistent and sustained activity was indistinguishable from spontaneous activity. Population results also suggested that onset and sustained rates for /d/ and /g/ were significantly degraded for the impaired population relative to the normal population, particularly in the appropriate CF region for the consonants (Figs 4.7D-4.7G). For example, although onset rate to /g/ was slightly reduced (p=0.06; unpaired *t-test*) for AN fibers with CF > 2 kHz, it was significantly reduced (p<.0001; unpaired *t-test*) for fibers with CF < 2 kHz, frequency region where /g/ has substantial energy (Fig 4.7F).

Next, response onsets were compared in FFRs recorded from animals in the two groups. Onset strength was quantified as the peak-to-peak FFR amplitude in an onset window. FFR onset was significantly reduced for /d/ following NIHL (Fig 4.7H), consistent with the

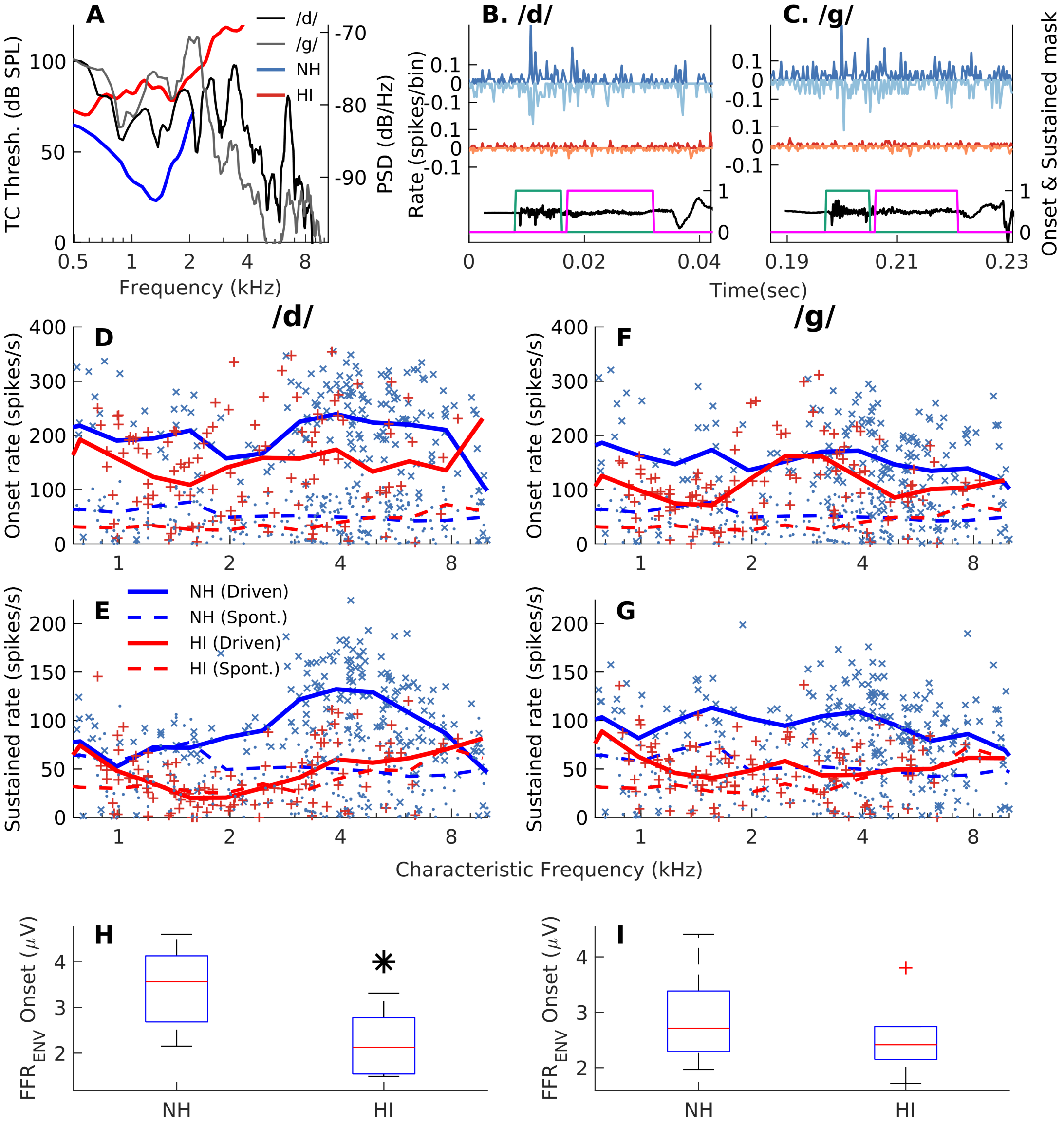

Figure 4.7. Driven rate for low-intensity stop consonants (/d/ and /g/) was also diminished despite compensating for ABR-based audibility loss. (A) Right Y-axis: spectra for /d/ (dark gray) and /g/ (light gray). Left Y-axis: FTCs for a normal (blue) and an impaired (red) AN fiber. Responses of these AN fibers are representative of their groups. (B and C) Responses of the two fibers considered in A to /d/ (B) and /g/ (C). Same format as Fig 3A. PSTH bin width = 200 µs. (D-G) For both stop consonants, onset and sustained rates for impaired AN fibers were significantly reduced compared to rates for normal-hearing AN fibers. Same format as Fig 4A. (H and I) Distributions of FFR peak-to-peak onset amplitudes for chinchillas in both groups for /d/ (H) and /g/ (I). Onset amplitude was significantly reduced for the impaired group in response to /d/, and to a lesser extent, in response to /g/.

universal reduction in onset rate across the whole CF range for impaired neurons (Fig 4.7D). Onset response to /g/, however, was only slightly reduced (not significant) for the hearing-impaired chinchillas compared to the normal-hearing chinchillas (Fig 4.7I). This reduction in effect size in the FFR is likely due the following two reasons: (1) high-frequency neurons contribute more to FFRs (Shinn-Cunningham et al., 2013), and (2) the onset rate to /g/ was only slightly reduced in high-frequency (>2 kHz) neurons following NIHL (Fig 4.7F).

These results, taken together with results from previous sections, show that a flat amplification targeted at compensating for loss of audibility (based on thresholds) restored driven rate for voiced sections. However, this amplification strategy was inadequate at restoring driven onset and sustained rates for low-intensity consonants, like fricatives and plosives, for majority of AN fibers.

4.3.8 Background noise exacerbates degraded fricative coding following NIHL, particularly for AN fibers with lower spontaneous rate

As previously described, listeners with hearing loss often struggle in noisy environments. Confusions are substantially more common for consonants than for vowels. Consonant perception, particularly in noise, is mediated by a number of acoustic cues, such as spectral peaks, spectral edges, formant transitions, voice onset time, and more. Here, we investigated how the fricative /s/ was represented in the presence of background noise for both groups. It was assumed that all spectral cues were captured by response envelope (i.e., fluctuation in driven rate). Plosive coding in noise was not considered for this envelope-based analysis because plosives were completely masked by noise, even at the highest SNR (i.e., 0 dB) used here. Instead, inferences about plosive coding in noise based on formant transitions is considered in the Discussion.

When signal is mixed with noise to achieve a particular SNR, including negative SNRs, the resultant signal often has scattered spectrotemporal regions that have a favorable SNR. The regions with favorable SNRs likely mediate robust speech perception in noise (Cooke, 2006). For example, after mixing speech and speech-shaped-noise to achieve an overall SNR of 0 dB, the fricative portion of the speech sentence has a positive SNR above 3 kHz (Fig 4.8A). Normal AN fibers, which are narrowly tuned to frequencies near this high SNR region,

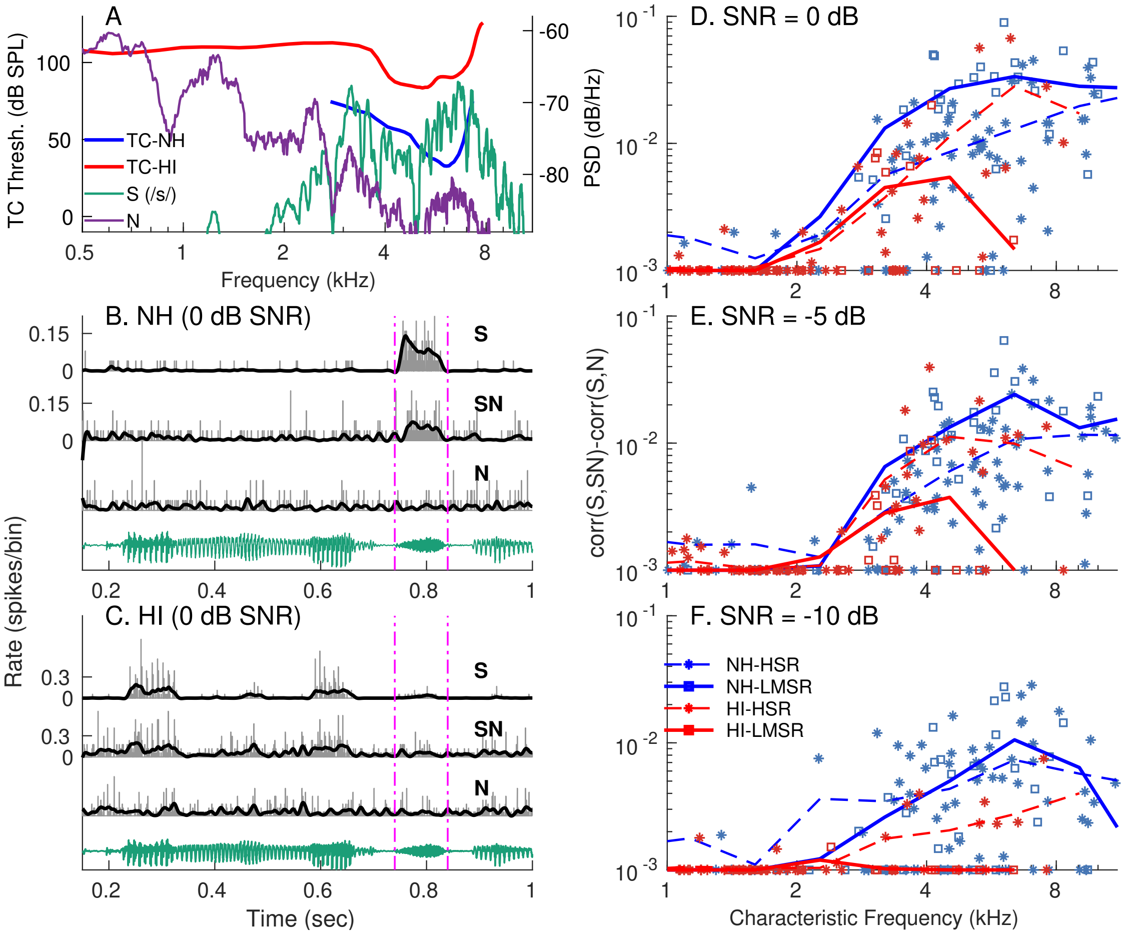

Figure 4.8. Effects of distorted tonotopy severely disrupted responses to fricative in noise, particularly for AN fibers with low and medium SR. (A) Right Y-axis: spectra for /s/ (green) and the concurrent noise segment (purple, for 0 dB overall SNR condition). Even though overall SNR was 0 dB, SNR in the high frequency region (>3 kHz) was greater than 0 dB. Same was the case for -5 dB and -10 dB SNR conditions (not shown). Left Y-axis: FTCs of two example AN fibers that demonstrate the increased deleterious effect of noise following NIHL. (B and C) PSTHs in response to speech-alone, noisy-speech, and noise-alone for the normal (B, SR=0.2/s) and impaired (C, SR=1.1/s) fibers considered in A. PSTH bin width = 200 μs. Thick black lines represent response envelopes, which were obtained by low-pass filtering PSTHs at 32 Hz (fourth-order zero-phase IIR filter). (D-F) Corrected correlation (see Experimental Design) between responses to speech-alone and noisy-speech for the /s/ portion. Squares and asterisks correspond to AN fibers with low/medium SR (<20 /s) and high SR (>20 /s), respectively. Normal AN fibers, particularly those with low/medium SR that are known to robustly encode speech features in noise, maintained good fricative coding even in the -10 dB SNR condition. In contrast, impaired AN fibers showed lower correlation values, particularly those with low/medium SR.

responded selectively to the fricative energy (e.g., Fig 4.8B). In contrast, impaired neurons show reduced tip sensitivity relative to FTC tails (e.g., Fig 4.8A). As a result, these neurons tuned to higher frequencies were found to respond poorly to fricative energy and strongly to low-frequency energy in either speech or noise (e.g., Fig 4.8C).

To evaluate the neural representation of /s/ in noise, correlation [corr(S,SN)] of the slowly varying envelope was quantified between responses in speech-alone and noisy-speech conditions during the fricative segment. To confirm that these correlation values were not spurious correlations between speech and noise, corr(S,SN) was corrected by subtracting the correlation [corr(S,N)] between response envelopes of speech-alone and noise-alone conditions for the same fricative window. For the normal-hearing population, these correlation values were robust for AN fibers with CF $>$ 3 kHz at SNRs down to -5 dB, and to a lesser extent at -10 dB (Figs 4.8D-4.8F). Responses of AN fibers with low or medium spontaneous rate were particularly resistant to noise, a finding that has been reported previously (Silkes and Geisler, 1991). On the other hand, corrected correlation values for low and medium SR fibers were significantly lower for the impaired population. This degradation in consonant coding in noise following NIHL is an expected outcome based on the effects of distorted tonotopy as shown in the two example AN fibers (Figs 4.8A-4.8C).

4.4 Discussion

The present study elucidates numerous ways in which distorted tonotopy, in addition to classical indicators of SNHL such as broader tuning and threshold elevation, potentially contributes to the clinically prevalent phenomenon of “I can hear you, but I cannot understand you”, particularly in noise. These results suggest that perceptual deficits experience by listeners with NIHL primarily stem from the physiological alterations at the level of the inner ear following acoustic trauma. In particular, threshold elevation in ABRs (and audiograms) may underestimate threshold changes at a population level (Fig 4.1). At suprathreshold levels, neuronal responses maintain near-normal rate responses (Fig 4.4A) and precision (Fig 4.2) but suffer from changes in spectral profiles (Figs 4.3 and 4.4). As a result, informative near-CF features are diminished at a single-neuron level. Instead, low-frequency energy of speech (leads to overrepresentation of minimally informative features) and/or noise (increased noise susceptibility) are enhanced in the response (Figs 4.3 and 4.4). Unlike driven rate to voiced segments, driven rate in response to a fricative and two stop consonants was not restored despite compensating for audibility loss (Figs 4.6 and 4.7). Finally, responses of AN fibers with low and medium SR, which exhibited more noise-resistant responses, were more adversely affected by noise following NIHL.

4.4.1 Distorted tonotopic changes in the periphery lead to substantial speech coding degradation following NIHL

Early evidence of distorted tonotopy comes from neurophysiological studies that showed abnormal changes in FTC following NIHL such as a reduced tip and a hypersensitive tail (Liberman and Dodds, 1984b), and confirmed a modest lowering of best frequency using anatomical labeling methods (Liberman, 1984; Liberman and Dodds, 1984b). These distortions in tonotopic mapping hold important implications for complex sound processing. For example, recordings from AN fibers show a lowering of effective best frequencies when white Gaussian noise is used as stimulus (Henry et al., 2016). Similar effects have been reported using stationary synthesized vowels, where the first formant was overrepresented in the responses of AN fibers across a wide CF range (Miller et al., 1997). In contrast, second and third formants were severely degraded. We recently showed that the degree of distorted tonotopy is even more deleterious for signals with negatively sloping spectra (Chapter 7). Very low-frequency energy (<250 Hz) in the same speech stimulus used here was overrepresented in the central nervous system based on frequency following responses. Present results confirm that those FFR-based findings were a direct manifestation of tonotopic changes happening at the most peripheral level (i.e., the inner ear).

4.4.2 Audiograms shifts with NIHL obscure more significant shifts in sensitivity of suprathreshold AN fibers

Overwhelming evidence suggests that audiograms are insensitive in capturing suprathreshold perceptual deficits of HI listeners, especially in complex environments. Similarly, ABR thresholds, which serve as proxy for audiograms, can also be insensitive. These near-threshold metrics reflect the activity of only a few sensitive neurons, and thus, can obscure the changes in threshold at a population level. For example, the difference in 10th-percentile threshold for normal and impaired neurons was 26 dB, which is similar to the average ABR threshold shift (Fig 4.1A). In contrast, the median threshold shift for the impaired population was elevated by 36 dB SPL. The shift in 10 and 50 percentile points of the threshold distribution following NIHL was further rigorously quantified in a CF-specific manner for chinchilla data in the present study as well as cat data from Heinz et al., 2005 (Figs 4.1E and 4.1F, respectively). The shift in 10 percentile point was lower than the shift in 50 percentile point for both chinchillas and cats for all CF bands. This alteration in threshold distribution of AN fiber following NIHL holds important implication for hearing-aid algorithms. In particular, linear gain aimed at mapping the output intensity range from audiogram threshold to pain threshold may skew the distribution of sensation levels of neurons, and thus, not be optimal. Instead, an optimal strategy may include mapping intensities such that distribution of sensation levels of neurons closely matches the threshold distribution of normal neurons. Future studies exploring the exact nature of such dynamic range alterations following NIHL will provide additional insight.

4.4.3 Changes in neural coding of voiced segments were primarily in the spectral profile and not the driven rate

Driven rates in response to voiced speech, which could be treated as markers for audibility, were comparable between the two groups. However, the spectral content in the responses differed substantially between the two groups. These results highlight the need to develop novel metrics that quantify “tonotopic audibility”, which are not only affected by driven rate along the tonotopic axis but also consider the tonotopic appropriateness of such neural activity. Such tonotopic audibility metrics could be useful in devising gain profiles for individual patients to enhance perceptually informative audio features and suppressing deleterious spectral regions.

4.4.4 Unlike voiced segments, audibility for (low intensity) consonants was not restored

Of the many features that influence consonant perception, two key features (driven rate and formant transitions) were considered here. Representation of formant transitions for higher formants (F2 and F3) were substantially degraded for the impaired population, especially in noise. As formant transitions are important for speech-in-noise perception, especially for low-intensity stop consonants that can be easily masked by noise, degradation of these transitions likely contributes to the perceptual deficits of hearing-impaired listeners and is consistent with the results from psychoacoustic studies that report a reduced ability to use formant transitions following NIHL (Zeng and Turner, 1990).

Unlike for voiced segments, driven rates for several low-intensity consonants were not restored. Such divergent audiometric effects are consistent with psychoacoustic studies that report a similar difference in audibility for consonants and vowels even after compensating for audibility (Phatak and Grant, 2014). For the fricative /s/, while the normal population encoded information via both onset and sustained rates, the representation for the impaired population was limited to sustained rates. This reduced redundancy following NIHL likely underlies fricative confusions, which are notoriously common for hearing-impaired listeners (Bilger and Wang, 1976; Owens et al., 1972; Turner and Robb, 1987). Similarly, stop consonant confusions for hearing-impaired listeners, also extremely prevalent, are likely mediated by reduced neural activity as a result of insufficient audibility compensation (Bilger and Wang, 1976; Owens et al., 1972; Turner and Robb, 1987). These results highlight the importance of adaptive amplification strategies for hearing aids that boost the audibility of low-intensity consonants without making the intensity of voiced segments too loud.

4.4.5 Noise-resistant ability of low and medium SR fibers was substantially degraded following NIHL

Hearing-impaired listeners particularly struggle in noisy environments, which may result from their inability to take advantage of spectrotemporal glimpses with high SNR, either in the audio domain (Cooke, 2006) or in the modulation domain (Jørgensen et al., 2013). To test this hypothesis, a temporal segment (/s/) was identified that had favorable SNR

(>0 dB) in a limited spectral band (above 3 kHz) even though the overall (across the whole duration) SNR was 0 dB or lower. Normal neurons with low and medium SR maintained robust noise-resistant representation, some even at -10 dB SNR. In contrast, counterpart impaired neurons were highly susceptible to noise (Fig 4.8). Overall, these results highlight the importance of noise-reduction algorithms, particularly for hearing aids to mitigate the effects of increased noise-susceptibility.

4.4.6 Precision clinical diagnostics will improve clinical outcomes

As mentioned earlier, there is strong consensus among researchers regarding the inadequacies of the audiogram to account for real-life perceptual deficits. Individual variability in speech perception likely stems from a variety of suprathreshold deficits, which are hidden from audiograms and vary across patients. For example, two subjects with similar degree but different etiologies of hearing loss may show widely varying neural coding profiles (Henry et al., 2019). Development of sensitive and specific diagnostics to identify underlying pathophysiology, such as synaptopathy or distorted tonotopy, and implementing individualized intervention strategies are hypothesized to improve clinical outcomes of hearing aids based on the variety of suprathreshold deficits elucidated here.

5. DEGRADED SPEECH-IN-NOISE ENVELOPE CODING IN THE AUDITORY NERVE FOLLOWING NOISE-INDUCED HEARING LOSS

Summary[1]

Speech-intelligibility (SI) models aim to predict human perceptual performance when speech is subjected to a range of acoustic manipulations. SI models not only test our understanding of how the brain processes speech but can also assess hearing-aid and cochlear-implant signal-processing strategies. To date, a single SI model that can predict all of the normal-hearing effects of the wide range of tested manipulations remains elusive. Furthermore, listeners with hearing loss continue to struggle in noisy environments despite state-of-the-art hearing-aid strategies. These problems highlight two key gaps. Firstly, most SI models are implemented in the audio domain (i.e., by processing the audio signal) and rely on numerous assumptions. Unfortunately, there is a scarcity of published neural data to evaluate these model assumptions. Secondly, it is not straightforward for audio-domain SI models to predict the effects of hearing impairment (beyond audibility). This gap is primarily due to our limited understanding of the variety of effects of cochlear hearing impairment on speech (especially speech-in-noise) coding. For example, two subjects with similar audiograms can have very different speech-understanding ability. Animal models of cochlear hearing impairment provide a useful opportunity to collect data to directly address these SI-model gaps. Here, we collected spike trains from auditory-nerve fibers in anesthetized chinchillas in response to speech, spectrally matched stationary noise, and noisy-speech mixtures at perceptually relevant SNRs (-10, -5 and 0 dB). Overall speech intensity was fixed at 65 dB SPL for normal-hearing and 80 dB SPL for hearing-impaired animals. Alternating-polarity peristimulus-time histograms constructed from spike trains were filtered to extract response envelopes in different modulation bands. Our data reveal that the correlation between AN-fiber response envelopes of noisy-speech and noise-alone is increased for hearing-impaired fibers, suggesting a greater potential degree of distraction from inherent envelope fluctuations following cochlear hearing loss. Interestingly, our data predict a stronger detrimental

[1] Manuscript *in prep*

effect of fluctuating noise compared to stationary noise, as has been reported in psychoacoustic studies. This novel neural finding is significant given the emphasis recent SI models have placed on the importance of inherent envelope fluctuations in predicting noisy-speech perception. These results highlight significant effects of cochlear hearing loss on the strength of inherent noise fluctuations in noisy speech, which are likely to be important in addition to speech-coding fidelity when predicting hearing-impaired SI.

5.1 Introduction

While normal-hearing (NH) listeners demonstrate robust speech-perception ability even in adverse listening conditions, listeners with hearing impairment (HI) often struggle to understand speech, even in moderately challenging backgrounds. In fact, listening difficulty in noisy environments is the number one complaint in audiology clinics, despite the use of state-of-the-art hearing aids (Souza, 2016). Clinical outcomes of hearing aids can be improved by selectively enhancing the most informative speech features. However, what these features are, is a matter of active debate. It is common to use envelope and temporal fine structure as a dichotomy of information present in a signal and has been quite popular in auditory research (Heinz and Swaminathan, 2009; Saberi and Haftert, 1995; Shamma and Lorenzi, 2013; Smith et al., 2002; Voelcker, 1966a; Xu and Pfingst, 2003). Psychoacoustic studies have highlighted the importance of envelope for speech perception (Crouzet and Ainsworth, 2001; Drullman et al., 1994a, 1994b; Houtgast and Steeneken, 1985; Shannon et al., 1995; Smith et al., 2002). What is not clear is the relative importance of envelope and fine structure on speech perception in the presence of noise.

Psychoacoustic studies hypothesize that speech-in-noise envelope representation in the hearing-impaired population is degraded more than the normal-hearing population, which contributes to perceptual speech-in-noise deficits that listeners with hearing loss experience (Moore and Glasberg, 1993; Moore et al., 1999). The importance of envelope is further supported by the success of envelope-based SI models (Goldsworthy and Greenberg, 2004; Houtgast et al., 1980; Jørgensen et al., 2013; Relaño-Iborra et al., 2016; Scheidiger et al., 2018). Recently, one study showed a modest contribution of enhanced masker envelope coding to degraded speech intelligibility in noise (Millman et al., 2017). While these electrophysiological results are promising, the place specificity of these far-field responses is not well understood (Bones and Plack, 2015; Dau, 2003; Gardi and Merzenich, 1979). Spike trains recorded from auditory-nerve (AN) fibers in response to speech provide an alternate way to quantify the relative degradation of envelope versus fine structure. The AN fiber response conveys the internal representation of speech. As there is general consensus that speech is faithfully represented in a tonotopic coding framework and because single-unit recordings are place specific (Delgutte and Kiang, 1984c; Palmer et al., 1986; Sachs and Young, 1980; Young and Sachs, 1979), SI models based on single neuron spike trains offer an unique advantage over SI models based on scalp recordings. While many studies have looked at speech coding in the AN, these studies have usually been limited to short and synthesized speech tokens (see Young, 2008 for review). These synthesized signals lack the rich dynamic spectrotemporal information that natural speech carries (Greenberg et al., 2003). In addition, there has been limited focus on the neural characterization of speech-in-noise, even for normal-hearing animal models; no study has looked at speech coding in noise for hearing-impaired animals.

In order to address these gaps, we collected spike-train data from individual AN fibers in response to natural speech at different SNRs for both normal and hearing-impaired chinchillas. A particularly important SI model is the speech envelope power spectrum model, which hypothesizes that the signal to noise ratio in the envelope domain (SNR_{ENV}) is the primary determinant of SI (Jørgensen and Dau, 2011). This hypothesis is important because both animal studies as well as psychoacoustic studies have shown that envelope coding is generally enhanced following hearing loss (Anderson et al., 2013; Goossens et al., 2018; Henry et al., 2014; Kale and Heinz, 2010; Lai and Bartlett, 2018; Millman et al., 2017; Presacco et al., 2016; Wilding et al., 2012; Zhong et al., 2014), although exceptions exist (Ananthakrishnan et al., 2016; Schoof and Rosen, 2016). It is hypothesized that stronger envelope coding is not necessarily advantageous for speech perception since broadband maskers can be over-represented in the envelope domain compared to speech, causing a reduction in SNR_{ENV} for the hearing-impaired population (Kale and Heinz, 2010; Moore and Glasberg, 1993). However, this hypothesis has not been tested neurophysiologically.

In this work, we use the multi-resolution envelope correlation metric proposed by Relaño-Iborra et al., 2016 to directly quantify the strength of speech coding and noise coding in AN fiber responses to noisy-speech mixtures for both normal and hearing-impaired chinchillas. We evaluated the enhanced envelope hypothesis in the framework of fluctuating masking release (FMR). FMR refers to the phenomenon that normal-hearing listeners show better speech-recognition ability in fluctuating backgrounds compared to stationary backgrounds. In contrast, hearing-impaired listeners do not show this benefit (Festen and Plomp, 1990). We hypothesized that the response to noisy-speech will be more correlated with the response to noise-alone for the impaired fibers compared to normal fibers, and that this effect will be more severe for a fluctuating masker compared to a stationary masker.

5.2 Methods

Surgical preparation, noise exposure, neurophysiological recording procedure are described in Secs 2.1, 2.2, and 2.3.

5.2.1 Stimuli

A speech sentence, described in Sec 2.4, was presented in two types of noises: (1) stationary speech-shaped noise (SSN) and (2) 8-Hz sinusoidally amplitude modulated fluctuating noise (FLN). Both SSN and FLN were spectrally matched to ten sentences spoken by the same talker as the speech sentence. Spike trains were collected in response to speech-alone, noise-alone, and noisy-speech mixtures at three different SNRs (-10, -5, and 0 dB). Intensity of the speech-alone stimulus was fixed at 65 dB SPL for normal-hearing animals and 80 dB SPL for hearing impaired animals to play speech at similar sensation levels. Driven rates to speech-alone stimulus were similar between the two groups at these intensities (Fig 4.4A). To create a noisy-speech mixture at a desired SNR, noise was scaled and added to the speech-alone stimulus. Both positive and negative onset polarities were presented for all stimuli.

5.2.2 Data Analysis

Peristimulus time histograms (PSTH, see Fig 5.1) were constructed using recorded spike trains (bin width = 0.5 ms). PSTHs were filtered using a modulation filter bank to extract

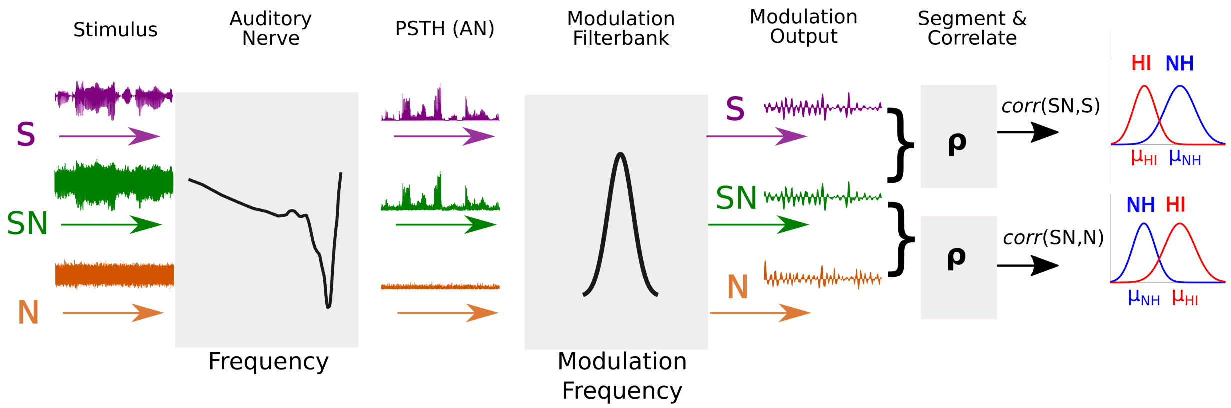

Figure 5.1. The framework used to estimate noisy-speech correlation with speech and with noise for NH and HI. The framework consists of three stages. The first stage represents processing by AN fibers with an audio signal (speech-alone [S], noise-alone [N], or noisy-speech mixture [SN]) as the input and spike-train data as the output. The second stage represents envelope extraction by a modulation filter bank that mimics properties of midbrain neurons. Peristimulus time histograms constructed using the AN spike trains are used as input to this modulation filter bank. The output is the extracted response envelope. In the third stage, these extracted envelopes are divided into segments in a multi-resolution framework (segments of $3/f_m$ duration for a modulation filter with f_m center frequency). For each segment, the correlation coefficient between S and SN as well as between N and SN are computed. Correlation values are averaged across segments to compute a single correlation value for a given neuron. These correlation values are used to construct probability density functions for $corr(SN, S)$ and $corr(SN, N)$.

response envelopes in different frequency bands (Parida et al., 2020). The center frequency of the modulation filters (octave-wide, zero-phase, 8-th order, IIR) spanned from 4 to 128 Hz (octave spaced). The output of each modulation filter was segmented in a multi-resolution framework using a segment length of $3/f_{mod}$ for a modulation filter with center frequency of f_{mod}.

For a given AN fiber in a single condition (based on noise type, SNR, and modulation filter), cross-correlation coefficient [$corr(SN, S)$] was computed between response envelopes for speech-alone (S) and noisy-speech (SN) for each segment. For that AN fiber, these correlation values were averaged across segments to generate a single correlation value. This

$corr(SN, S)$ metric served to quantify speech-coding fidelity for a single AN fiber. Distribution of $corr(SN, S)$ was estimated for both NH and HI groups by pooling $corr(SN, S)$ values across all AN fibers in a group (Fig 5.1). This procedure was repeated for all conditions (i.e., both noise types, all modulation filters, and all SNRs). Distributions were also obtained for correlation between response envelopes for noisy-speech (SN) and noise-alone (N) in the same multi-resolution framework [$corr(SN, N)$ in Fig 5.1]. This $corr(SN, N)$ metric was used to quantify the distracting nature of noise in noisy-speech representations.

Next, for a correlation metric [e.g., $corr(SN, S)$], the distance between group distributions was computed using a sensitivity metric (d') as follows.

$$d' = \frac{\mu_{HI} - \mu_{NH}}{\frac{1}{2}\left(\sigma_{HI}^2 + \sigma_{NH}^2\right)}, \tag{5.1}$$

where μ_{NH} is the distribution mean for the normal group and μ_{HI} is the distribution mean for the impaired group. σ_{NH} and σ_{HI} denote the standard deviations for normal and impaired groups, respectively. d' was computed for both $corr(SN, S)$ and $corr(SN, N)$.

Neural speech-intelligibility (SI_{neural}) metrics were constructed by averaging $corr(SN, S)$ values across four modulation bands for each group:

$$SI_{neural} = \frac{1}{4} \sum_{f_{mod}} \left[\frac{1}{K} \sum_{k=1}^{K} corr_k(SN, S) \right]^2, \tag{5.2}$$

where K denotes the number of AN fibers in a group, $corr_k(SN, S)$ denotes $corr(SN, S)$ for k-th fiber in that group. f_{mod}, which denotes the center frequency for modulation filters, was limited to 8, 16, 32, and 64 Hz to evaluate slow speech-related envelopes in speech. The 128-Hz modulation filter was excluded from neural SI estimation because this frequency band largely represented temporal-fine-structure-based pitch coding, particularly for impaired AN fibers.

Similarly, noise-related distraction (ND_{neural}) was quantified by averaging $corr(SN, N)$ values across modulation bands for each group:

$$ND_{neural} = \frac{1}{4} \sum_{f_{mod}} \left[\frac{1}{K} \sum_{k=1}^{K} corr_k(SN, N) \right]^2 , \tag{5.3}$$

where K denotes the number of AN fibers in a group, $corr_k(SN, N)$ denotes $corr(SN, N)$ for k-th fiber in that group. f_{mod} was limited to the same four (8, 16, 32, and 64 Hz) frequency bands that were used to compute SI_{neural}.

5.3 Results

Fig 5.2 shows an example of PSTHs constructed using spike trains from one normal AN fiber (CF = 5.1 kHz, high SR) in response to speech-alone (S), noise-alone (N), and noisy-speech (SN). In this example, the noise type was SSN. SNR for noisy-speech was 0 dB.

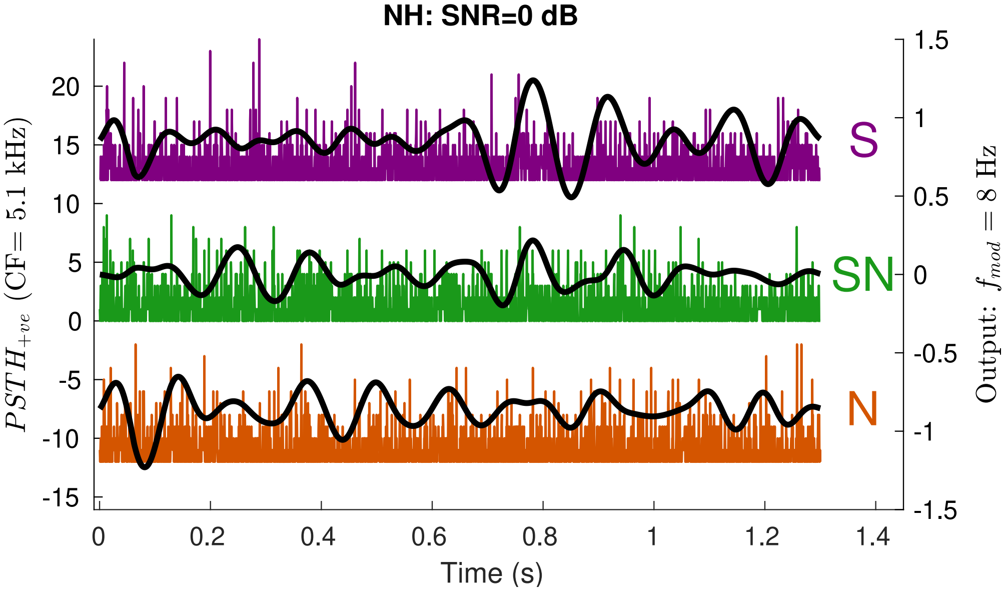

Figure 5.2. Exemplar PSTH (colored) and modulation filter output (black) for a normal fiber. PSTHs for a normal AN fiber (CF = 5.1 kHz, high SR) in response to speech-alone (purple), noisy-speech (green), and noise-alone (red) are plotted on the left Y-axis. Black traces (right Y-axis) represent output of processing PSTHs through a modulation filter (center frequency = 8 Hz, octave-wide).

Thick black lines overlaying PSTHs are the output of the modulation filter centered at 8 Hz. Qualitatively, noisy-speech output was well correlated with speech-alone output during certain segments of the stimulus. For example, response to a high-frequency fricative /s/ (near 0.8 sec) was still robust in this case for this high-frequency AN fiber.

Fig 5.3 shows data from two exemplar AN fibers (left, normal; right, impaired) for the SSN masker at 0 dB SNR. PSTHs are shown as colored lines. The response envelopes based on one modulation filter (center frequency = 8 Hz) are shown in black. Qualitatively, response envelopes were similar between S and SN for the normal AN fiber (e.g., near the fricative at 0.8 s). In contrast, response envelopes for N and SN were similar for the impaired AN fiber.

In order to quantify similarities between response envelopes, distributions of $corr(SN, S)$ and $corr(SN, N)$ were obtained for both groups in a multi-resolution framework as described in *Methods* (Sec 5.2.2). For each modulation filter (say center frequency = f_m), response envelopes were divided into segments of $3/f_m$ duration. For example, consider responses of a

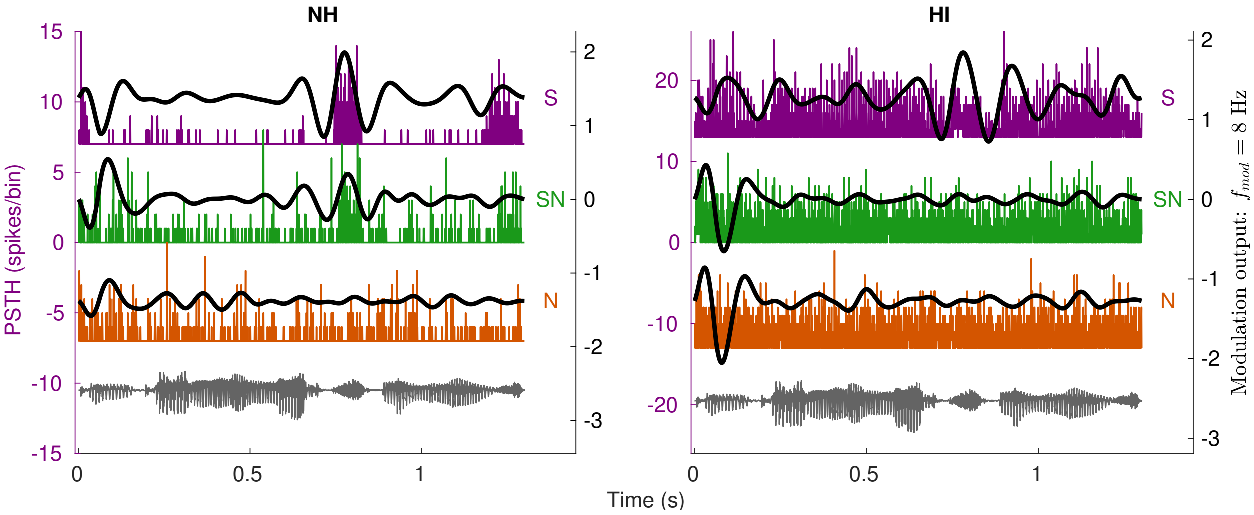

Figure 5.3. Example responses show more noise-resistant speech coding for a normal AN fiber compared to a HI AN fiber. Left and right panels show responses for a normal (CF = 6.4 kHz) AN fiber and an impaired (CF = 6.1 kHz) AN fiber, respectively. While the NH AN fiber primarily responds to the fricative portion (near 0.8s) of the speech stimulus, the HI AN fiber responds rather non-selectively to all portions of the stimuli. Qualitatively, SN response is more like N response for the impaired AN fiber.

single AN fiber. For a 8-Hz modulation filter, response envelopes were divided into segments of 375 ms. Correlation values were estimated for each segment, and then averaged across all segments. This procedure was repeated for all AN fibers to construct correlation distributions for each group. Fig 5.4 represents an exemplar scatter plot of these correlation values for the 8-Hz modulation filter at -5 dB SNR SSN condition for both groups. Distributions of correlation values for normal (blue) and impaired (red) AN fibers indicate a decrease in $corr(SN, S)$ and an increase in $corr(SN, N)$ for the HI group compared to NH group.

Fig 5.5 shows population statistics of d' for SSN for all SNR and modulation-filter conditions. Positive values on the y-axis indicate higher correlation for impaired AN fibers compared to normal AN fiber. d' for $corr(SN, S)$ (purple) was negative at lower ($\leq$ 32 Hz) modulation frequencies, which indicates degraded speech-coding fidelity in noisy-speech representations for the impaired fibers at those frequency bands. At 128 Hz (and to a lesser degree, at 64 Hz), the effect was opposite likely due to over-representation of the fundamental frequency as the impaired AN fibers responded to pitch energy via frequency-tuning-curve tail responses (Sec 4.3.4). More importantly, d' for $corr(SN, N)$ (green) was positive at all SNR and modulation-frequency conditions, suggesting a greater representation of noise in noisy-speech representations for impaired AN fibers.

Fig 5.6 shows responses from two exemplar AN fibers (left, normal; right, impaired) for speech-alone, noise-alone and noisy-speech at 0 dB SNR for FLN. Both fibers had high SR and similar CF. Qualitatively, envelope representation of the 8-Hz modulated noise was enhanced for the impaired AN fiber compared to the normal fiber in the response envelope for both N and SN. Moreover, there was substantial cycle-to-cycle variation in the response envelope of the normal AN fiber, likely due to its sharp frequency tuning. In contrast, the response envelope from cycle to cycle was less variable for the impaired AN fiber, likely due to its broader bandwidth (which makes response envelope similar to the broadband FLN envelope). These results were typical of normal and impaired AN fibers.

Fig 5.7 shows d' values for $corr(SN, S)$ and $corr(SN, N)$ for FLN. Similar to Fig 5.5, $corr(SN, S)$ was substantially affected at lower ($\leq$ 32 Hz) modulation frequencies for the impaired population (negative numbers indicate lower correlation values for impaired AN fibers). $corr(SN, N)$ was positive for all SNRs and all modulation frequencies. One striking difference between Fig 5.5 and Fig 5.7 is that the deleterious effect of FLN as quantified by d' for $corr(SN, N)$ was much stronger (worse) compared to SSN.

Finally, these neural correlation values were combined across modulation frequencies to estimate speech intelligibility (Figs 5.8A and 5.8B) and noise distractions (Figs 5.8C and

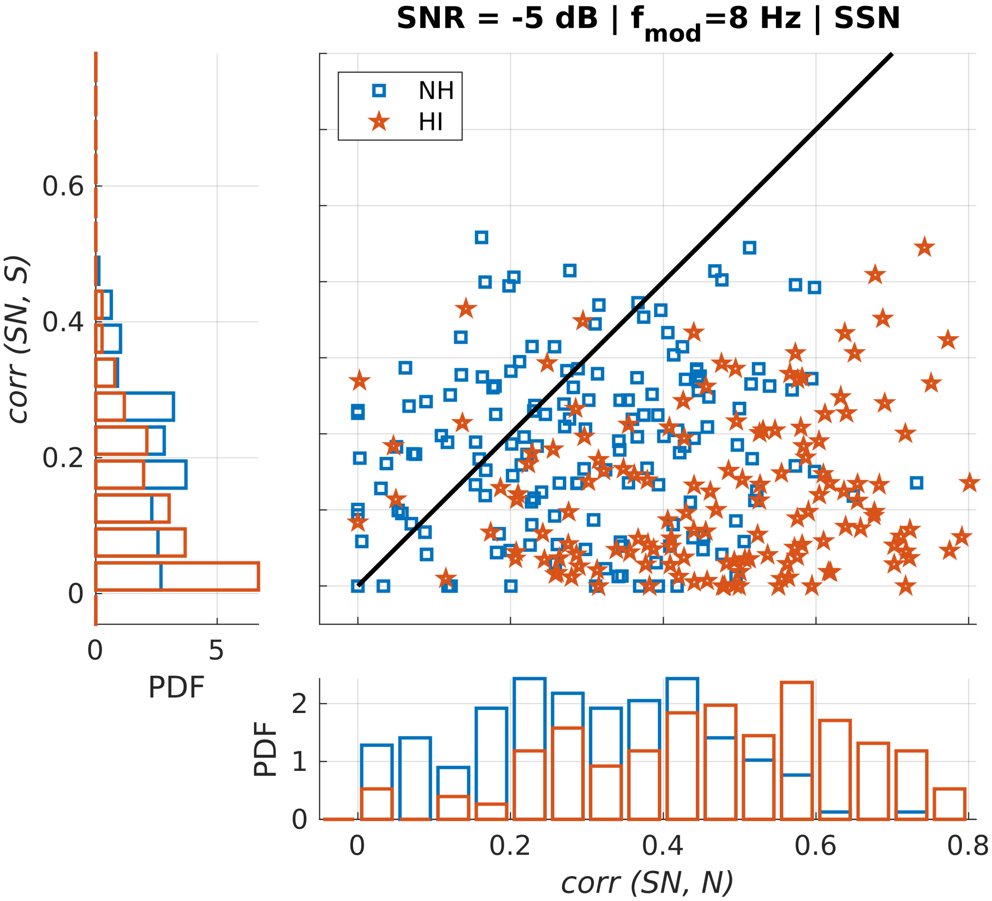

Figure 5.4. Example response-correlation distributions show that degradation in speech-coding fidelity and enhancement of noise representation in SN responses was more severe for impaired AN fibers compared to normal AN fibers. Distribution of $corr(SN, S)$ on *Y-axis* and $corr(SN, N)$ on *X-axis* for the 8-Hz modulation filter at -5 dB SNR for the SSN condition. Histograms show marginal Probability Density Functions for the NH (blue) and HI (red) groups. Unity slope line is shown in black. For impaired AN fibers, the distribution of $corr(SN, S)$ was shifted towards 0, indicating poorer speech representation in SN responses. In contrast, the distribution of $corr(SN, N)$ is shifted towards 1, which suggests noise-related distractions were enhanced in the peripheral representation following hearing loss.

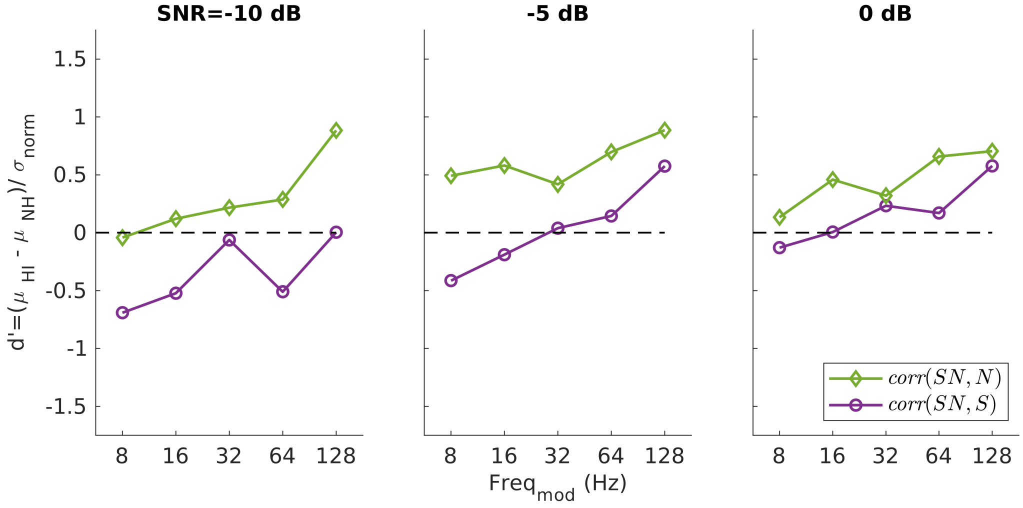

Figure 5.5. Background stationary noise was more detrimental for the impaired population compared to the normal population. d' plots for the stationary noise for all SNRs and modulation filters. Black line indicates d'=0. Note that positive values denote higher correlation for HI fibers.

5.8D) for each masker at different SNRs as described in Sec 5.2.2. Not surprisingly, neural speech intelligibility, as quantified by SI_{neural}, monotonically increased with SNR for both maskers for both groups (Figs 5.8A and 5.8B). While SI_{neural} was slightly greater for FLN than SSN for the HI population, it was substantially greater for the NH population. These neurophysiological results mirror psychoacoustic results from fluctuating masking release studies (e.g., Festen and Plomp, 1990). Furthermore, noise-related distractions, as quantified by ND_{neural}, decreased with increasing SNR for all but one conditions (except for the HI group in the presence of the stationary masker). The reasons for this anomaly remain unknown. ND_{neural} was greater for FLN than SSN for both groups. This increase was substantially greater for the HI population compared to the NH group, which highlights the potentially severe distracting nature of fluctuating background maskers. Note that SI_{neural} was an order of magnitude lower than ND_{neural}, which is expected because of the SNRs (≤ 0 dB) employed in this study. Overall, these results demonstrate the importance of considering both speech-coding fidelity as well as noise-related distractions to thoroughly characterize the effects of background maskers on speech coding, especially for hearing-impaired listeners.

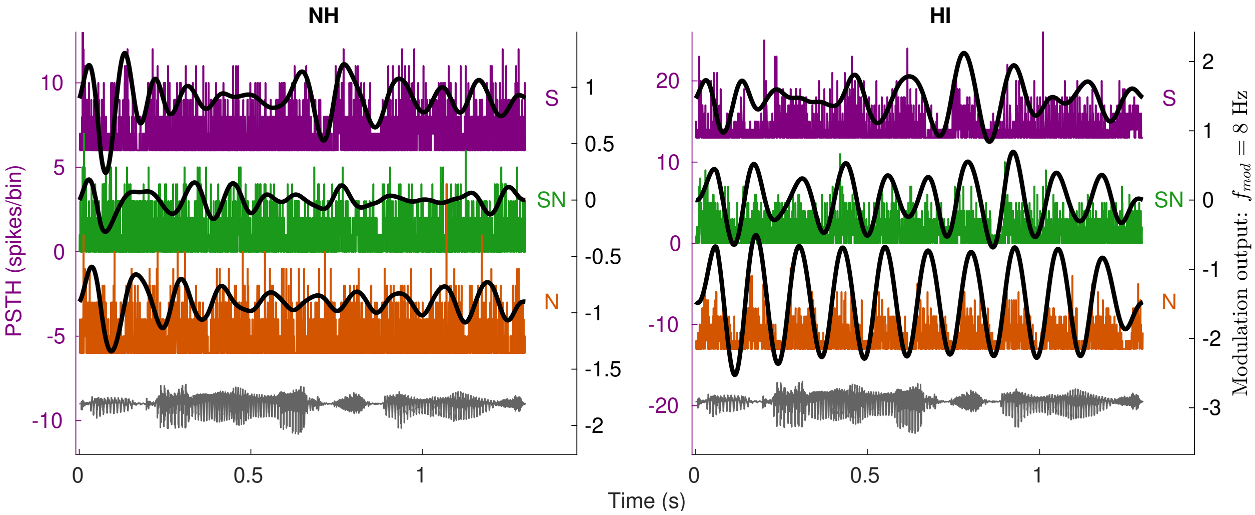

Figure 5.6. Example responses show that fluctuating masker severely affects HI SN responses. Same as Fig 5.3 but for a fluctuating noise condition. Notice similarity between envelopes of N and SN for the impaired AN fiber.

5.4 Discussion

5.4.1 The effect of background noise on speech-envelope-coding fidelity was more severe for the HI group

One of the main objectives of this study was to quantify the effects of background noise on speech-coding fidelity on the peripheral representation of natural speech. Speech-coding fidelity was quantified using the correlation between response envelopes for speech-alone and noisy-speech. As expected, speech-coding fidelity was degraded in noisy backgrounds (Fig 5.8). This degradation was more severe for the HI group compared to the NH group for both stationary (Fig 5.5) and fluctuating (Fig 5.7) maskers for most ($\leq$ 32 Hz) modulation bands. At higher (128 Hz) modulation bands, pitch frequency was overrepresented in the hearing-impaired population due to distorted tonotopy (e.g., Sec 4.3.4). Therefore, envelope-representation was enhanced for the HI group compared to the NH group for both maskers at this modulation frequency. However, this enhancement is minimally informative because of a lack of tonotopic coding. In fact, the degradation in speech-envelope coding in noise for the HI group reported here is likely an underestimate of the deficits underlying HI speech perception because the representation of speech-alone is already degraded to begin with

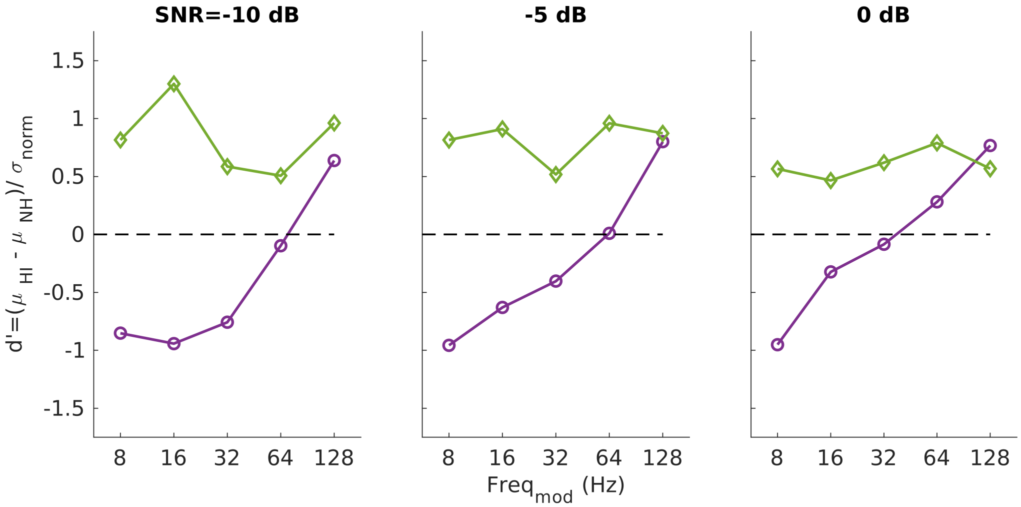

Figure 5.7. Compared to the stationary masker, the fluctuating masker had a more severe effect on SN responses for the impaired population. Same format as Fig 5.5 but for the fluctuating masker. d' for corr(S,SN) was reduced for these fluctuating-masker conditions than stationary-masker conditions. More importantly, d' corr(N,SN) was substantially enhanced. These results suggest greater speech-coding degradation and enhanced noise-related distraction for the HI population in the presence of fluctuating maskers.

because of distorted tonotopy. Overall, these results demonstrate the detrimental effects of background noise on speech envelope coding for the HI population.

5.4.2 Intrinsic fluctuations in noise had a greater detrimental effect on the HI representation

Noise-coding strength in the peripheral representation of noisy-speech was also evaluated to assess the potentially distracting nature of noise. In particular, noise-related distraction was quantified using the correlation between response envelopes for noise-alone and noisy-speech. Similar to degradation in speech coding, noise-related distraction worsened with decreasing SNR for all but one conditions tested (Fig 5.8). Interestingly, this noise overrepresentation was greater for the HI group compared to NH group for both stationary (Fig 5.5) and fluctuating (Fig 5.7) maskers. Overall, these results suggest that noise-related distraction, in addition to speech-coding degradation, contribute to perceptual difficulties that listeners with hearing impairment experience.

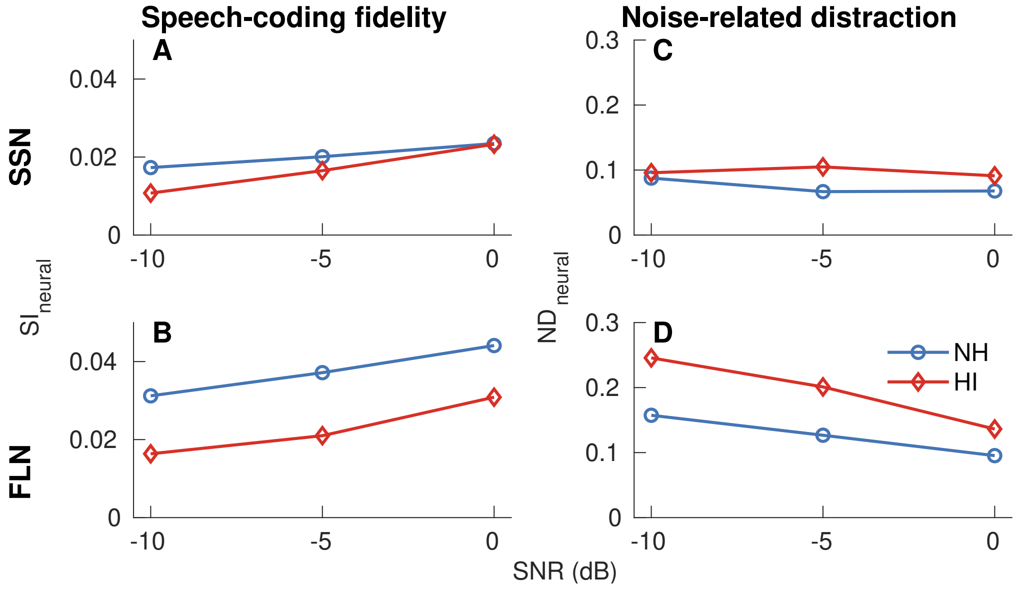

Figure 5.8. Neural estimates for speech intelligibility and noise-related distractions highlight the detrimental effects of noise for the HI population, especially for fluctuating maskers. (A-B) Speech-coding fidelity as quantified by *corr(S,SN)* (based on 4 to 64 Hz modulation frequencies) for the stationary (A) and fluctuating (B) maskers. (C-D) Similarly, noise-related distractions as quantified by *corr(N,SN)* (based on 4 to 64 Hz modulation frequencies) for the stationary (C) and fluctuating (D) maskers. Panels A and B are consistent with psychoacoustic results of why listeners with HI do not show fluctuating masking release (Festen and Plomp, 1990). Similarly, panels C and D highlight the enhanced masking effect of noise (especially of the fluctuating masker) for the HI population.

Distorted tonotopy and broader bandwidth likely contribute to the enhanced noise representation for the HI group. Normal AN fibers, due to their sharp frequency tuning, primarily respond to near-CF energy in the masker. Therefore, their response envelope reflects the spurious energy distribution in noise in a narrow frequency band. On the other hand, impaired AN fibers have a broader bandwidth due to distorted tonotopy. Therefore, responses of impaired AN fibers are dominated by noise energy over a broader (low-pass) band. As a result, normal AN fibers likely benefit from advantageous SNRs in the short-term due to the spectrotemporal sparsity of speech compared to noise (Cooke, 2006). In contrast, impaired AN fibers are potentially unable to make use of this sparsity due to broader filters. Note that a masker with speech-shaped spectrum is potentially more detrimental compared to a masker with white spectrum because of distorted tonotopy.

5.4.3 FLN had a more detrimental effect than SSN on the envelope representation of noisy-speech for the hearing-impaired group

Psychoacoustic studies have shown that normal-hearing listeners perform better when maskers with fluctuating envelope are used compared to when maskers with stationary envelope are used (Festen and Plomp, 1990). In contrast, listeners with hearing impairment do not show this benefit. Our neural data are consistent with these psychoacoustic results. Specifically, speech-coding strength, quantified by SI_{neural}, was greater in FLN conditions than in SSN conditions for the normal-hearing population (Figs 5.8A and 5.8B). In contrast, SI_{neural} values in the two masker conditions were similar for the HI group. The agreement between neural envelope-based speech representation and psychoacoustic studies support the importance of neural envelope for speech intelligibility, particularly for broadband noise maskers (Swaminathan, 2010).

In addition to considering speech-coding fidelity, it is also important to consider the effect of background noise as intrinsic fluctuations in the background noise can distract listeners from focusing on the target speech (Brungart, 2001; Takahashi and Bacon, 1992). In order to quantify noise-related distractions in the neural representation, response envelope correlation between noise-alone and noisy-speech was used to construct the final metric, ND_{neural}. ND_{neural} was greater in FLN conditions than for SSN conditions for both groups, suggesting that fluctuating noise can be more distracting than stationary noise (Figs 5.8C and 5.8D). More importantly, this increased noise representation in FLN conditions was substantially greater for the HI group. This enhancement of intrinsic noise modulations, particularly for fluctuating maskers, likely underlies perceptual difficulties in adverse environments that listeners with HI experience.

5.4.4 Implications for speech intelligibility modeling

A key motivation behind the present study was to understand how existing SI models can be extended to incorporate hearing-impaired data. While SI models based on data recorded from preclinical animal models of hearing loss is invaluable, it is not a convenient alternative to audio-domain SI models, which are easy to implement. To this end, results present in this study provide crucial insights to improve SI models for the HI population. First, our results support the inclusion of both speech-coding fidelity as well as noise-related distraction in SI models, as proposed by several modeling studies (Dubbelboer and Houtgast, 2008; Jørgensen and Dau, 2011). These are complementary factors that affect speech intelligibility. In particular, speech-coding fidelity relates to the available speech-related information in the neural representation. However, how well the central brain can use this information depends on the representation of intrinsic fluctuations in noise. These noise fluctuations are substantially enhanced in the HI representation, and therefore likely contribute to individual variability in SI in the HI population.

As mentioned earlier, correlation between response envelopes for speech-alone and noisy-speech for the HI group likely underestimates the actual deficits listeners with HI experience. That is because the neural representation of speech-alone is already degraded to begin with due to distorted tonotopic coding. As a result, the correlation between S and SN responses for the HI group is likely determined by low-frequency ($<$300 Hz) components in both responses, even for AN fibers with a high (up to 3 kHz) CF (Fig 4.4). Moreover, responses of HI neurons become significantly correlated across the cochlear axis (Heinz et al., 2010; Swaminathan and Heinz, 2011). Therefore, a metric like $corr(S, SN)$ does not account for this increased across-fiber dependence. These factors can be included in an SI model by modifying how response envelope correlations are computed. Specifically, for a given peripheral channel (e.g., an AN fiber) in the SI model, instead of using the response envelope for speech-alone as the template envelope (which is cross-correlated with the response envelope for noisy-speech), the bandlimited envelope of the speech stimulus can be used as the template. Speech can be bandlimited by band-pass filtering the stimulus in a band centered at the center frequency of the peripheral channel, and the filter's bandwidth can be the mean Q_{10} value for the NH population. This filtering approach restricts the contribution of low-frequency response components, which could otherwise contribute to misleadingly large $corr(S, SN)$ for the HI population.

6. REPRESENTATION OF VOICE PITCH IN DISCHARGE PATTERNS OF AUDITORY-NERVE FIBERS FOLLOWING NOISE-INDUCED HEARING LOSS

SUMMARY[1]

Pitch plays an important role in everyday communication, including speech and music perception. Broadly speaking, pitch is encoded in two neural mechanisms: (1) temporal and (2) temporal-place. The temporal mechanism is mediated by unresolved harmonics of the fundamental frequency (F_0) in a stimulus. The temporal-place mechanism is primarily mediated by resolved F_0 harmonics. Psychoacoustic studies have suggested that listeners with hearing impairment rely more on the temporal mechanism and less on the temporal-place mechanism to perceive pitch; however, these hypotheses have not been simultaneously tested neurophysiologically in animals with sensorineural hearing loss. To test these hypotheses, we recorded spike trains from single auditory-nerve fibers of anesthetized male chinchillas with either normal hearing or mild-moderate hearing loss due to acoustic trauma. A natural speech sentence was used as the stimulus. Temporal pitch coding was evaluated using correlograms and alternative-polarity peristimulus time histograms. Temporal-place pitch coding was evaluated using average-localized synchronized rate and cepstral analysis. Results showed an enhanced temporal representation for pitch for the hearing-impaired group, although this enhancement may be pathological because the representation was generally more complex (i.e., contained more frequency components). In contrast, temporal-place representation of pitch was degraded for the hearing-impaired group. Overall, these findings help to explain the neural basis underlying pitch perception deficits by listeners with hearing impairment for natural speech.

6.1 Introduction

Pitch plays an important role in speech and music perception in everyday life (Plack and Oxenham, 2005). For example, in most languages, pitch conveys prosodic informa-

[1] Manuscript *in prep*

tion as well as other non-linguistic information such as age, gender, and emotional state of the speaker (Moore and Carlyon, 2005). These cues are helpful in resolving individual sound sources in complex acoustic scenes (Bregman, 1994). While normal-hearing listeners perform seamlessly in extracting and using pitch cues, listeners with hearing impairment demonstrate poorer ability to use pitch information, even in quiet environments (Moore and Carlyon, 2005). The hypothesized neural correlates underlying poorer pitch perception following noise-induced hearing loss (NIHL) include poorer temporal coding, broader bandwidth, and reduced traveling-wave delay. However, these hypotheses have not been tested neurophysiologically. That is because, with the exception of one study (Kale et al., 2013), most neurophysiological pitch-coding studies have been carried out with normal-hearing animals and not with hearing-impaired animals (Cariani and Delgutte, 1996a, 1996b; Cedolin and Delgutte, 2005, 2010; Larsen et al., 2008; Miller and Sachs, 1984).

Broadly speaking, there are two leading theories of pitch coding: (1) temporal and (2) temporal-place (also called spectral) (De Cheveigne, 2005). The waveform of a stimulus with harmonic spectrum (Fig 6.1B), e.g., speech, is periodic in the time domain (Fig 6.1A). This periodicity is also preserved after peripheral filtering (Fig 6.1C) and shows up at the first few harmonics of the fundamental frequency, F_0 (Fig 6.1D). The temporal theory of pitch coding relies on this temporal periodicity for estimating pitch (green arrow). In contrast, the temporal-place theory assumes that individual F_0 harmonics are encoded by auditory nerve (AN) fibers with CF near that harmonic (purple arrow). The central processor is assumed to estimate pitch by integrating this temporal-place information across AN fibers with different characteristic frequencies (CFs). While one neurophysiological study has looked at the temporal representation of pitch following NIHL (Kale et al., 2013), no published study has investigated the temporal-place representation of pitch following NIHL. Moreover, pitch coding of natural speech following NIHL has not been explored.

Here, we recorded spike trains in response to a natural speech sentence from single AN fibers of anesthetized chinchillas with either normal hearing or NIHL. Alternative-polarity PSTHs and correlograms were used to evaluate temporal or temporal-place pitch-coding theories (Miller and Sachs, 1984; Parida et al., 2020). Temporal-place coding representation was evaluated using average-localized synchronized rate (ALSR) and cepstral analysis (Miller

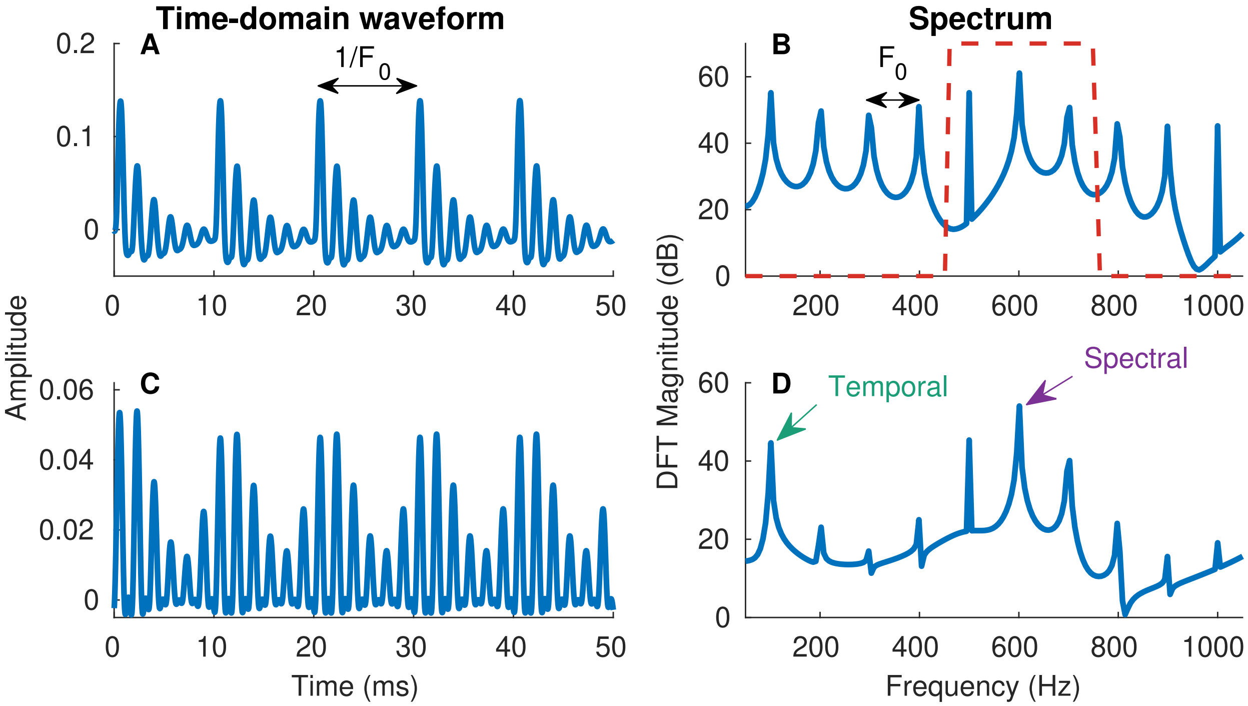

Figure 6.1. Illustration of temporal and temporal-place theories of pitch coding. (A) Time-domain waveform of a vowel (ʌ as in g<u>u</u>t). The waveform is periodic with pitch period ($1/F_0$). (B) DFT of the vowel used in A. The DFT has energy at multiples of F_0. (C) Time-domain response of a toy AN-fiber model. This toy model consists of a band-pass filter (dashed red line in panel B), half-wave rectification, and low-pass (fourth order, zero phase, IIR, cut-off = 1.5 kHz) filter. (D) Response spectrum for the toy model. The output of the toy model shows pitch periodicity (panel C), which also shows up as energy at F_0 in the spectral estimate (panel D, green arrow). This temporal periodicity is the basis underlying temporal pitch coding. In contrast, the output also shows F_0 harmonics that are within the bandwidth of the filter (in panel C). These spectral harmonics (purple arrow) in the response contribute to the temporal-place representation of pitch.

and Sachs, 1984). Our results show an enhanced temporal representation for the hearing-impaired group. However, this enhancement could be detrimental for pitch perception as this enhancement often led to more complex response waveforms (i.e., containing more frequency components), as hypothesized by previous studies (Rosen and Fourcin, 1986; Rosen, 1986). In contrast, the temporal-place representation was significantly degraded following NIHL. This degradation resulted primarily due to distorted tonotopic coding of individual F_0 harmonics so that fewer (low-frequency) harmonics were encoded in the hearing-impaired representation. These results provide new neural insights into pitch perception difficulties that listeners with hearing-impairment experience.

6.2 Materials and methods

The same speech stimulus and data were used as described in Chapter 2. Shuffled autocorrelograms were estimated and normalized using standard methods (Louage et al., 2004). SAC bin width was 100 μs.

Average-localized synchronized rate and cepstral analysis

ALSR was constructed using standard methods that are previously published (Miller and Sachs, 1984; Young and Sachs, 1979) except one difference. Instead of using the single-polarity PSTH, the difference PSTH ($d[n]$) was used to estimate ALSR because $d[n]$ minimizes rectifier-distortion artifacts (Parida et al., 2020). Briefly, Discrete Fourier Transform ($D_x[f]$) of $d_x[n]$ was estimated for each AN fiber (x).

$$D_x(f) = DFT(d_x[n]). \tag{6.1}$$

Next, a width (w) parameter was chosen to construct the ALSR. ALSR value at frequency f was estimated as the average DFT magnitude at frequency f for AN fibers with CF within w distance (log-scale) from f. ALSR bandwidth was limited to frequencies below 5 kHz.

$$ALSR[f] = \frac{\sum_{dist.[CF(x),f]<w} |D_x(f)|}{N[f]}, \tag{6.2}$$

where $dist.[\,*\,]$ denotes distance operator (in log-scale), $N[f]$ denotes the number of AN fibers that satisfy $dist.[CF(x), f] < w$, and $|\,*\,|$ denotes the magnitude operator.

The ALSR cepstrum was estimated as the inverse DFT (IDFT) of the logarithm of ALSR. As the ALSR only keeps the amplitude and disregards phase information, the ALSR cepstrum is real.

$$Cepstrum[q] = IDFT(ALSR[f]), \tag{6.3}$$

where q is the independent variable quefrency (in second). The cepstrum deconvolves the fast pitch-related harmonic component of a spectrum from the slow spectral envelope so that pitch-related information maps on to quefrencies at the pitch period and its harmonics. These methods have previously been applied to spike-train data from AN fibers for pitch estimation (Miller and Sachs, 1984).

6.3 Results

6.3.1 Temporal representation of voice pitch is enhanced following NIHL

First, the temporal representation of voice pitch in the responses of individual AN fibers was evaluated. For a stimulus with harmonic structure (fundamental frequency = F_0), AN-fiber responses can show energy at the first few harmonics of F_0. Two key sources contribute to energy at these low F_0 harmonics. First, responses of AN fibers are rectified; therefore, when a neuron responds to multiple F_0 harmonics, the PSTH is periodic with $1/F_0$ period (Young and Sachs, 1979). Second, AN fibers can respond to F_0 and its first few harmonics directly via frequency-tuning-curve tail responses (Kiang and Moxon, 1974).

Fig 6.2A shows responses from a normal and an impaired AN fiber for a 80-ms stimulus segment. The impaired AN fiber clearly followed the voice pitch. In contrast, the normal AN fibers showed fast spiking activity, which suggests that it primarily responded to high-frequency energy in the stimulus, and to a lesser extent, to F_0. These observations are also supported by PSTH spectra (Fig 6.2C). To quantify the total F_0-related power in the response, multitaper (two tapers) spectrum was estimated for both positive-polarity and negative-polarity PSTHs (i.e., PSTHs constructed using spike trains in response to positive and negative polarities of the stimulus, respectively). For each AN fiber, fractional power (i.e., band power normalized by total power) in 30-Hz bands centered at the first three F_0 harmonics was computed using the multitaper spectrum. For each AN fiber, the average of F_0 power estimates for both polarities was used as the final estimate. Results showed a significant increase in F_0 power for the impaired population ($p<.001$; unpaired *t-test*) for this example segment (Fig 6.2D). Repeating this analysis with the first two harmonics did not affect these results in any significant way.

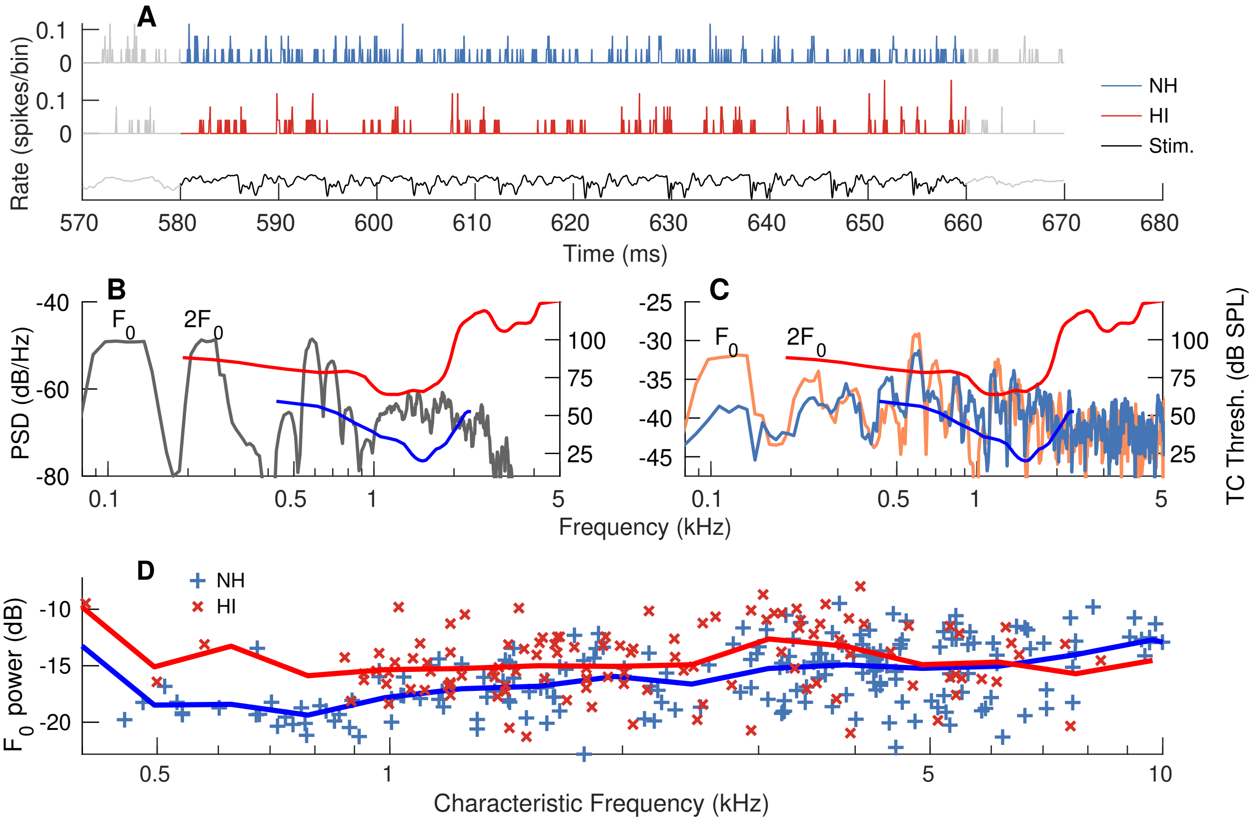

Figure 6.2. Temporal representation of F_0 was enhanced following NIHL for an example segment. (A) Positive-polarity PSTHs for a normal (blue) and an impaired (red) AN fiber in response to a quasi-stationary voiced segment (black). (B) Spectral estimate of the stimulus segment. Frequency tuning curves are also shown (right y-axis) for normal (blue) and impaired (red) AN fibers. (C) Multitaper (two tapers) spectral estimate for the PSTHs in A. While the normal fiber substantially responds primarily to near-CF energy, the impaired fiber responds to first few F_0 harmonics. (D) Fractional power metric used to quantify pitch coding strength for the segment in A for normal (blue) and impaired (red) AN fibers. Fractional power in three 30-Hz bands centered at the first three F_0 harmonics was used to quantify pitch-coding strength. Markers represent individual AN fiber data. Thick lines indicate two-third octave averages along CF. F_0 coding was significantly enhanced in the CF range 0.4 to 5 kHz ($p < .0001$; unpaired *t-test*) for the impaired group.

To quantify the strength of F_0 coding over the whole stimulus duration, fractional power was estimated in a 20-Hz band along the first three harmonics of the time-varying F_0 trajectory using frequency demodulation (Parida et al., 2020). This approach improves the spectral resolution of analysis. The fractional F_0 power was significantly (p<0.001; 0.4 kHz

< CF < 5 kHz; unpaired *t-test*) enhanced for the hearing-impaired population relative to the normal-hearing population (Fig 6.3).

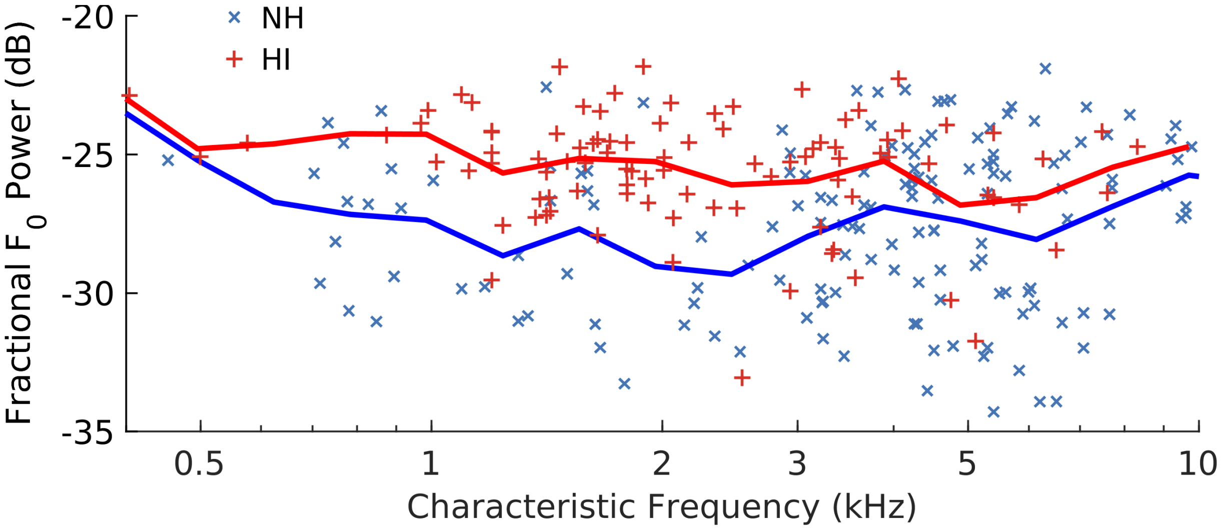

Figure 6.3. Temporal F_0 representation was significantly enhanced following NIHL (p<0.001; 0.4 kHz < CF < 5 kHz; unpaired *t-test*). Fractional power (used to quantify temporal coding strength of pitch) was estimated along 20-Hz bands centered at the first three time-varying F_0 harmonics using the harmonicgram (see Section 6.2). Same format as Fig 6.2D.

Although spectral analysis can provide power estimates in the desired bands, the spectral resolution of analysis becomes poorer at shorter duration (e.g., 40 ms). However, F_0 can significantly change over such a short duration, which the auditory system can follow (Miller and Sachs, 1984). Therefore, to estimate time-varying pitch and the strength of pitch coding in the temporal representation of AN-fiber responses, we used the pooled shuffled autocorrelogram approach (*SAC*, see Materials and Methods). Briefly, response was divided into 40-ms windows with 50% overlap. In each window, SAC was calculated for individual AN fibers. Finally, SACs were pooled across all fibers for each group to estimate the pooled SAC for that group (Figs 6.4A and 6.4B). For each window, delay corresponding to the maximum of the pooled SAC was used as the pitch period estimate. The value of the maximum was used as the strength of pitch coding. Search of SAC maximum was restricted to between 6 and 11 ms, which covers the dynamic range of the pitch period of the speech stimulus used. Ground truth F_0 trajectory was estimated using Praat (Boersma, 2001). Neural pitch estimation was robust for both normal (mean absolute deviation or MAD = 3.3 Hz) and impaired (MAD =

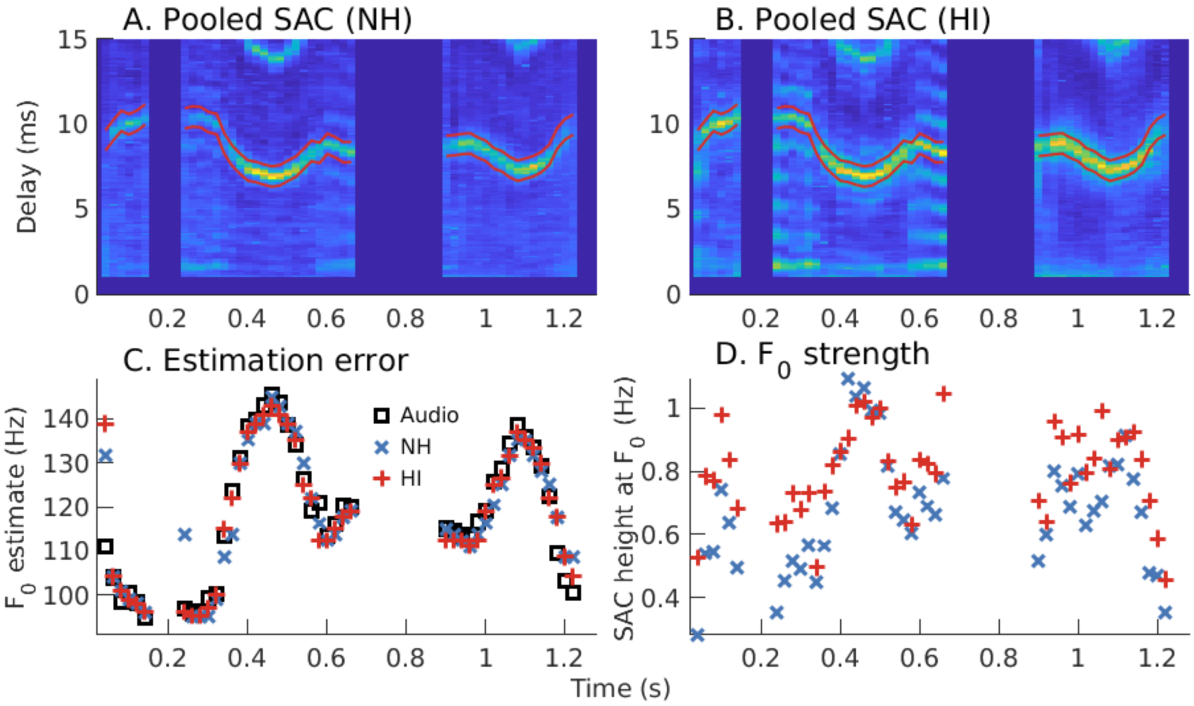

Figure 6.4. Temporal coding of F_0 was enhanced following NIHL. (A-B) Pooled SAC for normal (A) and impaired (B) groups. Ground truth pitch period, estimated using Praat, is highlighted within red boundaries. (C) Delay corresponding to SAC maximum was used as the F_0 estimate (inverse of the delay) for both groups. Both normal (blue crosses) and impaired (red pluses) groups showed robust pitch tracking. (D) Temporal pitch-coding strength, quantified using SAC-peak amplitude near $1/F_0$, was significantly enhanced ($p<.001$, unpaired *t-test*) for the impaired group.

2.5 Hz) groups (Fig 6.4A). In fact, pitch coding strength was significantly ($p<.001$; paired *t-test*) enhanced for the impaired group relative to the normal group (Fig 6.4D). Overall, these results show an enhanced temporal representation of voice pitch following NIHL.

An enhanced temporal representation of F_0 may result from broader filters as well as from distorted tonotopy. This enhancement may be detrimental to pitch perception because the output of auditory filters may result in complex patterns (Rosen and Fourcin, 1986). To test this hypothesis, we evaluated temporal coding for two 100-ms segments with different spectral properties (Figs 6.5A and 6.5B). The windows were divided into 40-ms segments with 50% overlap. To construct a single SAC for the 100-ms segment, SAC delay was normalized by pitch period and resulting SACs were averaged across windows (Figs 6.5C and 6.5D).

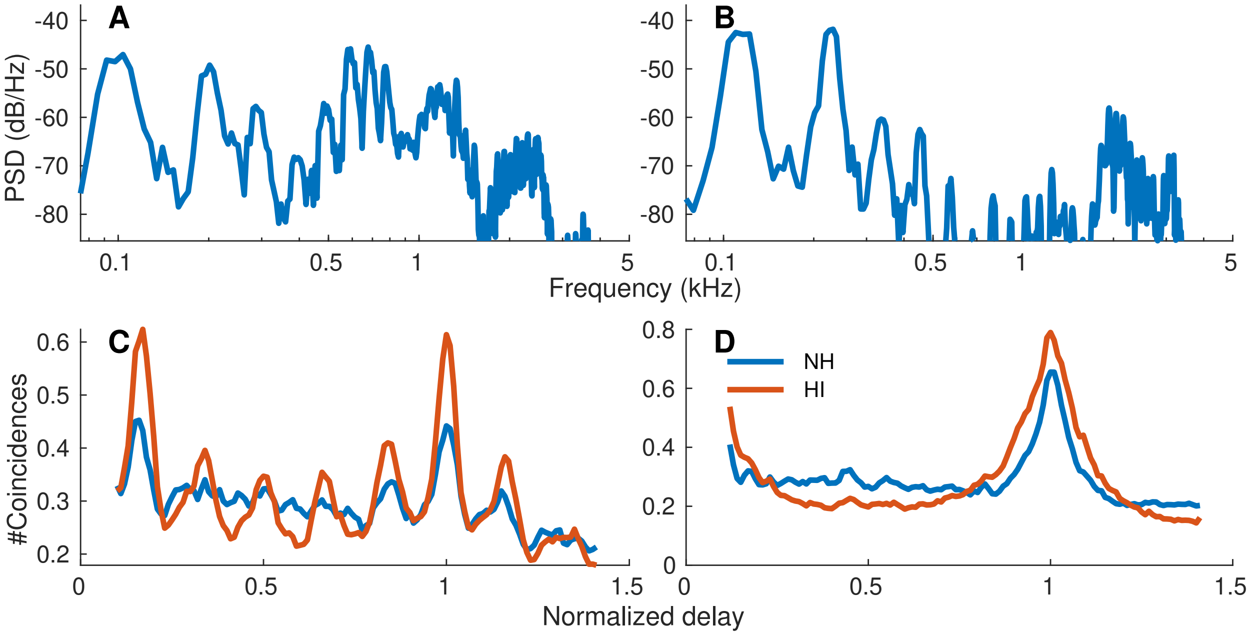

Figure 6.5. Distorted tonotopy can lead to increased SAC complexity. (A-B) Two exemplar segments with relatively flat (A) and severely sloping (B) spectrum. Both segments were 100-ms long. (C) Delay-normalized average SACs for the segment in A for both groups (see text). SAC complexity was increased, and pitch strength was enhanced for the impaired group for this segment due to distorted tonotopy. (D) Delay-normalized average SACs for the segment in B. F_0 coding strength was enhanced for the impaired group.

Qualitatively, for the segment with relatively flat spectrum (Fig 6.5A), normalized-SAC complexity for the impaired group was higher than that for the normal group. These results suggest distorted tonotopy plays a major role because impaired AN fibers tuned to higher (>1 kHz) CFs respond to multiple low-frequency (<1 kHz) harmonics. In contrast, for a segment with severely sloping spectrum, energy at F_0 drives tail responses of AN fibers in both groups, and as such, SAC for both groups showed a clear peak at F_0. However, the peak for the impaired group was larger likely because of distorted tonotopic effects.

6.3.2 Temporal-place (spectral) representation of voice pitch is degraded following NIHL

In order to evaluate the place-temporal representation of voice pitch, ALSR was constructed for both normal and impaired groups. For a given stimulus frequency, the ALSR representation only includes responses of AN fibers with CF near that frequency; thus, the ALSR considers both the place of AN fibers as well as their temporal response. Fig 6.6 shows the ALSR for an example stationary vowel (ʌ). The ALSR was constructed using spike-train data from 44 AN fibers from one chinchilla. To estimate the synchronized rate at a frequency, AN fibers with CF within one octave of that frequency were included. Individual F_0 harmonics, including the first two formants (F_1 and F_2), were clearly encoded in the ALSR (Fig 6.6B). The cepstrum of the ALSR shows a clear peak near the pitch period, i.e., at 10 ms (Fig 6.6C), as well as its second harmonic.

Although natural speech is nonstationary, it can be considered quasi-stationary at short (<50 ms) duration. Fig 6.7A shows the spectral estimate of one such 40-ms segment. ALSR

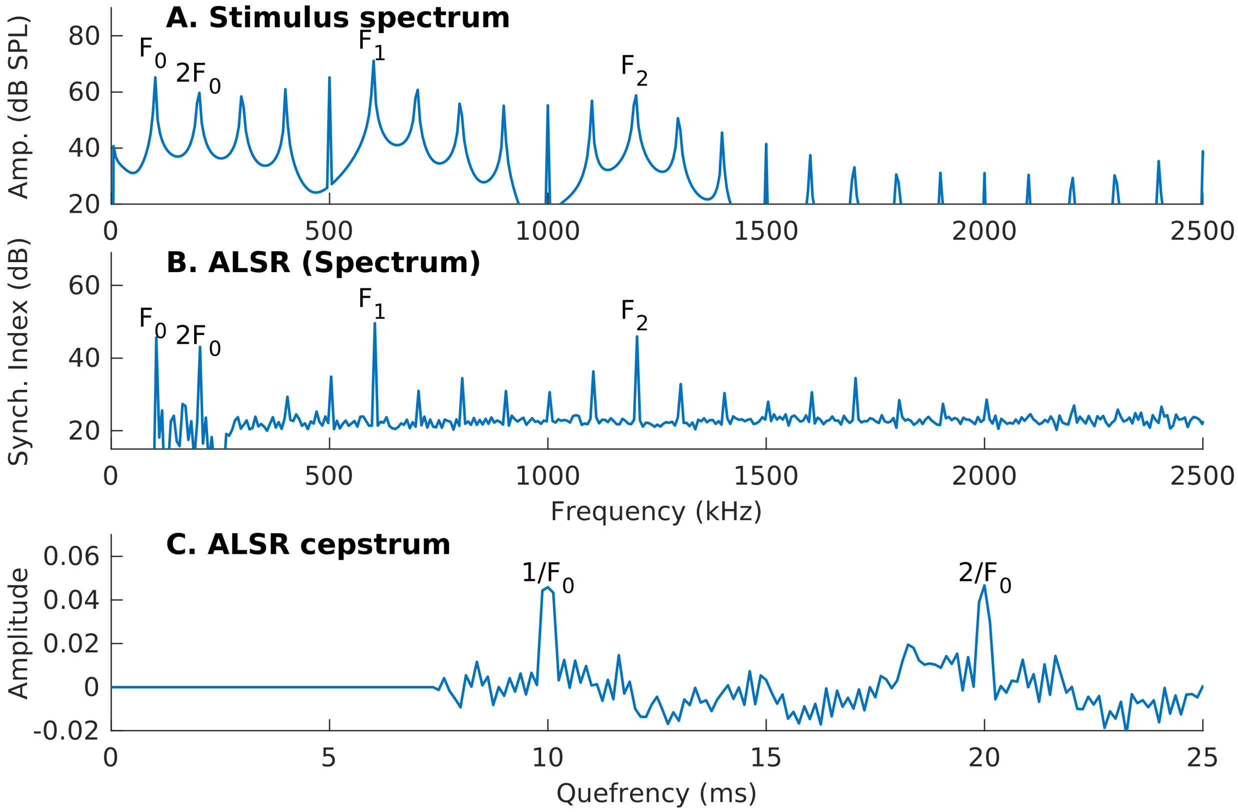

Figure 6.6. ALSR and cepstral analysis can be used to estimate F_0 from *apPSTH* data. (A) The DFT of the stationary vowel that was used as stimulus. (B) ALSR was constructed based on spike-train data recorded from AN fibers (N = 44). (C) Cepstrum of the ALSR in B shows a peak at the pitch period (10 ms). Cepstrum for quefrency below 7 ms is set to zero to highlight pitch-related components.

for the normal-hearing group shows robust representation of multiple F_0 harmonics, even up to 2 kHz (Fig 6.7B). In contrast, ALSR for the hearing-impaired group only captures the first few F_0 harmonics. As a result, cepstrum for the hearing-impaired group shows a weaker F_0 representation compared to the cepstrum for the normal-hearing group (Fig 6.7C).

Next, we validated whether ALSR and cepstral analysis can be employed to track the time-varying F_0 trajectory of the speech stimulus. ALSR was computed for 40-ms windows (with 50% overlap) for both normal and impaired groups. This analysis was restricted to only voiced portions of the stimulus. Three sets of ALSRs were constructed by varying the averaging CF width. For each ALSR, F_0 was estimated as the inverse of the quefrency corresponding to the maximum of the cepstrum. Quefrency range was restricted to above 6

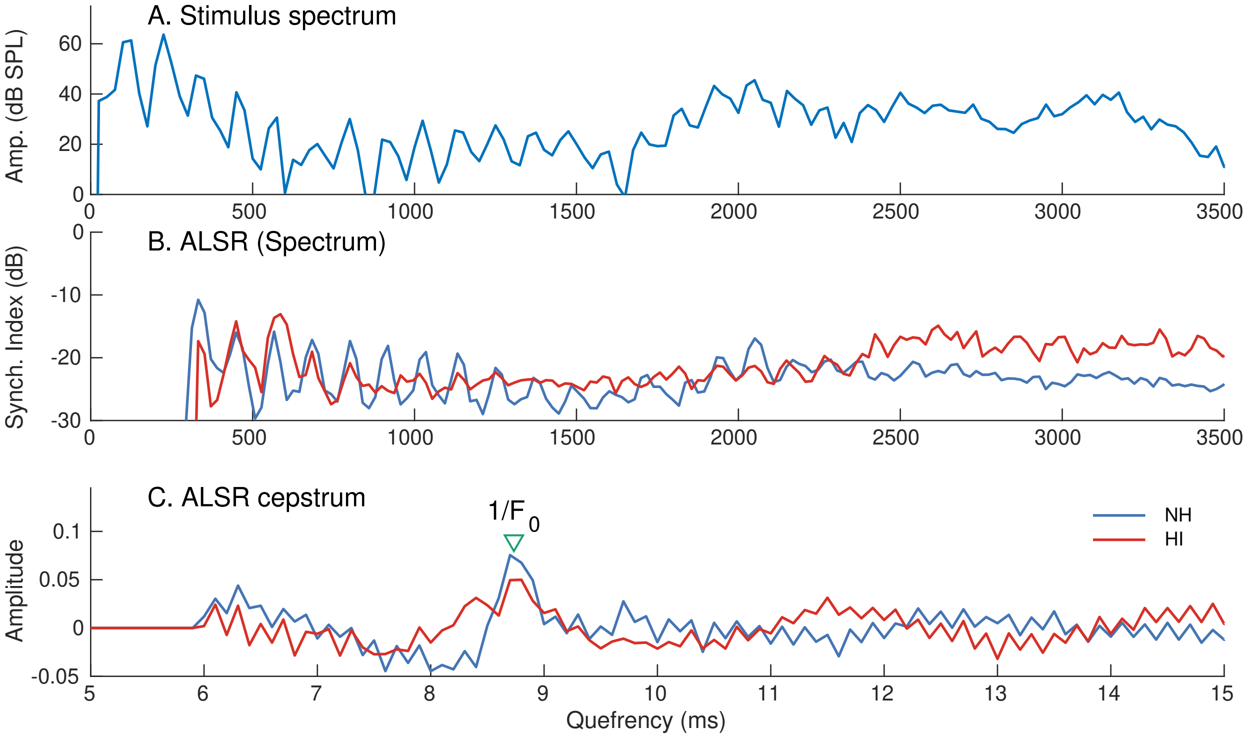

Figure 6.7. Temporal-place representation of pitch was degraded following NIHL for an example segment. (A) The DFT of an exemplar 40-ms long stimulus segment. (B) ALSR constructed using data from normal (blue) and impaired (red) AN fibers. ALSR for normal AN fibers showed response to multiple F_0 harmonics, whereas the ALSR for the impaired group responded to fewer F_0 harmonics. (C) Cepstrum for the normal group showed a peak at the expected pitch period (green triangle). In contrast, cepstrum for the impaired group showed a weak peak.

Table 6.1. The normal-hearing group had fewer errors and more robust pitch tracking compared to the hearing-impaired group. #Errors denotes the number of segments where the neural F_0 estimate deviated from the ground truth F_0 by more than 1 Hz (out of 42 total segments) for each group. MAD represents the mean absolute deviation of error across all segments.

ALSR Width (in octave)	# Errors		MAD (in Hz)	
	NH	HI	NH	HI
0.33	2	3	1.6	2.4
0.67	4	10	5.9	7.2
1.0	4	16	3.6	18

ms and below 11 ms to limit F_0 estimates within the dynamic range of the F_0 of the speech stimulus. As described previously, ground truth F_0 trajectory was estimated using Praat. For all averaging widths and most stimulus segments, neural F_0 estimates for the normal-hearing group robustly followed the ground truth F_0 trajectory (Fig 6.8). F_0 estimates for the impaired group also followed the ground truth F_0 trajectory, although there were more frequent errors compared to the normal-hearing group (Table 6.1). Similarly, the mean absolute deviation of the estimates (relative to ground truth) was higher for the hearing-impaired group relative to the normal-hearing group (Table 6.1).

Finally, we compared the strength of pitch coding between the two groups. Strength was quantified as the cepstrum amplitude of the local maximum near (within 0.5 quefrency) ground truth pitch period for individual segments. Cepstrum amplitude was significantly ($p<0.001$ for all widths; paired *t-test*) reduced in impaired fibers for all ALSR-averaging widths (Fig 6.9). Overall, these results show that temporal-place representation of voice pitch is substantially degraded following NIHL.

6.4 Discussion

The present study evaluated how neural coding of voice pitch for a natural speech sentence changes following NIHL. Two leading pitch-coding theories were considered, temporal and temporal-place (spectral). Temporal representation of F_0 carried sufficient information for

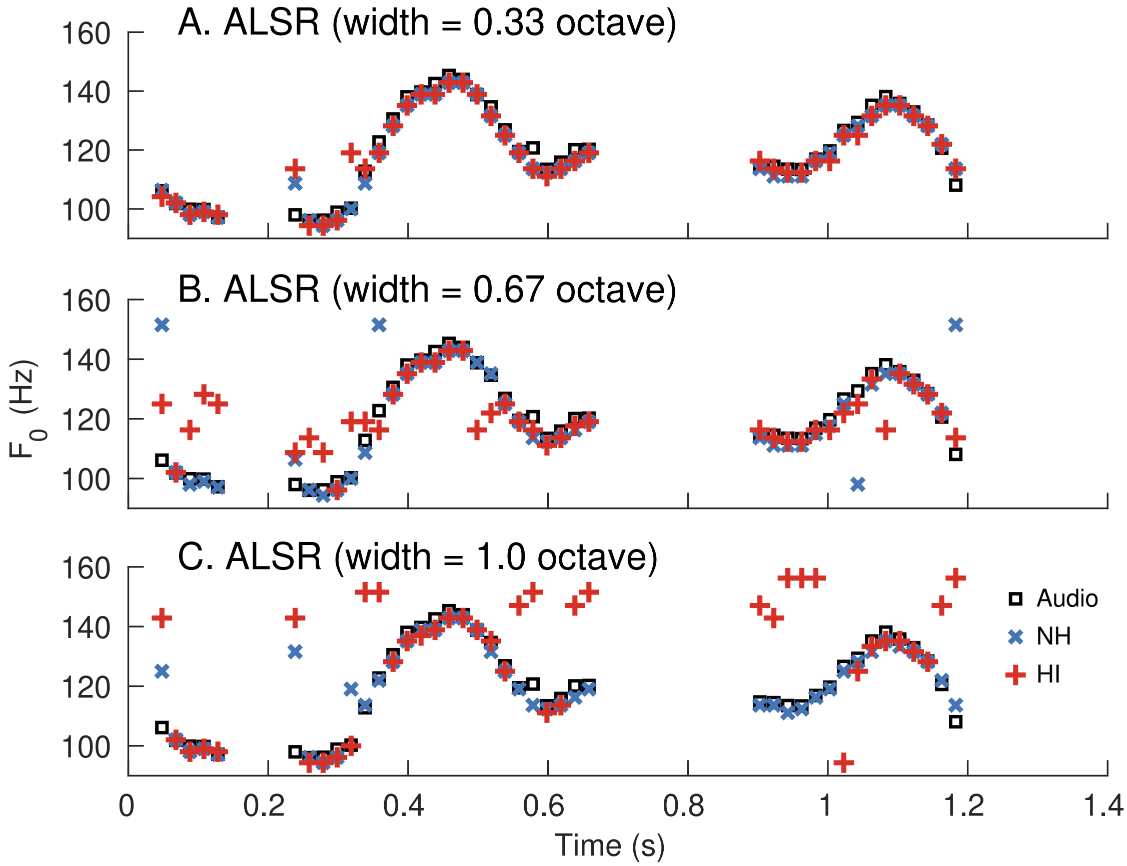

Figure 6.8. Cepstral analysis on ALSR was robust in predicting the time-varying F_0 trajectory. (A-C) Neural F_0 estimates using cepstrum of the ALSR constructed using spike-train from normal (blue crosses) and impaired (red pluses) AN fibers. ALSRs were constructed for 40-ms moving windows (50% overlap). Panels correspond to different CF widths for constructing ALSR. For each cepstrum, quefrency corresponding to the maximum below 11 ms was used as the pitch period estimate. Black squares in all panels represent F_0 (ground truth) of the audio signal estimated using Praat. F_0 estimates were robust for most segments for all CF widths. However, estimates deviated from the ground truth for a few segments. These deviations were more frequent for the hearing-impaired group compared to the normal-hearing group.

accurate F_0 tracking for both groups (Fig 6.4). However, F_0-coding strength was enhanced following NIHL as quantified by both harmonicgram (Fig 6.3) and shuffled autocorrelogram (Fig 6.4) analyses. This enhancement was often accompanied by increased SAC complexity (Fig 6.5).

Temporal-place pitch-coding theory was evaluated using ALSR and cepstral analysis. Temporal-place F_0 representation also conveyed sufficient information to track time-varying

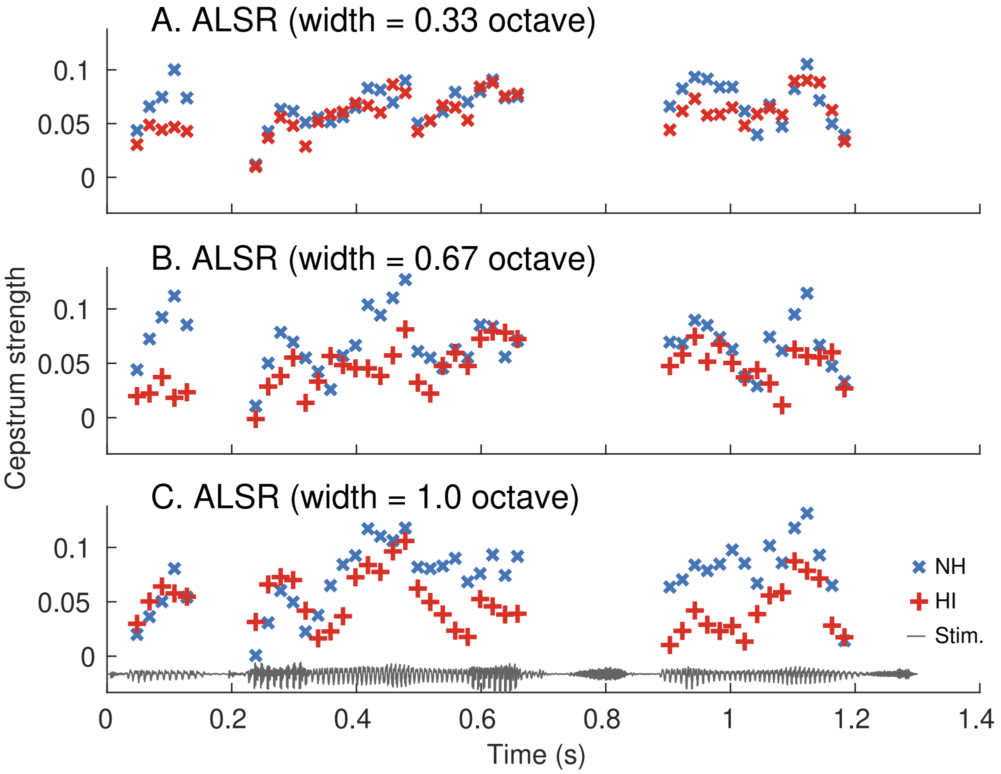

Figure 6.9. Temporal-place representation of pitch was degraded following NIHL. Panels correspond to different CF widths used to construct ALSR. For most CF-widths and segments, cepstrum amplitude was smaller for impaired (red pluses) compared to the normal (blue crosses) group. These differences were significant (p<.001) for CF-width conditions.

F_0 for both groups, although there were more frequent errors for the impaired group (Table 6.1). The strength of temporal-place F_0 coding was significantly reduced for the impaired group (Fig 6.9).

6.4.1 Temporal pitch representation was enhanced following NIHL

Enhancement in the temporal representation of pitch found in this study is consistent with earlier reports of enhanced temporal envelope in animal models following NIHL (Henry et al., 2016; Kale and Heinz, 2010). Although broader bandwidth is likely to contribute, distorted tonotopy offers a better account of these pathological enhancements. For example, low-frequency stimulus energy can dominate the response of basal AN fibers. As a result, response complexity of these AN fibers increases. These neural findings are consistent with results from psychoacoustic studies. For example, adding lower harmonics to a narrowband harmonic complex can lead to poorer pitch discrimination ability for hearing-impaired listeners (Moore and Peters, 1992). Distorted tonotopic coding also leads to more complex responses; this increased complexity may lead to poorer F_0 estimates by the central processor (Rosen and Fourcin, 1986).

6.4.2 Temporal-place pitch representation was degraded following NIHL

Unlike the temporal representation, the temporal-place representation of voice pitch was severely degraded following NIHL. These results can be directly attributed to distorted tonotopy (Chapter 4). While normal AN fibers primarily respond to near-CF F_0 harmonics, impaired AN fibers over-respond to low-frequency F_0 harmonics. As a result, higher F_0 harmonics are not well captured at the peripheral level (Chapter 4, also see Fig 6.7). Therefore, pitch amplitude in the ALSR-cepstrum is significantly reduced for the impaired group.

These divergent neural effects regarding temporal and temporal-place pitch representations provide insights into pitch perception deficits experienced by listeners with NIHL. Temporal-place pitch representation requires phase locking to individual F_0 harmonics. In fact, lower resolved harmonics (better temporal-place representation) produce more salient pitch percept for normal-hearing listeners compared to higher unresolved harmonics (better temporal representation) (Carlyon and Shackleton, 1994; Plack and Oxenham, 2005). In contrast, listeners with hearing-impairment rely more on high-frequency unresolved harmonics (Moore and Carlyon, 2005). Results presented in this report provide several neural bases underlying these psychoacoustic results. Future studies exploring the role of across-fiber phase information and the effect of background noise on these pitch-coding theories will be informative to advance our understanding of pitch coding following NIHL.

7. NON-INVASIVE MEASURES OF DISTORTED TONOTOPIC SPEECH CODING FOLLOWING NOISE-INDUCED HEARING LOSS

SUMMARY[1]

Animal models of noise-induced hearing loss (NIHL) show a dramatic mismatch between cochlear characteristic frequency (CF, based on place of innervation) and the dominant response frequency in single auditory-nerve-fiber responses to broadband sounds (i.e., distorted tonotopy, DT). This noise-trauma effect is associated with decreased frequency-tuning-curve (FTC) tip-to-tail ratio, which results from decreased tip sensitivity and enhanced tail sensitivity. Notably, DT is more severe for noise-trauma than for metabolic (e.g., age-related) losses of comparable degree, suggesting individual differences in DT may contribute to speech-intelligibility differences in patients with similar audiograms. Although DT has implications for many neural-coding theories for real-world sounds, it has primarily been explored in single-neuron studies that are not viable with humans. Thus, there are no noninvasive measures to detect DT. Here, frequency following responses (FFRs) to a conversational speech sentence were recorded in anesthetized male chinchillas with either normal hearing or NIHL. Tonotopic sources of FFR envelope and temporal fine structure (TFS) were evaluated in normal-hearing chinchillas. Results suggest FFR-envelope primarily reflects activity from high-frequency neurons, whereas FFR-TFS receives broad tonotopic contributions. Representation of low- and high-frequency speech power in FFRs was also assessed. FFRs in hearing-impaired animals were dominated by low-frequency stimulus power, consistent with oversensitivity of high-frequency neurons to low-frequency power. These results suggest that DT can be diagnosed noninvasively. A normalized DT metric computed from speech FFRs provides a potential diagnostic tool to test for DT in humans. A sensitive noninvasive DT metric could be used to evaluate perceptual consequences of DT and to optimize hearing-aid amplification strategies to improve tonotopic coding for hearing-impaired listeners.

[1] This chapter has been published in JARO as Parida and Heinz, 2020

7.1 Introduction

Hearing-impaired listeners often have widely varying speech-perception abilities that cannot be explained by audiograms. Peripheral differences, such as hair-cell specific dysfunction due to mechanical and/or metabolic insults, may contribute to individual variability in speech intelligibility. One peripheral factor that has not received much attention, but is likely to be important for complex-sound perception, is distorted tonotopy (DT). Distorted tonotopy is the over-representation of low-frequency power by basal auditory-nerve fibers tuned to higher frequencies, which effectively reduces the number of independent cochlear information channels. DT accompanies desensitized tip and hypersensitive tail responses in the FTC of auditory-nerve fibers (Liberman and Dodds, 1984b). Effects of DT on suprathreshold coding have been studied with a synthesized vowel (Miller et al., 1997), and were recently generalized using broadband noise and system-identification approaches (Henry et al., 2016). These degradations are much more severe for NIHL than for metabolic-based (e.g., age-related) loss (Henry et al., 2019) of similar degree. These findings suggest that the degree of DT differs across hearing-loss etiologies, which likely contributes to individual variability in speech perception. Currently, there are no measures to assess DT and its consequences on speech coding in humans. Here, the viability of frequency following responses (FFRs) was explored as a diagnostic tool to detect DT noninvasively.

Scalp-recorded FFRs have been used frequently to quantify the strength of phase-locking, modulation coding, and vowel-formant coding in normal and hearing-impaired subjects to explore the neural bases for psychoacoustic results (Aiken and Picton, 2008; Krishnan et al., 2005; Shinn-Cunningham et al., 2013). Studies have also used FFRs in animal models of sensorineural hearing loss to understand subsequent coding deficits, as well as the proper interpretation of these far-field scalp recordings (Dolphin and Mountain, 1993; Parthasarathy et al., 2010; Zhong et al., 2014). Histological assessments in animals provide direct anatomical correlates for physiological and perceptual deficits of hearing impairment (Liberman and Dodds, 1984a, 1984b). Numerous studies have looked at human speech FFRs, but only a handful of studies have looked at speech FFRs in animals (Wible et al., 2005), where the large individual variability often seen in humans can be reduced by controlling the type and degree of cochlear insult and environmental factors.

FFRs to complex stimuli are often studied by decomposing responses into slow envelope (ENV) and rapid temporal-fine-structure (TFS) components via polarity inversion (Aiken and Picton, 2008; Ananthakrishnan et al., 2016). Although quantification of envelope- and TFS-based FFR coding is informative, caution should be exercised regarding the place specificity of these FFR components while interpreting results from narrowband and broadband stimuli (Gockel et al., 2015; Shinn-Cunningham et al., 2013). Usually envelope responses related to fundamental frequency come from high-frequency cochlear regions and their afferents in the brainstem and midbrain since unresolved harmonics contribute to the polarity-tolerant part of the response (Wang et al., 2019; Zhu et al., 2013). Moreover, the cochlear traveling wave enters the cochlear base and travels rapidly, resulting in good synchrony (and thus large evoked responses) across basal locations. More-apical activity is less synchronized because the traveling wave slows down near the place of cochlear resonance. These synchrony profiles across cochlear frequency locations are preserved in central nuclei, e.g., the inferior colliculus (Langner et al., 1987), which is presumably the dominant generator of the FFR; however, place specificity of low-frequency power in the FFR-TFS remains a matter of controversy (Shinn-Cunningham et al., 2013). While some studies suggest that FFR-TFS is place specific (e.g., Dolphin and Mountain, 1993; Shinn-Cunningham et al., 2013), others argue that FFR-TFS arises from summed activity of high-frequency neurons even at moderate sound levels due to upward spread of excitation (e.g., Bones and Plack, 2015; Dau, 2003; Gockel et al., 2015).

Here, in experiment I, the relative contribution from low- and high-frequency neurons to envelope and TFS of speech-driven FFRs recorded from normal-hearing chinchillas was investigated using a high-pass masking paradigm. Results showed that low-frequency (<500 Hz) power in FFR-envelope comes primarily from high-frequency (>600 Hz) neurons, whereas FFR-TFS receives contributions from both low- and high-frequency neurons. In experiment II, the relative contributions of low- and high-frequency power in speech to FFR envelope and TFS power were explored for normal and noise-exposed chinchillas. Data showed increased susceptibility of FFR-TFS to low-frequency stimulus power in noise-exposed animals. These results suggest that high-frequency neurons over-respond to low-frequency speech power following NIHL, consistent with distorted tonotopy.

7.2 Materials and methods

7.2.1 Animals

See Chapter 2.1.

7.2.2 Noise exposure

See Chapter 2.2.

7.2.3 Stimuli

One speech sentence (practice list #1, sentence #3) from the Danish speech intelligibility test material (CLUE, Nielsen and Dau, 2009) was used as the stimulus in both experiments. The sentence was shortened to 1.3 s duration by truncating it at silence intervals to allow more stimulus repetitions (Fig 2.1). In Experiment I, the speech sentence was presented at ∼70 dB SPL in the presence of high-pass filtered (600-Hz cut-off frequency) pink masking noise at five signal-to-noise ratios (SNRs: -20, -10, 0, 10 dB and without noise). In Experiment II, the speech sentence was presented at ∼70 dB SPL for both normal-hearing (NH) and hearing-impaired (HI) animals. The number of repetitions per polarity of the speech stimulus were 300 in Experiment I, and 500 in Experiment II.

7.2.4 Data recordings

A microphone-transducer pair (Etymotic ER-10B, Etymotic ER-2, Etymotic Research, Elk Grove Village, IL, USA) was used for calibration and sound presentation, with a foam ear tip inserted into the external ear canal. The transducer speaker was calibrated in-ear at the beginning of all experiments. Sound was delivered monaurally to the right ear. Calibration data showed that frequency responses were flat (within ± 7 dB) up to 10 kHz, and were similar across animals. Results are reported without accounting for calibration;

however, accounting for calibration did not affect the conclusions drawn here. Auditory brainstem responses (ABR) and distortion product otoacoustic emissions (DPOAE) were used as screening measures both before and after noise exposure (i.e., following the two-week wait period). ABRs were recorded to short tone bursts (5 ms with 0.5 ms rise/fall time) at .5, 1, 2, 4, and 8 kHz, from 0 to 80 dB SPL in 10-dB steps. Another intensity (odd multiple of 5 dB) near preliminary threshold estimate was used to fine-tune the threshold estimate. ABR threshold was estimated as the intensity at which linear regression of waveform correlation of ABRs at intensities below 60 dB SPL with a template ABR (response waveform at 60 dB SPL) was three standard deviations above the correlation noise floor. Distribution of the correlation noise floor was constructed from the spurious peaks in the cross-correlation function between the template and physiological noise in no-response regions of the recordings (Henry et al., 2011). DPOAE amplitudes (at $2f_1 - f_2$) were recorded in response to sets of two tones (f_1 and f_2), ranging in frequency from $f_2 = .5$ kHz to 12 kHz, with a frequency ratio of $f_2/f_1 = 1.2$, played at 75 (f_1) and 65 (f_2) dB SPL.

ABRs and FFRs were recorded using subdermal needle electrodes in a vertical montage (mastoid to vertex, differential mode, with common ground placed near the nose, Henry et al., 2011). Both ABR and FFR signals were analog filtered through ISO-80 (.005 to 10 kHz, x100 gain; World Precision Instruments, Sarasota, FL), and Dagan (open filter, x200 gain; 2400A, Minneapolis, MN) amplifiers. ABRs were further analog filtered in the band 0.3 to 3 kHz (Krohn-Hite, 3550, Brockton, MA). FFRs were digitally band-pass filtered offline between .03 and 1 kHz with a 4th-order zero phase IIR filter. Recordings with and without coupling between speaker and ear canal confirmed that FFRs were not corrupted by electrical artifact (see Fig 7.1).

7.2.5 Experimental design and statistical analysis

Envelope [$ENV_{FFR}(t)$] and TFS [$TFS_{FFR}(t)$] temporal responses were extracted from the FFR as follows. Consider two signals: (1) $FFR_{+ve}(t)$, response to positive polarity of the stimulus, and (2) $FFR_{-ve}(t)$, response to negative polarity of the stimulus. Then,

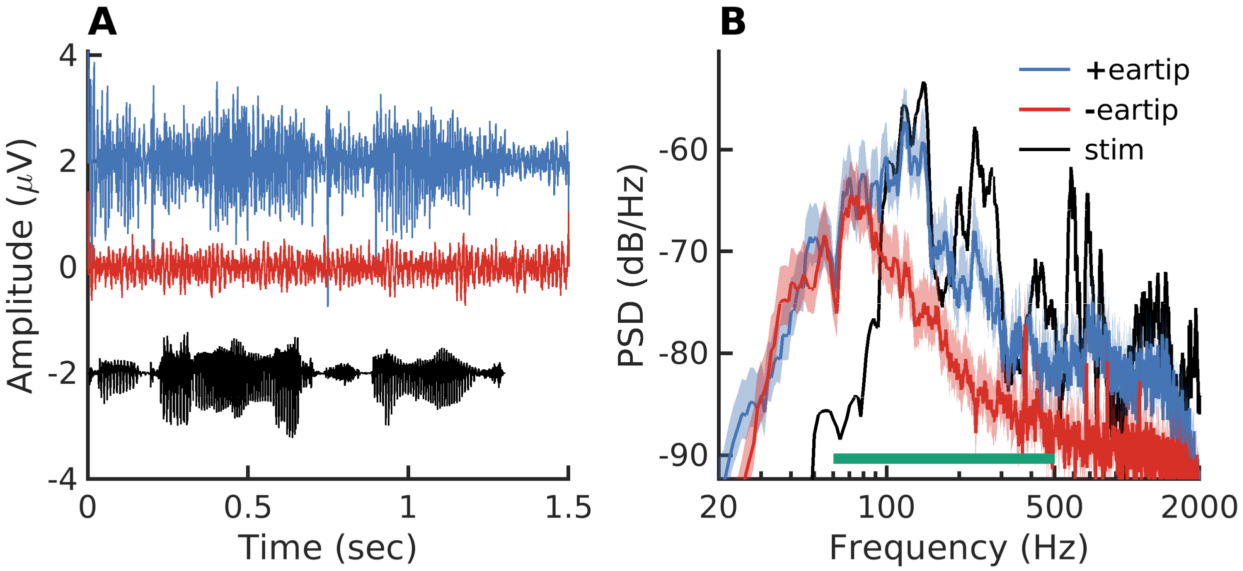

Figure 7.1. Speech-driven FFRs were well above the physiological noise floor for the frequency range of interest (60 to 500 Hz). FFRs with speaker ear tip inserted into the ear canal (blue, labeled +eartip, shifted up by 2μV for display), and when ear tip was not coupled to the ear but was inserted into a closed syringe in close proximity to the ear (red, labeled -eartip). FFRs are shown in response to positive onset polarity of the speech stimulus (black, labeled stim, downscaled and shifted down by 2μV for display). Results were similar for negative onset polarity of the stimulus (not shown). Data are shown for one normal-hearing animal, Q371. **A** Time-domain waveforms, and **B** FFR spectra. Spectral region of interest (60 to 500 Hz) is marked with a thick green horizontal line. Shading represents 95th percentile confidence interval. Spectrum and time-waveform examples of FFR envelope and TFS are shown in Figs. 7.3 and 7.6, respectively.

$$ENV_{FFR}(t) = \frac{FFR_{+ve}(t) + FFR_{-ve}(t)}{2}$$

and

$$TFS_{FFR}(t) = \frac{FFR_{+ve}(t) - FFR_{-ve}(t)}{2}$$

Prior to analysis, recorded FFRs were preprocessed to remove line harmonic components (Mitra and Pesaran, 1999). Signal spectrum was estimated using a multitaper approach (Thomson, 1982; Babadi and Brown, 2014). The spectra for $ENV_{FFR}(t)$ and $TFS_{FFR}(t)$

are referred to as $ENV_{FFR}(f)$ and $TFS_{FFR}(f)$, respectively. Data were analyzed using custom MATLAB (The MathWorks, Natick, MA) programs.

Linear regressions in Figs. 7.6 to 7.9 were performed in MATLAB. The hypotheses in Eqs. 7.2 to 7.4 were evaluated using a linear mixed-effect model in JMP14 (SAS Institute, Cary, NC). Either envelope (Eq. 7.2) or TFS (Eqs. 7.3 to 7.4) power in the FFR was the dependent variable. Both high- (0.5 to 5 kHz) and low-frequency (60 to 500 Hz) band stimulus power, hearing status, and their interactions were considered as fixed effects. Animal identifier was included in the mixed-model as a random effect. A p-value of 0.01 was used as the criterion for significance throughout.

7.3 Results

7.3.1 Mild-moderate noise-induced hearing loss model

Figure 7.2A shows pre- and post-noise exposure ABR thresholds for individual animals for the five frequencies tested. Post-exposure thresholds were elevated by 10 to 30 dB, consistent with mild-moderate hearing loss as per ASHA guidelines (Clark, 1981). DPOAE amplitude, plotted in Fig 7.2B, was similarly affected. Of the six hearing-impaired animals, one animal (Q369, light blue and orange in Fig 7.2) was not included in the group analysis (for hearing-impaired group), but was treated as a special case, because of its small threshold shift and abnormal changes in DPOAE amplitude profile following noise exposure.

7.3.2 Experiment I: Tonotopic sources of speech FFR in normal animals

Hypotheses. In order to evaluate the relative contributions of low-frequency (< 500 Hz) and high-frequency (> 600 Hz) neurons to speech-driven FFRs, running high-pass filtered pink noise was used as a masker at different SNRs. FFRs were collected from four normal-hearing chinchillas. As the FFR-passband was restricted to below 500 Hz, a cut-off frequency of 600 Hz was chosen for the pink masker. The reasoning was that if $TFS_{FFR}(t)$ is frequency specific, then power in $TFS_{FFR}(t)$ in the band below 500 Hz should be fairly resistant to masking by high-passed pink noise. Similarly, power in $ENV_{FFR}(t)$ in the band below 500 Hz was expected to be severely affected by the high-passed pink masker. The reasoning

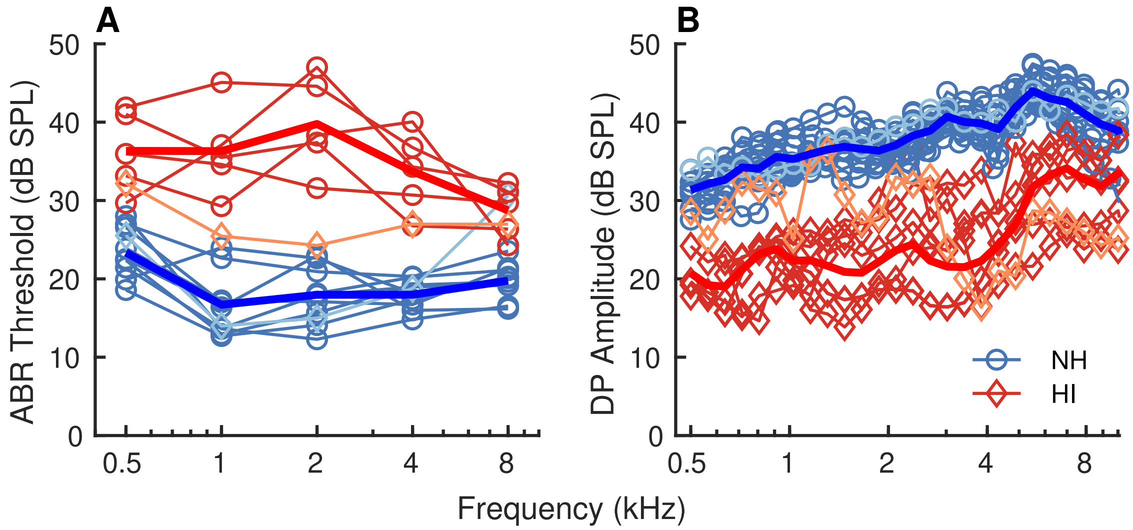

Figure 7.2. Screening confirmed mild-moderate hearing loss with outer-hair-cell dysfunction following noise exposure. **A** Auditory brainstem response (ABR) threshold to tone pips, and **B** distortion product (DP) otoacoustic emissions amplitude for two simultaneous tones, for normal-hearing (blue) and noise-exposed (red) chinchillas. Thin lines indicate individual animal data and thick lines show group averages. ABR thresholds for the noise-exposed group were elevated by up to 30 dB relative to the normal-hearing group. Post-exposure DP amplitudes were reduced by up to 20 dB, suggesting significant outer hair cell dysfunction. Animal Q369 (light blue/orange) is highlighted (and discussed in the text) as a special case for which post-exposure thresholds were still largely within the pre-exposure range.

was as follows. The phase-tolerant envelope component arises due to interaction between spectral components within the bandwidth of a filter (e.g., receptive field of a neuron). Since high-frequency neurons have larger bandwidths and show weak phase-locking to the carrier (i.e., weak phase-sensitive response component), they allow better envelope coding in their responses (Kale and Heinz, 2010). Thus energy in the response envelope is limited to frequencies less than the bandwidth of the filter even though the signal may not have physical energy in this frequency region. Furthermore, high-frequency neurons are better synchronized due to the compressed traveling-wave delay that boosts the envelope representation of these neurons in evoked responses, like the FFR.

Examples. Figure 7.3 shows example FFR spectra for 0 and -20 dB SNR conditions for one normal-hearing animal. Comparing the in-quiet response spectra to the speech (stimulus) spectrum confirms that the first few harmonics of the fundamental frequency were well preserved in both $ENV_{FFR}(f)$ and $TFS_{FFR}(f)$. For 0-dB SNR, $ENV_{FFR}(f)$ was significantly degraded, whereas $TFS_{FFR}(f)$ was largely unaffected. Results for +10 and -10 dB SNR conditions were similar to the 0 dB SNR condition. At -20 dB SNR, $ENV_{FFR}(f)$ showed little resemblance to the clean-speech response; $TFS_{FFR}(f)$ was significantly degraded, but with peaks near the first two harmonics remained discernible.

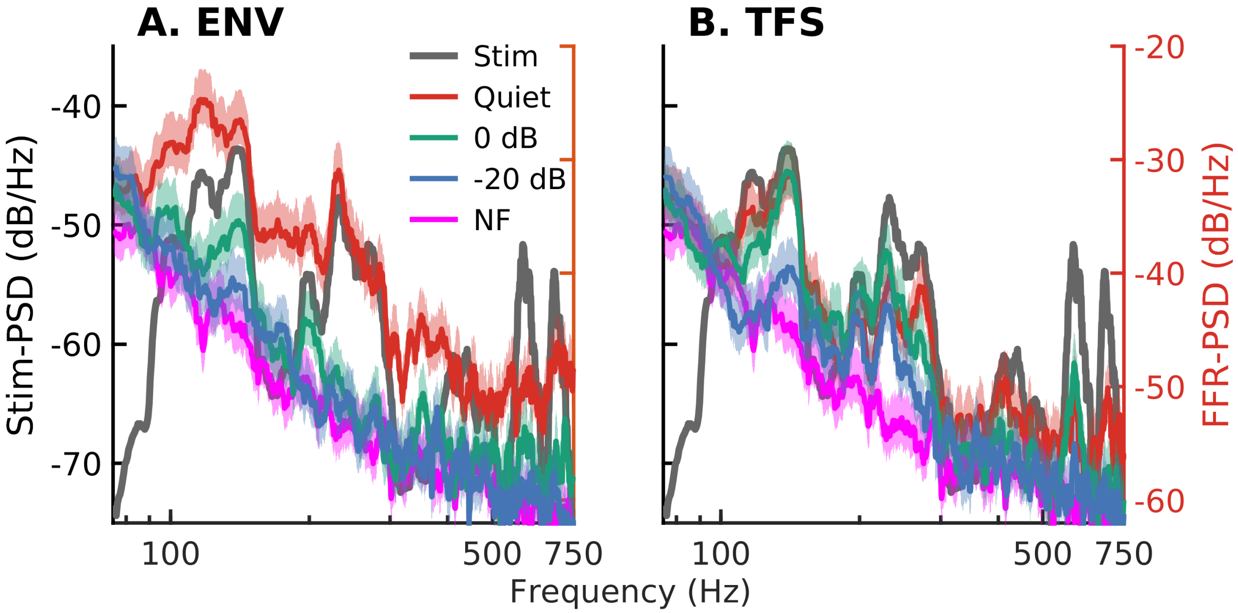

Figure 7.3. The high-pass filtered pink masker was much more detrimental to $ENV_{FFR}(f)$ than $TFS_{FFR}(f)$. Pink masker was high-pass filtered at 600 Hz. **A** Spectrum of stimulus (gray trace, labeled Stim, left y-axis) and $ENV_{FFR}(f)$ (right y-axis in panel **B** for speech in quiet (labeled Quiet, red), 0 dB SNR (green) and -20 dB SNR (blue) conditions. Noise floor (labeled NF, magenta) was estimated by first subtracting the even-indexed trials from odd-indexed trials for both polarities in response to speech-in-quiet and then averaging them. **B** Same as panel **A**, but for $TFS_{FFR}(f)$. Data are shown for a single normal-hearing animal, Q371. Spectra were computed over the whole duration of the stimulus. Note that for the 0 dB SNR condition (green), $TFS_{FFR}(f)$ was affected very little whereas $ENV_{FFR}(f)$ was affected significantly. Similarly, while $ENV_{FFR}(f)$ was almost entirely masked at -20 dB SNR, $TFS_{FFR}(f)$ still had discernible peaks, albeit weaker.

Population results. The effect of masking noise as a function of SNR on envelope and TFS coding was quantified by estimating the change in power in the FFR relative to the speech-in-quiet condition. Any temporal structure due to running pink noise was not expected in the FFR due to the stochastic nature of the noise. Figure 7.4A shows the total power in 60 to 500

Hz band for $ENV_{FFR}(t)$ and $TFS_{FFR}(t)$ at various SNRs. Figure 7.4B shows the reduction in power in $ENV_{FFR}(t)$ and $TFS_{FFR}(t)$ with decreasing SNR. Power in $ENV_{FFR}(f)$ was reduced by ~15 dB at 10 dB SNR, and by ~22 dB at -10 dB SNR, suggesting $ENV_{FFR}(t)$ primarily receives contributions from high-frequency neurons masked by the high-pass noise. Power in $TFS_{FFR}(t)$ was relatively resistant (less change in power) to masking compared to $ENV_{FFR}(t)$, suggesting higher contribution from place-specific (low-frequency) neurons. However, the fact that reduction in power did not saturate at the lowest two SNRs, but rather dropped linearly supports partial contributions to $TFS_{FFR}(t)$ from high-frequency neurons. In conclusion, $TFS_{FFR}(t)$ likely has contributions from both place-specific as well as high-frequency (> 600 Hz) neurons at conversational sound levels. It is noteworthy that power in $ENV_{FFR}(t)$ was ~ 12 dB higher than power in $TFS_{FFR}(t)$ for all animals in the speech-in-quiet condition despite the fact that speech has more power at low frequencies. This superiority of envelope coding is consistent with the idea that the FFR has a high-frequency bias (better synchrony across high-frequency neurons due to less dispersion at the cochlear base).

7.3.3 Experiment II: Effects of NIHL on the relative strengths of envelope and TFS representations in speech FFRs

Definitions. In this experiment, how power in $ENV_{FFR}(t)$ and $TFS_{FFR}(t)$ is related to power in the low- and high-frequency speech bands was evaluated, for both normal-hearing and hearing-impaired chinchillas (five animals per group). Two audio frequency bands were defined: (1) low-frequency band (LF; 60 to 500 Hz) and (2) high-frequency band (HF; .5 to 5 kHz). The low-frequency band contains power related to fundamental frequency and its first few harmonics, in addition to the first formant for some vowels. The high-frequency band contains most of the information related to vowel identity such as formants and their transitions (Lindblom and Studdert-Kennedy, 1967). The cut-off frequencies were chosen to capture the severity of DT effects (i.e., the masking by low-frequency energy that contains less information) on the neural coding of speech. Based on these definitions, the following notations are used hereafter.

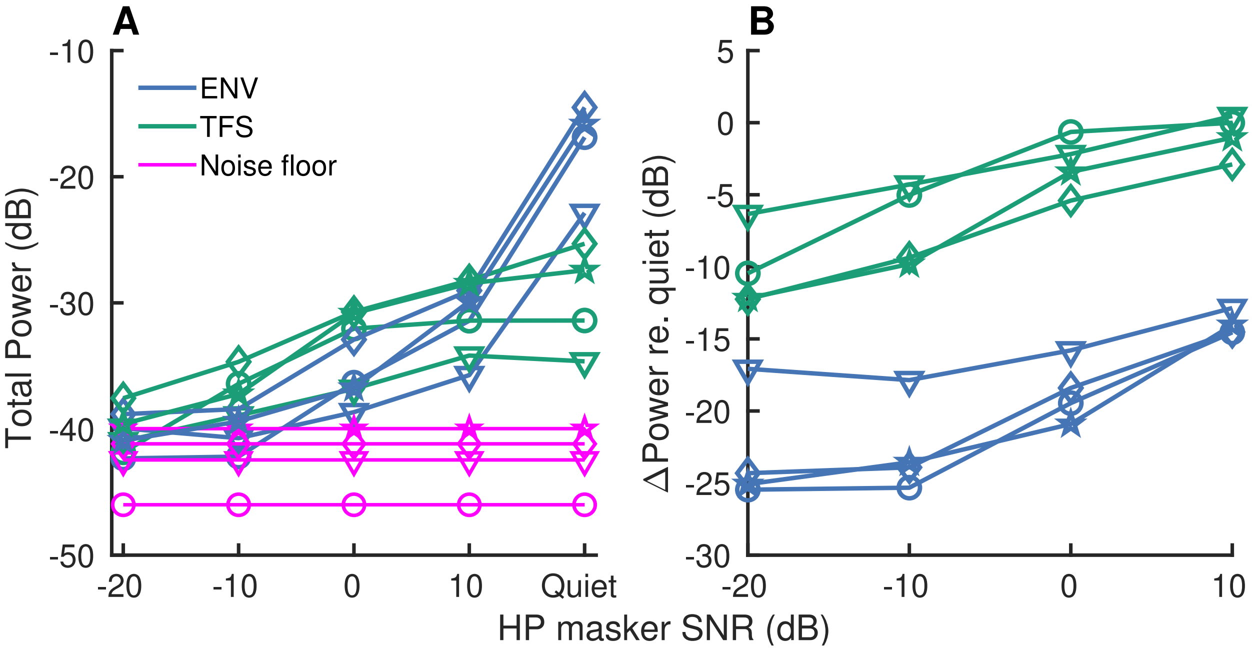

Figure 7.4. $ENV_{FFR}(t)$ reflects activity from neurons tuned to CFs > 600 Hz, while $TFS_{FFR}(t)$ receives contributions from both place-specific (low-frequency) and high-frequency neurons for speech stimuli. **A** Power in $TFS_{FFR}(t)$ (green) and $ENV_{FFR}(t)$ (blue) at the five SNRs tested. Noise floor (magenta) was estimated as described in Fig 7.3. **B** Change in power in $TFS_{FFR}(t)$ and $ENV_{FFR}(t)$ relative to the speech-in-quiet condition. Data are shown for four normal-hearing animals (indicated by different symbols). FFRs were band-pass filtered between 60 and 500 Hz.

$$TFS_{FFR}^{power} \triangleq \text{Power in } TFS_{FFR}(t) \text{ [60 to 500 Hz]}$$

$$ENV_{FFR}^{power} \triangleq \text{Power in } ENV_{FFR}(t) \text{ [60 to 500 Hz]}$$

$$LF_{stimulus}^{power} \triangleq \text{Stimulus power in LF band [60 to 500 Hz]}$$

$$HF_{stimulus}^{power} \triangleq \text{Stimulus power in HF band [0.5 to 5 kHz]}$$

Using these definitions, four specific hypotheses capturing the relations between the two stimulus powers and the two FFR powers are described first. Next, these hypotheses are used to derive normalized metrics to quantify relative TFS-to-envelope coding for both normal-hearing and hearing-impaired groups. Figure 7.5 depicts a schematic to illustrate the main DT hypotheses (#2 and #3).

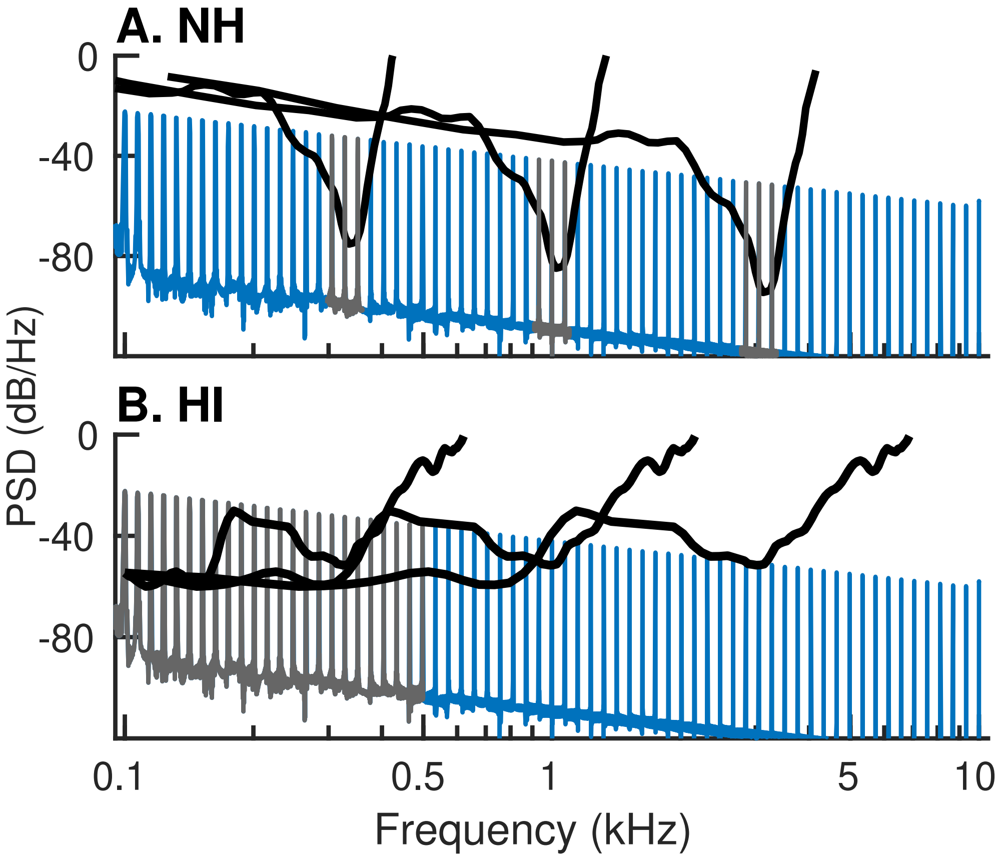

Figure 7.5. Schematic illustrating hypotheses(Eqs. 7.2 to 7.4) for relative TFS to envelope enhancement in the FFR following NIHL. Speech-like stimulus processing for normal (**A**) and impaired (**B**) auditory systems. Normal-hearing neurons will primarily respond to power near their CF (gray harmonics in panel **A**). Thus, as FFRs have a high-frequency bias, envelope coding in high-frequency neurons will be more saliently represented compared to TFS coding in FFRs recorded from normal-hearing animals. In contrast, hearing-impaired neurons will respond more to low-frequency stimulus power due to hypersensitive tuning-curve tails (gray harmonics in panel **B**), and less to near-CF stimulus power due to desensitized tuning-curve tips. Thus, FFRs recorded from hearing-impaired animals will be dominated by low-frequency stimulus power.

Hypothesis #1: ENV_{FFR}^{power} and $LF_{stimulus}^{power}$. Studies have shown that neural responses to low-frequency ($<$ 500 Hz) stimuli are dominated by the phase-sensitive (i.e., TFS) component and do not contain a large phase-tolerant (i.e., envelope) component (e.g., Kale and Heinz, 2010). Thus, ENV_{FFR}^{power} and $LF_{stimulus}^{power}$ were hypothesized to be uncorrelated (Eq. 7.1).

$$ENV_{FFR}^{power} \not\propto LF_{stimulus}^{power}, \text{ for NH and HI.} \tag{7.1}$$

Hypothesis #2: ENV_{FFR}^{power} *and* $HF_{stimulus}^{power}$. At the playback intensity (~70 dB SPL) used in this study, it is reasonable to assume that most auditory-nerve fibers in normal-hearing animals will be largely saturated (Sachs and Young, 1979), resulting in roughly level-independent envelope coding (Kale and Heinz, 2010). Although suprathreshold modulation coding is stronger at more central levels of the auditory system (e.g., inferior colliculus), there is not a significant level dependence well above threshold, except a slight reduction for low modulation frequencies (Joris et al., 2004). Finally, with a complex broadband stimulus like speech, spread of excitation effects that could increase evoked envelope responses are not expected. Thus, a level-dependence in envelope coding for normal-hearing animals is not expected. On the other hand, it is likely that neurons in hearing-impaired animals will not be saturated due to elevated thresholds (compare panels A and B in Fig 7.5), which allows envelope coding strength to rise with increasing stimulus intensity (Joris et al., 2004; Kale and Heinz, 2010). Thus, ENV_{FFR}^{power} was hypothesized to be correlated with $HF_{stimulus}^{power}$ for HI animals, but not for NH animals (Eq. 7.2).

$$ENV_{FFR}^{power} \begin{cases} \not\propto HF_{stimulus}^{power}, \text{ for NH,} & (7.2a) \\ \propto HF_{stimulus}^{power}, \text{ for HI.} & (7.2b) \end{cases}$$

Hypothesis #3: TFS_{FFR}^{power} *and* $LF_{stimulus}^{power}$. Central to the main focus of this study (i.e., DT), it was hypothesized that TFS_{FFR}^{power} will be correlated with $LF_{stimulus}^{power}$ for HI animals, more so than for NH animals. The reasoning was as follows. In Experiment I, power in $TFS_{FFR}(t)$ was found to be at least an order of magnitude lower compared to power in $ENV_{FFR}(t)$ for speech-in-quiet (Fig 7.4A), which in addition to the high-pass masking results (Fig 7.4B) suggests a modest contribution to $TFS_{FFR}(t)$ that comes from both the high-frequency and place-specific neurons. Since central neurons in the brainstem and midbrain will inherit degraded tuning properties from auditory-nerve fibers, following acoustic overexposure, high-frequency neurons will have DT (due to hypersensitive tails, desensitized tips, and lowered best frequencies), and thus will contribute disproportionately more to $TFS_{FFR}(t)$ (see Fig 7.5B). Thus, a stronger correlation between power in $TFS_{FFR}(t)$ and low-frequency stimulus power was hypothesized for the hearing-impaired chinchillas (Eq. 7.3),

$$TFS_{FFR}^{power} \begin{cases} \not\propto LF_{stimulus}^{power}, \text{ for NH (weak corr),} & (7.3a) \\ \propto LF_{stimulus}^{power}, \text{ for HI (strong corr),} & (7.3b) \end{cases}$$

in addition to an overall enhancement of TFS coding (Eq. 7.4).

$$TFS_{FFR}^{power} \text{ for HI} > TFS_{FFR}^{power} \text{ for NH.} \tag{7.4}$$

Hypothesis #4: TFS_{FFR}^{power} and $HF_{stimulus}^{power}$. Since $TFS_{FFR}(t)$ is filtered below 500 Hz, it should be affected by stimulus power below 500 Hz (i.e., $LF_{stimulus}^{power}$) and not by stimulus power above 500 Hz (i.e., $HF_{stimulus}^{power}$). Thus, TFS_{FFR}^{power} and $HF_{stimulus}^{power}$ were hypothesized to be uncorrelated (Eq. 7.5).

$$TFS_{FFR}^{power} \not\propto HF_{stimulus}^{power}, \text{ for NH and HI.} \tag{7.5}$$

Hypothesis for TFS coding relative to envelope. Hypotheses 7.2 and 7.3 can be combined by defining the following two normalized power metrics.

$$\text{FFR TFS-to-ENV power ratio} \triangleq \frac{TFS_{FFR}^{power}}{ENV_{FFR}^{power}}$$

$$\text{Stimulus LF-to-HF power ratio} \triangleq \frac{LF_{stimulus}^{power}}{HF_{stimulus}^{power}}$$

Normalized metrics, like those described above, help to minimize individual variability unrelated to neural coding of speech (e.g., overall FFR magnitude, differences in head size). Combining the hypotheses represented by Eqs. 7.2 and 7.3, we predict

$$\frac{TFS_{FFR}^{power}}{ENV_{FFR}^{power}} \begin{cases} \not\propto \dfrac{LF_{stimulus}^{power}}{HF_{stimulus}^{power}}, \text{ for NH,} & \text{(7.6a)} \\ \propto \dfrac{LF_{stimulus}^{power}}{HF_{stimulus}^{power}}, \text{ for HI.} & \text{(7.6b)} \end{cases}$$

In general, across-pair ratios of correlated random-variable pairs are not correlated. However, in the special case of nonnegative correlated random-variable pairs passing through the origin (e.g., the power metrics used here), across-pair ratios of correlated pairs are correlated (e.g., Eq. 7.6b). Distributions of powers (Eqs. 7.1 to 7.5) and power ratios (Eq. 7.6) were obtained by computing these metrics from the speech stimulus and FFR responses in sliding windows of 64 ms without overlap. Since the hypotheses rely on harmonic structure in the signal spectrum, the analysis window was restricted to voiced portions of the speech stimulus. Voiced segments were identified by low-pass filtering the stimulus at 250 Hz with zero phase, half-wave rectifying, low-pass filtering again at 32 Hz and then using a threshold criterion (see Fig 2.1B).

Normal-hearing example. Figures 7.6A and 7.6D show example data from one normal-hearing chinchilla. One example stimulus segment is highlighted within the red boundaries near 1050 ms of the speech sentence. The power spectral density (PSD) was estimated for $ENV_{FFR}(t)$, $TFS_{FFR}(t)$, and the audio signal in that window (Fig 7.6C shows the stimulus-segment PSD); these spectra were used to compute the metrics in Eqs. 7.1 to 7.6. In this example, the power ratios computed for this one segment correspond to the red circle in the scatter plot in Fig 7.6D. Note that $ENV_{FFR}(t)$ has higher power than $TFS_{FFR}(t)$ as reflected by a negative (∼-7 dB) TFS-to-envelope FFR power ratio; this was typical for NH animals. The scatter plot was populated by sliding the window as described previously. A linear regression model was fit to the data to test the main hypothesis represented in Eq. 7.6A. For this normal-hearing animal, ${}^{TFS_{FFR}^{power}}/_{ENV_{FFR}^{power}}$ was not correlated with ${}^{LF_{stimulus}^{power}}/_{HF_{stimulus}^{power}}$, as hypothesized.

Hearing-impaired example. Figures 7.6B and 7.6E show a comparable example for one hearing-impaired animal. One difference that stands out between Figs. 7.6A and 7.6B is the enhanced $TFS_{FFR}(t)$ for the hearing-impaired animal. For the same segment considered

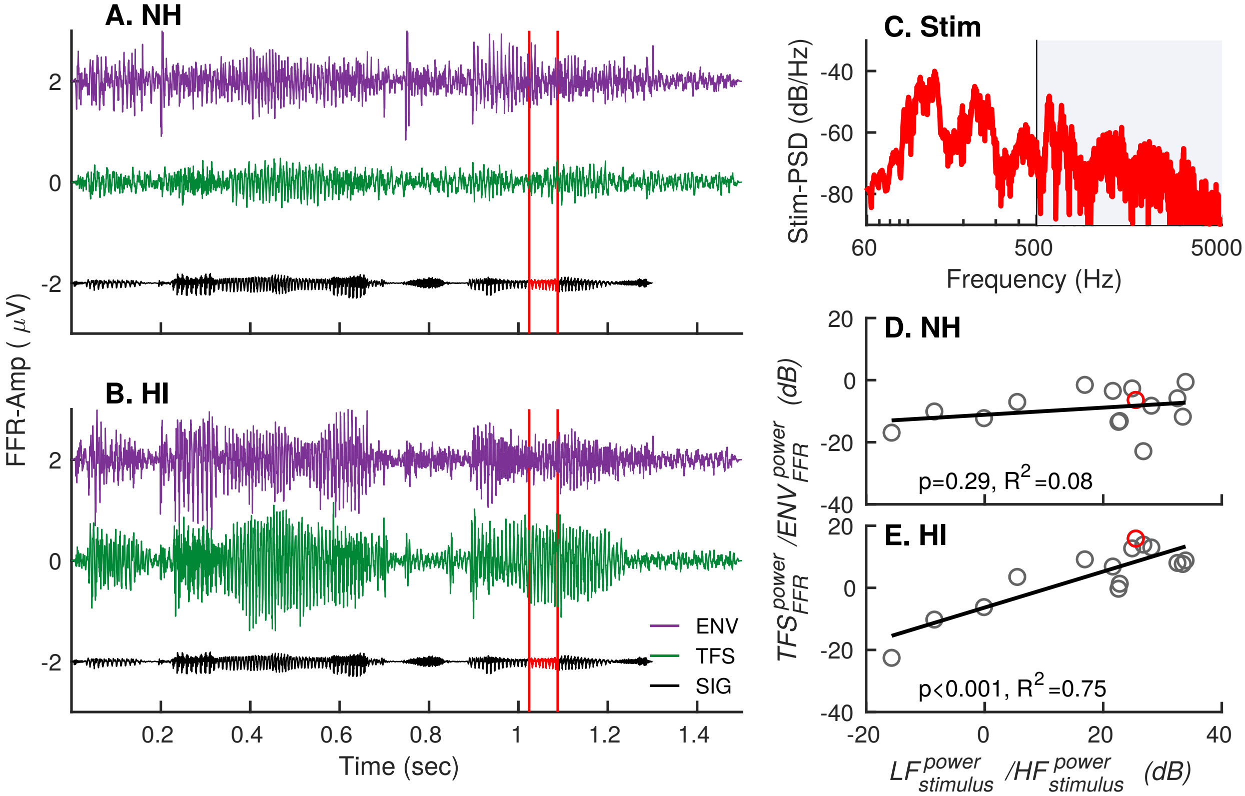

Figure 7.6. Examples support the main hypothesis by illustrating the effect of hearing loss on the dependence of relative (TFS to envelope) FFR responses on relative (low to high frequency) stimulus power. (**A** and **B**) Time waveforms: stimulus (black, downscaled and shifted down by 2μV for display), $ENV_{FFR}(t)$ (purple), and $TFS_{FFR}(t)$ (green) for a normal-hearing and hearing-impaired animal, respectively. $ENV_{FFR}(t)$ is shifted up by 2μV for display. **C** Spectrum of the 64-ms voiced stimulus segment shown in panel **A** within the red vertical lines. (**D** and **E**) Scatter plots between metrics in Eq. 7.6 for the normal-hearing and hearing-impaired animals, respectively. Lines show linear regression fits, with statistical parameters shown as in-panel text. The red circles correspond to the example stimulus segment within the red vertical boundaries in panels **A** and **B**. Notice that for the normal-hearing animal $TFS^{power}_{FFR}/ENV^{power}_{FFR}$ is negative (~ -7 dB) for this segment, indicating higher envelope power in the FFR relative to TFS. In contrast, for the hearing-impaired animal, $TFS^{power}_{FFR}/ENV^{power}_{FFR}$ is positive ($\sim +13$ dB) for the same segment, indicating higher TFS power in the FFR relative to envelope. Across all voiced stimulus segments, the ratio of low-frequency to high-frequency stimulus power was uncorrelated with FFR TFS-to-envelope power ratio for the normal-hearing animal **D**, but was correlated for the hearing-impaired animal **E**.

in Fig 7.6A (red lines), the TFS to envelope power ratio for this animal was ∼+13 dB (in contrast to ∼-7 dB for the normal-hearing animal). This enhancement in $TFS_{FFR}(t)$ was typical for all hearing-impaired animals. Note that this TFS "enhancement" is detrimental to speech coding as it represents a dramatic increase in the severity of upward spread of masking by low-frequency energy in speech for the hearing-impaired population.

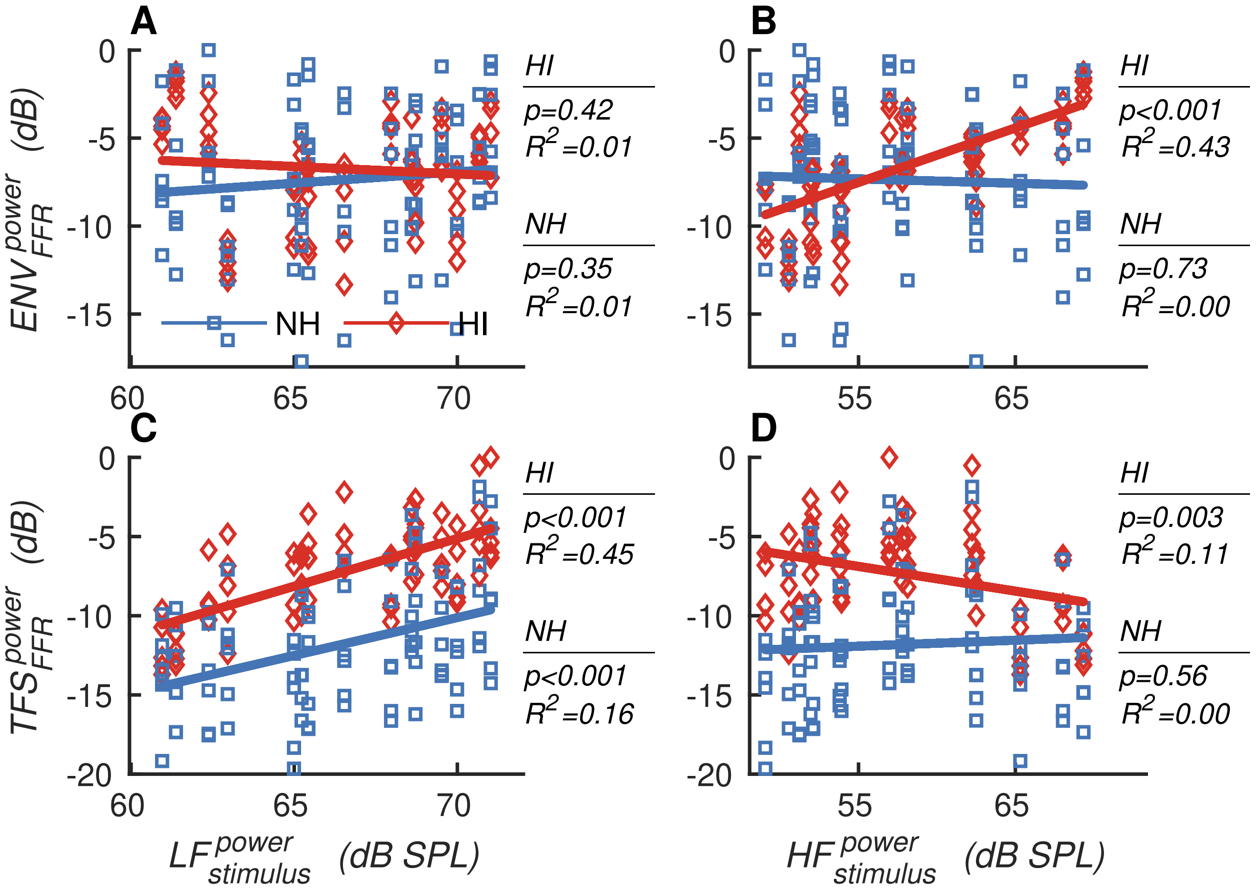

Figure 7.7. Pooled scatter plots of the four FFR-to-stimulus relations underlying the hypotheses in Experiment II (Eqs. 7.1 to 7.5). Panels **A** and **B** show the dependence of ENV_{FFR}^{power} on stimulus power; **C** and **D** show the TFS_{FFR}^{power} dependence. Panels **A** and **C** show the relation of low-frequency stimulus power to FFR responses; **B** and **D** show the relations with high-frequency stimulus power. Data are shown for five normal-hearing (NH, blue) and five hearing-impaired (HI, red) animals. Statistics for linear-regression models are shown on the right of each panel. Equations 7.1 to 7.5 were generally supported by these data.

Population results for envelope and TFS coding. Figure 7.7 shows the empirical data used to test the hypotheses represented by Eqs. 7.1 to 7.5, i.e., the relations between power in the FFR (both envelope and TFS) and power in the stimulus bands (both LF and HF). ENV_{FFR}^{power} and $LF_{stimulus}^{power}$ were not correlated for either group (Fig 7.7A; Eq. 7.1). ENV_{FFR}^{power}

was correlated with $HF^{power}_{stimulus}$ for HI animals, but not for NH animals (Fig 7.7B; Eq. 7.2). Similarly, TFS^{power}_{FFR} was strongly correlated with $LF^{power}_{stimulus}$ for HI animals, but moderately correlated for NH animals (lower R^2 value for the linear-regression model), consistent with DT (Fig 7.7C; Eq. 7.3). Importantly, the data in Fig 7.7C show that TFS^{power}_{FFR} is larger for HI than for NH, consistent with the hypothesis represented by Eq. 7.4. The correlation between TFS^{power}_{FFR} and $HF^{power}_{stimulus}$ was not significant for normal-hearing chinchillas, but it was significant (p=.005) for hearing-impaired chinchillas (Fig 7.7D; Eq. 7.5). However, residual plots of TFS^{power}_{FFR} versus $HF^{power}_{stimulus}$ were not normally distributed for either group (Fig 7.7D). In fact, the plot of $LF^{power}_{stimulus}$ versus $HF^{power}_{stimulus}$ followed a similar inverted-U shape. Thus, the unexpected significant negative correlation for HI in Fig 7.7D may arise due to the interaction between stimulus statistics (e.g., an inverse relation between high- and low-frequency stimulus power) and the more consistent relation of TFS^{power}_{FFR} with low-frequency stimulus power for HI animals (Fig 7.7C). The idea that this unexpected result does not actually represent an inverse linear relationship between TFS^{power}_{FFR} and $HF^{power}_{stimulus}$ is also supported by the linear mixed-model for TFS^{power}_{FFR} described next. Overall, these results are in good agreement with the hypotheses described in Eqs. 7.1 to 7.5.

To test statistically the hypotheses represented in Eqs. 7.1 to 7.5, two separate linear mixed-effect models were constructed with ENV^{power}_{FFR} and TFS^{power}_{FFR} as dependent variables. Fixed effects for both the models included $LF^{power}_{stimulus}$, $HF^{power}_{stimulus}$, hearing status, and their interaction terms. Animal identifier was used as a random effect. As described earlier, a p-value of .01 was used as the criterion for significance. For ENV^{power}_{FFR}, the effect of $HF^{power}_{stimulus}$ ($F_{(1,134.0)} = 16.9$, $p < .0001$) was significant, whereas $LF^{power}_{stimulus}$ ($F_{(1,134.0)} = 3.2$, $p = .08$) was not; these results support the idea that FFR envelope is influenced directly by $HF^{power}_{stimulus}$ at conversational speech sound levels. Other significant effects included hearing status ($F_{(1,122.7)} = 8.3$, $p = .0046$) and $HF^{power}_{stimulus}\times$ hearing status ($F_{(1,134.0)} = 37.1$, $p < .0001$); these significant effects support the hypothesis in Eq. 7.2 that ENV^{power}_{FFR} is correlated with $HF^{power}_{stimulus}$ for hearing-impaired chinchillas and not for normal-hearing chinchillas. All other effects were not significant for ENV^{power}_{FFR}. For TFS^{power}_{FFR}, the main effect of $LF^{power}_{stimulus}$ ($F_{(1,135.2)} = 121.4$, $p < .0001$) was significant, whereas $HF^{power}_{stimulus}$ ($F_{(1,135.2)} = 0.25$, $p = .62$) was not significant. These results confirmed that TFS^{power}_{FFR} was specifically dependent

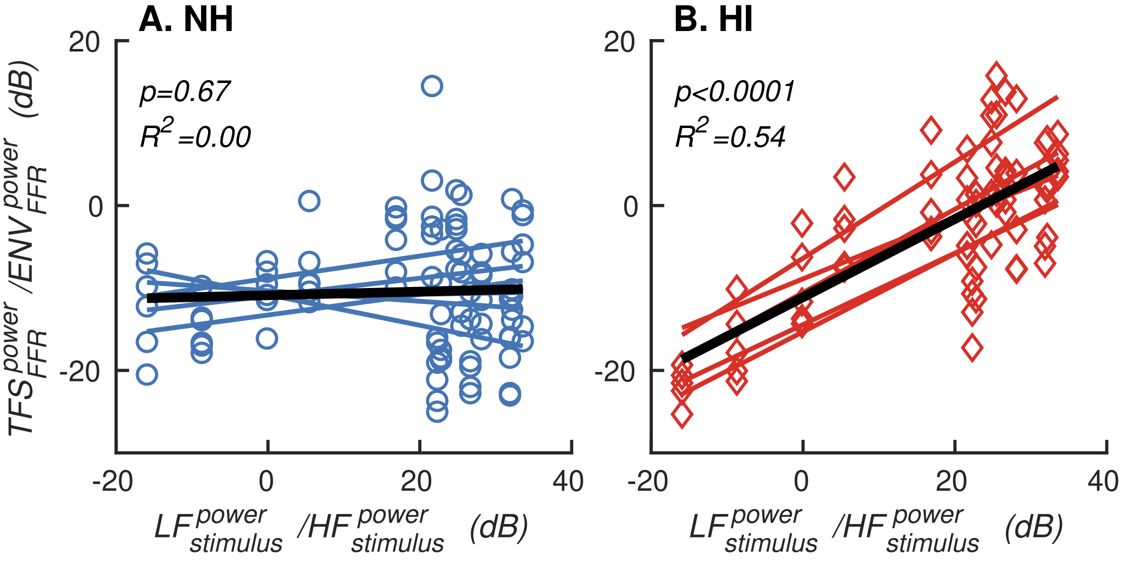

Figure 7.8. Normalized FFR-TFS power was correlated with normalized low-frequency stimulus power for HI animals, but not for NH animals. Pooled scatter plots show a consistent lack of correlation for NH (**A**) and consistent correlations for HI animals (**B**). Each group had five animals. Thin colored lines are regression lines for individual animals. Thick black lines are population regression lines per group (*i.e.*, linear model fit to data pooled across all animals for each group). Statistics are reported in figure panels. The hypothesis represented by Eq. 7.6 was strongly supported by these data.

on $LF_{stimulus}^{power}$, and not on $HF_{stimulus}^{power}$. The main effect of hearing status ($F_{(1,127.8)} = 38.5$, $p < .0001$) was also significant for TFS_{FFR}^{power}, confirming that TFS coding was significantly enhanced in the hearing-impaired group. Overall, these results reinforce the idea that the effects seen in these data are not merely a reflection of audibility-driven FFR responses, but rather specific effects regarding the relative strength of envelope and TFS coding, as hypothesized in Eqs. 7.1 to 7.5.

Population results for TFS coding relative to envelope. In order to reduce the variability related to signal strength (due to electrode placement, animal differences, *etc.*), power in $TFS_{FFR}(t)$ was normalized by power in $ENV_{FFR}(t)$ in Eq. 7.6. The relation between this normalized FFR power ratio and the stimulus power ratio in Eq. 7.6 was evaluated. Figure 7.8 shows pooled regression plots for all normal-hearing and hearing-impaired animals. $TFS_{FFR}^{power}/ENV_{FFR}^{power}$ and $LF_{stimulus}^{power}/HF_{stimulus}^{power}$ were correlated for all hearing-impaired animals, but not for any normal-hearing animal, consistent with the hypothesis represented by Eq. 7.6.

The consistency of these results suggest that the slope of this relation between normalized FFR-TFS responses and normalized low-frequency stimulus power may be a useful metric for the non-invasive diagnosis of DT.

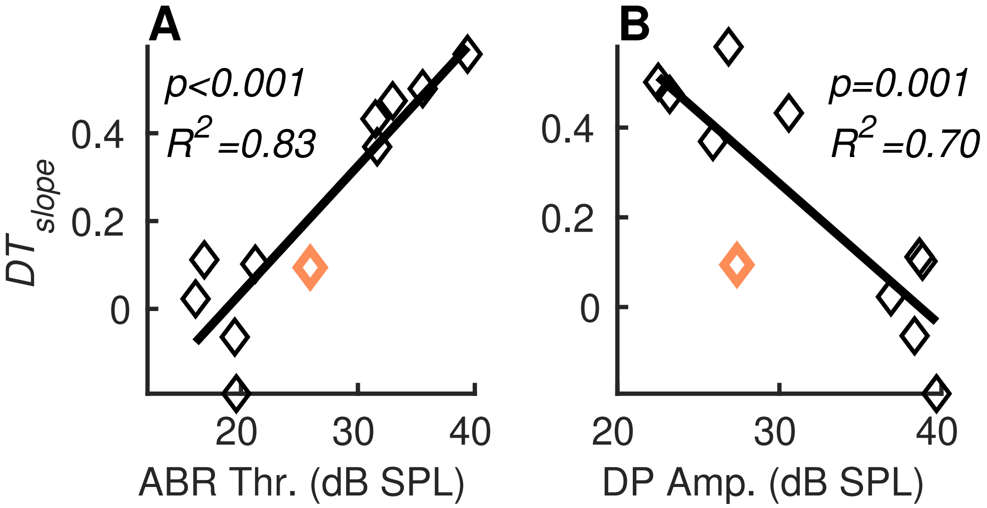

Figure 7.9. Normalized DT slopes from Fig 7.8 are correlated with the degree of noise-induced hearing loss, as reflected by ABR and DPOAE measures. Slope in Fig 7.8 as a function of **A** mean ABR threshold (Thr.) and **B** mean DPOAE amplitude (Amp.) between 1 and 8 kHz for all animals irrespective of hearing status (N=11). Lines represent linear-regression fits, with statistical parameters reported in each panel. Notice Q369 (shown in orange), which was considered a special case because of its near-normal ABR threshold following noise-exposure, had near-normal DT_{slope}, despite being noise exposed.

Relation between the normalized DT metric and auditory screening measures. The degree of disruption in speech FFR coding due to distorted tonotopy was quantified by the normalized metric DT_{slope}, which was defined as the slope between the power ratios in Eq. 7.6 (also see Fig 7.8). To evaluate the relation between noise-induced hearing loss and the degree of distorted tonotopy, a linear model was fit with auditory screening measures (i.e., ABR threshold and DPOAE amplitude) as independent variables and DT_{slope} as the dependent variable. As seen in Fig 7.9, DT_{slope} was significantly correlated with both mean ABR threshold and DPOAE amplitude. All animals were included in this analysis irrespective of hearing status. Chinchilla Q369 (shown in orange), the animal that was excluded from the hearing-impaired group (due to minimal hearing loss after noise exposure, i.e., a tough ear), was included here. DT_{slope} for Q369 was largely within normal limits, consistent with its near-normal ABR thresholds following noise exposure. However, Q369's DPOAE amplitudes were reduced, which is not consistent with its ABR thresholds or DT_{slope} for unknown reasons. Overall, the FFR-based DT_{slope} metric is significantly correlated with the degree of noise-induced hearing loss from ABRs. This finding is consistent with single-unit results showing that the degree of distorted tonotopy (i.e., elevated TFS coding in basal fibers) is correlated with auditory-nerve-fiber threshold shifts following noise exposure (Fig 7B in Henry et al., 2016).

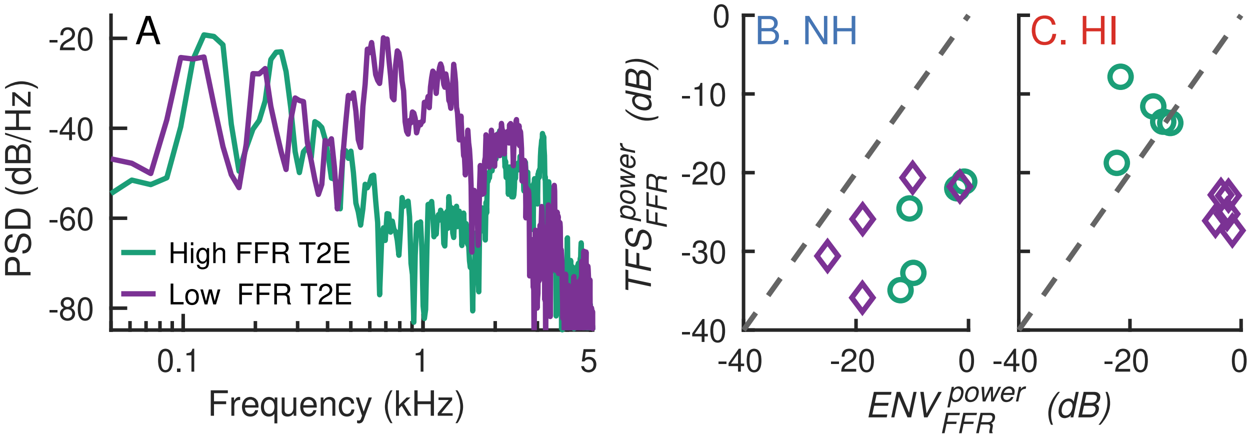

Figure 7.10. Effects of DT are more severe for speech segments with a negative-sloping spectrum than for those with a flatter spectrum. **A** The green trace represents the spectrum of the stimulus segment for which the FFR TFS-to-envelope (T2E) ratio, $TFS^{power}_{FFR}/ENV^{power}_{FFR}$, is largest for the HI group relative to the NH group (higher stimulus-power ratios in Fig 7.8). Note the negative spectral tilt and high power concentration in the low-frequency region. The purple trace represents the spectrum of the stimulus segment for which the difference in $TFS^{power}_{FFR}/ENV^{power}_{FFR}$ is maximally negative between the HI and NH groups (lower stimulus-power ratios in Fig 7.8). Note that this stimulus spectrum is relatively flat up to 1500 Hz, with significant energy between 600 and 1500 Hz. **B** Scatter plot for TFS^{power}_{FFR} and ENV^{power}_{FFR} for the two segments in panel **A** for all the NH animals. Green circles correspond to the high TFS-to-ENV segment. Purple diamonds correspond to the low TFS-to-ENV segment. Gray dashed line shows equality between TFS^{power}_{FFR} and ENV^{power}_{FFR}. Note that for the NH chinchillas, all points lie in the lower diagonal, where $TFS^{power}_{FFR} < ENV^{power}_{FFR}$. **C** Same as panel **B**, but for HI animals. Note the differential effects of enhanced TFS for the green segment and enhanced envelope for the purple segment (relative to NH). Symbols in **B** and **C** represent the five animals in each group.

Separating DT (enhanced TFS) from enhanced envelope. It can be seen from Fig 7.8 that the consistent differences in the DT_{slope} metric between NH and HI groups arise because the FFR TFS-to-envelope ratio for the HI group is differentially affected at low and high stimulus-power ratios. For the segment with the lowest stimulus-power ratio, $^{LF^{power}_{stimulus}}/_{HF^{power}_{stimulus}}$, the HI group has a lower FFR ratio, $^{TFS^{power}_{FFR}}/_{ENV^{power}_{FFR}}$, than the NH group. In contrast, for segments with higher stimulus-power ratios ($>$ 20 dB), the FFR TFS-to-envelope ratios for the HI group are significantly higher than for the NH group.

To develop insight into the factors underlying the sensitivity of the DT_{slope} metric, stimulus spectra were compared between these two extremes of the data in Fig 7.8. The green spectrum in Fig 7.10A represents the stimulus segment for which the HI group showed the largest FFR TFS-to-envelope ratio relative to the NH group, whereas the purple spectrum represents the smallest FFR ratio relative to NH. Note that the green spectrum has a negative spectral tilt and a high concentration of energy at very low frequencies. For the segment with this negative-sloping spectrum (green circles in Figs. 7.10B and 7.10C), neurons from HI animals have small tip-to-tail ratios and are likely to suffer from severe upward spread of masking by the very low-frequency energy in the stimulus. Note that no values of TFS^{power}_{FFR} are greater than -20 dB for NH (Fig 7.10B), in contrast to HI for which all values are greater than -20 dB (Fig 7.10C). The enhanced TFS in the hearing-impaired FFRs seen in this study likely reflects this drastic spread of upward excitation from low-frequency stimulus energy (i.e., DT).

A separate factor contributing to the sensitivity of the DT_{slope} metric for HI is highlighted by the spectrum of the segment for which the FFR TFS-to-envelope ratio is most reduced in the HI group (Fig 7.10A). The purple spectrum is markedly different from the green spectrum in that it has significant energy from 600 to 1500 Hz in addition to the strong low-frequency energy from 70 to 200 Hz (i.e., the spectrum is almost flat up to 1500 Hz). For such spectra, the high-frequency energy prevents a dramatic upward spread of low-frequency energy such that TFS responses are not significantly over-represented in the FFR (see comparable TFS^{power}_{FFR} values for purple diamonds in Figs. 7.10B and 7.10C). Without over-represented TFS responses, the effects of enhanced envelope responses following HI that have been previously reported (e.g., Henry et al., 2016; Kale and Heinz, 2010; Zhong et al.,

2014) are emphasized (see higher HI ENV_{FFR}^{power} values for purple diamonds, Figs. 7.10B and 7.10C). Note that ENV_{FFR}^{power} for the NH animals is higher for the green segment relative to the purple segment, even though the purple segment has more high-frequency energy. This reduction in ENV_{FFR}^{power} could be partly due to reduced envelope coding at high stimulus intensities (Joris and Yin, 1992; Kale and Heinz, 2010) and destructive interference between responses from different spectral bands (Wang et al., 2019).

Overall, the present results demonstrate that noise-induced hearing loss creates a much larger variation in the FFR TFS-to-envelope ratio across a sentence (Fig 7.8B), compared to the relatively constant value observed for NH (Fig 7.8A). Figure 7.10 illustrates that this greater variation across a sentence arises from both enhanced TFS responses (due to DT in speech segments for which low-frequency stimulus power is prominent) and enhanced envelope responses (in segments for which significant high-frequency energy prevents the dramatic upward spread of excitation associated with DT). The perceptual consequences of this abnormally large variation in relative TFS-to-envelope coding in responses to natural speech following HI remains to be determined, but non-invasive metrics such at DT_{slope} provide the potential to explore these issues in humans.

7.4 Discussion

7.4.1 Profiles of tonotopic sources for the FFR change following NIHL

The tonotopic origin of the FFR, especially that of $TFS_{FFR}(t)$, has been a topic of debate for years (Dolphin and Mountain, 1993; Wang et al., 2019). Understanding the sources of the FFR is critical for drawing appropriate scientific conclusions and advancing the use of FFRs as a diagnostic tool for issues like hidden hearing loss (Bharadwaj et al., 2014; Shaheen et al., 2015; Verhulst et al., 2018), cochlear hearing-loss (Kuwada et al., 1986), learning disabilities (Wible et al., 2004), and more. The focus of Experiment I was to evaluate the tonotopic origins of $ENV_{FFR}(t)$ and $TFS_{FFR}(t)$ in normal-hearing animals in response to speech at a conversational sound level. Consistent with previous studies, results from Experiment I suggest that $ENV_{FFR}(t)$ primarily reflects activity from high-frequency ($>$ 600 Hz) neurons (Fig. 7.4) (Wang et al., 2019; Zhu et al., 2013). On the other hand, $TFS_{FFR}(t)$ received partial contributions from place-specific ($<$ 600 Hz) as well as high-frequency ($>$ 600 Hz) neurons (Fig. 7.4). More importantly, power in $TFS_{FFR}(t)$ was significantly lower than power in $ENV_{FFR}(t)$ in responses to speech-in-quiet (Fig. 7.4A), supporting the high-frequency bias of FFRs, as neurons with higher CF are better encoders of envelope than TFS (Joris and Yin, 1992; Kale and Heinz, 2010). Thus, in normal-hearing animals, envelope representation is stronger than TFS in FFRs to speech stimuli in quiet.

The primary objective of this study was to diagnose distorted tonotopic mapping in animals with noise-induced hearing loss. In order to characterize DT noninvasively, Experiment II took advantage of speech spectrum statistics, auditory-nerve-fiber physiological properties, and the high-frequency bias of FFRs to form hypotheses (Eqs. 7.1 to 7.6). Studies have shown that the strength of single-fiber envelope coding positively correlates with stimulus intensity near threshold and drops off at suprathreshold intensities (Joris and Yin, 1992; Kale and Heinz, 2010). As high-frequency neurons in hearing-impaired animals are not as sensitive as normal-hearing animals (Fig. 7.5), it was hypothesized that high-frequency stimulus power should be correlated with $ENV_{FFR}(t)$ power for hearing-impaired animals and not for normal-hearing animals in this equal-SPL paradigm (Eq. 7.2); this hypothesis was supported by the data (Fig. 7.7B). More importantly, power in $TFS_{FFR}(t)$ was hypothesized to be higher for hearing-impaired animals, in addition to having a stronger correlation with low-frequency stimulus power (Eq. 7.3 and 7.4). The reasoning was that high-frequency neurons will respond to low-frequency stimulus power due to the effects of DT in hearing-impaired animals, allowing a stronger FFR due to better synchronization at the cochlear base. These hypotheses were also supported by the data (Fig. 7.7C). The linear model for predicting TFS_{FFR}^{power} from $LF_{stimulus}^{power}$ accounted for only 18% of the variance in normal-hearing animals, whereas this model explained 45% of the variance for hearing-impaired animals. Overall, Eqs. 7.1 to 7.5 represent the predicted relations between the pairs ($LF_{stimulus}^{power}$, TFS_{FFR}^{power}) and ($HF_{stimulus}^{power}$, ENV_{FFR}^{power}), and the predicted independence between the pairs ($LF_{stimulus}^{power}$, ENV_{FFR}^{power}) and ($HF_{stimulus}^{power}$, TFS_{FFR}^{power}), which were largely supported by the data (Fig. 7.7). Figure 7.5 depicts a schematic illustrating the underlying mechanisms for normal-hearing and hearing-impaired animals that underlie these predictions. Normal-hearing neurons (Fig. 7.5A) primarily respond to near-tip power that facilitates strong envelope coding in the FFR.

In contrast, desensitization of tips for hearing-impaired neurons (Fig. 7.5B) results in diminished envelope coding in the FFR, with an over-representation of low-frequency stimulus power due to increased tail sensitivity of high-frequency neurons (also see Fig. 7 in Henry et al., 2019). When the analyses in Fig. 7.7 and 7.8 were repeated using 250 Hz or 700 Hz as the boundary between low-frequency and high-frequency bands (instead of 500 Hz), results were largely unaltered (not shown) suggesting the first two harmonics of the fundamental were driving the hearing-impaired $TFS_{FFR}(t)$ (consistent with Fig. 7.10). The importance of the first two harmonics reflects the high occurrence of very low-frequency ($<$ 300 Hz) first formant energy in the speech sentence used here. Overall, these results show the increased susceptibility of the hearing-impaired FFR to upward spread of masking by low-frequency energy at the expense of near-tip envelope coding, suggesting that the key effects of DT are captured in the FFR.

7.4.2 Benefits of using normalized metrics

Noninvasive scalp recordings like the FFR can be affected by various extraneous factors, such as electrode placement, speaker positioning, and head-size differences, which are unrelated to the issues under investigation (Bharadwaj et al., 2019). Variability related to these extraneous sources can be reduced by using normalized metrics. The benefit of normalization in Eq. 7.6 over Eq. 7.3 can be quantified by looking at the standard deviation of model slopes for both groups in Figs. 7.7C and 7.8. Following normalization, the standard deviation reduced from 0.21 to 0.13 (40%) for normal-hearing animals, and from 0.11 to 0.08 (29%) for hearing-impaired animals.

A potential confound that can contribute to normal-hearing and hearing-impaired coding differences is the so-called "central-gain hypothesis" (Chambers et al., 2016; Salvi et al., 2000). Studies have shown that evoked responses in the inferior colliculus, presumably a major contributor to the FFR, are enhanced following acoustic trauma (Salvi et al., 1990). The use of a normalized TFS-to-envelope metric in the same spectral band (Eq. 7.6) attempts to reduce the effect of an increase in central gain following noise exposure, which is expected to increase the absolute power in both TFS and envelope. It is possible that the effect of central gain could enhance $ENV_{FFR}(t)$ slightly more than $TFS_{FFR}(t)$ because central gain enhances responses of the caudal generators that contribute more to $ENV_{FFR}(t)$ than $TFS_{FFR}(t)$ (King et al., 2016); however, this relative effect is expected to be much smaller than the absolute effect on both TFS and envelope. Nonetheless, results in this study show the opposite trend, i.e., $TFS_{FFR}(t)$ was enhanced over $ENV_{FFR}(t)$, providing strong support for the idea that FFRs are systematically influenced by a disruption in tonotopic mapping and thus have diagnostic potential.

7.4.3 Diagnostic potential of the speech-based DT_{slope} metric

Effects of DT on complex-sound coding are considerably under-explored and likely to be affected by an array of physiological changes. It is possible that neural metrics based on simple tones such as FTC threshold and tip-to-tail ratio are inadequate to capture the full range of DT-related deficits for complex sounds. For example, degradation in suppressive nonlinearities is more severe for noise-exposed animals than for aged animals with similar threshold shifts (Schmiedt et al., 1990). Similarly, rate-level functions to broadband stimuli and tonal stimuli show systematic differences between normal-hearing and noise-exposed animals (Heinz and Young, 2004). The list of other nonlinear factors that may be important for speech coding, but that are not captured in FTCs, includes cochlear compression and efferent feedback systems like the middle-ear muscle reflex and the medial olivocochlear reflex. Thus, it is clear that a thorough characterization of DT and its consequences warrants the use of natural complex stimuli.

Significant individual variability in speech intelligibility remains across hearing-impaired listeners even after compensating for audibility loss. A recent study found clear differences in the neural coding of suprathreshold complex stimuli at the level of the auditory nerve between animal models of noise-induced and age-related hearing loss, the two most common hearing-loss etiologies (Henry et al., 2019). Even when the two groups had a similar degree of hearing loss, the effects of DT were more severe for the noise-exposed group. Thus, variations in DT may play a role in individual variability in speech perception.

The DT_{slope} metric (Fig. 7.8; Eq. 7.6) provides a noninvasive means to assess the degradation of cochlear tonotopic mapping. DT_{slope} was both highly sensitive and specific to hearing status in the group of animals studied here (Fig. 7.8). Moreover, DT_{slope} was highly correlated with ABR threshold and DPOAE amplitude (Fig. 7.9), demonstrating the potential merit of DT_{slope} in predicting the degree of noise-induced hearing loss. Based on data from Henry et al., 2019, it is reasonable to speculate that DT_{slope} for age-related hearing loss would fall in between the normal-hearing and noise-exposed groups included in the current study.

While understanding changes in neural coding of speech following hearing loss is important, it may not be an optimal stimulus for diagnosing distorted tonotopy. To this end, spectral properties of segments that maximally separated the normal-hearing and hearing-impaired groups were identified (Fig. 7.10), which can guide the development of more optimal diagnostic stimuli (*e.g.*, multi-band harmonic complexes). Interestingly, two distinct neural coding profiles emerged separating the mechanisms underlying enhanced envelope and enhanced TFS following acoustic trauma. For segments with more high-frequency energy concentration (e.g., purple spectrum in Fig. 7.10), neural envelope was significantly enhanced relative to TFS for the hearing-impaired data. On the other hand, for stimuli with low-frequency energy concentration (e.g., green spectrum in Fig. 7.10), neural TFS was significantly enhanced relative to envelope for the hearing-impaired group. These results supplement the results of Henry et al., 2019 in that the effects of distorted tonotopy on neural coding of complex signals not only depend on the system (*e.g.*, normal, noise-exposed, or metabolic) but also on the spectral properties of the stimulus used (*e.g.*, pink or white spectrum).

This study was motivated by tonotopic changes known to occur in the auditory nerve following noise-induced hearing loss, and the assumption that neurons in the brainstem and midbrain, which contribute significantly to the FFR, inherit their tuning from auditory-nerve fibers. Future studies investigating neural coding following noise-induced hearing loss at the level of the inferior colliculus (both at the single-neuron and gross-potential levels) will be helpful in directly understanding the effects of distorted tonotopy in the midbrain. Similarly, it should also be noted that evidence to date for DT is based on anesthetized-animal data following single discrete noise exposures (*e.g.*, chinchillas here; cats in Miller et al., 1997), and has not been studied directly in awake humans with a variety of sensorineural hearing loss (SNHL) etiologies. If DT were to be validated in humans with SNHL, metrics such as the DT_{slope} would provide new avenues for characterizing the perceptual consequences of DT, as well as for developing and evaluating hearing-aid approaches to optimally reduce the deleterious effects of DT. Results from the current study predict that enhancing perceptually informative envelope cues and minimizing low-frequency TFS features would improve speech-intelligibility outcomes; however, these predictions remain to be tested.

8. SUMMARY AND PERSPECTIVES

Hearing loss still affects the daily life of many individuals despite state-of-the-art interventions. A key factor contributing to the limited benefit of these interventions is their primary dependence on the audiogram. The audiogram, which indicates hearing sensitivity in noise-attenuating sound booths, reflects the activity of only the most sensitive group of neurons. Perception of complex loud sounds in noisy environments is likely mediated by information integration across many more neurons compared to those needed to detect tones in quiet environments at near-threshold intensities. Therefore, we need a better understanding of the suprathreshold effects of hearing loss on complex sound coding.

Here, we used a chinchilla model of mild-moderate hearing loss to study neural-coding changes of speech in noise following NIHL. First, a quantitative framework was developed in Chapter 3 that permits direct comparison of invasive spike-train data with noninvasive evoked data. Thus, this framework boosts the translational aspects of animal models. In addition, several spectrally specific analyses were developed in this framework to characterize responses to nonstationary signals. For example, unlike spectrogram-based analyses, frequency-demodulation-based analyses, such as the harmonicgram, facilitate the same spectral resolution to analyze stationary and nonstationary signals. In Chapter 4, this framework was applied to study the effects of hearing loss on the coding of natural speech in noise. In particular, the coding of vowels and consonants were evaluated in terms of TFS and envelope responses, respectively. Our data contradicted several existing hypotheses (e.g., our data did not show any compression of ANF thresholds or reduction in spike-timing precision). Instead, our results showed that vowel coding deficits primarily stem from distorted tonotopic mapping following acoustic trauma without a significant change in driven rate or spike-timing precision. In contrast, neural coding deficits for low-intensity consonants reflected uncompensated audibility for most neurons despite the linear gain of 15 dB, which was based on ABR thresholds. These effects on vowel and consonant coding were also robustly captured in the FFR responses. Adding background maskers had a more deleterious effect on envelope and TFS coding of speech for the hearing-impaired group compared to the normal group, for both stationary and fluctuating maskers; this neural overrepresentation of broadband masking noise likely contributes to the increased noise-distractibility that many listeners with hearing loss experience. In Chapter 6, the temporal and temporal-place representation of pitch were evaluated based on AN data from both groups. The temporal representation for impaired neurons showed more complex patterns. The temporal-place representation was significantly degraded for the impaired group. Overall, these findings show that increased susceptibility to masking noise on top of these severe degradations of envelope, TFS, and pitch coding for speech following NIHL result primarily from distorted tonotopy and, to a lesser extent, from broader auditory filters. These deficits potentially underlie perceptual deficits that many listeners with hearing impairment frequently experience. In the following, these issues are discussed in more detail.

8.1 Benefits of using *apPSTHs*

The effects of SNHL has been extensively studied using invasive anatomical and physiological data collected from animal models of SNHL (Henry et al., 2016; Kale and Heinz, 2010; Liberman, 1984; Liberman and Dodds, 1984a, 1984b; Liberman and Klang, 1984; Miller et al., 1997). In contrast, only a handful of studies have investigated the effects of SNHL on noninvasive data collected from these animals (Salvi et al., 1990; Trautwein et al., 1996; Zhong et al., 2014). Although these studies have provided great insights into the physiological changes following SNHL, the relation between invasive and noninvasive data is not always clear as these data are often collected from different animal cohorts. Establishing this relation will allow us to draw inferences regarding the anatomical and physiological state of human subjects based on noninvasive data. However, as the binary-valued invasive spike-train data and discrete-valued discrete-time noninvasive far-field data are essentially of different data types (even though the responses themselves are related), there is a major disconnect between the temporal-coding analyses employed for these different responses.

To address this gap, we proposed a unifying quantifying framework that allows the same spectrotemporal analyses to be applied to spike-train data and evoked far-field data (Chapter 3). The framework was extensively applied to spike-train data from auditory-nerve fibers and evoked frequency following responses in Chapters 4 to 7. In this framework, spike trains are analyzed using PSTHs, which are essentially discrete-valued and discrete-time signals, similar to evoked data. By using both positive and negative polarities of the stimulus, a set of PSTHs can be derived (collectively called *alternative-polarity PSTHs* or *apPSTHs*). These PSTHs analytically relate to popular temporal-coding metrics, such as the vector strength and the correlogram. In fact, correlograms can be more efficiently computed using *apPSTHs* instead of using the conventional spike-time tallying method. Furthermore, *apPSTHs* can be used to derive multitaper spectral estimates, which yield optimal (in terms of reducing bias and variance) estimates of the power spectral density. This approach is particularly important for spike-train data, which are finite and inherently stochastic.

The use of *apPSTHs* opens up the possibility for major advancement in speech-intelligibility modeling, particularly for listeners with hearing-impairment. Existing speech-intelligibility models, which are largely based on audio-domain signal processing, work well for normal-hearing listeners and can predict speech intelligibility for a wide variety of speech manipulations. However, it is not clear how to extend these models to account for speech perception of listeners with hearing loss, as we lack a clear understanding of the various effects of sensorineural hearing loss on speech coding. Preclinical animal models of well-defined pathophysiology can be used to collect spike-train data in response to manipulated speech. This approach can be used to validate existing speech-intelligibility models as well as to identify neural correlates underlying speech perception following SNHL. In this regard, *apPSTHs* outperform correlograms and PSTHs as a wider range of audio-domain speech-intelligibility models can be realized using *apPSTHs*, as *apPSTHs* retain response-phase information as well. The *apPSTH*-based framework can also be used with computational models of AN fibers that have the option to include different etiologies of sensorineural hearing loss.

Temporal coding metrics in the neural domain should be spectrally specific as there is great scope for these metrics to be corrupted, biased, and be variable. Key factors that contribute to these issues include the finite nature of the data, stochasticity of neural systems, and rectifier distortions involved in the inner-hair-cell transduction process. Bias and variance can be optimally reduced by using a multitaper approach as described previously. However, the multitaper spectrum (and correlogram spectrum) are corrupted by rectifier distortions, which is likely an artifact of analysis and does not have any perceptual relevance. Therefore, any broadband metric, like correlogram peak-height, will be biased. Instead, these temporal metrics should be band-limited to include only the desired frequency bands and should exclude rectifier distortions. Restricting the bandwidth of these temporal metrics also helps minimize the effects of neural stochasticity. For example, to compare envelope-coding strength of two harmonic complexes with different fundamental frequencies, power in narrow frequency bands centered at the first few harmonics of the fundamental frequency should be considered (as opposed to total power in the response spectral estimate).

Another benefit of *apPSTHs* is that the (true) phase response can be estimated using the Hilbert-phase PSTH, $\phi(t)$. The difference PSTH, $d(t)$, has been traditionally used as the TFS response. However, as shown in Chapter 3, $d(t)$ contains modulation (envelope) information as it represents the coding of the whole signal (both envelope and TFS, and not just TFS). In contrast, $\phi(t)$ represents a better account of the response TFS in isolation. As the role of envelope and TFS in noisy-speech perception is an active topic of debate (Ding et al., 2014; Shamma and Lorenzi, 2013; Swaminathan and Heinz, 2012), $\phi(t)$ can be used to study the role of TFS in the proposed framework. $\phi(t)$ also holds important implications for cochlear-implant research involving stimulation strategies, as $\phi(t)$ closely relates to the stimulus zero-crossing component (Logan, 1977; Voelcker, 1966b; Wiley, 1981), which is the signal cochlear implants often rely on to provide TFS information (Chen and Zhang, 2011; Grayden et al., 2004).

A major benefit of using *apPSTHs* is the application of frequency demodulation to improve the spectral resolution of temporal analysis. As discussed in section 3.6.1, power along any desired spectrotemporal trajectory can be computed by using frequency demodulation, followed by low-pass filtering. This approach can be used to derive the harmonicgram ($\mathcal{HG}$), a spectrally compact representation for signals with harmonic spectrum. For nonstationary signals, the $\mathcal{HG}$ consists of power along harmonics of the fundamental frequency (F_0), which can vary over time. An example application of the $\mathcal{HG}$ was to derive formant power in spike-train responses in Fig 4.5. This method of only including the harmonics of F_0 to evaluate natural voiced-speech coding is similar to the previously used methods for evaluating formant coding for stationary synthesized vowels (Miller et al., 1997; Young and Sachs, 1979). Overall, $\mathcal{HG}$-based metrics have better spectral specificity, and therefore these metrics are less affected by neural stochasticity.

8.2 Distorted tonotopy significantly affects the coding of several speech features

The key effects of hearing loss have been traditionally thought to include threshold elevation, broader tuning, and reduced ability to phase lock (Salvi et al., 1983; Young, 2012). It is clear that threshold is elevated, and tuning gets broader following NIHL; these effects are also evident in the data presented in this dissertation. However, the majority of neural data do not support the hypothesis that phase locking is reduced following NIHL (Harrison and Evans, 1979; Kale and Heinz, 2010). When spike-timing precision in speech responses was quantified using the across-trial Victor-Purpura distance, which inversely relates to precision, there was no degradation in spike-time precision for the impaired population (Fig 4.2). Thus, the fundamental ability of AN fibers to precisely encode stimulus envelope and TFS features was not degraded following NIHL. Similarly, to ensure that speech-coding deficits for voiced speech are not due to uncompensated audibility, a linear gain of 15 dB (flat across frequency) was used for the hearing-impaired group. Comparison of voiced-speech-driven spike count across both groups confirmed that this linear gain was sufficient to restore driven rate for the impaired population.

The major physiological change that accounts for most of the neural-coding degradations for speech was distorted tonotopy. The effects of distorted tonotopy include reduced sensitivity to near-CF energy, oversensitivity to low-frequency energy (via FTC tail) and lowering of the effective best frequency relative to the characteristic frequency appropriate for cochlear place. These effects have been previously characterized using White Gaussian Noise, which has a flat spectrum across frequency (Henry et al., 2016). Similar effects were also found using a stationary synthesized vowel as the stimulus that had comparable energy at the first three formants (i.e., flat spectrum up to 3 kHz) (Miller et al., 1997). However, natural speech has a negatively sloping spectrum, which could potentially make the effects of distorted tonotopy even more severe. In fact, our results show dramatic changes in the spectral profile of responses, where very low-frequency (< 300 Hz) speech energy dominates the responses of AN fibers with much higher frequency (> 1 kHz). This shift in best frequency is much higher than the one octave shifts reported previously for white Gaussian noise (Henry et al., 2016). Similarly, although broadening of auditory filters likely plays a role in these results, this broadening does not account for the overrepresentation of very low-frequency stimulus energy by impaired neurons (Fig 4.4). Therefore, these spectral distortions in neural responses primarily stem from distorted tonotopy.

From a feature-coding perspective, distorted tonotopy reduces the weight of high-frequency features (e.g., F_2 and F_3) at the expense of overrepresented low-frequency features (e.g., F_1 and the first few F_0 harmonics). However, psychoacoustic studies have shown that speech content within 1 to 4 kHz frequency band is highly informative for speech perception (French and Steinberg, 1947; Pavlovic, 1987). In fact, recent studies have shown that extended high-frequency (up to 10 kHz) information significantly improves speech intelligibility (Levy et al., 2015; Monson et al., 2019; Moore, 2016). However, as a direct result of distorted tonotopy, information above 1 kHz is underrepresented at the peripheral auditory system, which leads to a significant reduction in the channel (information) capacity of the auditory nerve.

8.3 Acoustic overexposure causes ANF threshold distribution expansion, not compression, which leads to coding deficits for low-intensity consonants

Psychoacoustic studies have hypothesized that ANF threshold distributions get compressed following acoustic trauma (Moore et al., 1985; Zeng and Turner, 1990). This compression has been thought to contribute to abnormal growth of masking functions (loudness recruitment) for patients with hearing loss. However, auditory-nerve data are inconsistent with this hypothesis (Heinz et al., 2005). In fact, current and previous published data show an expansion of the ANF threshold distribution as reflected by an increase in standard deviation for these distributions for the impaired group (see Table 4.1). This threshold distribution expansion has important consequences for speech coding as described below.

Even though speech was played at a fixed overall intensity for each group, short-term intensity for speech can vary over a wide (up to ∼30 dB) range. Therefore, the effects of hearing loss may differ across speech tokens that have different intensities. For example, applying a flat 15-dB gain for the impaired population equalized the driven rate across both groups in response to voiced speech, which typically has higher intensity. In contrast, the driven rate for low-intensity unvoiced speech was significantly lower for the noise-exposed population. These differential effects for voiced and unvoiced speech likely reflect threshold distribution changes following noise exposure. In particular, ABR threshold likely reflects the neural activity of the most sensitive group of neurons and may not capture population-level threshold changes. For example, when post-exposure shifts in the 10th and 50th percentiles were quantified based on ANF threshold distribution, the shift in the 50th percentile point was always greater than the shift in the 10th percentile point for all frequency bands tested, for both chinchilla data and cat data. Therefore, any audiometric assay that is based on the lower tail of ANF threshold distributions is likely to underestimate population level audiometric deficits.

8.4 Susceptibility to masking noise increases following NIHL

The number one complaint in the audiology clinic is regarding difficulty understanding speech in noisy environments (Chung, 2004; Souza, 2016). To understand the neural basis underlying these perceptual difficulties in noise, we recorded spike-train data in response to speech in the presence of stationary or fluctuating masking noise at several SNRs. Voiced-speech coding was evaluated by quantifying the strength of the trajectory of the first three formants in the response. The first three formants carry sufficient information to support robust vowel identification (Fant, 1973; Klatt, 1982). Furthermore, formant trajectories of F_2 and F_3 are informative to understand preceding and following speech tokens due to coarticulation effects (Lindblom and Studdert-Kennedy, 1967; Mann, 1980). In the hearing-impaired population, the representation of F_2 and F_3 was already degraded for speech alone (i.e., without noise). Adding stationary noise at 0 dB SNR significantly affected F_3 coding for the hearing-impaired population, but not for the normal-hearing population. Similarly, stationary noise at -5 dB SNR significantly affected F_2 coding for the hearing-impaired population. Both distorted tonotopy and broader tuning likely contribute to this increased susceptibility to noise for the impaired population. However, distorted tonotopy is likely the dominant factor because the noise, which was speech-shaped, had substantial energy at low frequency.

The effect of noise on neural coding was also quantified for the fricative /s/. This fricative had a favorable SNR even when the stationary masker was mixed with speech to generate an overall SNR of -10 dB. For the normal-hearing population, low and medium SR AN fibers demonstrated robust noise-resistant responses at all SNR conditions. This is consistent with results from previous studies that show low and medium SR fibers selectively respond to high-frequency speech energy without responding to speech-shaped noise in any significant way (Geisler and Silkes, 1991; Silkes and Geisler, 1991). These previous studies used only nonnegative SNRs. Our results extend these findings to show that the noise-resistance responses of low and medium SR AN fibers are substantially better than high SR fibers even at -5 dB SNR. In contrast to the normal-hearing population, the low and medium SR AN fibers in the hearing-impaired population showed significantly poorer fricative coding in quiet and in noise. Distorted tonotopy and adaptation likely contributed to these effects. Impaired high-frequency neurons responded less to fricative energy due to elevated FTC tips and overresponded to low-frequency energy in speech and in background noise as a result of distorted tonotopy. Similarly, as these high-CF impaired AN fibers were responding to preceding voiced speech and noise, it is likely that these AN fibers were affected by adaptation and responded less robustly to the fricative.

Envelope is critically important for speech intelligibility (Shannon et al., 1995; Smith et al., 2002). Envelope coding in AN-fiber responses is enhanced following NIHL (Henry et al., 2016; Kale and Heinz, 2010). However, this enhancement is deemed pathological and is hypothesized to adversely affect speech perception in noisy backgrounds (Füllgrabe et al., 2003; Kale and Heinz, 2010; Moore et al., 1996). In particular, this enhanced noise envelope is thought to contribute to a lack of masking release for fluctuating maskers for listeners with hearing impairment. The phenomenon of fluctuating masking release refers to the improved speech perception by normal-hearing listeners when the masker has fluctuations in the envelope (compared to stationary noise with a flat envelope) (Festen and Plomp, 1990). To test the effect of NIHL on the envelope coding of speech, response envelopes were extracted from AN fiber PSTHs using a multiresolution modulation filter bank for speech-alone, noisy speech, and noise-alone conditions. This approach is a neural-domain extension of well-established speech-intelligibility models that are based on speech-signal processing (Dau et al., 1999; Jørgensen et al., 2013; King et al., 2019). Speech coding fidelity, quantified using response-envelope correlation for speech-alone and noisy-speech, was significantly reduced for the impaired population for both stationary and fluctuating maskers. More importantly, noise representation, quantified using response-envelope correlation for noise-alone and noisy-speech, was enhanced following NIHL in both stationary and fluctuating masker conditions. This increased noise representation highlights the more distracting nature of noise for listeners with hearing impairment. Noise representation was significantly enhanced for the fluctuating masker compared to the stationary masker. Any benefit in perception due to release of masking is substantially reduced because this enhanced noise representation potentially leads to greater distraction by the fluctuating masker.

8.5 Implications

8.5.1 Diagnosing distorted tonotopy noninvasively

Based on the results in this dissertation, distorted tonotopy could be an important factor contributing to individual variability in speech perception. For example, the degree of distorted tonotopic mapping can vary across hearing-loss etiologies even though the degree of hearing loss is similar (Henry et al., 2019). NIHL leads to more severe distorted tonotopy compared to metabolic hearing loss. However, whether distorted tonotopy occurs in humans or not is currently unknown. The exact degree of distorted tonotopy depends on the specific pattern of hair cell and/or stereocilia damage, which can vary across individuals based on noise-exposure history (Liberman, 1984; Liberman and Dodds, 1984b). Therefore, we need a noninvasive metric to diagnose DT in humans.

Frequency-following responses (FFRs) hold excellent promise to diagnose DT noninvasively primarily for two reasons. First, FFRs receive significant contributions from neurons tuned to high frequencies. The effects of DT are dramatic at these high frequencies. For normal AN fibers, responses are dominated by near-CF-energy-driven envelope components more so than TFS because phase-locking is weaker at high frequencies. In contrast, neurons demonstrating DT effects encode low-frequency stimulus energy via FTC tails responses, which shows up primarily as TFS responses. These differences in the weight of envelope and TFS components of high-frequency neurons makes FFR an attractive tool. A second reason that could also improve the sensitivity of the FFR as a diagnostic tool for DT is the compression of AN fiber latency profiles following NIHL. Because of this latency compression, the range of high-frequency neurons that are mutually synchronous extends to even lower CFs. As a result, a larger number of neurons with DT are recruited, which likely results in larger FFR TFS responses (e.g., Chapter 7).

The interaction between DT and the stimulus timbre (spectral slope) can be leveraged to form a sensitive DT diagnostic metric as described in Chapter 7. For stimuli with white spectrum, FFR responses for hearing-impaired animals are dominated by polarity-tolerant envelope components. This is because, for white noise filtered through FTCs affected by DT, there is near-equal representation of a broad range of frequencies, which leads to enhanced envelope. In contrast, stimuli with negatively sloping spectrum excite the tails of FTCs affected by DT substantially more than the tip because of high stimulus energy concentration at these tail frequencies. Future experiments that systematically study the effect of spectral slope and stimulus intensity on the representation of DT in FFRs will be helpful in developing an optimal diagnostic metric for DT based on FFRs.

8.5.2 Hearing-aid strategies

Even though our noise-exposure protocol yielded a flat-frequency hearing loss, neural representations of high-frequency speech features (such as F_2 and F_3 for voiced speech and spectral peaks in the fricative /s/) were significantly degraded compared to low-frequency features (such as F_0 and F_1). In fact, these low-frequency features were overrepresented in the responses of impaired AN fibers due to DT. Therefore, to counter these DT effects, amplifying high-frequency features and limiting the masking ability of low-frequency features are hypothesized to improve the neural coding of speech. These processing schemes apply specifically to hearing-impaired populations that are diagnosed to have DT.

Our results showed differential audibility effects for high-intensity and low-intensity speech tokens. Short-term intensity for speech can vary over a ∼30-dB range. In our data, neural activity was restored for high-intensity voiced speech but not for low-intensity unvoiced speech. In fact, these results are consistent with psychoacoustic studies that have reported uncompensated audibility for consonants at a gain that is sufficient for vowel perception (Phatak and Grant, 2014; Phatak et al., 2009). As described earlier, expansion of ANF threshold distributions potentially contributes to these effects. These results highlight the importance of adaptive gain control strategies, where gain can be computed based on short-term spectral properties to mitigate these differential effects.

Restoring overall audibility may not be sufficient to support near-normal (unvoiced) consonant perception for listeners with hearing impairment. That is because neural coding of consonants (e.g., stop consonants or fricatives) are mediated by sustained as well as onset cues for normal AN fibers (Delgutte, 1980; Heil, 2003). In contrast, onset responses were significantly reduced for impaired neurons, more so than sustained responses (e.g., for /s/). Adaptation as a result of overresponding to preceding voiced speech segments could have contributed to these onset reductions. Promisingly, a previous study has shown that onset rate is similar (or even enhanced) for impaired neurons when tones are played at equal sensation level for both (normal and impaired) groups (Scheidt et al., 2010). Therefore, it could be possible to enhance onset responses by using higher intensities. Another way onset responses could be enhanced is by using a steeper rise time for these low-intensity consonants. Although neural evidence for enhanced onset with steeper rise time is limited to interaural time difference encoding by neurons in the inferior colliculus (Dreyer and Delgutte, 2006; Griffin et al., 2005), it is likely that onset coding of consonants presented monaurally can also be improved in a similar way. Future research investigating these effects of envelope shape on consonant perception will be informative.

Psychoacoustic studies have highlighted the importance of high-frequency (4 to 10 kHz) information in speech although most hearing-aids rarely go over 4 kHz (Boyd-Pratt and Donai, 2020; Levy et al., 2015; Monson et al., 2019; Moore, 2016). Results presented here demonstrate that usable speech information is encoded by auditory-nerve fibers with CF at these high frequencies, even at -10 dB SNR. In particular, AN fibers with low and medium spontaneous rate show noise-robust representations. Therefore, it may be worthwhile to provide these extended high-frequency cues if listeners have sufficient residual high-frequency hearing.

8.5.3 Both speech-coding fidelity and noise-related distractions are important factors to consider for speech perception

Speech perception in the presence of broadband (speech-shaped) noise could be adversely affected by two related but distinct factors. These factors are: (1) reduced speech-coding fidelity and (2) enhanced noise representation. Speech-coding fidelity refers to the available speech-related information, which is already reduced for clean-speech (i.e., without background noise) following NIHL. Enhanced noise representation emerges when background noise is present. This noise-enhancement can be distracting for listeners to make use of the available speech-related information. Noise representation was enhanced for the impaired AN-fiber population as a direct result of distorted tonotopy and broader tuning because low-frequency energy in speech-shaped broadband noise dominated the response of impaired fibers. These effects reflect the reduced ability of impaired AN fibers to use high-SNR spectrotemporal glimpses because their broader tuning allows substantial noise energy to pass through the filter. These results are consistent with perceptual results in that speech perception by listeners with hearing impairment is substantially affected by background noise. Overall, these results highlight the need for robust noise-reduction algorithms for any communication channel, e.g., hearing aids and telephone lines, to mitigate the effects of this increased susceptibility to masking noise. Furthermore, design of these communication applications as well as any acoustic space (e.g., auditoriums and lecture halls) should consider both maximizing speech-coding fidelity and minimizing noise-related distractions.

8.5.4 Implications for modeling studies

Speech-intelligibility (SI) models quantify the effects of acoustic manipulations on speech intelligibility. More importantly, SI models can be used to optimize intervention strategies for listeners with hearing impairment in an objective framework (Heinz, 2015). However, existing SI models work well for normal-hearing listeners but not for listeners with hearing impairment because of our limited understanding of the various effects of sensorineural hearing loss on speech coding. Our results offer two key insights into how these SI models can be improved and extended to account for hearing-loss effects on SI. First, SI models should consider the effects of both speech-coding fidelity and noise-related distractions.

Second, the front-end of these SI models should accommodate physiological differences across individuals. Physiological effects of sensorineural hearing loss can dramatically vary across hearing-loss etiologies. For example, inner- and outer-hair-cell-specific hearing loss of a similar degree can lead to widely different neural representations (Axe, 2017). Similarly, even at the hair cell level, damage to tip link or stereocilia can result in different neural coding profiles (Clark and Pickles, 1996). Therefore, these physiological factors should be included in SI models. A few existing SI models have the provision for inner- versus outer-hair-cell damage (Scheidiger et al., 2018; Zaar et al., 2019); however, including more precise fine-grained options (e.g., stereocilia versus tip link damage) will likely lead to more accurate SI models. Future research establishing the various effects of these different physiological insults on speech coding and developing noninvasive assays to diagnose these different trauma types will improve modeling outcomes.

8.6 Limitations

8.6.1 Limitations of the dataset

As with any neurophysiological study, our experimental design was based on recordings from a limited number of AN fibers for either group. As a result, sampling of CFs and SRs may be biased based on the exact electrode placement in individual experiments. To minimize this bias, AN-fiber data were pooled across animals, and multiple sequential electrode placements were used for each animal. Despite these efforts, these data may still be biased. For example, CF and SR distributions were different between the two groups. SR distribution for the impaired group was shifted to lower SRs; this effect likely reflects damage to inner hair cell stereocilia (resulting in reduced transduction currents) more so than sampling differences (Liberman and Dodds, 1984a). Similar to SR distributions, CF distributions were also different between the two groups. We attempted to minimize these differences by comparing CF-band averages across groups for most neural metrics. Another limitation was the relatively sparse sampling of very low-frequency (<500 Hz) AN fibers.

Therefore, it is difficult to confidently comment on speech-coding degradations for fibers at these low frequencies. However, this is a minor limitation because of two reasons. First, speech information is maximally concentrated between the 0.5 to 4 kHz band, which was well represented in our dataset. Second, clinically common hearing-loss etiologies include greater damage at high frequency compared to at very low frequency (Parthasarathy et al., 2020).

Another limitation is the lack of histological assessments to confirm the nature of hair-cell damage in our dataset. Instead, physiological inferences were made as a proxy to anatomical labeling methods (Liberman, 1984). For example, CFs were estimated based on the high-frequency-side slopes of FTCs (Liberman, 1984). Reduction in DPOAE levels and reduced Q_{10} values were used as a confirmation for significant OHC damage.

Data were collected in response to a single speech sentence and two (stationary and 8-Hz sinusoidally amplitude modulated noise) frozen maskers. The exact phase relationship across these stimuli could have systematic effects on the neural responses. This is particularly true for speech in the presence of the fluctuating masker. However, these phase effects are not expected to systematically vary across groups because the same set of stimuli was used for both groups. Future studies documenting the effects of several maskers and speech sentences will be helpful in replicating and generalizing these results.

The number of intensities used per stimulus was limited due to AN-fiber recording-time limits ($\sim$15 min/fiber on average). Therefore, our data lacks a thorough characterization of the effects of intensity variation. Another potential confound could be the difference in intensities used for different groups (i.e., 65 dB SPL for the normal-hearing group and 80 dB SPL for the hearing-impaired group). A higher intensity could partly contribute to the distorted-tonotopic effects for the impaired population. However, this intensity difference is unlikely to play a significant role because of the following reasons. First, FFR data collected at an equal-intensity (70 dB) paradigm for both groups showed distorted tonotopy effects for the impaired population, similar to AN data. Second, distorted tonotopy is evident in impaired AN-fiber responses when stimuli are presented only 10 dB above threshold (Henry et al., 2016). Finally, temporal responses of normal AN fibers show distorted tonotopic effects only at very high (>100 dB SPL) intensities and are relatively narrowband below 100 dB SPL (Wong et al., 1998). Therefore, the use of fixed intensities and the difference in these intensities between groups should not affect the conclusions presented in this dissertation in any significant way.

8.6.2 Anesthetized animals as a model of the (awake) human auditory system

While single unit responses and FFRs data reported here provide crucial insights into the neural-coding alterations following noise-induced hearing impairment, they do not reflect other central or feedback (efferent) processes that may be important for speech understanding. One such feedback process is the medial olivocochlear reflex, which can dynamically change the input-output function of single nerve fibers (Guinan, 2006; Liberman and Guinan, 1998). However, in our anesthetized-chinchilla animal model, medial olivocochlear reflex effects are expected to be small (Aedo et al., 2015). Similarly, the effects of MEMR are likely to be significantly reduced. Another important factor is the role of attention, which can critically affect speech perception performance and perceptual tasks in general (Knudsen, 2007; Lu et al., 2017; Shinn-Cunningham and Best, 2008). However, the analyses used in this dissertation were limited to quantifying the information available in the periphery and midbrain and comparing this information between groups. Ways to overcome these limitations include using single-/multi-unit recordings or evoked recordings such as EEG and FFR from awake animals, using ecologically relevant stimuli for animals, or training the animal to discriminate speech tokens.

8.6.3 Other physiological factors affecting neural coding

Several other physiological factors are important to consider to thoroughly characterize the effects of sensorineural hearing loss on speech coding. However, the effects of these factors, such as suppression, adaptation, and latency, were not systematically quantified in this dissertation. It should be noted that the effects of these factors were likely present in our data, unlike the effects of efferent systems, which are substantially reduced due to anesthesia. Suppression is reduced following acoustic trauma (Miller et al., 1997; Sayles et al., 2016; Schmiedt et al., 1990). In fact, the strength of suppression can vary across hearing loss etiologies (Schmiedt et al., 1990). For example, reduction in suppression is more severe for NIHL than for age-related hearing loss. Temporal dynamics (i.e., onsets/offsets) can also differentially affect the responses from the two groups because AN fibers become less frequency selective following NIHL. Compared to normal AN fibers, impaired AN fibers responded (nonselectively) to a greater portion of the speech stimulus (e.g., Fig 4.8). Therefore, the effects of temporal dynamics or adaptation could be substantially different between the two groups. Similarly, acoustic trauma reduces latency of AN fiber responses (Henry et al., 2014; Scheidt et al., 2010). These changes in latency profiles hold important implications for across-fiber coding schemes for speech (Heinz et al., 2010). Future studies and analyses investigating the role of these physiological factors on speech coding following sensorineural hearing loss will be insightful.

8.6.4 Biological feasibility of the analyses employed

The approach adopted in this dissertation was to quantify information available in the periphery for normal-hearing and hearing-impaired animals. A variety of metrics (from simple rate count to more sophisticated cepstral analysis) were used to quantify the coding strength of several stimulus features. All these methods may not be strictly biologically feasible. Nevertheless, the brain may have alternative neural representations of these features that correlate with the metrics used in this dissertation. For example, a modulation filter bank was used to estimate central envelope representation. Although the exact filters may not exist in the brain, a related representation may exist in the responses of modulation-sensitive neurons in the midbrain (Joris et al., 2004; Krishna and Semple, 2000; Nelson and Carney, 2007). Similarly, spectral (phase-locking) information may be extracted by monaural coincidence detection mechanisms (Young, 2008). Although there is no evidence to support that neural circuits are able to perform cepstral analysis, a metric based on simple across-CF coincidence detection could be proportional to cepstral peak height of the ALSR.

8.7 Future Work

This dissertation has made three broad significant contributions. First, it has established a framework where invasive spike-train data and noninvasive evoked responses can be analyzed using the same advanced signal processing tools. Second, by applying this framework to invasive and noninvasive data in response to speech in noise, this dissertation has identified neural correlates of several perceptual deficits that listeners with hearing impairment experience during everyday communication. In particular, DT may have dramatic implications for human speech perception. Finally, this dissertation has developed a FFR-based metric to noninvasively diagnose DT in humans. These contributions pave the way for several potential avenues of future research as described below.

8.7.1 New research avenues using the unifying framework

A key benefit of the quantitative framework presented in Chapter 3 is that by using *apPSTHs* to analyze spike-train data, the same metrics can be used to analyze both spike-train data and evoked responses. These *apPSTHs* include existing [e.g., $d(t)$] as well as new [e.g., $\phi(t)$] PSTHs. The study of response TFS has traditionally been carried out using $d(t)$, or using the *difcor*, which is related to the autocorrelation function of $d(t)$. However, $d(t)$ and *difcor* may quantify the total coding (i.e., both envelope and TFS) of the signal and not just the TFS, particularly for low-frequency stimuli (see Chapter 3). In contrast to $d(t)$, $\phi(t)$ represents the Hilbert-phase of the response, and therefore is a more appropriate representation of the response TFS. As recent studies have highlighted the role of response TFS on speech perception in noisy environments (Ding et al., 2014; Shamma and Lorenzi, 2013), $\phi(t)$ would be useful in accurately quantifying TFS components in neural responses.

Another example of an advanced signal processing approach possible because of using *apPSTHs* is the use of frequency demodulation and filtering to estimate the strength of dynamic spectrotemporal components (e.g., formants in voiced speech) at high spectral resolution. Narrower spectral resolution leads to more specific spectral estimates. Similarly, the harmonicgram represents a spectrally compact representation for signals with a harmonic spectrum (e.g., voiced speech). These spectrally specific metrics are advantageous to use compared to previous broadband metrics for studying the neural representation of natural and complex stimuli, in healthy and diseased auditory systems. Several such examples were presented throughout this dissertation. Examples of future research questions include studying the coding of frequency-modulated tones following inner-hair-cell-specific damage with high spectral resolution (unlike previous studies that used spectrogram-like approaches (Axe, 2017). Psychoacoustic studies have hypothesized that inner-hair-cell-specific damage can lead to neural-coding deficits for frequency-modulated tones. Using the harmonicgram (or the frequency demodulation) approach, representation of frequency-modulated tones can be studied at high spectral resolution.

8.7.2 Role of distorted tonotopy on speech perception

Although several dramatic effects of distorted tonotopy were reported in this dissertation, whether distorted tonotopy affects humans or not is currently unknown. Therefore, a timely research question is to test for distorted tonotopy in human listeners. Primary evidence for distorted tonotopy is based on invasive spike-train data recorded from AN fibers (Henry et al., 2016; Miller et al., 1997). As similar spike-train data are unlikely to be recorded from human listeners in the near future (Verschooten et al., 2019), we have to resort to noninvasive assays to diagnose distorted tonotopy. In this regard, a metric similar to the DT_{slope} metric described in Chapter 7 will prove useful for testing DT in humans.

While speech was an ideal signal to quantify neural coding deficits due to distorted tonotopy, it may not be the most optimal stimulus for a diagnostic metric. Instead, multi-band harmonic complexes could be useful to develop an efficient diagnostic metric. For example, consider a stimulus, $x[n]$, which is a mixture of two harmonic complexes, $x_1[n]$ and $x_2[n]$. Say $x_1[n]$ has a fundamental frequency of $F0_1$, and it only contains lower-order harmonics such that the spectral power of $x_1[n]$ is restricted to below a frequency, F_{co}. Similarly, say $x_2[n]$ has a fundamental frequency of $F0_2$, and it only contains higher-order harmonics such that its spectral bandwidth is restricted to above F_{co}. Spectral properties of $x_1[n]$ and $x_2[n]$ can be independently manipulated. Such a multiband harmonic complex ($x[n] = x_1[n] + x_2[n]$) can be an efficient diagnostic stimulus to record FFRs. FFRs from listeners without distorted tonotopy are expected to show large $F0_2$-based envelope components because of the high-frequency bias of FFRs. In contrast, listeners with distorted tonotopy should show enhanced $F0_1$-based TFS components in the FFR response. The exact parameters (e.g., $F0_1$, $F0_2$, and F_{co} frequencies and intensity associated with individual harmonic complexes) can be adjusted to develop an optimal diagnostic metric.

If DT were to be demonstrated in humans, it would be important to document the role of DT on speech perception. Based on the results presented in this dissertation, DT could have dramatic effects on speech perception as it severely degrades any temporal-place coding scheme. However, whether this hypothesis holds true for humans remains to be seen.

8.7.3 Disentangling DT effects from broader-tuning effects

Traditionally, threshold elevation and broader bandwidth have been thought to be the major factors affecting speech perception for listeners with hearing loss (Shrivastav, 2012; Young, 2012). Data presented in this dissertation suggest that distorted tonotopy could have an even greater adverse effect of speech perception. For example, the dramatic reduction in near-CF power relative to low-frequency power in the response spectrum (Fig 4.4) is unlikely due to bandwidth broadening alone. While a broader bandwidth can reduce near-CF power, it does not explain the increase in low-frequency power. These qualitative differences between the effects of DT and broader bandwidth on speech coding can be formally quantified in at least two ways as described below.

First, information theoretic analysis can be used to disentangle the effects of these two factors by estimating the channel capacity of the auditory nerve. Approximate models of the auditory filter could be used to capture important tuning properties with and without distorted tonotopy. For example, tip-to-tail ratio of AN fiber FTCs can be varied to control the degree of DT. The effects of broader bandwidth can be thought of as to reduce the number of independent channels in the system. For example, consider three channels with CF = 1 kHz, 1.25 kHz, and 1.5 kHz. All these channels could be independent in a normal-hearing system. For a system with broader bandwidth, channels with CF = 1 kHz and 1.25 kHz may become dependent whereas channels with CF = 1 kHz and 1.5 kHz may still be independent. In contrast to broader bandwidth, DT leads to significant across-fiber dependency (e.g., between AN fibers with CF = 1 kHz and 1.5 kHz) that spans many octaves of frequencies, particularly for stimuli with pink spectrum (e.g., speech). Therefore, DT would likely have a substantially more severe effect on the channel capacity compared to broader bandwidth.

Second, modeling and machine learning approaches can be combined to further investigate the differences between the effects of DT and broader tuning on phoneme confusion. Models of DT and broader bandwidth, similar to those described in the previous paragraph, can be used to derive peripheral representations in response to several phonemes. Classifiers can be constructed for both systems using the same training and validation datasets (e.g., phonemes in the background of stationary noise and reverberation). A novel dataset (e.g., phonemes masked by multi-talker babble) can be used to test the efficiency of individual classifiers. Results presented in this dissertation suggest that a classifier based on the DT-affected system will perform worse than a classifier based on the system with broader tuning.

8.7.4 Effects of different hearing-loss etiologies on speech in noise coding

Substantial variability in speech perception remains despite compensating for audibility loss. A major source of this variability is hypothesized to be peripheral differences. In this dissertation, speech coding deficits were explored in a specific model of hearing loss, which is a chinchilla model of noise-induced hearing loss due a two-hour long discrete noise exposure. The profiles of inner- and outer-hair-cell damage could potentially make this model a good model for some human listeners. However, these damage profiles vary widely depending on parameters of the noise exposure and species (Kujawa and Liberman, 2019). The exact pattern of hair-cell/stereocilia damage for a human listener with a history of life-long exposure to moderate sounds may be quite different from the pattern observed for chinchillas exposed to a single discrete loud noise. Investigating speech-coding deficits for a variety of hearing-loss etiologies (e.g., blast exposed, hair-cell-specific damage, metabolic hearing loss) and relating physiological data with anatomical and evoked data will be critical in improving our understanding of the effects of peripheral factors on speech-intelligibility variability. In this regard, establishing the effects of specific hearing-loss protocols on the severity of distorted tonotopy and ANF threshold distribution will be particularly important.

REFERENCES

Aedo, C., Tapia, E., Pavez, E., Elgueda, D., Delano, P. H., & Robles, L. (2015). Stronger efferent suppression of cochlear neural potentials by contralateral acoustic stimulation in awake than in anesthetized chinchilla. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00021

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hear Res*, *245*(1), 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Allen, J. B., & Li, F. (2009). Speech perception and cochlear signal processing [Life Sciences] [Conference Name: IEEE Signal Processing Magazine]. *IEEE Signal Processing Magazine*, *26*(4), 73–77. https://doi.org/10.1109/MSP.2009.932564

Ananthakrishnan, S., Krishnan, A., & Bartlett, E. (2016). Human Frequency Following Response: Neural Representation of Envelope and Temporal Fine Structure in Listeners with Normal Hearing and Sensorineural Hearing Loss. *Ear Hear*, *37*(2), e91–e103. https://doi.org/10.1097/AUD.0000000000000247

Anderson, S., Parbery-Clark, A., White-Schwoch, T., Drehobl, S., & Kraus, N. (2013). Effects of hearing loss on the subcortical representation of speech cues. *The Journal of the Acoustical Society of America*, *133*(5), 3030–3038. https://doi.org/10.1121/1.4799804

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—a review. *Int J Audiol*, *42*(sup2), 17–20. https://doi.org/10.3109/14992020309074639

Axe, D. R. (2017). *The effects of hair-cell specific dysfunction on neural coding in the auditory periphery* (Doctoral dissertation). Purdue University. West Lafayette, IN.

Babadi, B., & Brown, E. N. (2014). A Review of Multitaper Spectral Analysis. *IEEE Trans Biomed Eng*, *61*(5), 1555–1564. https://doi.org/10.1109/TBME.2014.2311996

Bandyopadhyay, S., & Young, E. D. (2004). Discrimination of Voiced Stop Consonants Based on Auditory Nerve Discharges. *The Journal of Neuroscience*, *24*(2), 531–541. https://doi.org/10.1523/JNEUROSCI.4234-03.2004

Bharadwaj, H. M., Mai, A. R., Simpson, J. M., Choi, I., Heinz, M. G., & Shinn-Cunningham, B. G. (2019). Non-Invasive Assays of Cochlear Synaptopathy – Candidates and Considerations. *Neuroscience*, *407*, 53–66. https://doi.org/10.1016/j.neuroscience.2019.02.031

Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., & Shinn-Cunningham, B. G. (2014). Cochlear neuropathy and the coding of supra-threshold sound. *Front Syst Neurosci*, *8*. https://doi.org/10.3389/fnsys.2014.00026

Bilger, R. C., & Wang, M. D. (1976). Consonant confusions in patients with sensorineural hearing loss. *Journal of Speech and Hearing Research*, *19*(4), 718–748.

Billings, C. J., Bologna, W. J., Muralimanohar, R. K., Madsen, B. M., & Molis, M. R. (2019). Frequency following responses to tone glides: Effects of frequency extent, direction, and electrode montage [tex.ids: billings_frequency_2019-1]. *Hearing Research*, *375*, 25–33. https://doi.org/10.1016/j.heares.2019.01.012

Billings, S. A., & Zhang, H. (1994). Analysing non-linear systems in the frequency domain–II. The phase response. *Mechanical Systems and Signal Processing*, *8*(1), 45–62. https://doi.org/10.1006/mssp.1994.1004

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, *66*(4), 1001–1017. https://doi.org/10.1121/1.383319

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345.

Bones, O., & Plack, C. J. (2015). Subcortical representation of musical dyads: Individual differences and neural generators. *Hearing Research*, *323*, 9–21. https://doi.org/10.1016/j.heares.2015.01.009

Bourk, T. R. (1976). *Electrical responses of neural units in the anteroventral cochlear nucleus of the cat.* (PhD Thesis). Massachusetts Institute of Technology.

Boyd-Pratt, H. A., & Donai, J. J. (2020). The perception and use of high-frequency speech energy: Clinical and research implications. *Perspectives of the ASHA Special Interest Groups*, *5*(5), 1347–1355.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound.* MIT press.

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2002). The Time-Rescaling Theorem and Its Application to Neural Spike Train Data Analysis. *Neural Computation*, *14*(2), 325–346. https://doi.org/10.1162/08997660252741149

Bruce, I. C., Erfani, Y., & Zilany, M. S. A. (2018). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hearing Research*, *360*, 40–54. https://doi.org/10.1016/j.heares.2017.12.016

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. https://doi.org/10.1121/1.1345696

Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *76*(3), 1698–1716. https://doi.org/10.1152/jn.1996.76.3.1698

Cariani, P. A., & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *76*(3), 1717–1734. https://doi.org/10.1152/jn.1996.76.3.1717

Carlyon, R. P., & Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *95*(6), 3541–3554. https://doi.org/10.1121/1.409971

Carney, L. H., & Geisler, C. D. (1986). A temporal analysis of auditory-nerve fiber responses to spoken stop consonant–vowel syllables. *The Journal of the Acoustical Society of America*, *79*(6), 1896–1914. https://doi.org/10.1121/1.393197

Cedolin, L., & Delgutte, B. (2005). Pitch of Complex Tones: Rate-Place and Interspike Interval Representations in the Auditory Nerve. *Journal of Neurophysiology*, *94*(1), 347–362. https://doi.org/10.1152/jn.01114.2004

Cedolin, L., & Delgutte, B. (2010). Spatio-Temporal Representation of the Pitch of Harmonic Complex Tones in the Auditory Nerve. *J Neurosci*, *30*(38), 12712–12724. https://doi.org/10.1523/JNEUROSCI.6365-09.2010

Chambers, A. R., Resnik, J., Yuan, Y., Whitton, J. P., Edge, A. S., Liberman, M. C., & Polley, D. B. (2016). Central Gain Restores Auditory Processing following Near-Complete Cochlear Denervation. *Neuron*, *89*(4), 867–879. https://doi.org/10.1016/j.neuron.2015.12.041

Chen, F., & Zhang, Y.-T. (2011). Zerocrossing-based nonuniform sampling to deliver low-frequency fine structure cue for cochlear implant. *Digital Signal Processing*, *21*(3), 427–432. https://doi.org/10.1016/j.dsp.2010.12.002

Chintanpalli, A., & Heinz, M. G. (2007). Effect of auditory-nerve response variability on estimates of tuning curves. *The Journal of the Acoustical Society of America*, *122*(6), EL203–EL209. https://doi.org/10.1121/1.2794880

Chung, K. (2004). Challenges and Recent Developments in Hearing Aids: Part I. Speech Understanding in Noise, Microphone Technologies and Noise Reduction Algorithms. *Trends in Amplification*, *8*(3), 83–124. https://doi.org/10.1177/108471380400800302

Clark, J. A., & Pickles, J. O. (1996). The effects of moderate and low levels of acoustic overstimulation on stereocilia and their tip links in the guinea pig. *Hearing Research*, *99*(1), 119–128. https://doi.org/10.1016/S0378-5955(96)00092-5

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *Asha*, *23*(7), 493–500.

Clinard, C. G., & Cotter, C. M. (2015). Neural representation of dynamic frequency is degraded in older adults [tex.ids: clinard_neural_2015-1]. *Hearing Research*, *323*, 91–98. https://doi.org/10.1016/j.heares.2015.02.002

Clinard, C. G., Tremblay, K. L., & Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: Evidence from human frequency-following response recordings. *Hearing Research*, *264*(1), 48–55. https://doi.org/10.1016/j.heares.2009.11.010

Colburn, H. S., Carney, L. H., & Heinz, M. G. (2003). Quantifying the Information in Auditory-Nerve Responses for Level Discrimination. *JARO*, *4*(3), 294–311. https://doi.org/10.1007/s10162-002-1090-6

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

Crouzet, O., & Ainsworth, W. A. (2001). On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation. *Workshop on Consistent and Reliable Cues for Sound Analysis*.

Dau, T. (2003). The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *J Acoust Soc Am*, *113*(2), 936–950. https://doi.org/10.1121/1.1534833

Dau, T., Verhey, J., & Kohlrausch, A. (1999). Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. *The Journal of the Acoustical Society of America*, *106*(5), 2752–2760. https://doi.org/10.1121/1.428103

Davis, A. C., & Hoffman, H. J. (2019). Hearing loss: Rising prevalence and impact. *Bull World Health Organ*, *97*(10), 646–646A. https://doi.org/10.2471/BLT.19.224683

Dawes, P., Emsley, R., Cruickshanks, K. J., Moore, D. R., Fortnum, H., Edmondson-Jones, M., McCormack, A., & Munro, K. J. (2015). Hearing Loss and Cognition: The Role of Hearing Aids, Social Isolation and Depression [Publisher: Public Library of Science]. *PLOS ONE*, *10*(3), e0119616. https://doi.org/10.1371/journal.pone.0119616

De Cheveigne, A. (2005). Pitch perception models. *Pitch* (pp. 169–233). Springer.

Delgutte, B., Hammond, B. M., & Cariani, P. A. (1998). Neural coding of the temporal envelope of speech: Relation to modulation transfer functions. *Psychophysical and physiological advances in hearing*, 595–603.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *68*(3), 843–857. https://doi.org/10.1121/1.384824

Delgutte, B. (1997). Auditory neural processing of speech. *The handbook of phonetic sciences*, 507–538.

Delgutte, B., & Kiang, N. Y. S. (1984a). Speech coding in the auditory nerve: I. Vowel-like sounds. *The Journal of the Acoustical Society of America*, *75*(3), 866–878. https://doi.org/10.1121/1.390596

Delgutte, B., & Kiang, N. Y. S. (1984b). Speech coding in the auditory nerve: III. Voiceless fricative consonants. *The Journal of the Acoustical Society of America*, *75*(3), 887–896. https://doi.org/10.1121/1.390598

Delgutte, B., & Kiang, N. Y. S. (1984c). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *The Journal of the Acoustical Society of America*, *75*(3), 897–907. https://doi.org/10.1121/1.390599

Denes, P. B. (1963). On the statistics of spoken English. *The journal of the Acoustical Society of America*, *35*(6), 892–904.

Deng, L., & Geisler, C. D. (1987). Responses of auditory-nerve fibers to nasal consonant–vowel syllables. *The Journal of the Acoustical Society of America*, *82*(6), 1977–1988. https://doi.org/10.1121/1.395642

Diehl, R. L. (2008). Acoustic and auditory phonetics: The adaptive design of speech sound systems. *Philos Trans R Soc Lond B Biol Sci*, *363*(1493), 965–978. https://doi.org/10.1098/rstb.2007.2153

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, *88*, 41–46. https://doi.org/10.1016/j.neuroimage.2013.10.054

Dolphin, W. F., & Mountain, D. C. (1993). The envelope following response (EFR) in the Mongolian gerbil to sinusoidally amplitude-modulated signals in the presence of simultaneously gated pure tones. *J Acoust Soc Am*, *94*(6), 3215–3226. https://doi.org/10.1121/1.407227

Dreyer, A., & Delgutte, B. (2006). Phase Locking of Auditory-Nerve Fibers to the Envelopes of High-Frequency Sounds: Implications for Sound Localization [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *96*(5), 2327–2341. https://doi.org/10.1152/jn.00326.2006

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, *95*(5), 2670–2680. https://doi.org/10.1121/1.409836

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, *95*(2), 1053–1064. https://doi.org/10.1121/1.408467

Dubbelboer, F., & Houtgast, T. (2008). The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. *The Journal of the Acoustical Society of America*, *124*(6), 3937–3946. https://doi.org/10.1121/1.3001713

Dubno, J. R., Dirks, D. D., & Langhofer, L. R. (1982). Evaluation of hearing-impaired listeners using a nonsense-syllable test II. Syllable recognition and consonant confusion patterns. *Journal of Speech, Language, and Hearing Research*, *25*(1), 141–148.

Elliott, T. M., & Theunissen, F. E. (2009). The Modulation Transfer Function for Speech Intelligibility. *PLoS Comput Biol*, *5*(3), e1000302. https://doi.org/10.1371/journal.pcbi.1000302

Fant, G. (1970). *Acoustic theory of speech production*. Walter de Gruyter.

Fant, G. (1973). *Speech sounds and features*. The MIT Press.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736. https://doi.org/10.1121/1.400247

French, N. R., & Steinberg, J. C. (1947). Factors Governing the Intelligibility of Speech Sounds. *The Journal of the Acoustical Society of America*, *19*(1), 90–119. https://doi.org/10.1121/1.1916407

Füllgrabe, C., Meyer, B., & Lorenzi, C. (2003). Effect of cochlear damage on the detection of complex temporal envelopes. *Hearing Research*, *178*(1), 35–43. https://doi.org/10.1016/S0378-5955(03)00027-3

Galambos, R., & Davis, H. (1943). The response of single auditory-nerve fibers to acoustic stimulation. *Journal of Neurophysiology*, *6*(1), 39–57. https://doi.org/10.1152/jn.1943.6.1.39

Gardi, J., & Merzenich, M. (1979). The effect of high-pass noise on the scalp-recorded frequency following response (FFR) in humans and cats. *The Journal of the Acoustical Society of America*, *65*(6), 1491–1500. https://doi.org/10.1121/1.382913

Geisler, C. D., & Silkes, S. M. (1991). Responses of "lower-spontaneous-rate" auditory-nerve fibers to speech syllables presented in noise. II: Glottal-pulse periodicities. *The Journal of the Acoustical Society of America*, *90*(6), 3140–3148. https://doi.org/10.1121/1.401422

Goblick, T. J., & Pfeiffer, R. R. (1969). Time-Domain Measurements of Cochlear Nonlinearities Using Combination Click Stimuli. *The Journal of the Acoustical Society of America*, *46*(4B), 924–938. https://doi.org/10.1121/1.1911812

Gockel, H. E., Krugliak, A., Plack, C. J., & Carlyon, R. P. (2015). Specificity of the Human Frequency Following Response for Carrier and Modulation Frequency Assessed Using Adaptation. *Journal of the Association for Research in Otolaryngology*, *16*(6), 747–762. https://doi.org/10.1007/s10162-015-0533-9

Goldberg, J. M., & Brown, P. B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization. *Journal of Neurophysiology*, *32*(4), 613–636. Retrieved September 10, 2015, from http://jn.physiology.org/content/32/4/613

Goldsworthy, R. L., & Greenberg, J. E. (2004). Analysis of speech-based speech transmission index methods with implications for nonlinear operations. *The Journal of the Acoustical Society of America*, *116*(6), 3679–3689.

Goman, A. M., & Lin, F. R. (2016). Prevalence of Hearing Loss by Severity in the United States. *Am J Public Health*, *106*(10), 1820–1822. https://doi.org/10.2105/AJPH.2016.303299

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2018). Neural envelope encoding predicts speech perception performance for normal-hearing and hearing-impaired adults. *Hearing Research*, *370*, 189–200. https://doi.org/10.1016/j.heares.2018.07.012

Gopinath, B., Wang, J. J., Schneider, J., Burlutsky, G., Snowdon, J., McMahon, C. M., Leeder, S. R., & Mitchell, P. (2009). Depressive symptoms in older adults with hearing impairments: The Blue Mountains Study [Publisher: Wiley Online Library]. *Journal of the American Geriatrics Society*, *57*(7), 1306–1308.

Grayden, D., Burkitt, A., Kenny, O., Clarey, J., Paolini, A., & Clark, G. (2004). A cochlear implant speech processing strategy based on an auditory model. *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004.*, 491–496. https://doi.org/10.1109/ISSNIP.2004.1417510

Greenberg, S., & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum. *Seventh European Conference on Speech Communication and Technology*.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics*, *31*(3-4), 465–485.

Greenwood, J. A., & Durand, D. (1955). The Distribution of Length and Components of the Sum of n Random Unit Vectors. *Ann. Math. Statist.*, *26*(2), 233–246. https://doi.org/10.1214/aoms/1177728540

Griffin, S. J., Bernstein, L. R., Ingham, N. J., & McAlpine, D. (2005). Neural Sensitivity to Interaural Envelope Delays in the Inferior Colliculus of the Guinea Pig [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *93*(6), 3463–3478. https://doi.org/10.1152/jn.00794.2004

Guinan, J. J., & Peake, W. T. (1967). Middle-Ear Characteristics of Anesthetized Cats. *The Journal of the Acoustical Society of America*, *41*(5), 1237–1261. https://doi.org/10.1121/1.1910465

Guinan, J. J. (2006). Olivocochlear Efferents: Anatomy, Physiology, Function, and the Measurement of Efferent Effects in Humans: *Ear and Hearing*, *27*(6), 589–607. https://doi.org/10.1097/01.aud.0000240507.83072.e7

Hagiwara, S. (1954). ANALYSIS OF INTERVAL FLUCTUATION OF THE SENSORY NERVE IMPULSE. *The Japanese Journal of Physiology*, *4*, 234–240. https://doi.org/10.2170/jjphysiol.4.234

Halliday, L. F., Rosen, S., Tuomainen, O., & Calcus, A. (2019). Impaired frequency selectivity and sensitivity to temporal fine structure, but not envelope cues, in children with mild-to-moderate sensorineural hearing loss. *The Journal of the Acoustical Society of America*, *146*(6), 4299–4314. https://doi.org/10.1121/1.5134059

Harrison, R. V., & Evans, E. F. (1979). Some aspects of temporal coding by single cochlear fibres from regions of cochlear hair cell degeneration in the guinea pig. *Archives of oto-rhino-laryngology*, *224*(1), 71–78. https://doi.org/10.1007/BF00455226

Heil, P. (2003). Coding of temporal onset envelope in the auditory system. *Speech Communication*, *41*(1), 123–134. https://doi.org/10.1016/S0167-6393(02)00099-7

Heil, P., & Peterson, A. J. (2015). Basic response properties of auditory nerve fibers: A review. *Cell Tissue Res*, *361*(1), 129–158. https://doi.org/10.1007/s00441-015-2177-9

Heinz, M. G. (2015). Neural modelling to relate individual differences in physiological and perceptual responses with sensorineural hearing loss. *1*, *5*, 137–148. Retrieved June 9, 2019, from https://proceedings.isaar.eu/index.php/isaarproc/article/view/2015-16

Heinz, M. G., Issa, J. B., & Young, E. D. (2005). Auditory-Nerve Rate Responses are Inconsistent with Common Hypotheses for the Neural Correlates of Loudness Recruitment. *Journal of the Association for Research in Otolaryngology*, *6*(2), 91–105. https://doi.org/10.1007/s10162-004-5043-0

Heinz, M. G., & Swaminathan, J. (2009). Quantifying Envelope and Fine-Structure Coding in Auditory Nerve Responses to Chimaeric Speech. *J Assoc Res Otolaryngol*, *10*(3), 407–423. https://doi.org/10.1007/s10162-009-0169-8

Heinz, M. G., Swaminathan, J., Boley, J. D., & Kale, S. (2010). Across-Fiber Coding of Temporal Fine-Structure: Effects of Noise-Induced Hearing Loss on Auditory-Nerve Responses. In E. A. Lopez-Poveda, A. R. Palmer, & R. Meddis (Eds.), *The Neurophysiological Bases of Auditory Perception* (pp. 621–630). Springer New York. https://doi.org/10.1007/978-1-4419-5686-6_56

Heinz, M. G., & Young, E. D. (2004). Response Growth With Sound Level in Auditory-Nerve Fibers After Noise-Induced Hearing Loss. *J Neurophysiol*, *91*(2), 784–795. https://doi.org/10.1152/jn.00776.2003

Henry, K. S., & Heinz, M. G. (2013). Effects of sensorineural hearing loss on temporal coding of narrowband and broadband signals in the auditory periphery. *Hearing Research*, *303*, 39–47. https://doi.org/10.1016/j.heares.2013.01.014

Henry, K. S., Kale, S., & Heinz, M. G. (2016). Distorted Tonotopic Coding of Temporal Envelope and Fine Structure with Noise-Induced Hearing Loss. *J Neurosci*, *36*(7), 2227–2237. https://doi.org/10.1523/JNEUROSCI.3944-15.2016

Henry, K. S., Kale, S., Scheidt, R. E., & Heinz, M. G. (2011). Auditory brainstem responses predict auditory nerve fiber thresholds and frequency selectivity in hearing impaired chinchillas. *Hear Res*, *280*(1–2), 236–244. https://doi.org/10.1016/j.heares.2011.06.002

Henry, K. S., Sayles, M., Hickox, A. E., & Heinz, M. G. (2019). Divergent auditory-nerve encoding deficits between two common etiologies of sensorineural hearing loss. *J Neurosci*, 6879–6887. https://doi.org/10.1523/JNEUROSCI.0038-19.2019

Henry, K. S., Kale, S., & Heinz, M. G. (2014). Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers. *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00020

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research*.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*, *105*(6), 3509–3523. https://doi.org/10.1121/1.424676

Houtgast, T., & Steeneken, H. J. M. (1973). The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility. *Acta Acustica united with Acustica*, *28*(1), 66–73.

Houtgast, T., & Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, *77*(3), 1069–1077. https://doi.org/10.1121/1.392224

Houtgast, T., Steeneken, H. J. M., & Plomp, R. (1980). Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics. *Acta Acustica united with Acustica*, *46*(1), 60–72.

Jakobson, R. F., & Fant, G. (1963). *G., & Halle, M.(1963). Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *The Journal of the Acoustical Society of America*, *68*(4), 1115–1122. https://doi.org/10.1121/1.384982

Jørgensen, S., & Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, *130*(3), 1475–1487. https://doi.org/10.1121/1.3621502

Jørgensen, S., Ewert, S. D., & Dau, T. (2013). A multi-resolution envelope-power based model for speech intelligibility. *The Journal of the Acoustical Society of America*, *134*(1), 436–446. https://doi.org/10.1121/1.4807563

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural Processing of Amplitude-Modulated Sounds. *Physiol. Rev*, *84*(2), 541–577. https://doi.org/10.1152/physrev.00029.2003

Joris, P. X. (2003). Interaural Time Sensitivity Dominated by Cochlea-Induced Envelope Patterns. *J. Neurosci.*, *23*(15), 6345–6350. Retrieved September 30, 2015, from http://www.jneurosci.org/content/23/15/6345

Joris, P. X., Louage, D. H., Cardoen, L., & van der Heijden, M. (2006). Correlation Index: A new metric to quantify temporal coding. *Hearing Research*, *216–217*, 19–30. https://doi.org/10.1016/j.heares.2006.03.010

Joris, P. X., & Yin, T. C. T. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *J Acoust Soc Am*, *91*(1), 215–232. https://doi.org/10.1121/1.402757

Kale, S., & Heinz, M. G. (2010). Envelope Coding in Auditory Nerve Fibers Following Noise-Induced Hearing Loss. *JARO*, *11*(4), 657–673. https://doi.org/10.1007/s10162-010-0223-6

Kale, S., Micheyl, C., & Heinz, M. G. (2013). Effects of Sensorineural Hearing Loss on Temporal Coding of Harmonic and Inharmonic Tone Complexes in the Auditory Nerve. *Basic Aspects of Hearing* (pp. 109–118). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-1590-9_13

Karpa, M. J., Gopinath, B., Beath, K., Rochtchina, E., Cumming, R. G., Wang, J. J., & Mitchell, P. (2010). Associations between hearing impairment and mortality risk in older persons: The Blue Mountains Hearing Study [Publisher: Elsevier]. *Annals of epidemiology*, *20*(6), 452–459.

Kiang, N. Y. S., & Moxon, E. C. (1974). Tails of tuning curves of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *55*(3), 620–630. https://doi.org/10.1121/1.1914572

Kiang, N. Y. S., Moxon, E. C., & Levine, R. A. (1970). Auditory-Nerve Activity in Cats with Normal and Abnormal Cochleas. *Ciba Foundation Symposium - Sensorineural Hearing Loss* (pp. 241–273). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470719756.ch15

Kiang, N. Y. S., Watanabe, T., Thomas, E., & Clark, L. (1965). Discharge patterns of single fibers in the cat's auditory nerve. *MIT, Cambridge, MA*, *1*(1), 104–105. https://doi.org/10.1016/0021-9924(67)90048-2

King, A., Hopkins, K., & Plack, C. J. (2016). Differential Group Delay of the Frequency Following Response Measured Vertically and Horizontally. *JARO*, *17*(2), 133–143. https://doi.org/10.1007/s10162-016-0556-x

King, A., Varnet, L., & Lorenzi, C. (2019). Accounting for masking of frequency modulation by amplitude modulation with the modulation filter-bank concept. *The Journal of the Acoustical Society of America*, *145*(4), 2277–2293. https://doi.org/10.1121/1.5094344

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America*, *67*(3), 971–995. https://doi.org/10.1121/1.383940

Klatt, D. H. (1982). Prediction of perceived phonetic distance from critical-band spectra: A first step. *ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, *7*, 1278–1281. https://doi.org/10.1109/ICASSP.1982.1171512

Klumpp, R. G., & Eady, H. R. (1956). Some Measurements of Interaural Time Difference Thresholds [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *28*(5), 859–860. https://doi.org/10.1121/1.1908493

Knudsen, E. I. (2007). Fundamental Components of Attention. *Annual Review of Neuroscience*, *30*(1), 57–78. https://doi.org/10.1146/annurev.neuro.30.051606.094256

Kraus, N., Anderson, S., & White-Schwoch, T. (2017). The Frequency-Following Response: A Window into Human Communication. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_1

Krishna, B. S., & Semple, M. N. (2000). Auditory Temporal Processing: Responses to Sinusoidally Amplitude-Modulated Tones in the Inferior Colliculus. *Journal of Neurophysiology*, *84*(1), 255–273. https://doi.org/10.1152/jn.2000.84.1.255

Krishnan, A., & Parkinson, J. (2000). Human Frequency-Following Response: Representation of Tonal Sweeps. *AUD*, *5*(6), 312–321. https://doi.org/10.1159/000013897

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res*, *25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004

Kryter, K. D. (1962). Methods for the Calculation and Use of the Articulation Index. *The Journal of the Acoustical Society of America*, *34*(11), 1689–1697. https://doi.org/10.1121/1.1909094

Kujawa, S. G., & Liberman, M. C. (2009). Adding Insult to Injury: Cochlear Nerve Degeneration after "Temporary" Noise-Induced Hearing Loss. *J. Neurosci.*, *29*(45), 14077–14085. https://doi.org/10.1523/JNEUROSCI.2845-09.2009

Kujawa, S. G., & Liberman, M. C. (2019). Translating animal models to human therapeutics in noise-induced and age-related hearing loss. *Hearing Research*, *377*, 44–52. https://doi.org/10.1016/j.heares.2019.03.003

Kuwada, S., Batra, R., & Maher, V. L. (1986). Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hear Res*, *21*(2), 179–192. https://doi.org/10.1016/0378-5955(86)90038-9

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Lai, J., & Bartlett, E. L. (2018). Masking Differentially Affects Envelope-following Responses in Young and Aged Animals. *Neuroscience*, *386*, 150–165. https://doi.org/10.1016/j.neuroscience.2018.06.004

Langner, G., Schreiner, C., & Merzenich, M. M. (1987). Covariation of latency and temporal resolution in the inferior colliculus of the cat. *Hear Res*, *31*(2), 197–201. https://doi.org/10.1016/0378-5955(87)90127-4

Larsen, E., Cedolin, L., & Delgutte, B. (2008). Pitch Representations in the Auditory Nerve: Two Concurrent Complex Tones. *Journal of Neurophysiology*, *100*(3), 1301–1319. https://doi.org/10.1152/jn.01361.2007

Lesica, N. A. (2018). Why Do Hearing Aids Fail to Restore Normal Auditory Perception? *Trends in neurosciences*.

Levy, S. C., Freed, D. J., Nilsson, M., Moore, B. C. J., & Puria, S. (2015). Extended high-frequency bandwidth improves reception of speech in spatially separated masking speech. *Ear and hearing*, *36*(5), e214–e224. https://doi.org/10.1097/AUD.0000000000000161

Liberman, M. C., & Guinan, J. J. (1998). Feedback control of the auditory periphery: Anti-masking effects of middle ear muscles vs. olivocochlear efferents. *Journal of communication disorders*, *31*(6), 471–82, quiz 483, 553. https://doi.org/10.1016/S0021-9924(98)00019-7

Liberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber [tex.ids: liberman_auditory-nerve_1978]. *The Journal of the Acoustical Society of America*, *63*(2), 442–455. https://doi.org/10.1121/1.381736

Liberman, M. C. (1984). Single-neuron labeling and chronic cochlear pathology. I. Threshold shift and characteristic-frequency shift. *Hearing Research*, *16*(1), 33–41. https://doi.org/10.1016/0378-5955(84)90023-6

Liberman, M. C., & Dodds, L. W. (1984a). Single-neuron labeling and chronic cochlear pathology. II. Stereocilia damage and alterations of spontaneous discharge rates. *Hear Res*, *16*(1), 43–53. https://doi.org/10.1016/0378-5955(84)90024-8

Liberman, M. C., & Dodds, L. W. (1984b). Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves. *Hear Res*, *16*(1), 55–74. https://doi.org/10.1016/0378-5955(84)90025-X

Liberman, M. C., & Klang, N. Y. .-S. (1984). Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions. *Hearing Research*, *16*(1), 75–90. https://doi.org/10.1016/0378-5955(84)90026-1

Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the Rôle of Formant Transitions in Vowel Recognition. *J Acoust Soc Am*, *42*(4), 830–843. https://doi.org/10.1121/1.1910655

Logan, B. F. (1977). Information in the Zero Crossings of Bandpass Signals. *Bell System Technical Journal*, *56*(4), 487–510. https://doi.org/10.1002/j.1538-7305.1977.tb00522.x

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *PNAS*, *103*(49), 18866–18869. https://doi.org/10.1073/pnas.0607364103

Louage, D. H. G., Heijden, M. v. d., & Joris, P. X. (2004). Temporal Properties of Responses to Broadband Noise in the Auditory Nerve. *Journal of Neurophysiology*, *91*(5), 2051–2065. https://doi.org/10.1152/jn.00816.2003

Lu, K., Xu, Y., Yin, P., Oxenham, A. J., Fritz, J. B., & Shamma, S. A. (2017). Temporal coherence structure rapidly shapes neuronal interactions. *Nature Communications*, *8*, 13900. https://doi.org/10.1038/ncomms13900

Lucchetti, F., Deltenre, P., Avan, P., Giraudet, F., Fan, X., & Nonclercq, A. (2018). Generalization of the primary tone phase variation method: An exclusive way of isolating the frequency-following response components [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *144*(4), 2400–2412. https://doi.org/10.1121/1.5063821

Lybarger, S. F. (1978). SELECTIVE AMPLIFICATION—A REVIEW AND EVALUATION. *Ear and Hearing*, *3*(6), 258.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412. https://doi.org/10.3758/BF03204884

Mardia, K. V. (1972). A Multi-Sample Uniform Scores Test on a Circle and its Parametric Competitor [_eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1972.tb00891.x]. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(1), 102–113. https://doi.org/10.1111/j.2517-6161.1972.tb00891.x

Martin, L. (2012). Chapter 43 - Chinchillas as Experimental Models. In M. A. Suckow, K. A. Stevens, & R. P. Wilson (Eds.), *The Laboratory Rabbit, Guinea Pig, Hamster, and Other Rodents* (pp. 1009–1028). Academic Press. https://doi.org/10.1016/B978-0-12-380920-9.00043-2

May, B. J. (2003). Physiological and psychophysical assessments of the dynamic range of vowel representations in the auditory periphery. *Speech Communication*, *41*(1), 49–57. https://doi.org/10.1016/S0167-6393(02)00092-4

McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them? *International Journal of Audiology*, *52*(5), 360–368. https://doi.org/10.3109/14992027.2013.769066

Mener, D. J., Betz, J., Genther, D. J., Chen, D., & Lin, F. R. (2013). Hearing Loss and Depression in Older Adults. *Journal of the American Geriatrics Society*, *61*(9), 1627–1629. https://doi.org/10.1111/jgs.12429

Micheyl, C., & Oxenham, A. J. (2010). Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. *Hearing Research*, *266*(1), 36–51. https://doi.org/10.1016/j.heares.2009.09.012

Middelweerd, M. J., Festen, J. M., & Plomp, R. (1990). Difficulties with Speech Intelligibility in noise in spite of a normal pure-tone audiogram: Original papers. *Audiology*, *29*(1), 1–7.

Miller, M. I., & Sachs, M. B. (1983). Representation of stop consonants in the discharge patterns of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *74*(2), 502–517.

Miller, M. I., & Sachs, M. B. (1984). Representation of voice pitch in discharge patterns of auditory-nerve fibers. *Hearing research*, *14*(3), 257–279.

Miller, R. L., Calhoun, B. M., & Young, E. D. (1999). Discriminability of vowel representations in cat auditory-nerve fibers after acoustic trauma. *The Journal of the Acoustical Society of America*, *105*(1), 311–325.

Miller, R. L., Schilling, J. R., Franck, K. R., & Young, E. D. (1997). Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers. *The Journal of the Acoustical Society of America*, *101*(6), 3602–3616.

Millman, R. E., Mattys, S. L., Gouws, A. D., & Prendergast, G. (2017). Magnified Neural Envelope Coding Predicts Deficits in Speech Perception in Noise. *Journal of Neuroscience*, *37*(32), 7727–7736. https://doi.org/10.1523/JNEUROSCI.2722-16.2017

Mitra, P. P., & Pesaran, B. (1999). Analysis of Dynamic Brain Imaging Data. *Biophys J*, *76*(2), 691–708. https://doi.org/10.1016/S0006-3495(99)77236-X

Møller, A. R. (1970). The Use of Correlation Analysis in Processing Neuroelectric Data. *Progress in Brain Research* (pp. 87–99). Elsevier. https://doi.org/10.1016/S0079-6123(08)62444-9

Monson, B. B., Rock, J., Schulz, A., Hoffman, E., & Buss, E. (2019). Ecological cocktail party listening reveals the utility of extended high-frequency hearing. *Hearing Research*, *381*, 107773. https://doi.org/10.1016/j.heares.2019.107773

Moore, B. C. J. (2007). *Cochlear hearing loss: Physiological, psychological and technical issues*. John Wiley & Sons.

Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people [Publisher: Springer]. *Journal of the Association for Research in Otolaryngology*, *9*(4), 399–406.

Moore, B. C. J. (2014). *Auditory Processing Of Temporal Fine Structure: Effects Of Age And Hearing Loss* [Google-Books-ID: h_62CgAAQBAJ]. World Scientific.

Moore, B. C. J. (2016). A review of the perceptual effects of hearing loss for frequencies above 3 kHz. *International Journal of Audiology*, *55*(12), 707–714. https://doi.org/10.1080/14992027.2016.1204565

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch: Neural Coding and Perception* (pp. 234–277). Springer New York. https://doi.org/10.1007/0-387-28958-5_7

Moore, B. C. J., & Glasberg, B. R. (1993). Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. *The Journal of the Acoustical Society of America*, *94*(4), 2050–2062. https://doi.org/10.1121/1.407478

Moore, B. C. J., Glasberg, B. R., Hess, R. F., & Birchall, J. P. (1985). Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *77*(4), 1505–1513. https://doi.org/10.1121/1.392045

Moore, B. C. J., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *The Journal of the Acoustical Society of America*, *91*(5), 2881–2893. https://doi.org/10.1121/1.402925

Moore, B. C. J., Peters, R. W., & Stone, M. A. (1999). Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *The Journal of the Acoustical Society of America*, *105*(1), 400–411. https://doi.org/10.1121/1.424571

Moore, B. C. J., Wojtczak, M., & Vickers, D. A. (1996). Effect of loudness recruitment on the perception of amplitude modulation. *The Journal of the Acoustical Society of America*, *100*(1), 481–489. https://doi.org/10.1121/1.415861

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *80*(5), 1297–1308. https://doi.org/10.1121/1.394433

Nelson, P. C., & Carney, L. H. (2007). Neural Rate and Timing Cues for Detection and Discrimination of Amplitude-Modulated Tones in the Awake Rabbit Inferior Colliculus. *Journal of Neurophysiology*, *97*(1), 522–539. https://doi.org/10.1152/jn.00776.2006

Ngan, E. M., & May, B. J. (2001). Relationship between the auditory brainstem response and auditory nerve thresholds in cats with hearing loss. *Hearing Research*, *156*(1–2), 44–52. https://doi.org/10.1016/S0378-5955(01)00264-7

Nielsen, J. B., & Dau, T. (2009). Development of a Danish speech intelligibility test. *International Journal of Audiology*, *48*(10), 729–741. https://doi.org/10.1080/14992020903019312

Olhede, S., & Walden, A. (2005). A generalized demodulation approach to time-frequency projections for multicomponent signals [Publisher: Royal Society]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *461*(2059), 2159–2179. https://doi.org/10.1098/rspa.2005.1455

Oppenheim, A. V. (1999). *Discrete-time signal processing*. Pearson Education India.

Owens, E., Benedict, M., & D. Schubert, E. (1972). Consonant Phonemic Errors Associated with Pure-Tone Configurations and Certain Kinds of Hearing Impairment. *Journal of Speech and Hearing Research*, *15*(2), 308–322. https://doi.org/10.1044/jshr.1502.308

Paliwal, K. K., & Alsteris, L. (2003). Usefulness of Phase Spectrum in Human Speech Perception. *Eighth European Conference on Speech Communication and Technology*, 4.

Palmer, A. R. (1990). The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *The Journal of the Acoustical Society of America*, *88*(3), 1412–1426. https://doi.org/10.1121/1.400329

Palmer, A. R., & Russell, I. J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research*, *24*(1), 1–15. https://doi.org/10.1016/0378-5955(86)90002-X

Palmer, A. R., Winter, I. M., & Darwin, C. J. (1986). The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons. *The Journal of the Acoustical Society of America*, *79*(1), 100–113. https://doi.org/10.1121/1.393633

Paraouty, N., Stasiak, A., Lorenzi, C., Varnet, L., & Winter, I. M. (2018). Dual Coding of Frequency Modulation in the Ventral Cochlear Nucleus [tex.ids: paraouty_dual_2018-1]. *The Journal of Neuroscience*, *38*(17), 4123–4137. https://doi.org/10.1523/JNEUROSCI.2107-17.2018

Parida, S., Bharadwaj, H., & Heinz, M. G. (2020). Spectrally specific temporal analyses of spike-train responses to complex sounds: A unifying framework [Publisher: Cold Spring Harbor Laboratory Section: New Results]. *bioRxiv*, 2020.07.17.208330. https://doi.org/10.1101/2020.07.17.208330

Parida, S., & Heinz, M. G. (2020). Noninvasive Measures of Distorted Tonotopic Speech Coding Following Noise-Induced Hearing Loss. *JARO*. https://doi.org/10.1007/s10162-020-00755-2

Parthasarathy, A., Cunningham, P. A., & Bartlett, E. L. (2010). Age-Related Differences in Auditory Processing as Assessed by Amplitude-Modulation Following Responses in Quiet and in Noise. *Front Aging Neurosci*, *2*. https://doi.org/10.3389/fnagi.2010.00152

Parthasarathy, A., Romero Pinto, S., Lewis, R. M., Goedicke, W., & Polley, D. B. (2020). Data-driven segmentation of audiometric phenotypes across a large clinical cohort. *Scientific Reports*, *10*(1), 6704. https://doi.org/10.1038/s41598-020-63515-5

Pavlovic, C. V. (1987). Derivation of primary parameters and procedures for use in speech intelligibility predictions. *The Journal of the Acoustical Society of America*, *82*(2), 413–422. https://doi.org/10.1121/1.395442

Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for physical applications*. cambridge university press.

Perkel, D. H., Gerstein, G. L., & Moore, G. P. (1967a). Neuronal Spike Trains and Stochastic Point Processes: I. The Single Spike Train. *Biophysical Journal*, *7*(4), 391–418. https://doi.org/10.1016/S0006-3495(67)86596-2

Perkel, D. H., Gerstein, G. L., & Moore, G. P. (1967b). Neuronal Spike Trains and Stochastic Point Processes: II. Simultaneous Spike Trains. *Biophysical Journal*, *7*(4), 419–440. https://doi.org/10.1016/S0006-3495(67)86597-4

Phatak, S. A., & Grant, K. W. (2014). Phoneme recognition in vocoded maskers by normal-hearing and aided hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *136*(2), 859–866. https://doi.org/10.1121/1.4889863

Phatak, S. A., Yoon, Y.-s., Gooler, D. M., & Allen, J. B. (2009). Consonant recognition loss in hearing impaired listeners. *The Journal of the Acoustical Society of America*, *126*(5), 2683–2694.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37*, 5S. https://doi.org/10.1097/AUD.0000000000000312

Pickett, J. M. (1999). *The acoustics of speech communication: Fundamentals, speech perception theory, and technology*. Allyn and Bacon Boston.

Plack, C. J., & Oxenham, A. J. (2005). The Psychophysics of Pitch. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch: Neural Coding and Perception* (pp. 7–55). Springer New York. https://doi.org/10.1007/0-387-28958-5_2

Plomp, R. (1994). Noise, amplification, and compression: Considerations of three main issues in hearing aid design [Publisher: LWW]. *Ear and hearing*, *15*(1), 2–12.

Presacco, A., Simon, J. Z., & Anderson, S. (2016). Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *Journal of Neurophysiology*, *116*(5), 2346–2355. https://doi.org/10.1152/jn.00372.2016

Rallapalli, V. H., & Heinz, M. G. (2016). Neural Spike-Train Analyses of the Speech-Based Envelope Power Spectrum Model: Application to Predicting Individual Differences with Sensorineural Hearing Loss. *Trends in Hearing*, *20*, 2331216516667319. https://doi.org/10.1177/2331216516667319

Rangayyan, R. M. (2015). *Biomedical signal analysis* (Vol. 33). John Wiley & Sons.

Rees, A., & Palmer, A. R. (1989). Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise. *The Journal of the Acoustical Society of America*, *85*(5), 1978–1994. https://doi.org/10.1121/1.397851

Relaño-Iborra, H., May, T., Zaar, J., Scheidiger, C., & Dau, T. (2016). Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain. *The Journal of the Acoustical Society of America*, *140*(4), 2670–2679. https://doi.org/10.1121/1.4964505

Robles, L., & Ruggero, M. A. (2001). Mechanics of the Mammalian Cochlea. *Physiological Reviews*, *81*(3), 1305–1352. https://doi.org/10.1152/physrev.2001.81.3.1305

Rodieck, R. W. (1967). Maintained activity of cat retinal ganglion cells. [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *30*(5), 1043–1071. https://doi.org/10.1152/jn.1967.30.5.1043

Rodieck, R., Kiang, N. Y. S., & Gerstein, G. (1962). Some Quantitative Methods for the Study of Spontaneous Activity of Single Neurons. *Biophysical Journal*, *2*(4), 351–368. https://doi.org/10.1016/S0006-3495(62)86860-X

Rose, J. E., Brugge, J. F., Anderson, D. J., & Hind, J. E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of Neurophysiology*, *30*(4), 769–793. https://doi.org/10.1152/jn.1967.30.4.769

Rosen, S., & Fourcin, A. (1986). Frequency selectivity and the perception of speech [Publisher: Academic London]. *Frequency selectivity in hearing*, 373–488.

Rosen, S. (1986). Monaural Phase Sensitivity: Frequency Selectivity and Temporal Processes. In B. C. J. Moore & R. D. Patterson (Eds.), *Auditory Frequency Selectivity* (pp. 419–427). Springer US. https://doi.org/10.1007/978-1-4613-2247-4_45

Rothauser, E. H. (1969). IEEE recommended practice for speech quality measurements. *IEEE Trans. on Audio and Electroacoustics*, *17*, 225–246.

Saberi, K., & Haftert, E. R. (1995). A common neural code for frequency- and amplitude-modulated sounds. *Nature*, *374*(6522), 537. https://doi.org/10.1038/374537a0

Sachs, M. B., Bruce, I. C., Miller, R. L., & Young, E. D. (2002). Biological basis of hearing-aid design. *Annals of Biomedical Engineering*, *30*(2), 157–168.

Sachs, M. B., & Young, E. D. (1979). Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. *J Acoust Soc Am*, *66*(2), 470–479. https://doi.org/10.1121/1.383098

Sachs, M. B., & Young, E. D. (1980). Effects of nonlinearities on speech encoding in the auditory nerve. *The Journal of the Acoustical Society of America*, *68*(3), 858–875. https://doi.org/10.1121/1.384825

Sadjadi, S. O., & Hansen, J. H. L. (2011). Hilbert envelope based features for robust speaker identification under reverberant mismatched conditions. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5448–5451. https://doi.org/10.1109/ICASSP.2011.5947591

Salvi, R. J., Saunders, S. S., Gratton, M. A., Arehole, S., & Powers, N. (1990). Enhanced evoked response amplitudes in the inferior colliculus of the chinchilla following acoustic trauma. *Hear Res*, *50*(1), 245–257. https://doi.org/10.1016/0378-5955(90)90049-U

Salvi, R. J., Henderson, D., Hamernik, R., & Ahroon, W. A. (1983). Neural Correlates of Sensorineural Hearing Loss. *Ear & Hearing*, *4*(3), 115–129.

Salvi, R. J., Wang, J., & Ding, D. (2000). Auditory plasticity and hyperactivity following cochlear damage. *Hear Res*, *147*(1), 261–274. https://doi.org/10.1016/S0378-5955(00)00136-2

Sayles, M., & Heinz, M. G. (2017). Afferent Coding and Efferent Control in the Normal and Impaired Cochlea. *Understanding the Cochlea* (pp. 215–252). Springer, Cham. https://doi.org/10.1007/978-3-319-52073-5_8

Sayles, M., Stasiak, A., & Winter, I. M. (2015). Reverberation impairs brainstem temporal representations of voiced vowel sounds: Challenging "periodicity-tagged" segregation of competing speech in rooms. *Front. Syst. Neurosci.*, *8*. https://doi.org/10.3389/fnsys.2014.00248

Sayles, M., Walls, M. K., & Heinz, M. G. (2016). Suppression Measured from Chinchilla Auditory-Nerve-Fiber Responses Following Noise-Induced Hearing Loss: Adaptive-Tracking and Systems-Identification Approaches. *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing* (pp. 285–295). Springer, Cham. https://doi.org/10.1007/978-3-319-25474-6_30

Sayles, M., & Winter, I. M. (2008). Reverberation Challenges the Temporal Representation of the Pitch of Complex Sounds. *Neuron*, *58*(5), 789–801. https://doi.org/10.1016/j.neuron.2008.03.029

Scharenborg, O. (2007). Reaching over the gap: A review of efforts to link human and automatic speech recognition research. *Speech Communication*, *49*(5), 336–347. https://doi.org/10.1016/j.specom.2007.01.009

Scheidiger, C., Carney, L. H., Dau, T., & Zaar, J. (2018). Predicting Speech Intelligibility Based on Across-Frequency Contrast in Simulated Auditory-Nerve Fluctuations. *Acta Acustica united with Acustica*, *104*(5), 914–917.

Scheidt, R. E., Kale, S., & Heinz, M. G. (2010). Noise-induced hearing loss alters the temporal dynamics of auditory-nerve responses. *Hearing Research*, *269*(1), 23–33. https://doi.org/10.1016/j.heares.2010.07.009

Schilling, J. R., Miller, R. L., Sachs, M. B., & Young, E. D. (1998). Frequency-shaped amplification changes the neural representation of speech with noise-induced hearing loss. *Hearing research*, *117*(1-2), 57–70.

Schmiedt, R. A., Mills, J. H., & Adams, J. C. (1990). Tuning and suppression in auditory nerve fibers of aged gerbils raised in quiet or noise. *Hear Res*, *45*(3), 221–236. https://doi.org/10.1016/0378-5955(90)90122-6

Schoof, T., & Rosen, S. (2016). The Role of Age-Related Declines in Subcortical Auditory Processing in Speech Perception in Noise. *Journal of the Association for Research in Otolaryngology*, *17*(5), 441–460. https://doi.org/10.1007/s10162-016-0564-x

Shaheen, L. A., Valero, M. D., & Liberman, M. C. (2015). Towards a Diagnosis of Cochlear Neuropathy with Envelope Following Responses. *J Assoc Res Otolaryngol*, *16*, 727–45. https://doi.org/10.1007/s10162-015-0539-3

Shamma, S. A. (1985). Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *The Journal of the Acoustical Society of America*, *78*(5), 1622–1632. https://doi.org/10.1121/1.392800

Shamma, S. A., & Lorenzi, C. (2013). On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *The Journal of the Acoustical Society of America*, *133*(5), 2818–2833. https://doi.org/10.1121/1.4795783

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, *270*(5234), 303–304. Retrieved September 18, 2018, from http://www.jstor.org/stable/2888543

Shinn-Cunningham, B. G., & Best, V. (2008). Selective Attention in Normal and Impaired Hearing. *Trends Amplif*, *12*(4), 283–299. https://doi.org/10.1177/1084713808325306

Shinn-Cunningham, B. G., Ruggles, D. R., & Bharadwaj, H. (2013). How Early Aging and Environment Interact in Everyday Listening: From Brainstem to Behavior Through Modeling. In B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, & H. E. Gockel (Eds.), *Basic Aspects of Hearing* (pp. 501–510). Springer New York.

Shrivastav, M. N. (2012). Suprathreshold Auditory Processing in Noise-Induced Hearing Loss. In C. G. L. Prell, D. Henderson, R. R. Fay, & A. N. Popper (Eds.), *Noise-Induced Hearing Loss* (pp. 137–150). Springer New York. https://doi.org/10.1007/978-1-4419-9523-0_7

Silkes, S. M., & Geisler, C. D. (1991). Responses of "lower-spontaneous-rate" auditory-nerve fibers to speech syllables presented in noise. I: General characteristics. *The Journal of the Acoustical Society of America*, *90*(6), 3122–3139. https://doi.org/10.1121/1.401421

Sinex, D. G., & Geisler, C. D. (1981). Auditory-nerve fiber responses to frequency-modulated tones. *Hearing Research*, *4*(2), 127–148. https://doi.org/10.1016/0378-5955(81)90001-0

Sinex, D. G., & Geisler, C. D. (1983). Responses of auditory-nerve fibers to consonant–vowel syllables. *The Journal of the Acoustical Society of America*, *73*(2), 602–615. https://doi.org/10.1121/1.389007

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear Hear*, *31*(3), 302–324. https://doi.org/10.1097/AUD.0b013e3181cdb272

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87. https://doi.org/10.1038/416087a

Smits, R., ten Bosch, L., & Collier, R. (1996). Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment. *The Journal of the Acoustical Society of America*, *100*(6), 3852–3864.

Souza, P. (2016). Speech Perception and Hearing Aids. In G. R. Popelka, B. C. J. Moore, R. R. Fay, & A. N. Popper (Eds.), *Hearing Aids* (pp. 151–180). Springer International Publishing. https://doi.org/10.1007/978-3-319-33036-5_6

Stagray, J. R., Downs, D., & Sommers, R. K. (1992). Contributions of the Fundamental, Resolved Harmonics, and Unresolved Harmonics in Tone-Phoneme Identification [Publisher: American Speech-Language-Hearing Association]. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1406–1409. https://doi.org/10.1044/jshr.3506.1406

Stilp, C. E., Kiefte, M., Alexander, J. M., & Kluender, K. R. (2010). Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentencesa). *The Journal of the Acoustical Society of America*, *128*(4), 2112–2126. https://doi.org/10.1121/1.3483719

Swaminathan, J. (2010). *The role of envelope and temporal fine structure in the perception of noise degraded speech* (Ph.D.). Purdue University. United States – Indiana. Retrieved June 27, 2018, from https://search.proquest.com/docview/859202778/abstract/8A6AC5854A324C37PQ/1

Swaminathan, J., & Heinz, M. G. (2011). Predicted effects of sensorineural hearing loss on across-fiber envelope coding in the auditory nerve. *The Journal of the Acoustical Society of America*, *129*(6), 4001–4013. https://doi.org/10.1121/1.3583502

Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological Analyses Demonstrate the Importance of Neural Envelope Coding for Speech Perception in Noise. *J. Neurosci.*, *32*(5), 1747–1756. https://doi.org/10.1523/JNEUROSCI.4493-11.2012

Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(7), 2125–2136. https://doi.org/10.1109/TASL.2011.2114881

Takahashi, G. A., & Bacon, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1410–1421.

Temchin, A. N., Recio-Spinoso, A., van Dijk, P., & Ruggero, M. A. (2005). Wiener Kernels of Chinchilla Auditory-Nerve Fibers: Verification Using Responses to Tones, Clicks, and Noise and Comparison With Basilar-Membrane Vibrations. *Journal of Neurophysiology*, *93*(6), 3635–3648. https://doi.org/10.1152/jn.00885.2004

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proc IEEE*, *70*(9), 1055–1096. https://doi.org/10.1109/PROC.1982.12433

Tobias, J. V. (1959). Relative Occurrence of Phonemes in American English. *The Journal of the Acoustical Society of America*, *31*(5), 631–631. https://doi.org/10.1121/1.1907766

Trautwein, P., Hofstetter, P., Wang, J., Salvi, R., & Nostrant, A. (1996). Selective inner hair cell loss does not alter distortion product otoacoustic emissions. *Hearing Research*, *96*(1), 71–82. https://doi.org/10.1016/0378-5955(96)00040-8

Tremblay, K. L., Billings, C. J., Friesen, L. M., & Souza, P. E. (2006). Neural Representation of Amplified Speech Sounds. *Ear and Hearing*, *27*(2), 93–103. https://doi.org/10.1097/01.aud.0000202288.21315.bd

Trevino, M., Lobarinas, E., Maulden, A. C., & Heinz, M. G. (2019). The chinchilla animal model for hearing science and noise-induced hearing loss. *The Journal of the Acoustical Society of America*, *NIHLNS2019*(1), 3710–3732. https://doi.org/10.1121/1.5132950@jas.2019.NIHLNS2019.issue-1

Turner, C. W., & Robb, M. P. (1987). Audibility and recognition of stop consonants in normal and hearing-impaired subjects. *The Journal of the Acoustical Society of America*, *81*(5), 1566–1573. https://doi.org/10.1121/1.394509

Van de Grift Turek, S., Dorman, M. F., Franks, J. R., & Summerfield, Q. (1980). Identification of synthetic /bdg/ by hearing-impaired listeners under monotic and dichotic formant presentation. *The Journal of the Acoustical Society of America*, *67*(3), 1031–1040. https://doi.org/10.1121/1.384070

van Hemmen, J. L. (2013). Vector strength after Goldberg, Brown, and von Mises: Biological and mathematical perspectives. *Biol Cybern*, *107*(4), 385–396. https://doi.org/10.1007/s00422-013-0561-7

Vasilkov, V., & Verhulst, S. (2019). *Towards a differential diagnosis of cochlear synaptopathy and outer-hair-cell deficits in mixed sensorineural hearing loss pathologies* (preprint). Otolaryngology. https://doi.org/10.1101/19008680

Verhulst, S., Ernst, F., Garrett, M., & Vasilkov, V. (2018). Suprathreshold psychoacoustics and envelope-following response relations: Normal-hearing, synaptopathy and cochlear gain loss. *Acta Acust Acust*, *104*(5), 800–803.

Verschooten, E., & Joris, P. X. (2014). Estimation of Neural Phase Locking from Stimulus-Evoked Potentials. *Journal of the Association for Research in Otolaryngology*, *15*(5), 767–787. https://doi.org/10.1007/s10162-014-0465-9

Verschooten, E., Shamma, S. A., Oxenham, A. J., Moore, B. C. J., Joris, P. X., Heinz, M. G., & Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing Research*, *377*, 109–121. https://doi.org/10.1016/j.heares.2019.03.011

Victor, J. D., & Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: A metric-space analysis. *Journal of Neurophysiology*, *76*(2), 1310–1326. https://doi.org/10.1152/jn.1996.76.2.1310

Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., & Pennartz, C. M. A. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage*, *55*(4), 1548–1565. https://doi.org/10.1016/j.neuroimage.2011.01.055

Viswanathan, V., Bharadwaj, H. M., Shinn-Cunningham, B. G., & Heinz, M. G. (2019). Evaluating human neural envelope coding as the basis of speech intelligibility in noise [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *145*(3), 1717–1717. https://doi.org/10.1121/1.5101298

Voelcker, H. B. (1966a). Demodulation of Single-Sideband Signals Via Envelope Detection. *IEEE Transactions on Communication Technology*, *14*(1), 22–30. https://doi.org/10.1109/TCOM.1966.1089285

Voelcker, H. B. (1966b). Toward a unified theory of modulation—Part II: Zero manipulation. *Proceedings of the IEEE*, *54*(5), 735–755. https://doi.org/10.1109/PROC.1966.4843

Wang, L., Bharadwaj, H., & Shinn-Cunningham, B. G. (2019). Assessing Cochlear-Place Specific Temporal Coding Using Multi-Band Complex Tones to Measure Envelope-Following Responses. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2019.02.003

Westerman, L. A., & Smith, R. L. (1988). A diffusion model of the transient response of the cochlear inner hair cell synapse. *The Journal of the Acoustical Society of America*, *83*(6), 2266–2276. https://doi.org/10.1121/1.396357

Wible, B., Nicol, T., & Kraus, N. (2004). Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biological Psychology*, *67*(3), 299–317. https://doi.org/10.1016/j.biopsycho.2004.02.002

Wible, B., Nicol, T., & Kraus, N. (2005). Encoding of complex sounds in an animal model: Implications for understanding speech perception in humans. *Auditory Cortex: Towards a Synthesis of Human and Animal Research. Lawrence Erlbaum Associates, Oxford*, 241–254.

Wilding, T., McKay, C., Baker, R., & Kluk, K. (2012). Auditory Steady State Responses in Normal-Hearing and Hearing-Impaired Adults: An Analysis of Between-Session Amplitude and Latency Repeatability, Test Time, and F Ratio Detection Paradigms. *Ear and hearing*, *33*(2), 267–278. https://doi.org/10.1097/AUD.0b013e318230bba0

Wiley, R. (1981). Approximate FM Demodulation Using Zero Crossings [Conference Name: IEEE Transactions on Communications]. *IEEE Transactions on Communications*, *29*(7), 1061–1065. https://doi.org/10.1109/TCOM.1981.1095091

Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B., & Young, E. D. (1998). Effects of high sound levels on responses to the vowel /ε/ in cat auditory nerve. *Hearing research*, *123*(1-2), 61–77.

Woolf, N. K., Ryan, A. F., & Bone, R. C. (1981). Neural phase-locking properties in the absence of cochlear outer hair cells. *Hearing Research*, *4*(3), 335–346. https://doi.org/10.1016/0378-5955(81)90017-4

Wu, P. Z., Liberman, L. D., Bennett, K., de Gruttola, V., O'Malley, J. T., & Liberman, M. C. (2019). Primary Neural Degeneration in the Human Cochlea: Evidence for Hidden Hearing Loss in the Aging Ear. *Neuroscience*, *407*, 8–20. https://doi.org/10.1016/j.neuroscience.2018.07.053

Xu, L., & Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception (L). *The Journal of the Acoustical Society of America*, *114*(6), 3024–3027. https://doi.org/10.1121/1.1623786

Yin, P., Johnson, J. S., O'Connor, K. N., & Sutter, M. L. (2010). Coding of Amplitude Modulation in Primary Auditory Cortex. *Journal of Neurophysiology*, *105*(2), 582–600. https://doi.org/10.1152/jn.00621.2010

Young, E. D. (2008). Neural representation of spectral and temporal information in speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 923–945. https://doi.org/10.1098/rstb.2007.2151

Young, E. D. (2012). Neural Coding of Sound with Cochlear Damage. In C. G. Le Prell, D. Henderson, R. R. Fay, & A. N. Popper (Eds.), *Noise-Induced Hearing Loss: Scientific Advances* (pp. 87–135). Springer New York. https://doi.org/10.1007/978-1-4419-9523-0_6

Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *66*(5), 1381–1403. https://doi.org/10.1121/1.383532

Zaar, J., Dau, T., & Carney, L. H. (2019). Predicting speech intelligibility in normal-hearing and hearing-impaired listeners based on a physiologically inspired model of the auditory periphery. *Proceedings of the 23rd International Congress on Acoustics*.

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., & Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2293–2298. https://doi.org/10.1073/pnas.0406460102

Zeng, F.-G., & Turner, C. W. (1990). Recognition of voiceless fricatives by normal and hearing-impaired. *EL128EL134, 2007. Publications*, 338–346.

Zhong, Z., Henry, K. S., & Heinz, M. G. (2014). Sensorineural hearing loss amplifies neural coding of envelope information in the central auditory system of chinchillas. *Hear Res*, *309*, 55–62. https://doi.org/10.1016/j.heares.2013.11.006

Zhu, L., Bharadwaj, H., Xia, J., & Shinn-Cunningham, B. G. (2013). A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. *J Acoust Soc Am*, *134*(1), 384–395. https://doi.org/10.1121/1.4807498

VITA

Satyabrata Parida
Ph.D.
Weldon School of Biomedical Engineering,
Purdue University

🖂 715 Clinic Drive,
West Lafayette, IN-47907
🖃 parida@purdue.edu
☞ sites.google.com/view/satyabrataparida/

Education

- **Purdue University** West Lafayette, IN, US
  Ph.D., Biomedical Engineering *Aug 2015 - Present*
  - Grade Point Average: 4/4
- **Indian Institute of Technology** Kharagpur, WB, India
  Bachelor of Technology, Electrical Engineering *July 2011 - May 2015*
  - Grade Point Average: 8.84/10 (3.54/4)

Research Experience

- **Neuro- and electro-physiological modeling and experiments to quantify speech intelligibility** Aug 2015 - Dec 2020
  Graduate Research Assistant, Supervisor: Michael G. Heinz, Ph.D.
  - Developed a unifying quantitative framework to analyze invasive spike-train data and noninvasive evoked responses at high spectral specificity
  - Applied spectral, wavelet, and information-theoretic analyses to quantify speech coding fidelity in normal and impaired auditory-nerve fiber responses and frequency-following responses
  - Predicted speech intelligibility (SI) in the presence of steady-state and fluctuating noise maskers from recorded spike train responses in the framework of existing psychoacoustic SI models
- **A custom MATLAB analysis tool for multichannel EMG signal processing** Apr 2016 - Dec 2018
  With Georgia Malandraki, Ph.D., CCC-SLP, BCS-S
  - Created a MATLAB GUI for preprocessing and correlation-based analysis of multichannel EMG data collected from subjects while performing various speech and swallowing tasks
  - Future application of this tool includes biofeedback to improve treatment quality for children with unilateral brain damage
- **Visiting researcher at Technical University of Denmark** May 2017 - June 2017
  With Torsten Dau, Ph.D.

  - Extended multi-resolution speech-based envelope power spectrum model to neural domain in collaboration with the Hearing Systems Group

- **Information theoretic analysis of receptive fields of nonlinear auditory neurons** July 2014-July 2015

  Undergraduate Thesis, Advisor: Sharba Bandyaypadhyay, Ph.D.

  - Implemented a gradient ascent algorithm to find maximally informative spectral dimensions for a high dimensional non-Gaussian stimulus space that resembles natural stimuli for highly nonlinear dorsal cochlear neurons

- **Vocal tract trajectory estimation from Electromagnetic Articulography (EMA)** May 2013-Dec 2014

  Research Project at IISc Bangalore. Advisor: Prasanta Kumar Ghosh, Ph.D

  - Applied dynamic programming aided affine transformation for spatiotemporal alignment of real-time MRI and EMA to interpolate spatially rich articulator trajectories with the high temporal resolution and created a GUI for on the fly generation of midsagittal video from audio using acoustic-to-articulatory inversion

Professional Experience

- **Internship at Texas Instruments, Bangalore, India** May 2014 - July 2014

  EMI Characterization of FRAM based Metering Board

  - Designed schematics of power and communication components to match PCB specifications and executed functional Electromagnetic analysis using nWave
  - Interfaced various components with FRAM-microcontroller to achieve faster and power economic energy meter

Teaching/Mentoring Experience

- Teaching Assistant for *BME-511 Biomedical Signal Processing* for graduate students
  Purdue University, Fall 2020 and Fall 2018. Professor: Michael Heinz, Ph.D.

  - Responsibilities include designing and lecturing two weeks of the course on *frequency domain characterization of signals*, all grading, and weekly office hours

- Teaching Assistant for *BME-595 Neural Surgery and Instrumentation for Systems Neuroscience* for graduate students, Purdue University, Fall 2017. Professor: Mark Sayles, M.D., Ph.D.

  - Responsible for laboratory experiments, helping students with neural data analysis, all grading, and weekly office hours

- Mentor in Summer Undergraduate Research Fellowship Program at Purdue University
  - Involved in designing the project and guiding student throughout the project and final presentation

Journal Publications

- ***Parida, Satyabrata.***, Michael Heinz. "Noninvasive measures of distorted tonotopic speech coding following noise-induced hearing loss." *Journal of the Association for Research in Otolaryngology. 2020.*
- ***Parida, Satyabrata.***, Hari Bharadwaj, Michael G. Heinz. "Spectrally specific temporal analyses of spike-train responses to complex sounds: A unifying framework." bioRxiv (2020)

Oral Presentations

- ***Parida, Satyabrata.***, Michael G. Heinz. "Degradation of Speech-In-Noise Coding in Auditory-Nerve Fibers Following Cochlear Hearing Loss: Insights from Spectro-Temporal and Information-Theoretic Approaches." Abstract 43rd Midwinter Meeting. Association for Research in Otolaryngology, 2020, San Jose, CA
- ***Parida, Satyabrata.***, Michael G. Heinz. "Evidence for distorted tonotopy following noise-induced hearing loss using speech frequency following responses." Abstract Midwest Auditory Research Conference, 2019, Springfield, IL
- ***Parida, Satyabrata.***, Michael G. Heinz. "Effects of noise-induced hearing loss on speech-in-noise envelope coding: Inferences from single-unit and non-invasive measures in animals" Abstract 177th Meeting. Acoustical Society of America, 2019, **Invited**

Poster Presentations

- ***Parida, Satyabrata.***, Michael G. Heinz. "Effects of Noise-Induced Hearing Loss on Speech-In-Noise Envelope Coding." Abstract 42nd Midwinter Meeting. Association for Research in Otolaryngology, 2019, Baltimore, MD, PS 581
- ***Parida, Satyabrata.***, Michael G. Heinz. "Neurophysiological Evaluation of Speech Masking Release Based on the Envelope Power Spectrum Model." Abstr. 41st Midwinter Meeting. Assoc. for Research in Otolaryngology, 2018, San Diego, CA, PS 588
- ***Parida, Satyabrata.***, Michael G. Heinz. "Neurophysiological Evaluation of the Speech based Envelope Power Spectrum Model." Abstract 40th Midwinter Meeting. Association for Research in Otolaryngology, 2017, Baltimore, MD, PS 488

- ***Parida, Satyabrata.***, Ashok Kumar Pattem, and Prasanta Kumar Ghosh. “Estimation of the Air-Tissue Boundaries of the Vocal Tract in the Mid-Sagittal Plane from Electromagnetic Articulograph Data.” Sixteenth Annual Conference of the International Speech Communication Association. 2015.

Scholastic Achievements

- 2020, 2018 Neuroscience Research Travel Award, Purdue Institute for Integrative Neuroscience (PIIN)
- 2020 Tier-1 Purdue Graduate Student Government Travel Award
- 2019 Midwest Auditory Research Conference Travel Grant
- 2019 Purdue College of Engineering Conference Travel Award
- 2019 Acoustical Society of America Travel Award
- 2019 Association for Research in Otolaryngology Travel Award
- 2017 Joe Bourland Graduate Student Travel Reimbursement Award, Purdue University
- Finalist for Best Student Papers, INTERSPEECH-2015
- Qualified for IIT-JEE-2011 (All India Rank 894), AIEEE-2011 (All India Rank 264)
- Ranked 7/60k in Higher Secondary (2011) and 8/0.4 Million in Secondary (2009) Board Exams, Odisha State, India

Technical Skills

Programming Languages	C/C++/Python
Operating Systems	Ubuntu-Linux, Microsoft Windows
Other Applications	GITHUB, LaTeX, MATLAB, Eclipse

Relevant Courses

Information Theory
Biomedical Signal Processing
Probability and Random Variables
Psychophysics
Signals and Networks
Biomedical Instrumentation
Probability and Stochastic Processes
Estimation Theory
Digital Signal Processing
Machine Learning
Neural Mechanisms in Health and Disease
Programming and Data Structure
Matrix Algebra
Digital Image Processing