

2014

# Individual differences in supra-threshold auditory perception - mechanisms and objective correlates

---

<https://hdl.handle.net/2144/10939>

*Downloaded from DSpace Repository, DSpace Institution's institutional repository*

BOSTON UNIVERSITY  
COLLEGE OF ENGINEERING

Dissertation

**INDIVIDUAL DIFFERENCES IN SUPRA-THRESHOLD AUDITORY  
PERCEPTION - MECHANISMS AND OBJECTIVE CORRELATES**

by

**HARI M. BHARADWAJ**

B.Tech., Indian Institute of Technology, Madras, 2006  
M.S., University of Michigan, Ann Arbor, 2008

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

2014

© 2014 by  
HARI M. BHARADWAJ  
All rights reserved except  
Chapter 2, © by Elsevier B.V.

Approved by

First Reader



---

Barbara G. Shinn-Cunningham, PhD  
Professor of Biomedical Engineering

Second Reader



---

H. Steven Colburn, PhD  
Professor of Biomedical Engineering

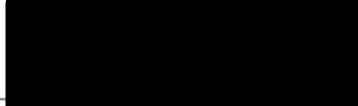
Third Reader



---

Adrian K. C. Lee, ScD  
Assistant Professor of Speech and Hearing Sciences,  
University of Washington

Fourth Reader



---

Cara E. Stepp, PhD ✓✓  
Assistant Professor of Speech, Language and Hearing Sciences

## Acknowledgments

First and foremost, I would like to express my gratitude to Barbara Shinn-Cunningham, my adviser and mentor for her guidance throughout my time as a graduate student. Her mentorship was paramount in providing a well-rounded experience consistent with my long-term career goals. By being a pillar of unfailing support, Barb gave me the freedom and confidence to pursue unfamiliar areas of inquiry. I have truly enjoyed myself along the way and could not be prouder of my academic roots.

I would like to thank Adrian KC Lee for introducing me to Barb and for vouching for me to help get my graduate career started. I have learnt a lot, both scientifically and professionally, from observing KC in action and I am thankful for that.

I would like to extend special thanks to Steve Colburn for his warm presence and guidance throughout my time at BU and for being a role model who I can only hope to emulate. Further, I would like to thank Steve for heading the Hearing Research Center and for creating a stimulating environment for honest and open exchange of ideas.

I would also like to thank my other examiners, Jason Ritt and Cara Stepp for thoughtful, encouraging and constructive feedback throughout the development and solidification of the thesis.

To my comrades at the Auditory Neuroscience Lab, thank you for welcoming me as a friend and for making my time here most enjoyable. I am particularly grateful for your collaborative spirit and for the many insightful discussions on various aspects of hearing research.

I thank Dorea Ruggles, Sharon Kujawa, Charles Liberman and Luke Shaheen for paving the way for my thesis through their research efforts.

Finally, I would like to thank my family starting with my wife Vibha. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past few years of my life have been built. I thank my parents for instilling in me all those qualities and moral values that have aggregated to constitute my identity.

**INDIVIDUAL DIFFERENCES IN SUPRA-THRESHOLD AUDITORY  
PERCEPTION - MECHANISMS AND OBJECTIVE CORRELATES**

(Order No.                    )

**HARI M. BHARADWAJ**

Boston University, College of Engineering, 2014

Major Professor: Barbara G. Shinn-Cunningham, PhD,

Professor of Biomedical Engineering

**ABSTRACT**

To extract content and meaning from a single source of sound in a quiet background, the auditory system can use a small subset of a very redundant set of spectral and temporal features. In stark contrast, communication in a complex, crowded scene places enormous demands on the auditory system. Spectrotemporal overlap between sounds reduces modulations in the signals at the ears and causes masking, with problems exacerbated by reverberation. Consistent with this idea, many patients seeking audiological treatment seek help precisely because they notice difficulties in environments requiring auditory selective attention. In the laboratory, even listeners with normal hearing thresholds exhibit vast differences in the ability to selectively attend to a target. Understanding the mechanisms causing these supra-threshold differences, the focus of this thesis, may enable research that leads to advances in treating communication disorders that affect an estimated one in five Americans.

Converging evidence from human and animal studies points to one potential source of these individual differences: differences in the fidelity with which supra-threshold sound is encoded in the early portions of the auditory pathway. Electrophysiological measures of sound encoding by the auditory brainstem in humans and animals support the idea that the temporal precision of the early auditory neural representation can be poor even when hearing thresholds are normal. Concomitantly, animal studies show that noise exposure

and early aging can cause a loss (cochlear neuropathy) of a large percentage of the afferent population of auditory nerve fibers innervating the cochlear hair cells without any significant change in measured audiograms.

Using behavioral, otoacoustic and electrophysiological measures in conjunction with computational models of sound processing by the auditory periphery and brainstem, a detailed examination of temporal coding of supra-threshold sound is carried out, focusing on characterizing and understanding individual differences in listeners with normal hearing thresholds and normal cochlear mechanical function. Results support the hypothesis that cochlear neuropathy may reduce encoding precision of supra-threshold sound, and that this manifests as deficits both behaviorally and in subcortical electrophysiological measures in humans. Based on these results, electrophysiological measures are developed that may yield sensitive, fast, objective measures of supra-threshold coding deficits that arise as a result of cochlear neuropathy.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Sensory encoding - Electrophysiological measures . . . . .	3
1.2	Cochlear neuropathy - Evidence, consequences for supra-threshold sensory coding, and predictions for subcortical steady responses . . . . .	4
1.3	Relationship between cochlear mechanics, early neural encoding and behavior	6
<b>2</b>	<b>Rapid acquisition of subcortical steady-state responses with multichannel recordings</b>	<b>7</b>
2.1	Introduction . . . . .	8
2.2	Methods . . . . .	11
2.2.1	Complex principal component analysis (cPCA) . . . . .	12
2.2.2	Simulations . . . . .	16
2.2.3	EEG Data . . . . .	17
2.3	Results . . . . .	20
2.3.1	Simulations . . . . .	20
2.3.2	Human EEG Data . . . . .	23
2.4	Discussion . . . . .	29
2.4.1	Clinical use . . . . .	29
2.4.2	Set-up time versus recording length . . . . .	30
2.4.3	The role of raw-signal narrowband SNR and other sources of variability	30
2.4.4	Source separation versus cPCA . . . . .	33
2.4.5	Optimal recording configuration . . . . .	35



2.5	Conclusions . . . . .	35
<b>3</b>	<b>Cochlear neuropathy and the coding of supra-threshold sound</b>	<b>37</b>
3.1	Introduction . . . . .	38
3.2	Coding of supra-threshold sound . . . . .	41
3.2.1	The diversity of auditory nerve fibers . . . . .	41
3.2.2	Temporal coding and its importance for auditory perception . . . . .	43
3.2.3	Consequences of cochlear neuropathy for temporal coding . . . . .	45
3.2.4	Objective measures of subcortical temporal coding . . . . .	46
3.3	Evidence for cochlear neuropathy . . . . .	52
3.3.1	Neuropathy and selective loss of lower-SR fibers in animals . . . . .	52
3.3.2	Human data consistent with the neuropathy hypothesis . . . . .	55
3.4	Diagnosing cochlear neuropathy . . . . .	58
3.4.1	Measuring brainstem coding: ABRs versus SSSRs . . . . .	59
3.4.2	Emphasizing the contribution of lower-SR auditory nerve fibers to the envelope following response . . . . .	61
3.4.3	Isolating cochlear neuropathy . . . . .	64
3.5	Future Experiments . . . . .	66
3.5.1	Accounting for individual differences in cochlear mechanical function	66
3.5.2	Developing quantitative models of envelope following responses . . .	67
3.5.3	Using envelope following responses to assess supra-threshold coding fidelity . . . . .	68
3.6	Summary and Conclusions . . . . .	71
<b>4</b>	<b>Individual differences in supra-threshold auditory perception are consis-</b> <b>tent with early effects of noise-exposure and aging</b>	<b>73</b>
4.1	Introduction . . . . .	74
4.2	Methods . . . . .	78
4.2.1	Participants . . . . .	78

4.2.2	Characterization of cochlear-mechanical function . . . . .	80
4.2.3	Behavioral measures of temporal coding . . . . .	81
4.2.4	Electrophysiological measure of temporal coding . . . . .	83
4.2.5	ITD based attention task . . . . .	85
4.3	Results . . . . .	86
4.3.1	Correlates of peripheral processing . . . . .	86
4.3.2	Correlates of temporal coding fidelity . . . . .	87
4.3.3	Envelope following response results . . . . .	88
4.3.4	Relationship between attention task performance and correlates of cochlear amplification and supra-threshold temporal coding . . . . .	89
4.4	Discussion . . . . .	90
4.5	Conclusions . . . . .	93
<b>5</b>	<b>Conclusions and Future Work</b>	<b>94</b>
5.1	Summary of findings . . . . .	94
5.2	Significance . . . . .	95
5.3	Future research . . . . .	96
	<b>List of Journal Abbreviations</b>	<b>98</b>
	<b>Bibliography</b>	<b>101</b>
	<b>Curriculum Vitae</b>	<b>116</b>

## List of Tables

4.1	Data acquisition summary . . . . .	79
-----	------------------------------------	----

## List of Figures

2.1	Phase disparities between steady-state responses recorded at different scalp locations . . . . .	12
2.2	Performance of the frequency domain complex-PCA on simulated steady-state responses . . . . .	21
2.3	Relationship between estimated and true between-channel phase variations using complex-PCA . . . . .	22
2.4	Performance of complex-PCA on real human recordings and comparison to traditional single-channel approach . . . . .	25
2.5	Comparison of signal-to-noise ratios obtained using the multichannel complex-PCA and traditional single-channel approaches . . . . .	27
2.6	<i>A posteriori</i> analysis of SNR gain as a function of number of channels when using the complex-PCA approach to extract subcortical steady-state responses	28
2.7	Simulation results illustrating the signal-to-noise ratio dependent bias introduced in the estimated phase-locking-value . . . . .	32
3.1	Synapses of the inner hair cell . . . . .	47
3.2	Diversity of auditory nerve fibers, spontaneous rates and associated properties	48
3.3	Early neural coding of temporal information and relationship to listening in a complex scene . . . . .	53
3.4	Subcortical steady-state responses and phase-locking value . . . . .	57
3.5	A simple model of brainstem population responses to amplitude-modulated stimuli . . . . .	65

4.1	Low-spontaneous-rate selective cochlear neuropathy in noise-exposure and aging . . . . .	81
4.2	Simple model of brainstem population response to amplitude-modulated stimuli suggesting that high sound levels and shallow modulations may maximize the contributions of low-SR fibers . . . . .	83
4.3	Spatial attention task setup . . . . .	86
4.4	DPOAEs obtained from listeners with normal hearing thresholds and hearing loss . . . . .	87
4.5	Behavioral and electrophysiological measures of temporal coding and their relationships . . . . .	90

## List of Abbreviations

2AFC .....	Two alternatives, forced choice
AM .....	Amplitude modulation
AMF .....	Amplitude modulation following response
ANF .....	Auditory nerve fiber
APD .....	Auditory processing disorder
ASSR .....	Auditory steady-state response
CN .....	Cochlear nucleus
CAPD .....	Central auditory processing disorder
cPCA .....	complex principal component analysis
DPOAE .....	Distortion product oto-acoustic emission
DPSS .....	Discrete prolate-spheroidal sequence
EEG .....	Electroencephalogram
EFR .....	Envelope following response
ENV .....	Envelope
ERB .....	Equivalent rectangular bandwidth
FM .....	Frequency modulation
FFR .....	Frequency following response
HL .....	Hearing level
IC .....	Inferior colliculus
ILD .....	Interaural level difference
ITD .....	Interaural time difference
MCMC .....	Monte-carlo Markov chain

NHT .....	Normal hearing thresholds
OAD .....	Obscure auditory dysfunction
PCA .....	Principal component analysis
PLV .....	Phase-locking value
PTA .....	Pure-tone audiogram
SNR .....	Signal-to-noise ratio
SPL .....	Sound pressure level
SR .....	Spontaneous (discharge) rate
SSSR .....	Subcortical steady-state response
TFS .....	Temporal fine structure

## Chapter 1

### Introduction

The ability to attend to a sound source of interest while ignoring irrelevant sounds is vital for successfully navigating the complex acoustic environments that we find ourselves in on an everyday basis. Yet, aging and even modest hearing loss impair the ability to direct selective attention leading to a disproportionately severe handicap in navigating crowded, noisy settings such as restaurants, cocktail parties and busy streets (Dubno, 1984; Gatehouse and Noble, 2004). In addition, some listeners, despite their clinical normal hearing status and seemingly sufficient cognitive prowess, seek audiological help precisely because they find it difficult communicating in such noisy environments (Hind et al., 2011; Dawes and Bishop, 2009; Chermak and Musiek, 1997; Saunders and Haggard, 1992; Catts et al., 1996).

Clinically, normal hearing is defined based on pure-tone threshold audiometry. Hearing threshold is the level of the faintest tone that a listener can detect at a given frequency. A diagnosis of hearing loss is given only when a listener's thresholds exceed normal levels by a criterion amount (typically 20 dB HL). However, being able to listen effectively in a complex environment conceivably requires a lot more than just being able to detect faint tones. Thus, it is not surprising that having normal thresholds does not guarantee the ability to successfully communicate in everyday settings. In laboratory setting, this is manifested as large individual differences in tasks that mimic everyday listening conditions (Ruggles and Shinn-Cunningham, 2011; Ruggles et al., 2011, 2012).

Unlike listening to a single source of sound in a quiet background, selective attention is a complex neuropsychological process that places enormous demands on the auditory



system (Cherry, 1953). First, acoustic information about both the target sounds and the irrelevant signals in the environment has to be robustly encoded by the auditory periphery and the early neural pathway. This encoded neural information must then be effectively parsed and used to form coherent perceptual “objects” (Shinn-Cunningham, 2008; Shinn-Cunningham and Best, 2008). Throughout this process, resources such as prior experience, working memory, language ability, etc., are continuously drawn upon. Consequently, differences across individuals can arise at one or more stages of this processing hierarchy. Indeed, individual differences in general cognitive ability, working memory, familiarity with the target sounds etc., can contribute to individual differences in selective attention performance (e.g., see Conway et al., 2001; Colflesh and Conway, 2007; Kidd et al., 2007; Surprenant and Watson, 2001; Drennan and Watson, 2001). However, individual differences in auditory sensory encoding may also lead to individual differences in selective attention ability, because, deficient encoding places a fundamental limit on performance (Ruggles and Shinn-Cunningham, 2011; Ruggles et al., 2011, 2012).

Though sensory encoding is in general not a phenomenon of purely “bottom-up” processes, it is useful to consider its role separately (Gilbert and Sigman, 2007; Pessoa et al., 2003; Shinn-Cunningham, 2008). A permanent threshold shift, that is identified clinically as hearing loss, is an example of deficiency in sensory encoding that impacts listening in complex environments. However, the seemingly high prevalence of difficulties in noisy environments among listeners with normal hearing thresholds (NHTs) and the apparent disproportionate difficulties among middle-aged and older listeners with NHTs or modest threshold elevations (Hind et al., 2011; Chermak and Musiek, 1997; Saunders and Haggard, 1992; Fitzgibbons and Gordon-Salant, 2010; Snell et al., 2002; Catts et al., 1996), suggest that sensory encoding deficits may exist in the general population beyond the mechanisms leading to clinical hearing loss and threshold elevation.

This thesis is dedicated to the study of individual differences among NHT listeners and documents “supra-threshold” differences in electrophysiological measures (1.1) of sensory encoding of acoustic information and its contribution to individual differences in listening

in a complex scene. Taken together, the results of the various experiments and modeling efforts provide evidence that a dominant contribution to individual differences in supra-threshold perceptual abilities of NHT listeners comes from differences in a very early *neural* component of sensory encoding. In addition, these differences are consistent with cochlear neuropathy that has been shown to occur in noise-exposure and early aging in animal models.

Each chapter is written in the format of a journal article with sufficient background and discussion to stand on its own. The remaining sections of this introductory chapter describe the organization and contents of the chapters to follow.

## 1.1 Sensory encoding - Electrophysiological measures

A significant portion of this thesis is devoted to the development and application of electrophysiological measures of brainstem responses to sounds. While different perceptual attributes of sound are related to different spectro-temporal acoustic features, many depend on temporal information. As a result of cochlear filtering, each auditory nerve fiber (ANF) is essentially driven by a narrow frequency band of sound energy. Thus, the temporal information encoded by the ANFs can be logically separated into two parts; the temporal fine-structure (TFS), corresponding to the timing of the nearly sinusoidal narrowband carrier fluctuations, and the slower temporal envelope of that carrier, whose temporal fluctuations are limited by the bandwidth of the corresponding cochlear filter. For low-frequency cochlear channels, ANFs convey both TFS and envelope information; neural spikes are phase-locked to the carrier and the instantaneous firing rate, in steady-state approximately follows the envelope. At higher frequencies, ANFs do not phase lock to the TFS; however, responses convey temporal information by phase locking to envelope fluctuations. Several aspects of sound perceptions, such as perception of speech (Zeng et al., 2005), source location (Blauert, 1997), grouping of acoustic constituents into objects (Elhilali et al., 2009), and release from various kinds of maskers (Moore, 2008; Christiansen

et al., 2013), depend on using TFS and ENV cues.

Subcortical steady state responses (SSSRs), frequently referred to as frequency following responses (FFRs), are the scalp-recorded responses originating from sub-cortical portions of the auditory nervous system and provide an objective correlate of temporal encoding. These responses phase lock to periodicities in the acoustic waveform and to periodicities induced by cochlear processing (Glaser et al., 1976). The responses specifically phase locked to the envelopes of amplitude modulated sounds are sometimes called amplitude modulation following responses (AMFRs) or envelope following responses (EFRs) (Dolphin and Mountain, 1992; Kuwada et al., 2002). Responses to amplitude-modulated sounds originating from both the sub-cortical and cortical portions of the auditory pathway are also collectively referred to as auditory steady-state responses (ASSR) (Rees et al., 1986). In contrast to auditory brainstem responses (ABRs; the stereotypical responses to sound onsets and offsets; Jewett et al., 1970), SSSRs are the sustained responses to ongoing sounds. SSSRs have been used extensively in basic neurophysiologic investigation of auditory function and sound encoding (e.g., Aiken and Picton, 2008; Kuwada et al., 1986; Gockel et al., 2011, also see Chandrasekaran and Kraus, 2010; Krishnan, 2006; Picton et al., 2003a, for reviews). Given the frequency specificity possible with SSSRs, they have also been recommended for objective clinical audiometry (Lins et al., 1996).

A detailed examination of temporal coding at the level of individual listeners using SSSRs requires that SSSRs be acquired for several stimulus manipulations. This is the subject of Chapter 2. This chapter was published as a peer-reviewed journal article in *Clinical Neurophysiology* (Bharadwaj and Shinn-Cunningham, 2014).

## **1.2 Cochlear neuropathy - Evidence, consequences for supra-threshold sensory coding, and predictions for subcortical steady responses**

The much-studied component of sensory encoding deficiencies is the degradation in the mechanics and associated changes in the functioning of the so called “cochlear amplifier”.

This type of “cochlear hearing loss” is often associated with the interrelated phenomena of elevated thresholds, loss of cochlear gain at low sound levels, loss of cochlear-compression at mid-levels, loudness recruitment, loss of frequency selectivity, etc. (see Moore, 2007 for an in-depth treatment of the subject). Correlates of cochlear mechanical function include psychophysical measures of gain and tuning (Oxenham and Shera, 2003; Glasberg and Moore, 1990) and otoacoustic emissions (Shera and Guinan Jr, 1999; Neely et al., 2003). Deficits in basilar membrane function resulting in cochlear mechanical dysfunction and hearing loss are often associated with the loss of outer hair cells of the organ of Corti while the inner hair cells, the site of the origin afferent neural process are spared (Stebbins et al., 1979). However, following the work of Kujawa and Liberman (2009), it has come to light that even when the inner hair cells and afferent nerve cells are themselves intact, noise exposure and aging can result in the loss of (a significant proportion of) synapses and unmyelinated terminals of the cochlear nerve (“cochlear neuropathy”) innervating the inner hair cells without permanent elevation of thresholds. Further, these findings also suggest that the cell bodies in the spiral ganglia remain intact for a considerable period (about 10-20% of average lifespan) following terminal loss, further bringing to question the implications of the earlier observation that noise-exposure spares afferent nerve cells.

Chapter 3 is dedicated to review of the evidence for and implications of cochlear neuropathy on supra-threshold sensory coding and was published as a peer-reviewed journal article in *Frontiers in Neuroscience* (Bharadwaj et al., 2014b). Please refer to the preamble of Chapter 3 for the details regarding author contributions. Quantitative models of sound processing by the auditory nerve (Zilany and Bruce, 2006; Zilany et al., 2009, 2014) and brainstem neurons (Nelson and Carney, 2004) are combined to create models of population responses (Dau, 2003; Rønne et al., 2012). These models are used to design stimuli and make predictions about SSSRs in the context of cochlear neuropathy.

### **1.3 Relationship between cochlear mechanics, early neural encoding and behavior**

Chapter 4 presents the results of the detailed characterization of sensory coding differences across individual listeners. Objective and behavioral correlates of both cochlear mechanical function and sub-cortical temporal coding are obtained and the relative contributions of the two factors to individual differences in listening in a complex scene are evaluated. Distortion product otoacoustic emission measures (DPOAEs) are used as objective correlates of cochlear gain and compression across listeners. SSSRs are used to obtain an electrophysiological correlate of temporal coding. Stimulus regimes and manipulations are designed based on model predictions from Chapter 3.

Finally, a concluding chapter summarizes the major findings and suggests future work.

## Chapter 2

# Rapid acquisition of subcortical steady-state responses with multichannel recordings

### Preamble

This chapter develops a novel data acquisition and signal extraction approach to record subcortical steady-state responses (SSSRs) efficiently and discussed the implications both for clinical and basic neurophysiological research. The work was completed between Sep 2011 and Oct 2013 and published as a peer-reviewed journal article in *Clinical Neurophysiology* (Bharadwaj and Shinn-Cunningham, 2014). The documented software accompanying the manuscript was made publicly available at <http://nmr.mgh.harvard.edu/~hari/ANLffr/> and was used for all the analyses leading to the results presented in Chapter 4.

### Abstract

Auditory subcortical steady state responses (SSSRs), also known as frequency following responses (FFRs), provide a non-invasive measure of phase-locked neural responses to acoustic and cochlear-induced periodicities. SSSRs have been used both clinically and in basic neurophysiological investigation of auditory function. SSSR data acquisition typically involves thousands of presentations of each stimulus type, sometimes in two polarities, with acquisition times often exceeding an hour per subject. Here, we present a novel approach to reduce the data acquisition times significantly. Because the sources of the SSSR are deep compared to the primary noise sources, namely background spontaneous cortical activity,

the SSSR varies more smoothly over the scalp than the noise. We exploit this property and extract SSSRs efficiently, using multichannel recordings and an eigendecomposition of the complex cross-channel spectral density matrix. Our proposed method yields SNR improvement exceeding a factor of 3 compared to traditional single-channel methods. It is possible to reduce data acquisition times for SSSRs significantly with our approach. The proposed method allows SSSRs to be recorded for several stimulus conditions within a single session and also makes it possible to acquire both SSSRs and cortical EEG responses without increasing the session length.

## 2.1 Introduction

Subcortical steady state responses (SSSRs), frequently referred to as frequency following responses (FFRs), are the scalp-recorded responses originating from sub-cortical portions of the auditory nervous system. These responses phase lock to periodicities in the acoustic waveform and to periodicities induced by cochlear processing (Glaser et al., 1976). The responses specifically phase locked to the envelopes of amplitude modulated sounds are sometimes called amplitude modulation following responses (AMFRs) or envelope following responses (EFRs) (Dolphin and Mountain, 1992; Kuwada et al., 2002). Responses to amplitude-modulated sounds originating from both the sub-cortical and cortical portions of the auditory pathway are also collectively referred to as auditory steady-state responses (ASSR) (Rees et al., 1986). In contrast to auditory brainstem responses (ABRs; the stereotypical responses to sound onsets and offsets; Jewett et al., 1970), SSSRs are the sustained responses to ongoing sounds and include responses phase-locked to both the fine structure and the cochlear induced envelopes of broadband sounds. Since the term FFR, originally used to denote phase locked responses to pure tones, is suggestive of responses phase-locked to the fine-structure of narrowband or locally narrowband sounds, here we will use the term SSSR to describe the sustained responses originating from subcortical portions of the auditory pathway. This name distinguishes them from transient onset-

offset related responses and responses generated at the cortical level. SSSRs have been used extensively in basic neurophysiologic investigation of auditory function and sound encoding (e.g., Aiken and Picton, 2008; Kuwada et al., 1986; Gockel et al., 2011, also see Chandrasekaran and Kraus, 2010; Krishnan, 2006; Picton et al., 2003a, for reviews). Given the frequency specificity possible with SSSRs, they have also been recommended for objective clinical audiometry (Lins et al., 1996).

SSSRs are traditionally recorded with a single electrode pair placed in either a vertical or a horizontal montage (which differ in which underlying generators are emphasized; see Krishnan, 2006; Skoe and Kraus, 2010). To achieve an adequate signal-to-noise ratio (SNR) when measuring the SSSR, the stimulus is typically repeated thousands of times. Often, stimuli are presented in opposite polarities to separate the response components phase locked to the envelope from those phase locked to the fine structure of the acoustic waveform (Aiken and Picton, 2008; Ruggles et al., 2012). Since many studies require SSSR data acquisition for multiple conditions or with multiple stimuli, this often results in recording sessions exceeding an hour per subject.

Multichannel electroencephalography (EEG), which is widely used for the investigation of cortical processing, uses the same basic sensors as SSSR measurements, but requires many fewer trials because the cortical response generators are closer to the scalp and produce stronger electric fields. EEG systems with high-density arrays include as many as 64, 128, or sometimes even 256 scalp electrodes. Although the frequency response characteristics of some cortical EEG systems are not always optimized for picking up subcortical signals (which typically are at 80 Hz and above), these multi-electrode setups can nonetheless be used to record SSSR data from multiple scalp locations.

Given this, it is possible to simultaneously record subcortical and cortical processing of sounds with the high-frequency portions analyzed to yield SSSRs and the low-frequency portions representing cortical activity (Krishnan et al., 2012). The primary source of noise for the high-frequency SSSR portion of the recordings is background cortical activity (i.e., neural noise). Since the SSSR sources are deep compared to the dominant sources of



noise (in the cortex), the SSSR varies more smoothly over the scalp than the noise (for a discussion of the physics of the measurement process and how scalp fields relate to neural activity, see Hämäläinen et al., 1993). Scalp fields arising from cortical sources can cancel each other out if they are out of phase (Irimia et al., 2012). This can be exploited to help separate cortical and subcortical responses from the same EEG recordings by combining information obtained from a dense sensor array. Here, we propose and evaluate one method for combining measurements from multiple scalp channels to improve the SNR of SSSRs measured using cortical EEG arrays.

Although SSSRs can provide insight into auditory function and subcortical encoding, interpreting them can be a challenge. Multichannel recordings of brainstem responses have been used primarily in the analysis of the sources of the onset ABR, in which the activity from different generators can be temporally separated, into stereotypical responses known as waves I, III, and V (Grandori, 1986; Parkkonen et al., 2009; Scherg and Von Cramon, 1985). In contrast, since the SSSRs represent sustained activity, temporal separation of the activity from different generators is not possible. Moreover, in any narrow frequency band, particularly at high frequencies, multiple SSSR sources likely contribute to the aggregate measured response, each of which is a phase-locked response at a different phase. This notion is consistent with the observation that there are spectral notches and occasional phase discontinuities in the SSSR as a function of modulation frequency for amplitude modulated stimuli (Dolphin and Mountain, 1992; Kuwada et al., 2002; Purcell et al., 2004). This is also consistent with the observation that responses are attenuated but not eliminated in studies inducing isolated lesions of single auditory nuclei (Smith et al., 1975; Kiren et al., 1994). This multisource population activity produces scalp potentials that are different mixtures of the source activity at different scalp locations, depending on the geometry of the generators, the recording electrodes, and the volume conductor in between (Hubbard et al., 1971; Okada et al., 1997; Irimia et al., 2013). Consistent with this notion, the steady-state phase of the summed, observed response at a given frequency varies across different channels, as illustrated in Figure 2.1.

Unfortunately, time-domain methods to combine multichannel recordings, such as simple across electrode averaging or principal component analysis (PCA), assume that the signal is at the same phase across sensors. For instance, time-domain PCA involves recombination of multiple measurements with real-valued weights based on the covariance matrix. As a result, these methods lead to signal attenuation when the signal components in each sensor are not at the same phase. In other fields of analysis, complex principal component analysis (cPCA) in the frequency domain has been used to effectively combine multiple measurements when the signal components are correlated, but have phase differences (Brillinger, 2001; Horel, 1984). In contrast to traditional time-domain PCA, frequency domain cPCA recombines measurement channels using the complex-valued weights obtained by decomposing the complex cross-channel spectral density matrix. The weights thus include channel-specific magnitudes and phases in each frequency bin; the phases of each complex weight specifically adjust for phase differences between responses measured at different sites to optimally combine responses across multiple sensors. Here we apply cPCA to multichannel EEG recordings, thereby accounting for phase discrepancies across the scalp and extract SSSRs efficiently. We show that compared to single-channel recording, this approach reduces the data acquisition required to achieve the same SNR, both when applied to simulations and when analyzing real multichannel SSSR recordings.

## 2.2 Methods

First, we describe the steps involved in the cPCA method. Then, we describe our procedure to validate the method using simulated data. Finally, using EEG-data acquired from normal-hearing human listeners, we demonstrate how to apply the approach to extract SSSRs from multi-electrode recordings.

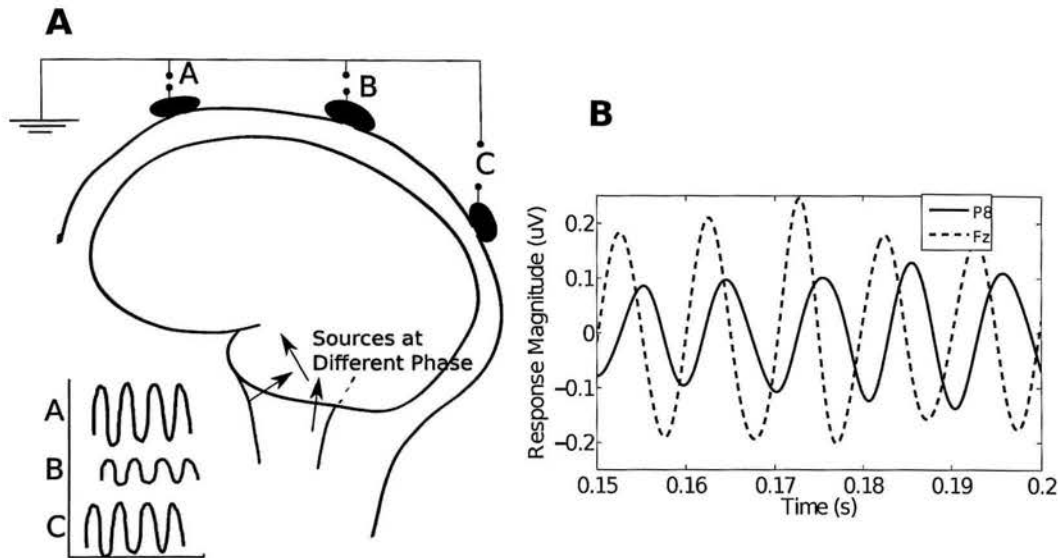


Figure 2.1: (A) A schematic illustration of the possible origin of phase differences of the SSSR recorded from different scalp electrodes. Each neural generator, shown as three different colored arrows, is phased-locked to the stimulus, but at a different unique phase. Moreover, the generators contribute different amounts to different scalp locations, as illustrated by the proportion of the ellipses shaded with the corresponding colors. This results in phase misalignment between the effective total response at different recording sites. (B) Real SSSR obtained from a typical subject from two distinct scalp locations (relative to the average potential between the two earlobes) showing phase differences in the response. The data is filtered between 90 and 110 Hz to emphasize the response at the fundamental stimulus frequency of 100 Hz.

### 2.2.1 Complex principal component analysis (cPCA)

Frequency-domain PCA can be used to effectively reduce the dimensionality of vector-valued time-series in the presence of between-component dependencies at delayed time intervals (Brillinger, 2001). As illustrated in Figure 2.1, for any frequency component, responses at different scalp locations occur with different effective phases. This is unlikely to be due to conduction delays between the recording site and the sources since the brain tissue and head together can be treated as a pure conductor (no capacitive effects) for frequencies below about 20 kHz. That is, the forward model that relates the measured potentials to the source currents can be treated as quasi-static (Hämäläinen et al., 1993). Because each subcortical source contributes a different amount to each scalp sensor, de-

pending on their geometry relative to the recording electrodes, the shapes and conductivity profiles of the different tissues in between, the choice of reference etc. (Hubbard et al., 1971; Okada et al., 1997; Irimia et al., 2013), the resultant signal in each sensor will have a different phase. For multivariate time-series, the cross-spectral density matrix captures up to second-order dependencies between the individual components of the time-series, and therefore has information we can exploit to account for these phase differences in the signal across the sensors. First, we apply a discrete prolate-spheroidal taper sequence,  $w_k(t)$ , to the recorded/simulated signals (Slepian, 1978). We then estimate the complex-cross channel spectral density matrix,  $M(f)$ , at each frequency bin, from which we estimate the principal eigenvalue,  $\lambda_k(f)$ , and the corresponding eigenweights,  $v_k(f)$ , using a diagonalization procedure. For a given frequency resolution, the use of the Slepian taper minimizes the bias introduced due leakage of frequency content from other frequencies outside the resolution bandwidth into the estimate of the passband content (Thomson, 1982). Consequently, the Slepian taper minimizes the bias in the estimates of the eigenvalues,  $\lambda(f)$ , which result from the spectra being colored (Brillinger, 2001). Thus, for recording epochs of duration  $T$ , we have:

$$X_i(f) = \sum_{t=0}^T w_k(t)x_i(t)\exp\{-j2\pi ft\} \quad (2.1)$$

$$M_{ij}(f) = \langle X_i(f)X_j^*(f) \rangle \quad (2.2)$$

where  $X_i(f)$  denotes the tapered Fourier transform of the data  $x_i(t)$  in the  $i^{th}$  recording channel,  $M_{ij}(f)$  denotes the cross-spectrum between channels  $i$  and  $j$ , corresponding to the  $ij^{th}$  element of the full cross-channel spectral density matrix  $M(f)$ , superscript  $*$  denotes complex-conjugate, and  $\langle \cdot \rangle$  denotes averaging over trials. The tapers  $w_k$ ,  $k = 1, 2, \dots, N_{tap}$  form an approximate basis for signals that are limited to a duration-bandwidth product of  $2TW$ , and satisfy an eigenvalue equation:

$$\sum_{s=1}^T \frac{\sin 2\pi W(t-s)}{\sin \pi(t-s)} w_k(s) = c_k w_k(s) \quad (2.3)$$

where the eigenvalues  $c_k \approx 1$  for  $k \leq N_{tap} = 2TW - 1$  are the concentrations of the tapers within the band  $-W \leq f \leq W$  (Slepian, 1978). The estimates of  $\lambda_k(f)$  obtained using the  $N_{tap}$  tapers (indexed by  $k$ ) are then averaged together to reduce the variance of the estimate without additional bias from spectral leakage from frequencies outside of the bandwidth  $W$ :

$$\lambda(f) = \sum_{k=1}^{N_{tap}} \lambda_k(f) \quad (2.4)$$

By construction,  $M(f)$  is Hermitian and positive semi-definite. Thus,  $M(f)$  has real, non-negative eigenvalues, and can be diagonalized using a Cholesky factorization procedure:

$$M(f) = Q(f)\Lambda(f)Q^H(f) \quad (2.5)$$

where  $Q(f)$  is the unitary matrix of complex eigenvectors of  $M(f)$ ,  $\Lambda(f)$  is the diagonal matrix of real eigenvalues and superscript  $H$  denotes conjugate-transpose. Note that a separate cross-channel spectral density matrix is estimated at each frequency bin and the eigendecomposition is also performed at each frequency bin separately. This is not redundant because, by the central limit theorem, for a stationary signal, the estimated frequency coefficients are uncorrelated and asymptotically Gaussian distributed. The principal eigenvalue  $\lambda(f) = \Lambda_{11}(f)$  estimates the power spectrum of the first principal component signal (Brillinger, 2001) and the phase of the corresponding eigenvector  $v(f) = Q_{\cdot 1}(f)$  estimates the phase delays that need to be applied to individual channels in order to maximally align them. Here  $Q_{\cdot 1}$  denotes the first column vector of  $Q$ , composed of the first element of all the rows of the matrix.

SSSRs are often appropriately analyzed in the frequency domain as they represent steady-state mixtures of subcortical source activity (Aiken and Picton, 2008; Gockel et al.,

2011; Krishnan, 1999, 2002). The eigendecomposition of the cross-spectral matrix is thus convenient for SSSR analysis in the sense that the principal eigenvalue directly provides a metric of the SSSR power at a given frequency without further processing being necessary. Moreover, the phase-locking value (PLV) (Lachaux et al., 1999) is a normalized, easily interpreted measure of across-trial phase locking of the SSSR at different frequencies and also has convenient statistical properties (Dobie and Wilson, 1993; Zhu et al., 2013). The use of the normalized cross-spectral density matrix  $C^{plv}(f)$  (see below) instead of  $M(f)$  allows for the direct estimation of PLV through the eigendecomposition. An analogous modification can be used to obtain estimates of inter-trial coherence (ITC) (see Delorme and Makeig, 2004; Dobie and Wilson, 1994) by using  $C^{itc}(f)$ , as defined below:

$$C_{ij}^{plv}(f) = \left\langle \frac{X_i(f)X_j^*(f)}{|X_i(f)||X_j(f)|} \right\rangle \quad (2.6)$$

$$C_{ij}^{itc}(f) = \frac{\langle X_i(f)X_j^*(f) \rangle}{\langle |X_i(f)| \rangle \langle |X_j(f)| \rangle} \quad (2.7)$$

where  $C_{ij}(f)$  denotes the  $ij^{th}$  element of the corresponding normalized cross-spectral density matrix  $C(f)$  and  $\langle \cdot \rangle$  denotes averaging over trials. The set  $p(f)$  of the largest eigenvalues of  $C(f)$  at each frequency bin then provides the PLV or the ITC of the corresponding first principal component directly. The variance of the PLV and ITC estimates depend only on the number of trials included in the estimation (Bokil et al., 2007; Zhu et al., 2013). In the absence of a phase-locked signal component, both the mean (bias) and the variance of the estimated PLV (i.e., the noise floor) are directly related to the number of trials. Taking advantage of this, in order to compare the SNR obtained using the cPCA method to the SNR from a single channel and from time-domain PCA, we normalize the PLV measure so that the noise floor is approximately normally distributed with a zero-mean and unit variance:

$$PLV_z(f) = \frac{PLV(f) - \mu_{noise}}{\sigma_{noise}} \quad (2.8)$$

where  $\mu_{noise}$  and  $\sigma_{noise}$  denote the sample mean and standard deviation of the noise floor estimated from the PLV or the ITC spectrum  $p(f)$  using a bootstrap procedure (Zhu et al., 2013). Alternately, similar estimates of noise can be obtained from  $p(f)$  by excluding the frequency bins that have stimulus-driven response components.  $PLV_z(f)$  gives the PLV as a function of frequency measured in  $z$ -scores relative to the noise floor and hence quantifies the SNR obtained using different methods, allowing them to be compared directly.

### 2.2.2 Simulations

Simulated SSSR recordings were produced by generating 32 channels of data with each containing a 200 ms burst of a 100 Hz sinusoid at a different randomly chosen phase (distributed uniformly around the circle). Background EEG-like noise was added to generate 200 simulated trials of raw EEG data. The noise had the same spectrum and spatial (between-channel) covariance as resting state EEG (note that the background cortical EEG activity is itself one of the primary sources of noise for SSSR measurements). The phase of the 100 Hz sinusoid (the SSSR signal of interest), though not aligned across channels, was kept constant across trials within each channel. The root-mean-squared (RMS) SNR for a single trial in each channel was set at -40 dB. This is comparable to typical SNRs for SSSRs obtained with our EEG setup, where the SSSR amplitude is typically on the order of hundreds of nanovolts, while the background, narrowband EEG amplitude is on the order of tens of microvolts. The SSSR was then extracted from the simulated data using traditional time-domain PCA as well as cPCA.

### 2.2.3 EEG Data

#### 2.2.3.1 Participants

Nine participants aged 20 to 40 were recruited from the Boston University community in accordance with procedures approved by the Boston University Charles River Campus Institutional Review Board and were paid for their participation. For all subject, pure-tone audiometric thresholds were measured from 250 Hz to 8000 Hz at octave intervals. All participants had hearing thresholds within 15 dB of normal hearing level in each ear at all tested frequencies, and none had any history of central or peripheral hearing deficits.

#### 2.2.3.2 Stimuli, Data Acquisition and Processing

Stimuli were generated offline in MATLAB (Natick, MA) and stored for playback using a sampling rate of 48,828 Hz. Each trial consisted of a train of 72  $\mu$ s-long clicks presented at a repetition rate of 100 Hz for a burst period of 200 ms. The inter-trial interval was random and uniformly distributed between 410 ms and 510 ms. This 100 ms jitter ensured that EEG noise that is not in response to the stimulus occurs at a random phase between  $-\pi$  and  $\pi$  for frequencies above 10 Hz. For eight of the nine participants, for a randomly chosen set of 500 out of the 1000 trials presented, the polarity of the click-trains was reversed. This allows responses phase-locked to the cochlear-induced envelopes to be separated from the responses phase-locked to the temporal fine structure of the acoustic input (Aiken and Picton, 2008; Ruggles et al., 2011). For one participant, the total number of trials was increased to 1500 (half presented in each polarity) to allow for a more detailed analysis of how the noise level varied with the number of trials for different analysis approaches (see section 2.2.3.3). Scalp responses to the click-train stimuli were recorded in 32 channels at a sampling rate of 16,384 Hz in a sound-shielded room using a BioSemi ActiveTwo EEG system. The measurements were then re-referenced offline to the average potentials recorded at the two earlobes using additional surface electrodes. An additional reference electrode was placed on the seventh cervical vertebra (C7) to allow for offline construction



of a vertical montage channel for comparison (Gockel et al., 2011; Krishnan, 2006; Marsh et al., 1975). The continuous recording from each electrode was high-pass filtered in MATLAB at 70 Hz using an FIR filter with zero group-delay to minimize signal contributions from cortical sources before epoching (Kuwada et al., 2002; Dolphin and Mountain, 1992; Herdman et al., 2002). Response epochs from  $-50$  ms to 250 ms relative to the stimulus onset time of each trial were segmented out from each channel with the epochs going from  $-50$  ms to 250 ms relative to the stimulus onset time of each trial, resulting in  $300ms$  long epochs. Epochs with signals whose dynamic range exceeded  $50\mu V$  in any channel were excluded from further analysis to remove movement and muscle activity artifacts.

### 2.2.3.3 Analysis

The epoched 32 channel data were processed using the cPCA method described above to provide estimates of PLV and  $PLV_z$ . In order to taper the  $300ms$  long epochs for frequency analysis, the time-bandwidth product was set to obtain a resolution  $2W = 6.66$  Hz in the frequency domain. This allowed the use of one Slepian taper that had a spectral concentration  $c \approx 1$ . The vertical-montage single channel (Fz - C7) was used for comparison. In addition to comparing the cPCA result to the single channel, we also combined the 32 channels using traditional time-domain PCA and estimated the PLV from the combined result. Two separate analyses were performed to (1) compare the SNR across single-channel, time-domain PCA and the cPCA methods and (2) to estimate the number of cPCA trials needed using the cPCA method to roughly obtain similar noise floor levels as the single-channel approach for individual subject results.

To compare the SNR across methods for a given number of trials, a bootstrapping procedure (Ruggles et al., 2011; Zhu et al., 2013) was used to generate estimated PLV distributions for each analysis approach that are approximately Gaussian distributed. This allowed simple, direct comparisons across the analysis methods. For each method (vertical montage, time-domain PCA and cPCA), 200 trials of each polarity were drawn at random 800 times with replacement and the PLV values estimated for each draw. The

PLV estimates from different draws were then averaged in order to make the result more normally distributed before transforming them to  $z$ -scores to yield  $PLV_z(f)$ . The value of the  $z$ -score at the fundamental frequency ( $F_0 = 100$  Hz) was used as a measure of SNR, given that the noise distributions were equalized across the different methods. We then systematically evaluated the effect of increasing the number of recording channels on the SNR of the extracted SSSRs.

In order to get an idea of the number of trials needed to obtain similar noise-floor levels as the more traditional, single-channel approaches at the level of an individual subject, we estimated the noise floor for a fixed pool of trials from the one subject for whom we measured responses to 1500 trials. For this analysis, we parametrically varied the number of trials we analyzed to determine how the noise floor varied with the trial pool size. The overall procedure is described step by step as follows:

1. Fix the analysis pool to the first  $N_{pool}$  trials acquired from the subject.
2. From the fixed pool of  $N_{pool}$  trials, draw  $N_{pool}$  trials with replacement.
3. For each draw, estimate the PLV spectrum using the cPCA method or for the vertical montage channel.
4. Repeat the drawing (with replacement) and PLV estimation procedure for a total of  $M$  draws with the same fixed pool of  $N_{pool}$  trials.
5. Estimate the variance of the noise floor  $\widehat{\sigma}^2(N_{pool})$ , for the fixed pool of trials using the plugin formula (Bickel and Freedman, 1981)

$$\widehat{\sigma}^2(N_{pool}) = \frac{1}{M-1} s^2 \quad (2.9)$$

where  $s^2$  is the sample sum of squared central deviations over the  $M$  draws.

6. Repeat the procedure for different pool sizes by progressively increasing  $N_{pool}$  to obtain the noise-floor estimate curves for the cPCA and the single-channel methods.

We performed this analysis while varying  $N_{pool}$  from 100 to 1500 in steps of 50, with the variance estimate calculated using  $M = 50$  draws for each trial pool. By using an individual subject’s data and by fixing the pool of trials at each stage, this procedure allows us to estimate how many trials are needed using a given analysis method to achieve a given amount of noise suppression. This analysis also allows us to compare the noise floor obtained using the traditional single-channel montage with that obtained using our multichannel approach, for a fixed number of trials.

## 2.3 Results

### 2.3.1 Simulations

Figure 2.2A show sample SSSRs obtained by averaging 200 trials on a single channel (top panel), using time-domain PCA (middle panel), and using the cPCA methods (bottom panel). Though the SSSR extracted by time-domain PCA shows some improvement in SNR relative to using a single electrode, the gain in SNR using the cPCA method is greater. This is further elucidated in Figure 2.3A, which shows the relationship between the phase delays estimated using the cPCA method and the original simulated phase delays for a typical simulation. The estimated signal phase at a given channel corresponds very closely to the true simulated phase of the signal in that channel. This shows that the complex eigenweights  $v(f)$  obtained from the cross-channel spectral density matrix  $M(f)$  capture the relative phase shifts between the individual channels. Figure 2.3B shows the relationship between the errors in the estimation of phase shift and the magnitude of the eigenweights across the different channels. The magnitude of the phase estimation error is inversely related to the channel weight magnitudes. This result shows that the channels with poor SNR have the largest phase estimation errors. Thus, using this method, the relative contribution of a given channel to the final extracted SSSR and estimate of PLV depends on the reliability of the channel; the channels that have poorly estimated phases contribute relatively little to the final signal estimate. Finally, an assessment of

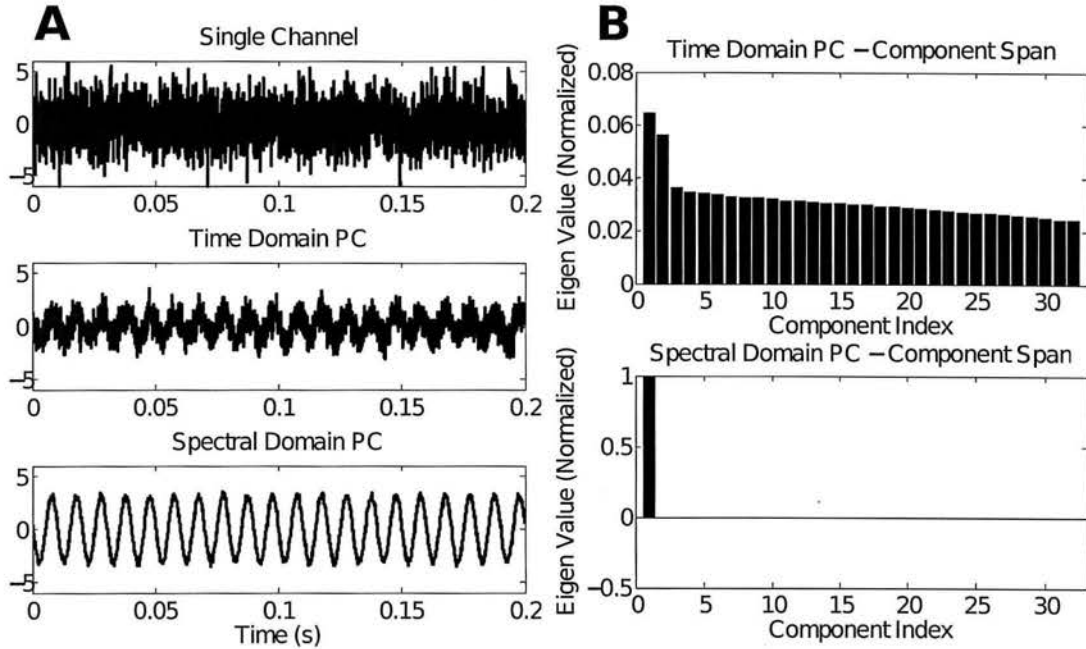


Figure 2.2: Simulation results: (A) The trial-averaged response at a single simulated channel (top panel), the extracted SSSR using time-domain PCA (middle panel), and the extracted SSSR using cPCA (bottom panel) are shown. Though the time-domain PCA has a greater SNR compared to any single channel, the cPCA method produces SSSRs of significantly higher SNR than does time-domain PCA. (B) The normalized eigenweights for the different principal components using traditional time-domain PCA (top panel) and cPCA (bottom panel) for the simulated EEG data. The cPCA method captures most of the signal energy in one component, showing that one weight vector accounts for both the magnitude and phase variations across channels.

the number of significant principal components shows that, in contrast to time-domain PCA, where multiple components are needed to capture all signal energy (top panel of Figure 2.2B), with the cPCA method, the majority of the signal energy is captured by the single extracted principal component (bottom panel of Figure 2.2B). This makes sense, given that the only parameters that distinguish between channels, namely the channel SNR and relative phase, are both accounted for by the complex weight vector obtained in the cPCA method. The magnitudes of the eigenweights account for the relative SNRs and between-channel correlations; the phase of the eigenweights accounts for the discrepancies in the phase alignment.

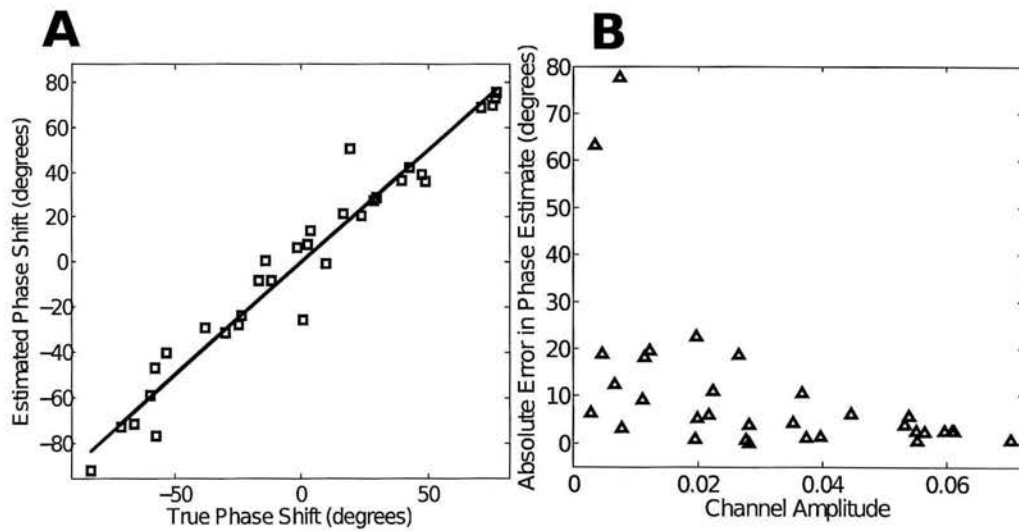


Figure 2.3: Simulation results: (A) Relationship between the true simulated phase and the phase shifts estimated using the cPCA method for a typical simulation. The cPCA method produces accurate estimates of the phase delay necessary to align the channels. (B) Phase estimation errors are inversely related to the channel weights (shown for a typical simulation). Specifically, the channels with larger phase estimation error have a lower relative weight, and hence contribute weakly to the final extracted SSSR, while channels with accurately estimated phases are weighted more strongly.

### 2.3.2 Human EEG Data

We applied the cPCA method to SSSR recordings obtained in response to click trains with a fundamental frequency of 100 Hz and harmonics up to 10 kHz. Estimates of the cross-channel spectral density matrices were obtained using 300 ms epochs and a time-bandwidth product of 2. This yielded an estimate of the complex cross-spectral density matrix  $M(f)$ , with a frequency resolution of 6.66 Hz. For a different, larger choice of frequency resolution, multiple, orthogonal tapers can be obtained that have the same time-bandwidth product, providing a multitapered estimate. Here, we used only a single taper, yielding the maximum possible frequency resolution.

The top, middle, and bottom panels of Figure 2.4A show the SSSR phase locking values obtained using the vertical montage channel, traditional time-domain PCA, and cPCA, respectively, for a representative subject. All three methods produced comparable phase-locking estimates. However, analogous to the simulation results, the variance of the noise floor (seen at the non-harmonic frequency bins where there is no signal) for the individual channels and the traditional PCA method were significantly higher than for the cPCA method. This is quantified for all nine subjects in Figure 2.4B, which shows the variance of the noise floor for the single-channel montage and for the multichannel estimate using the cPCA method. It is clear from visual inspection that for each subject, the cPCA method using 32 channels reduced the noise-floor variance, rendering the stimulus-related response peaks at the fundamental and harmonic frequencies more easily distinguishable from the noise floor.

This was tested statistically using a permutation procedure; the variances of the PLV estimates (one estimate per subject per method, i.e. three numbers per subject) obtained from the bootstrap PLV procedure for the different analysis methods (single-channel, time-domain PCA and cPCA; labelled *method*) were pooled together. For every subject, the *method* labels associated with the three variance estimates were randomly permuted. For each permutation, the within-subject difference in variance between pairs of methods was

calculated. These within subject variance differences between method pairs were pooled across subjects and across permutations to obtain null distributions for the differences in variance across the *methods*. The null distribution thus obtained is the non-parametric analog of the null distribution assumed in parametric within-subject tests (such as the paired t-test) and represents the variance differences that would have been obtained if the *methods* yielded the same variance on an average. The differences between the variances obtained from the correctly labelled *methods* were then compared to the generated null distribution to yield a p-value. For the same pool of trials, the cPCA method yielded significantly lower variance than both a single-channel analysis ( $p < 0.001$ ) and the time-domain PCA ( $p < 0.01$ ).

The statistical superiority of the cPCA method is illustrated further in Figure 2.5A, which shows, the PLVz estimates obtained using the noise-normalization procedure previously described for one representative subject. At each of the harmonics of 100 Hz, the z-scores are higher for the cPCA method than for traditional methods, indicating a gain in SNR. In order to quantify the gain in SNR further, the PLVz values from the 100 Hz bin are compared to the PLVz values obtained using the best channel for each subject by computing the ratio of the z-scores. This procedure allows us to quantify parametrically, the gain in SNR as more recording channels are included for both the traditional time-domain PCA and the cPCA methods. Figure 2.6 shows the comparison as the number of recording channels is increased from 1 to 32, averaged over 9 subjects. For this plot, the channels were ordered as follows: Channel 1 is the best channel for the individual subject. Channel 2 provides the maximum gain in SNR out of the remaining 31 channels when added to channel 1. Channel 3 provides the maximum SNR gain out of the remaining 30 channels when added to channels 1 and 2, and so on for each method. The theoretical gain in SNR that would be obtained when combining independent identically distributed measurements is shown in red for reference. The SNR increases as more and more channels are added for both the traditional PCA and the cPCA methods. Initially, the gain in SNR is rapid as more channels are included in the SSSR extraction, almost reaching the the-

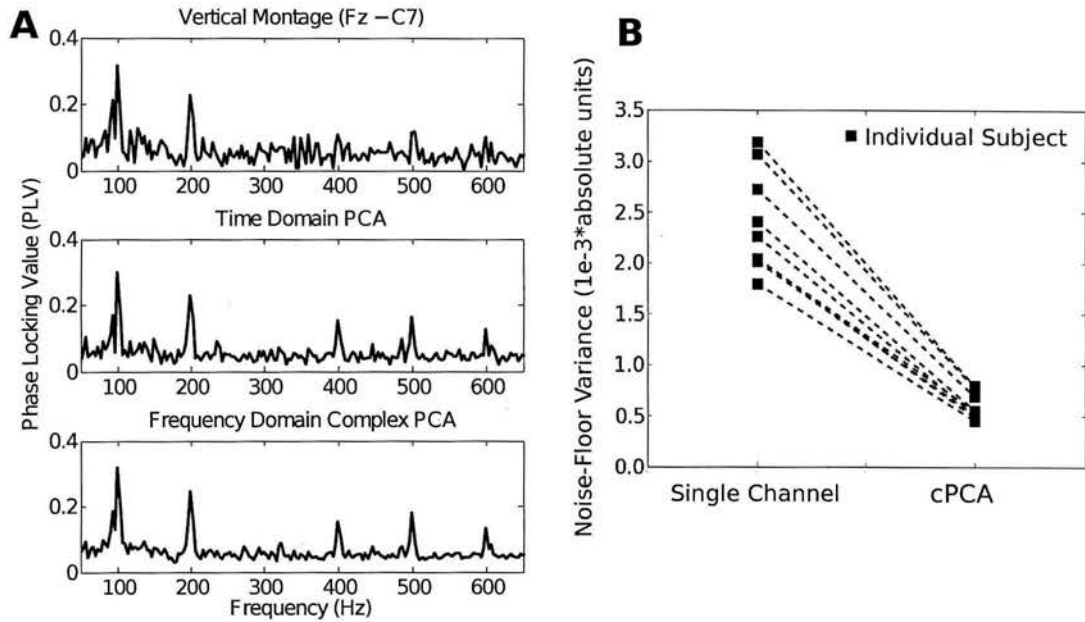


Figure 2.4: (A) Raw phase-locking value (PLV) scores obtained from a representative subject using a single channel (top), time-domain PCA (middle), and cPCA (bottom) for a 100 Hz click-train burst stimulus. The PLV obtained using the three methods are comparable at signal frequencies (multiples of 100 Hz), but differ in the variability of the noise floor. The cPCA method hence produces PLV values that are statistically more robust than the other methods. (B) The noise floor variance estimated using the bootstrap procedure is shown for each of the nine subjects for the single-channel montage and for the multichannel estimate using the cPCA method. It is clear from visual inspection, and confirmed using the permutation procedure, that the noise-variance was smaller using cPCA than for the other methods for every subject, rendering the responses more easily distinguishable from noise.



oretical maximum achievable when each channel has independent, identically distributed noise. However, the gain appears to plateau as the electrode-density on the scalp increases. From Figure 2.6, it is evident that the cPCA method outperforms time-domain PCA for all array sizes ( $p < 0.0001$ ; permutation test). Moreover, on average, SNR gains of about 3 can be obtained with as few as 10 sensors if the optimal arrangement was somehow known *a priori*. However, it has to be acknowledged that here, the sorting of channels is done *a posteriori*, optimally selecting each channel that is added. Thus in practice the gain in SNR that one would achieve by increasing the number of sensors is likely to be more gradual, with the plateau not being reached until larger array sizes. Nevertheless, to obtain a fixed SNR using multichannel recordings with typical EEG array sizes, the duration of the recording session could be significantly shorter when multiple recording channels are combined than for either single-channel recordings or using time-domain PCA.

To obtain a better understanding of the actual reduction in the number of trials that need to be presented to obtain similar noise suppression as 1000 presentations with a single-channel montage at the level of the individual subject, the variance estimation procedure with fixed data pools was employed as described in section 2.2.3.3. Figure 2.5B shows the results obtained. In both the cPCA and the vertical-montage cases, the variance drops inversely as the number of trials ( $N_{pool}$ ) increases. The best fitting  $1/N_{pool}$  functions are superposed to guide the eye. It is evident that the cPCA method needs only about 250 trials to reduce the noise floor to the same level as 1000 trials with the single channel montage. The procedure was repeated with the remaining 8 subjects to estimate the number of trials required with the cPCA procedure to achieve the same noise variance as the single channel montage. On average, the cPCA allowed for a 3.4 fold reduction in the number of trials needed. Thus, real data confirm the efficacy of using complex PCA with multichannel recordings to increase the SNR of SSSR recordings and significantly reduce data acquisition time.

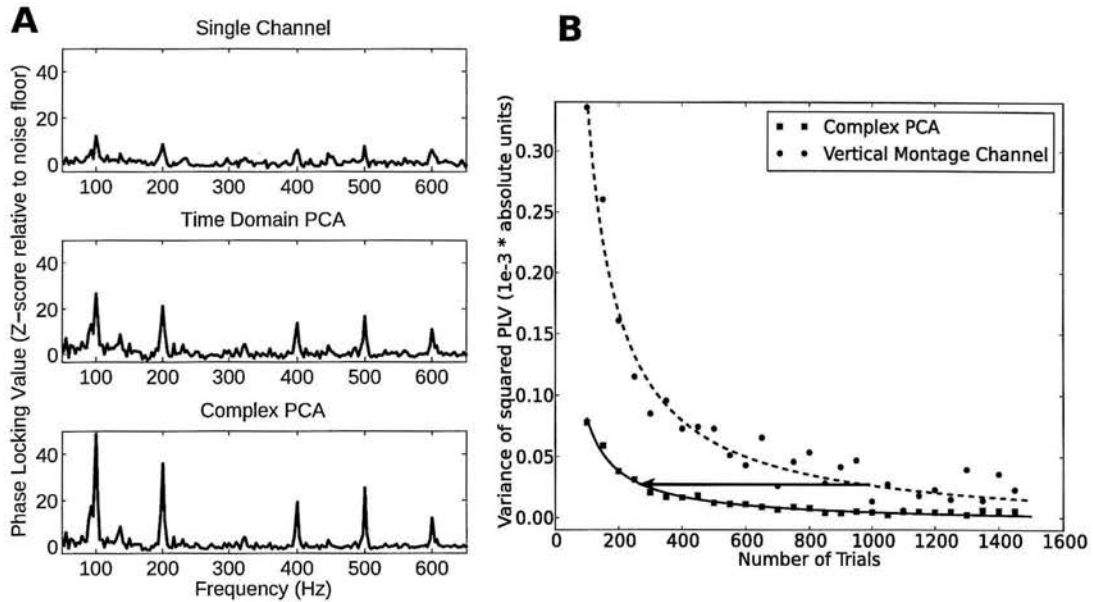


Figure 2.5: Individual subject results: (A) Z-scored PLV values obtained from a representative subject using a single channel (top), time-domain PCA (middle), and cPCA (bottom) for the 100 Hz click-train burst stimuli. Here the noise-floor in all three cases has been normalized to have a mean of zero and a variance of one (scaling the PLV into a z-score). The z-scores at the harmonics of 100 Hz thus indicate the SNR obtained using the three methods. The cPCA method has a significantly higher SNR than both a single-channel and the time-domain PCA. (B) Comparison of noise floor variance estimates as a function of the number of trials between the cPCA method and the traditional vertical montage channel from individual subject data. The arrow highlights the number of trials that may be required using the cPCA approach to obtain similar levels of noise suppression as from 1000 trials using the traditional approach.

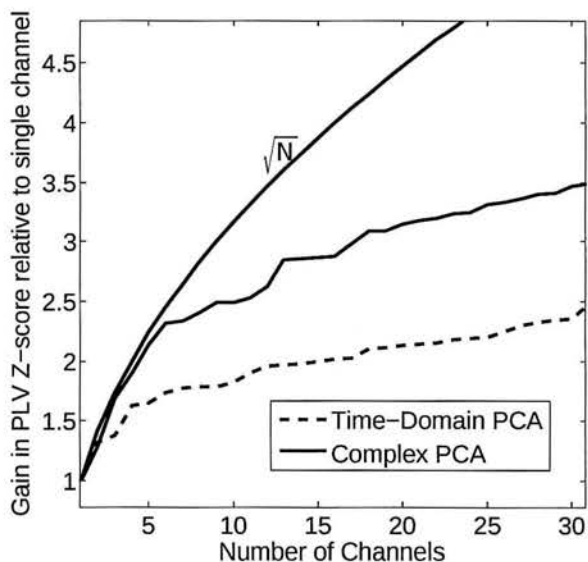


Figure 2.6: Real EEG results: The gain in SNR as the number of recording channels is increased, quantified as the gain in z-score relative to using a single channel, time-domain PCA, and cPCA methods. As more channels are added, both the time-domain PCA and the cPCA methods provide a gain in SNR, but the cPCA method produces larger improvements. The theoretical gain that would be obtained by combining independent, identically distributed measurements is shown in red for reference. Initially, the SNR gain approaches the reference curve, but then quickly plateaus. This suggests that the noise source activity captured in different channels are nearly independent when there are a small number of (optimally selected) channels included, but that as the electrode density increases, the noise in the different channels become correlated. Note, however that the rapid increase and subsequent plateau in SNR with increasing number of channels is obtained given the *a posteriori* knowledge of the best channels to select. In practice, the gain in SNR with increasing number of channels would be more gradual, since the channels would not be selected optimally from among a large set, but would instead be selected a priori.

## 2.4 Discussion

Brainstem steady state responses are increasingly being used to investigate temporal coding of sound in the auditory periphery and brainstem. Here, we demonstrate the advantages of multichannel acquisition of SSSRs, which are traditionally acquired with a single channel montage. Our novel approach combines the information from multiple channels to obtain a significant gain in response SNR using complex frequency domain principal component analysis. We illustrate the efficacy of the method using simulated data. In order to demonstrate that the method is applicable and advantageous in practice, we also apply the analysis to human EEG recordings from a cohort of nine subjects. The multichannel approach makes it possible to obtain significantly higher SNR for a given number of trials, or equivalently to significantly reduce the number of trials needed to obtain a fixed noise-level.

### 2.4.1 Clinical use

In addition to use in basic neurophysiological investigation of auditory function, the reduction in data acquisition time afforded by our multichannel approach renders the SSSR significantly more suitable for clinical use. ASSRs in general, and SSSRs in particular, have been suggested for clinical use for objective, frequency-specific assessment of the early auditory pathway including for assessment of hearing sensitivity, sensorineural hearing loss, and auditory neuropathy/dys-synchrony (see Picton et al., 2003a; Krishnan, 2006; Starr et al., 1996, for reviews). While the cortical-source 40-Hz ASSR amplitude depends on the state of arousal (e.g., it changes if the subject is asleep or under anesthesia), the higher-frequency SSSRs are relatively unaffected (Cohen et al., 1991; Lins et al., 1996; Picton et al., 2003b). This, along with the fact that SSSRs can be recorded passively, makes the SSSR suitable for objective clinical assessment of auditory function in special populations including infants and neonates (Rickards et al., 1994; Cone-Wesson et al., 2002). There appears to be an emerging consensus that the ASSR will play an important role in clinical

audiology in the future (Korczak et al., 2012).

#### 2.4.2 Set-up time versus recording length

The obvious downside to using multichannel recordings to improve the SNR for a given recording duration is the additional time required to place multiple scalp electrodes. For a trained graduate student setting up 32 recording channels with our Biosemi ActiveII EEG system, we find that set up takes about 15 minutes on average. On the other hand, the use of multichannel recordings with the cPCA method allows us to obtain stable PLV and ITC measurements (i.e., much smaller noise levels than obtained with the 1000 trials using single channel measurements) with about 7-10 minutes of recording for the typical stimuli we use (typically, 200-300 ms bursts of amplitude modulated tones, click trains, spoken syllables or the like, with inter-stimulus gaps of about 0.5 seconds), allowing us to obtain responses to as many as six different experimental manipulations within our typical recording session of 1 hour.

#### 2.4.3 The role of raw-signal narrowband SNR and other sources of variability

We have shown that the SNR of the extracted SSSR using the cPCA method is greater than when using a single channel or time-domain PCA. However, it is important to note that the SNR in the raw recordings (at each frequency bin) directly affects PLV estimates. For clarity, we shall refer to this raw-signal SNR in the frequency domain as the *narrowband SNR*. While the relationship between the *narrowband SNR* and conventional response analysis metrics such as time domain amplitude or spectral power is straightforward, metrics of phase locking such as PLV and ITC depend non-linearly on the *narrowband SNR*. However, since the distributions of the PLV and ITC only depend on the *narrowband SNR* and the number of trials used to calculate them, the effects are easy to simulate. To illustrate the effect of *narrowband SNR* on PLV at a particular frequency bin, the signal phase  $\phi_s$  in the frequency bin was modelled as coming from a von Mises distribution (a circular normal density) and the noise phase  $\phi_n$  as coming from a uniform distribution in  $(-\pi, \pi)$ .

Note that the use of a 100 ms jitter in stimulus presentation ensures that for the SSSR frequencies of interest, the noise phase is indeed distributed uniformly over the circle, as modeled here. 50 independent simulations were performed, each with 400 independent draws of signal and noise phase. The *narrowband SNR* ( $20\log_{10}A$ ) was set by adding the two phasors with the appropriate relative amplitude to obtain the simulated measurement,  $X_{sim}(f)$ , in the frequency bin:

$$X_{sim}(f) = Ae^{j\phi_s} + e^{j\phi_n} \quad (2.10)$$

The value of  $A$  was then systematically varied; the resulting growth of the PLV with *narrowband SNR* is shown in Figure 2.7. For a signal with phase  $\phi_s$  drawn from the von Mises density  $f(\theta|\mu, \kappa)$ , where  $\mu$  is the mean phase parameter and  $\kappa$  parametrizes the concentration of the phase distribution around the mean, the true PLV can be calculated analytically:

$$f(\theta|\mu, \kappa) = \frac{e^{\kappa\cos(\theta-\mu)}}{2\pi I_0\kappa} \quad (2.11)$$

$$PLV = |E(e^{j\theta})| \quad (2.12)$$

$$= \frac{I_1(\kappa)}{I_0(\kappa)} \quad (2.13)$$

where  $I_0$  and  $I_1$  are the  $0^{th}$  and the  $1^{st}$  order modified (hyperbolic) Bessel functions of the first kind and  $E(\cdot)$  is the expectation operator with respect to the density  $f(\theta|\mu, \kappa)$ . As seen in Figure 2.7, once the *narrowband SNR* is sufficiently large, the PLV quickly asymptotes to the true phase locking value and then becomes insensitive to the SNR. Moreover, we find that this behavior does not change if we draw the signal phase from distributions with higher skew or kurtosis. Thus, in this sense, the PLV estimates yield the “true” phase locking values for sufficiently high *narrowband SNR*. This observation reveals another benefit of using multichannel recordings along with the cPCA method. Since the

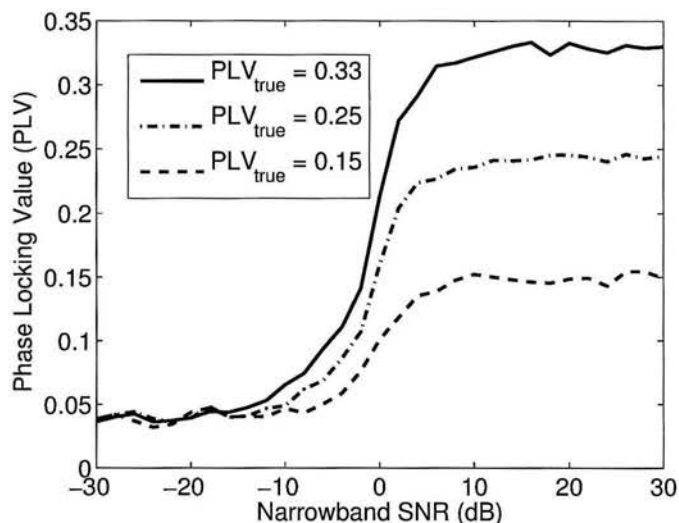


Figure 2.7: Simulations showing the effect of *narrowband SNR* in the raw recording on the non-linear relationship between the estimated PLV and the true PLV. At sufficiently high *narrowband SNR*, the PLV estimates converge to the true PLV. Since the cPCA method is more likely to push the *narrowband SNR* into this convergence region, the PLV calculated from the SSSR extracted using the cPCA method is more likely to represent the true PLV of the underlying response than are traditional methods.

cPCA method effectively combines channels optimally before the PLVs are computed, it is more likely to push the *narrowband SNR* of single trials into the saturation region of the PLV-*narrowband SNR* curve (Figure 2.7). As a result, the PLV estimate is more likely to lie closer to its true underlying value and be less biased by the noise in the measurements. This makes comparisons of phase-locking across conditions and individuals more reliable. In summary, not only does the cPCA method produce PLV estimates with a lower variance, it also increases the likelihood that these estimates are closer to the true underlying PLV. Further study is needed to assess if in practice, the *narrowband SNR* is indeed in the saturation region.

Another important factor to be considered in interpreting the efficacy of the cPCA method is the inherent session-to-session physiological variability of the SSSR itself. This can be accomplished by systematically studying the test-retest reliability of the PLV estimates for a given stimulus across multiple recording sessions. We are not aware of any

studies reporting the across-session variability of the SSSR. If the inherent session-to-session variability of the SSSR is very large, the improvement in SNR obtained using multichannel measurements might not be useful for studies comparing groups of subjects, since the improvement in SNR when extracting the SSSR from single-session data might be irrelevant in the face of large session-to-session variability that (if present) would undermine meaningful comparisons of measurements across subjects. On the other hand, for within-subject, across-condition comparisons, the improvement in SNR is likely to be very useful in three ways:

1. The cPCA method allows many more stimulus manipulations or conditions to be presented in a single session, thereby removing any across-session variability confounds that may otherwise reduce the power of across-stimulus comparisons.
2. By reducing the variance of the PLV estimates (within session but across different subsets of trials, i.e., primarily owing to background noise), within-subject differences across conditions can be more robustly compared.
3. By allowing for fewer trials to be presented, the cPCA method also helps to reduce any non-stationary effects of long-term adaptation and learning that are likely to be present when it is necessary to collect a large number of trials.

Indeed, in cases where cortical and subcortical data can be gathered simultaneously, the benefits of cPCA are likely to be particularly appreciated, reducing the number of trials necessary to estimate brainstem responses so that they can be obtained “for free” while cortical responses are gathered.

#### **2.4.4 Source separation versus cPCA**

Here, we combine recordings from multiple channels to yield a single SSSR and corresponding phase locking value estimates with low variance, so that comparisons across conditions are more reliable than traditional methods. However, when multiple generators are indeed



active, the physiological interpretation of what this SSSR represents is tricky. For frequencies in the range of 70 – 200 Hz, the group delay of the SSSR is consistent with a dominant generator coming from a neural population in the rostral brainstem/midbrain, likely the inferior colliculus (IC) (Dolphin and Mountain, 1992; Herdman et al., 2002; Smith et al., 1975; Sohmer et al., 1977; Kiren et al., 1994). Data from single-unit recordings of responses to amplitude-modulated sounds suggests that a transformation from a temporal to a rate code occurs as the signals ascend the auditory pathway, with the upper limit of phase-locking progressively shifting to lower modulation frequencies (Frisina et al., 1990; Joris et al., 2004; Joris and Yin, 1992; Krishna and Semple, 2000; Nelson and Carney, 2004). Because, for broadband sounds, the SSSRs are dominated by responses phase-locked to cochlear-induced envelopes (Gnanateja et al., 2012; Zhu et al., 2013), it is likely that the dominance of response generators higher up along the auditory pathway decreases at higher response frequencies. Thus, at higher modulation frequencies, more peripheral sources contribute appreciably to the SSSR, consistent with non-linear phase-response curves obtained at higher frequencies (Dolphin and Mountain, 1992).

One approach in SSSR data analysis would be to try and separate the multiple sources contributing to the SSSR at a given frequency. However, since the spatial resolution of EEG is poor, particularly for subcortical sources, separating the sources based on geometry alone is not feasible (Pascual-Marqui, 1999; Baillet et al., 2001). The source segregation problem is ill-posed in the sense that multiple source configurations can yield the same measured fields at the scalp level. Though it may be possible to sufficiently constrain the source estimation with the use of an elaborate generative model of the SSSR that takes into account the physiological properties of the neural generators along the auditory pathway (Dau, 2003; Rønne et al., 2012; Nelson and Carney, 2004), at present, not enough data is available from human listeners to specify such a model. Thus, we take the alternate approach of not trying to separate the sources contributing to the total observed signal. Instead, we combine measurements in order to extract a SSSR response that is robust and has low variance. This compound SSSR allows for more reliable comparisons across

stimulus manipulations than traditional acquisition/analysis approaches.

#### 2.4.5 Optimal recording configuration

As shown in Figure 2.6, as more recording channels are added, the SNR gain is initially steep and then plateaus. This suggests that the noise in the different electrodes are correlated when the number of channels is high (i.e., high channel density on the scalp). This begs the question as to what the best recording configuration would be in terms of the number of channels and their locations on the scalp. Answering this question involves consideration of two aspects of the measurements: (1) correlation of the noise between channels and (2) the variation of the signal strength itself across the channels. To appreciate a simple trade-off that exists between these two aspects, consider a pair of channels from distant scalp locations, with one channel having good sensitivity to the signal of interest and the other with poor or moderate sensitivity. When these two channels are combined with similar weights, though the noise is cancelled better, the signal would also be diluted by the inclusion of the channel with poor sensitivity. The sensitivity of different channels to the signal also depends on the choice of the reference and the tissue geometry of individual subjects, further complicating the discovery of an optimal recording configuration. Thus, though the results of the current study do not reveal an obvious recommendation for a subject-invariant, optimal configuration of electrodes for a small number of channels, typical EEG array sizes and configurations such as the standard 32 channel montage provide a large increase in SNR.

### 2.5 Conclusions

The cPCA approach to extracting SSSRs from multichannel measurements yields results that are significantly more reliable and robust than traditional single channel measurements. As a result, it is possible to record brainstem steady-state responses efficiently. This increased efficiency allows for SSSRs to be acquired simultaneously with cortical au-

ditory responses without a significant increase in the length of the recording session.

## Chapter 3

# Cochlear neuropathy and the coding of supra-threshold sound

### Preamble

This chapter presents a detailed review of the findings that noise-exposure and aging can result in the loss of auditory nerve fibers even when cochlear hair cells and audiometric thresholds are intact. The implications of these findings to supra-threshold sound coding are discussed and a connection between such cochlear neuropathy and individual differences in normal hearing listeners is hypothesised. In addition a quantitative model of subcortical steady state responses is developed and used to make predictions that are tested in Chapter 4. The work was completed between Sep 2011 and Oct 2013 and published as a peer-reviewed journal article in *Frontiers in Neuroscience* (Bharadwaj et al., 2014b). Co-authors Luke Shaheen and Charles Liberman wrote the sections reviewing the animal literature on cochlear neuropathy. Sarah Verhulst made important contributions to the writing of the sections involving otoacoustic emissions. Barbara Shinn-Cunningham made important contributions to the introduction and several other sections of the writing. The manuscript is reproduced here without revision.

### Abstract

Many listeners with hearing thresholds within the clinically normal range nonetheless complain of difficulty hearing in everyday settings and understanding speech in noise. Con-

verging evidence from human and animal studies points to one potential source of such difficulties: differences in the fidelity with which supra-threshold sound is encoded in the early portions of the auditory pathway. Measures of auditory subcortical steady-state responses in humans and animals support the idea that the temporal precision of the early auditory representation can be poor even when hearing thresholds are normal. In humans with normal hearing thresholds, behavioral ability in paradigms that require listeners to make use of the detailed spectro-temporal structure of supra-threshold sound, such as selective attention and discrimination of frequency modulation, correlate with subcortical temporal coding precision. Animal studies show that noise exposure and aging can cause a loss of a large percentage of auditory nerve fibers without any significant change in measured audiograms. Here, we argue that cochlear neuropathy may reduce encoding precision of supra-threshold sound, and that this manifests both behaviorally and in subcortical steady-state responses in humans. Furthermore, recent studies suggest that noise-induced neuropathy may be selective for higher-threshold, lower-spontaneous-rate nerve fibers. Based on our hypothesis, we suggest some approaches that may yield particularly sensitive, objective measures of supra-threshold coding deficits that arise due to neuropathy. Finally, we comment on the potential clinical significance of these ideas and identify areas for future investigation.

### **3.1 Introduction**

A significant number of patients seeking audiological treatment have normal hearing thresholds (NHT), but report perceptual difficulties in some situations, especially when trying to communicate in the presence of noise or other competing sounds (e.g., Hind et al., 2011). Such listeners are typically said to have central auditory processing disorders, more recently known simply as auditory processing disorders (CAPD/APD; Catts et al., 1996; Chermak and Musiek, 1997), a catchall diagnosis testifying to how little we know about the underlying causes.

In some ways, the fact that having NHTs does not automatically predict good performance in these conditions is not particularly surprising. Audiometric thresholds measure the lowest intensities that a listener can detect. In contrast, the ability to analyze the content of sound requires a much more precise sensory representation of acoustic features across a large dynamic range of sound intensities. Specifically, current audiometric screenings test the lowest level of sound listeners can hear at various frequencies, but they do not test whether they can make judgements about the spectral or temporal content of the sound, analogous to seeing an eye doctor and being asked whether you can tell that light is present, without worrying about whether or not you can tell anything about the object the light is coming from.

Consistent with the idea that analysis of supra-threshold sound differs amongst NHT listeners, many APD patients seek help precisely because they notice difficulties in situations requiring selective auditory attention (Demanez et al., 2003), which places great demands on the auditory system. Moreover, recent laboratory evidence suggests that the prevalence of NHT listeners with APD-like symptoms may be greater than one might predict based on the number of people seeking audiological treatment. Specifically, in the lab, NHT listeners have vastly different abilities on the types of tasks that typically frustrate APD listeners. One recent study shows that when NHT subjects are asked to report spoken digits from one direction amidst otherwise similar speech, performance ranges from chance levels to nearly 90% correct, with the bottom quartile of listeners falling below 60% correct (Ruggles and Shinn-Cunningham, 2011). Crucially, when subjects made errors, they almost always reported a digit coming from a non-target direction rather than an unspoken digit, suggesting that differences were unlikely due to higher-level deficits involving language such as differences in speech intelligibility. Instead, the errors appeared to be due to failing to select the target stream from amidst the maskers. Yet none of the listeners in the study complained of hearing difficulties, even those at the bottom of the distribution; moreover, none had entertained the idea of seeking audiological treatment.

Differences in higher-order processing clearly contribute to individual differences in

complex tasks such as the ability to selectively attend, process speech, or perform other high-level tasks (for instance see Surprenant and Watson, 2001). However, in this opinion paper, we focus on how low-level differences in the precision of spectro-temporal coding may contribute to differences in performance. We argue that poor sensory coding of supra-threshold sound is most likely to be revealed in complex tasks like those requiring selective attention, which helps to explain the constellation of symptoms that lead to APD diagnoses. Selective auditory attention hinges on segregating the source of interest from competing sources (object formation; see Bregman, 1990; Darwin and Carlyon, 1995; Alain et al., 2000; Carlyon, 2004), and then focusing on that source based on its perceptual attributes (object selection; see Shinn-Cunningham, 2008; Shinn-Cunningham and Best, 2008). Both object formation and object selection rely on extracting precise spectro-temporal cues present in natural sound sources, which convey pitch, location, timbre, and other source features. Given this, it makes sense that listeners with poor supra-threshold coding fidelity notice problems in crowded social settings, an ability that depends upon robust coding of supra-threshold sound features. Here, we argue that the fidelity with which the auditory system encodes supra-threshold sound is especially sensitive to the number of intact auditory nerve fibers (ANFs) encoding the input. In contrast, having NHTs likely depends only on having a relatively small but reliable population of ANFs that respond at low intensities. Indeed, one recent study shows that, in animals, audiometric thresholds can be normal even with only 10-20% of the inner hair cells (IHCs) of the cochlea intact (Lobarinas et al., 2013). Our hypothesis is that the convergence of multiple ANFs, while possibly redundant for detecting sound, is critical for analyzing supra-threshold sound.

In this paper, we first consider how supra-threshold sound content is normally encoded, focusing particularly on temporal coding. We then review animal evidence for cochlear neuropathy, a reduction in the number of ANFs responding to supra-threshold sound. We argue that this neuropathy can help explain why some listeners have difficulty performing selective attention and other supra-threshold tasks, despite having NHTs. We discuss evidence that lower-spontaneous rate ANFs (lower-SR ANFs; i.e., those with rates below

about 18 spikes/s) may be especially vulnerable to damage. We hypothesize that lower-SR ANFs may play a critical role in coding supra-threshold sound features, particularly under challenging conditions. We then discuss the use of the subcortical steady-state response (SSSR) to quantify temporal coding in the *Frontiers in Systems Neuroscience* early portions of the auditory pathway, including the challenges inherent in interpreting the SSSR and relating it to single-unit neurophysiology. With the help of simple models of brainstem responses, we suggest measures that may emphasize the effect of neuropathy on the SSSR. Using these ideas, we suggest future experiments to (1) test our hypothesis that cochlear neuropathy contributes to the supra-threshold coding deficits seen in some listeners; and (2) develop sensitive, objective correlates of such deficits that may be useful, clinically.

## 3.2 Coding of supra-threshold sound

### 3.2.1 The diversity of auditory nerve fibers

ANFs comprise the sole conduit for information about the acoustic environment, carrying spike trains from the cochlea to the central auditory system. As schematized in Figure 3.1A, each ANF contacts a single IHC via a single synapse. At each synapse, an electron-dense ribbon sits near the pre-synaptic membrane surrounded by a halo of glutamatergic vesicles. Sound in the ear canal leads to cochlear traveling waves that deflect IHC stereocilia, causing the opening of mechanoelectric transduction channels and a graded change in the IHC membrane potential. At the IHC's synaptic pole, this sound-driven receptor potential drives an influx of calcium causing an increased probability of fusion of synaptic vesicles with the IHC membrane in the region of the ribbon. Glutamate released into the synaptic cleft binds to the AMPA-type glutamate receptors at the post-synaptic active zone, causing depolarization and action potentials in the ANF. Between 10 and 30 ANFs synapse on each IHC, depending on species and cochlear location (Figure 3.1B), and there are roughly 3500 IHCs along the 35 mm cochlear spiral in humans. Thus, all the information we receive about our acoustic world is carried via the roughly 30,000 ANFs emanating from each cochlea.



ANFs in the mammalian inner ear can be subdivided into three functional groups. The classification is based on spontaneous discharge rate (SR; i.e., the spike rate in the absence of sound), because it is easy to quantify, but the key functional differences are in the sensitivity to sound. High-SR fibers have the lowest thresholds, low-SR have the highest thresholds, and medium SR thresholds are intermediate between the two (Figure 3.2A). The distribution of SRs is fundamentally bimodal (Figure 3.2B) with roughly 40% in the lower peak (SR < about 18 spikes/second), which includes both low-SR and medium-SR fibers (15% and 25% of all ANFs, respectively) and 60% in the higher peak (Liberman, 1978). In this paper, we shall use the term lower-SR ANFs to refer jointly to the low- and medium-SR groups, which are sometimes distinguished in the literature.

Anatomical studies suggest that all three ANF types can innervate the same IHC, however, lower-SR fibers have thinner axons, fewer mitochondria, and tend to synapse on the modiolar side of the IHC. In contrast, high-SR fibers have thicker axons, more mitochondria, and synapse on the pillar side (Liberman, 1982). There are also systematic differences in the sizes of presynaptic ribbons and post-synaptic glutamate-receptor patches (Liberman et al., 2011). All three ANF types send their central axons to the cochlear nucleus (CN), where they branch, sending collaterals to the anteroventral, posteroventral, and dorsal subdivisions. Although branches from all SR types are present in each CN subdivision, low- and medium-SR fibers give rise to more endings than high-SR fibers, especially in the small-cell cap of the anteroventral CN (Ryugo and Rouiller, 1988; Liberman, 1991). Hence, lower-SR fibers may have more downstream influence than suggested by the fact that they make up less than half of the population at the level of the auditory nerve (AN). The diversity of ANF threshold sensitivity is believed to be important in intensity coding in the auditory system, where level discrimination abilities are near-constant over a range of 100 dB or more (Florentine et al., 1987; Viemeister, 1988). This large dynamic range may be mediated, at least in part, by the differing dynamic ranges of low-, medium-, and high-SR fibers. As represented in Figure 3.2C, high-SR fibers, whose response thresholds are at or near behavioral detection threshold, likely determine the ability to detect sounds

in a quiet environment. However, 20-30 dB above threshold, their discharge rate saturates. By virtue of their higher thresholds and extended dynamic ranges, the lower-SR fibers may be particularly important for extending the dynamic range of hearing. Possibly more important is their contribution to hearing in a noisy environment. Activity of high-SR fibers is relatively easy to mask with continuous noise, as schematized in Figure 3.2D. Because they are so sensitive to sound, even near-threshold noise increases the background discharge rate of high-SR fibers. This continuous activation causes synaptic fatigue (i.e., vesicle depletion) and thus also decreases their maximum discharge rate to tone bursts or other transient signals that might be present (Costalupes et al., 1984; Costalupes, 1985). By virtue of their higher thresholds, the lower-SR fibers are more resistant to background noise. Thus with increasing levels of continuous broadband masking noise, lower-SR fibers likely become increasingly important to the encoding of acoustic signals, because they will increasingly show the largest changes in average discharge rate in response to transient supra-threshold stimuli (Figure 3.2D; also see Young and Barta, 1986).

### **3.2.2 Temporal coding and its importance for auditory perception**

As a result of cochlear filtering, each ANF is driven by a narrow frequency band of sound energy. Thus, the temporal information encoded by the ANFs can be logically separated into two parts; the temporal fine-structure (TFS), corresponding to the timing of the nearly sinusoidal narrowband carrier fluctuations, and the slower temporal envelope of that carrier, whose temporal fluctuations are limited by the bandwidth of the corresponding cochlear filter. For low-frequency cochlear channels, ANFs convey both TFS and envelope information; neural spikes are phase-locked to the carrier and the instantaneous firing rate follows the envelope. At higher frequencies, ANFs do not phase lock to the TFS; however, responses convey temporal information by phase locking to envelope fluctuations.

Although different perceptual attributes of natural sound are encoded by different spectro-temporal cues, many depend on reliable timing information. For instance, the computation of interaural time differences (ITD), important for spatial perception of

sound, requires temporal precision on the order of tens of microseconds (Blauert, 1997). While perceptually, TFS information in low-frequencies is the dominant perceptual cue determining perceived location (at least in anechoic conditions; Wightman and Kistler, 1992), for broadband and high-frequency sounds, ITDs can be conveyed by the envelope alone. Moreover, high-frequency envelope ITDs can be perceived nearly as precisely as low-frequency TFS ITDs (Bernstein and Trahiotis, 2002). In addition, envelopes may play a significant role in space perception in everyday settings such as rooms, where reverberant energy distorts TFS cues (Bharadwaj et al., 2013b; Dietz et al., 2013). The coherence of the temporal envelope across channels helps to perceptually bind together different acoustic constituents of an object in the auditory scene (Elhilali et al., 2009; Shamma et al., 2011). Coding of pitch and speech formants also may rely, at least in part, on both TFS and envelope temporal information, although the precision needed to convey this information is less than that needed to extract ITDs (see Plack et al., 2005, for a review). On an even slower time scale, speech meaning is conveyed by fluctuations in energy through time. Thus, a range of temporal features in both TFS and envelopes are necessary to enable a listener to parse the cacophonous mixture of sounds in which they commonly find themselves, select a sound source of interest, and analyze its meaning. Importantly, almost all of these tasks, when performed in everyday settings, require analysis of temporal information at supra-threshold sound intensities. To exacerbate matters, everyday settings typically contain competing sound sources and reverberant energy. Both degrade the temporal structure of the sound reaching a listener's ears, reducing the depth of signal modulations and interfering with the interaural temporal cues in an acoustic signal. If amplitude modulation is weakly coded in a listener with cochlear neuropathy, degradations in the input signal modulations due to competing sound and reverberant energy may render spatial information diffuse and ambiguous, pitch muddy, and speech less intelligible (e.g., see Stellmack et al., 2010; Jørgensen and Dau, 2011). TFS cues convey information important for speech intelligibility in noise (Lorenzi and Moore, 2008). Envelope cues are important for speech-on-speech masking release (Christiansen et al., 2013). Given all of

this, a listener with degraded coding of envelope and TFS is most likely to notice perceptual difficulties when trying to understand speech in challenging settings, even if they do not notice any other deficits and have no difficulty in quiet environments. Thus, we hypothesize that differences in the fidelity with which the auditory system encodes supra-threshold TFS and amplitude modulation accounts for some of the inter-subject differences that NHT listeners exhibit in tasks such as understanding speech in noise or directing selective auditory attention (also see Section 3.3.2). Based on this idea, we argue that a method for measuring supra-threshold temporal coding fidelity may have important clinical applications, enabling quantification of supra-threshold hearing deficits that affect how well listeners operate in everyday environments, but that are not commonly recognized today.

### 3.2.3 Consequences of cochlear neuropathy for temporal coding

One consequence of cochlear neuropathy (i.e., a reduction in the number of ANFs conveying sound) will be a reduction in the fidelity of temporal coding of supra-threshold sound. For instance, convergence of multiple, stochastic ANF inputs leads to enhanced temporal precision in the firing pattern of many CN cells (e.g., see Joris et al., 1994; Oertel et al., 2000). Thus, a reduction in the overall number of ANFs will reduce the precision with which both TFS and envelope temporal information are conveyed to higher centers (see also Lopez-Poveda and Barrios, 2013). While the importance of TFS coding for various aspects of sound perception cannot be overstated, we only briefly discuss TFS coding here. We focus primarily on the implications of cochlear neuropathy on the fidelity with which envelope information is conveyed. This focus is motivated particularly by recent data from guinea pigs and mice that suggest that noise-induced neuropathy preferentially damages the higher-threshold, lower-SR cochlear nerve fibers (Furman et al., 2013), rendering envelope coding especially vulnerable, as explained below.

Damage to lower-SR ANFs is likely to be especially detrimental to supra-threshold coding of sound envelopes, as high-SR fibers cannot robustly encode envelope timing cues

in sounds at comfortable listening levels. Specifically, the average firing rate of high-SR ANFs (ignoring the temporal pattern of the response) saturates at levels roughly 20-30 dB above threshold, around the sound level of comfortable conversation (see red solid line in Figure 3.2E). In addition, both measures of phase locking to the envelope (namely the modulated rate, which is the magnitude of the frequency domain representation of the post-stimulus time histogram of the ANF response, evaluated at the fundamental frequency of the input signal; see dashed red line in Figure 3.2E) and the synchronization index (also known as the vector strength, calculated as the modulated rate normalized by one half of the average rate; see red line in Figure 3.2F) of high-SR neurons drop off as sound levels approach and exceed comfortable listening levels. This drop off is particularly detrimental for relatively intense sounds with shallow modulation depths, where both the crests and troughs of the envelope of the signal driving the high-SR ANFs fall in the saturation range of intensities, resulting in relatively poor modulation in the temporal response of these fibers (Joris and Yin, 1992). In contrast, lower-SR fibers are more likely to encode these envelope fluctuations because they are likely to be at an operating point where the firing rate (in the steadystate) is still sensitive to fluctuations in the sound level. If noise exposure causes a selective neuropathy that preferentially affects lower-SR fibers, then the ability to analyze envelopes at conversational sound levels is likely to be impaired. Both theoretical simulations and preliminary experimental evidence from envelope-following responses (EFRs, described in Section 3.2.4) recorded in mice and humans are consistent with this reasoning, as discussed in Section 3.3.

### 3.2.4 Objective measures of subcortical temporal coding

Many psychophysical studies have been devoted to the development and discussion of behavioral measures to assess temporal coding in both NHT and hearing-impaired listeners (Moore, 2003; Strelcyk and Dau, 2009). On the other hand, SSSRs provide an objective window into how the subcortical nuclei of the ascending auditory pathway encode temporal information in sound. While behavioral characterizations are important indica-

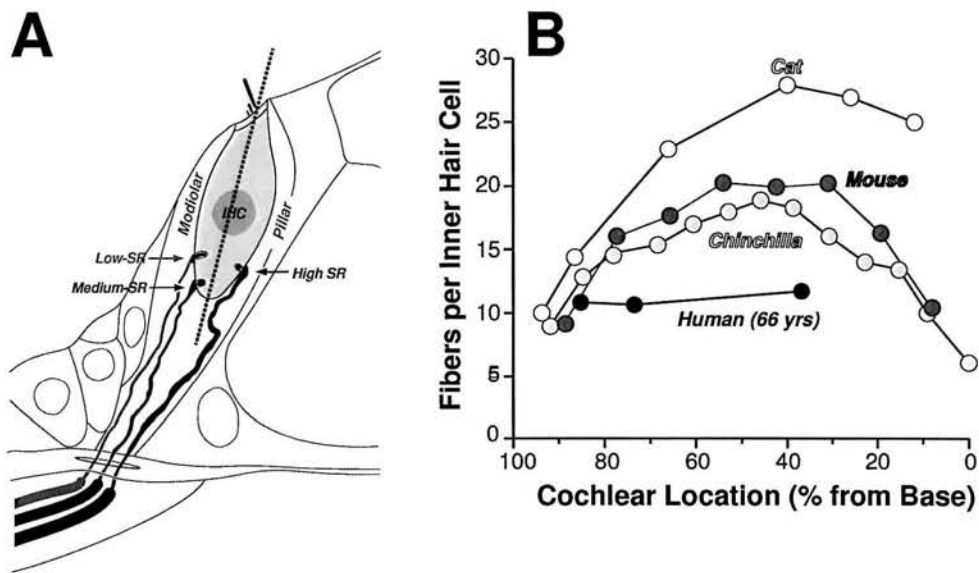


Figure 3.1: Innervation of the inner hair cell by terminals of the cochlear nerve. A: Schematic illustrating the spatial separation of the synaptic contacts of high- (SR > about 18 spikes/s) vs. medium- and low-SR fibers on the pillar vs. modiolar sides of the inner hair cell, respectively. B: Counts of cochlear nerve terminals per inner hair cell as a function of cochlear location from 4 mammalian species: cat (Lieberman et al., 1990), mouse (Maison et al., 2013), chinchilla (Bohne et al., 1982) and human (Nadol Jr, 1983)

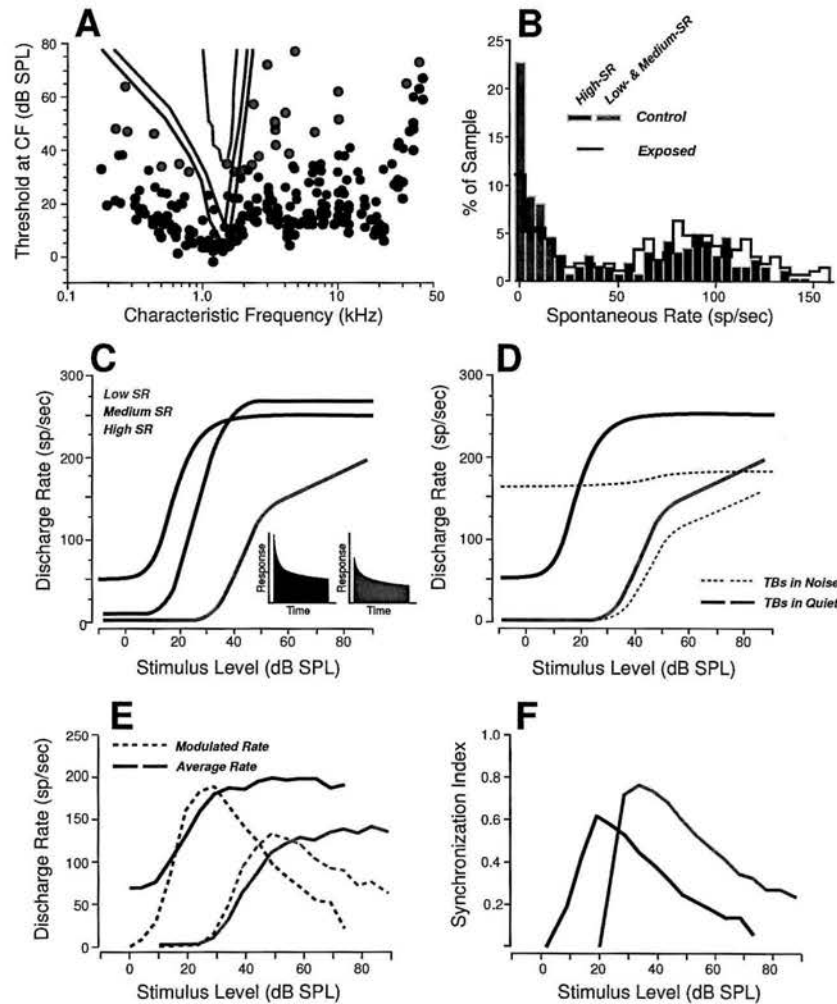


Figure 3.2: Response differences among cochlear nerve fibers of the three SR groups. A: Threshold tuning curves of a high- medium- and low-SR fiber (see key in C) are superimposed on a scatterplot of thresholds at the characteristic frequency for all the fibers sampled from one animal. Data from cat (Lieberman, 1978). B: Distribution of spontaneous rates in large samples of cochlear nerve fibers before (red and blue bars) vs. after (black line) a noise exposure causing a reversible elevation of thresholds. Data from guinea pig (Furman et al., 2013). C,D: Schematic rate-vs-level functions for high- medium- and low-SR fibers to tone bursts at the characteristic frequency, in quiet (C) and in continuous background noise at a fixed 0 dB spectrum level (D). Data from cat (Lieberman, 1978; Costalupes et al., 1984). The insets in panel C show schematic peri-stimulus time histograms of the response to a moderate-level tone burst: onset rates are higher in the high-SR fiber than in the low-SR fiber. E,F: Responses to SAM tones in high- vs low- SR fibers expressed as average rate and modulated rate (E) or average synchrony (F; see text for definitions). Responses are to carrier tones at the characteristic frequency, amplitude modulated at 100 Hz. Data from cat (Joris and Yin, 1992).

tors of everyday hearing ability, in order to limit the length and scope of this opinion paper and still provide substantial discussion, here we focus on objective, physiological measures that can quantify the temporal coding precision of supra-threshold sound in the individual listener. Such measures may also be helpful in identifying some of the mechanisms that lead to individual differences in behavioral ability.

SSSRs refer to the scalp-recorded responses originating from subcortical portions of the auditory nervous system. These responses phase lock both to periodicities in the acoustic waveform and to periodicities induced by cochlear processing (Glaser et al., 1976). SSSRs are related to auditory brainstem responses (ABRs; the stereotypical responses to sound onsets and offsets; Jewett et al., 1970); however, whereas ABRs are transient responses to sound onsets and offsets, SSSRs are sustained responses to ongoing sounds that can include responses phase locked to both the fine structure and the cochlear-induced envelopes of broadband sounds. SSSRs have been used extensively in basic neurophysiologic investigation of auditory function and sound encoding (e.g., Kuwada et al., 1986; Aiken and Picton, 2008; Gockel et al., 2011; also see Krishnan, 2006; Chandrasekaran and Kraus, 2010, for reviews). Given the frequency specificity possible with SSSRs, they have also been proposed as a potential tool for objective clinical audiometry (Lins et al., 1996). In addition, SSSRs have been shown to be sensitive to deafferentation in that IHC loss leads to degraded SSSRs, especially at moderate sound levels (Arnold and Burkard, 2002).

While there are many studies of SSSRs, confusingly, different branches of the scientific literature use different names to refer to the same kinds of measurements. Periodic responses to amplitude-modulated sounds originating from both the sub-cortical and cortical portions of the auditory pathway are often collectively referred to as auditory steady-state responses (ASSRs Galambos et al., 1981; Stapells et al., 1984; Rees et al., 1986). However, brainstem SSSRs can be distinguished from responses generated at the cortical level by virtue of their relatively high frequency content; practically speaking, cortical and SSSR responses can be extracted from the same raw scalp recordings by appropriate filtering (e.g., see Krishnan et al., 2012; Bharadwaj et al., 2014a). The responses that specifically phase



lock to the envelope of amplitude modulated (AM) sounds have been referred to as EFRs or amplitude modulation following responses (AMFRs, Dolphin and Mountain, 1992; Kuwada et al., 2002). In the recent literature, SSSRs are most commonly referred to as frequency following responses (FFRs), a term originally used to denote responses phase locked to pure tones (Marsh et al., 1975). Since the term FFR hints that responses are phase locked to the acoustic frequency content of input sound (i.e., the fine-structure of narrowband or locally narrowband sounds), here we will use the term SSSR to describe the sustained responses originating from subcortical portions (at frequencies  $> 80$  Hz or so in humans) of the auditory pathway. More specifically, we will focus on EFRs: SSSRs that are locked to the envelope.

While EFRs provide a convenient non-invasive measure of subcortical envelope coding, there are several difficulties in interpreting them. First, they represent neural activity that is the sum of a large population of neurons, filtered by layers of brain tissue, skull, and scalp. Depending on the stimulus parameters, thousands of neurons in each of multiple subcortical nuclei may contribute to the EFR (Kuwada et al., 2002). Neurons from several regions along the tonotopic axis could contribute to the EFR for high-level sounds due to spread of excitation, even for narrow-band sounds. Thus, relating EFR results to physiological responses of single neurons is not straightforward. ANF modulation frequency responses are uniformly low pass; high characteristic frequencies (CFs) fibers ( $> 10$  kHz) have cutoff frequencies around 1 kHz in cat (Joris and Yin, 1992). Below 10 kHz, cutoff frequency is dependent on CF, suggesting a limit imposed by an interaction between the content of the input signal and the bandwidths of cochlear filters (Joris and Yin, 1992). As signals ascend the auditory pathway, they are transformed from a temporal to a rate code, with the upper limit of phase locking progressively shifting to lower modulation frequencies (summarized in Figure 9 of Joris et al., 2004; see also Frisina et al., 1990; Joris and Yin, 1992; Krishna and Semple, 2000; Nelson and Carney, 2004). Modulation frequencies in the 70 to 200 Hz range elicit phase-locked responses in a cascade of subcortical auditory structures, from cochlear hair cells to inferior colliculus (IC) neurons, suggesting that many

sources can contribute to the EFRs in this frequency range. Luckily, compared to the IC, the more peripheral EFR generators generate relatively weak responses, both because they drive smaller synchronous neural populations and because they are more distant from the measurement site. Based on single-unit data, reversible inactivation studies, irreversible lesion studies, and studies analyzing EFR group delay, it has been argued that the dominant generators of the EFR move from caudal (AN and CN) to rostral (inferior colliculus or IC) as modulation frequency decreases (Sohmer et al., 1977; Dolphin and Mountain, 1992; Kiren et al., 1994; Herdman et al., 2002; Kuwada et al., 2002). These studies provide evidence that the IC dominates EFRs at modulation frequencies between about 70 and 200 Hz, in all species tested. Changes in the slope of the response phase vs. input modulation frequency can be used to calculate apparent latency of the sources and thereby infer changes in the relative strengths of different neural generators in the mixture (Kuwada et al., 2002); regions where the slope is constant indicate regions where the mixture of generators is constant. Above 200 Hz, the pattern of these changes varies across species, probably due to differing head sizes and shapes. Humans, rabbits, and mice exhibit regions of constant phase slopes out to 500, 700, and 1000 Hz, respectively (Kuwada et al., 2002; Purcell et al., 2004; Pauli-Magnus et al., 2007); in contrast, in gerbils, the phase slopes above 200 Hz are not constant (Dolphin and Mountain, 1992). These differences in phase slopes indicate that the specificity of EFRs is species-dependent. However, in all species it is clear that manipulation of modulation frequency can be used to bias responses towards more rostral or more caudal sources.

Despite these complications, all acoustic information is conveyed to the brain through the ANFs; moreover, deficiencies at the level of the ANF can be expected to have an effect downstream, in higher-order processing centers. Therefore, EFRs originating in the brainstem/mid-brain are likely to reflect the consequences of ANF neuropathy. Indeed, by using different stimuli, it may be possible to emphasize the contribution of different sub-cortical sources (by changing the modulation frequency of the input) or different portions of the cochlear partition (by changing the acoustic carrier of the signal). In particular,

metrics such as the phase-locking value (PLV) can be calculated to quantify the robustness of temporal coding in the EFR, akin to using the vector-strength to assess temporal coding in single-unit physiology studies (Joris et al., 2004).

When analyzing the temporal precision of signals, the PLV has a straightforward interpretation. The details of the PLV computation and its statistical properties are described in a number of previous studies (e.g., see Lachaux et al., 1999; Bokil et al., 2007; Ruggles et al., 2011; Zhu et al., 2013). Briefly, the PLV quantifies the consistency of the response phase across repetitions of the stimulus presentation (trials). For a given frequency bin, the response to each trial can be represented as a unit vector (phasor) in the complex plane whose phase equals the response phase. The PLV then equals the magnitude (length) of the vector average of the phasors, averaged across trials (Figure 3.3A). If the response is consistently at or near a fixed phase, then the resulting average has a magnitude near one and the PLV is high (top panel, Figure 3.3A). On the other hand, if the response phase relative to the stimulus is random over the unit circle, the phasors cancel, the resultant vector has a small magnitude, and the PLV is near zero (bottom panel of Figure 3.3A). An example of the PLV spectrum (computed for EFRs from 400 repetitions of a 100 Hz transposed tone at a carrier frequency of 4 kHz and 65 dB SPL) is shown in Figure 3.3C. Strong peaks are evident at the fundamental and harmonic frequencies of the envelope. The PLV thus is one way of assessing the temporal coding fidelity of the EFR, and of subcortical encoding of supra-threshold sound.

### 3.3 Evidence for cochlear neuropathy

#### 3.3.1 Neuropathy and selective loss of lower-SR fibers in animals

Recent studies in both mice and guinea pigs show that noise exposure that causes a temporary increase in threshold sensitivity (e.g., initial threshold elevations of as much as 40 dB that completely recover over 3-5 days) nevertheless can cause a rapid loss of 40-50% of the ANF synapses on IHCs as well as a slow death of the ANF cell bodies (spiral gan-

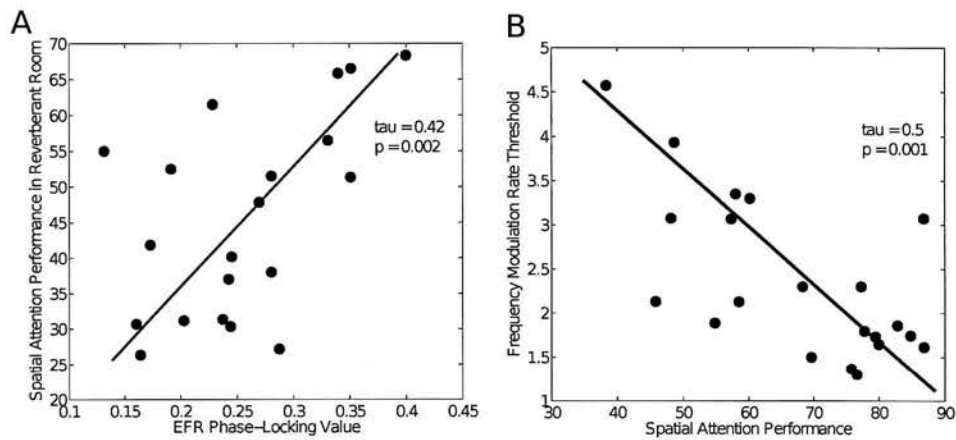


Figure 3.3: Human behavioral and EFR data (data from Ruggles et al., 2011, 2012) showing large variability in both performance and temporal coding fidelity among NHT participants. A: Relationship between spatial attention task performance in reverberation and EFR phase-locking value across NHT listeners. Task performance varied from chance levels (30%) to about 70% with a concomitant variation in EFR phase locking. Listeners with good temporal coding of envelopes as measured by the EFR PLV were able to spatially segregate the competing speech streams and performed well. B: Relationship between spatial attention task performance and frequency modulation (FM) detection thresholds (data from Ruggles et al., 2011), a task known to rely on robust encoding of temporal fine structure.

glion cells) and central axons (Kujawa and Liberman, 2009; Maison et al., 2013). Despite the extent of effects of such exposure on synapses and ganglion cells, it does not typically cause any loss of hair cells. Single-unit recordings in the guinea pig indicate that this noise-induced loss is selective for lower-SR fibers (Furman et al., 2013). Pharmacological studies suggest that this neuropathy is the result of a type of glutamate excitotoxicity, brought on by glutamate overload at particularly active synapses (Pujol et al., 1993). In the central nervous system, glutamate excitotoxicity is mediated by an increase in intracellular calcium concentration (Szydlowska and Tymianski, 2010). Since mitochondria comprise an important intracellular calcium buffering system, the relative paucity of mitochondria in the lower-SR fibers (Liberman, 1980) may contribute to their special vulnerability to glutamate excitotoxicity caused by noise exposure.

In aging mice, there is a steady degeneration of ANFs. Indeed, 30-40% of IHC synapses are lost by roughly 3/4 of the lifespan, an age at which threshold elevation is modest (typically less than 10 dB), but there is no significant loss of hair cells (Sergeyenko et al., 2013). Previous neurophysiological studies of age-related hearing loss in the gerbil suggest that this neurodegeneration is also selective for lower-SR fibers (Schmiedt et al., 1996). Unfortunately, relatively little is known about how aging impacts ANF synapses in humans. The only study that counted IHC synapses in the human inner ear (Figure 3.1B) found relatively low numbers of IHC synapses; however, this low count may reflect a significant degree of age-related neuropathy rather than a species difference, given that the tissue was obtained from a relatively old individual (63 years of age). Indeed, counts of spiral ganglion cells in an age-graded series of human temporal bones show degeneration of 30%, on average, from birth to death, even in cases with no hair cell loss (Makary et al., 2011). The marked delay between synaptic death and spiral ganglion cell death (12 years in mouse, and possibly much longer in humans) suggests that the loss of cochlear nerve synapses on IHCs is almost certainly significantly greater than 30%, on average, in the aged human ear.

Considering that only a small number of sensitive, intact ANFs may be needed for detection in quiet (Lobarinas et al., 2013), it seems likely that even considerable neuropathy

would not change thresholds for tones in quiet, and thus would not be detected by standard threshold audiometry. This is even more likely the case if the neuropathy is selective for ANFs with higher thresholds, which are not active near perceptual thresholds. It also seems likely that a loss of a large population of high-threshold ANFs could dramatically affect auditory performance on complex tasks that require analysis of supra-threshold sound content, such as those requiring the extraction of precise timing cues or extracting a signal in a noisy environment, as discussed above. Thus, we hypothesize that cochlear neuropathy in general - and possibly selective neuropathy of high threshold fibers in particular - is one of the reasons that aging often is found to degrade human performance on tasks requiring analysis of the content of supra-threshold sound.

### 3.3.2 Human data consistent with the neuropathy hypothesis

While there is no human data yet to directly support the neuropathy hypothesis, a series of studies from our lab are consistent with the hypothesis that cochlear neuropathy causes difficulties with coding of supra-threshold sound for humans and accounts for some of the individual variability seen in listeners with normal audiometric thresholds. NHT listeners exhibit marked differences in how well they can utilize precise temporal information to direct selective attention, from near chance levels to almost perfect performance (Ruggles and Shinn-Cunningham, 2011). As discussed in Section 3.2.3, cochlear neuropathy could result in degraded coding of both TFS and envelope information. In line with this hypothesis, differences in EFR phase locking accounts for some of this inter-subject variability in performance. Figure 3.4A shows the relationship between performance in a spatial attention task in reverberation and the PLV calculated from EFRs obtained separately (data from Ruggles et al., 2011, 2012). Pooled over age groups, listeners with higher EFR phase locking performed better in the selective attention task (Kendall  $\tau = 0.42$ ,  $p = 0.002$ ). Though age by itself did not correlate with performance in anechoic conditions, when temporal cues in the acoustic mixture were degraded by adding reverberation, middle-aged listeners showed a bigger drop in performance than younger listeners (Ruggles et al., 2012),

as if timing cues are encoded less robustly in middle-aged listeners than in young adults. In addition, as shown in Figure 3.4B, performance also correlated with thresholds for low-rate frequency modulation (FM) detection, a task known to rely on robust temporal coding of TFS (Kendall  $\tau = 0.5, p = 0.001$ , data from Ruggles et al., 2011, 2012). Crucially, all listeners in these studies had pure-tone audiometric thresholds of 15 dB HL or better at octave frequencies between 250 Hz and 8 kHz. The small differences in hearing threshold (within the NHT range) that did exist were not correlated with selective attention performance; similarly, reading span test scores (a measure of cognitive ability) were unrelated to performance. These results suggest that both TFS and envelope cues are important in everyday listening under challenging conditions, since individuals with poor TFS and envelope coding (as measured by FM detection thresholds and EFR phase locking respectively) perform poorly in a spatial attention task. (For a complete description of the spatial attention task, the FM detection task and the EFR measures, see Ruggles et al., 2011, 2012).

Several other studies have reported that some listeners with normal thresholds (particularly older participants) perform poorly on certain behavioral tasks, sometimes even on par with hearing-impaired subjects. Yet other studies show that temporal processing of both TFS and envelope degrades with aging and manifests independently of hearing loss (see Fitzgibbons and Gordon-Salant, 2010 for a review). In NHT listeners, sensitivity to ITD varies greatly across the population, with some listeners performing as poorly as older hearing-impaired subjects (see Grose and Mamo, 2010; Strelcyk and Dau, 2009). Recent studies have also demonstrated abnormal speech processing among hearing-impaired listeners even when the frequency content of the speech was limited to regions where thresholds are normal, pointing towards supra-threshold coding deficits (Horwitz et al., 2002; Lorenzi et al., 2009; Léger et al., 2012).

Older listeners also have been shown to exhibit deficits specific to envelope processing across a range of tasks, including speech recognition in the presence of modulated noise maskers (Dubno et al., 2003; Gifford et al., 2007) and temporal modulation sensi-

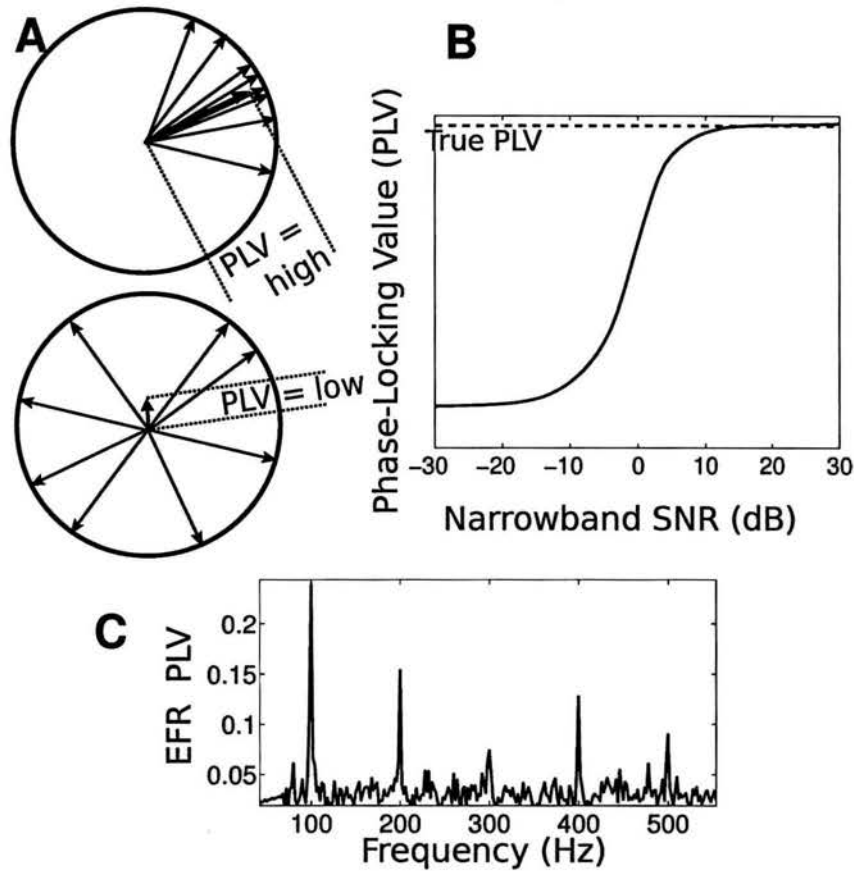


Figure 3.4: A: An illustration of the phase-locking value (PLV) metric computation. The SSSR from each trial is represented by a vector (phasor, shown as a black arrow) with unit magnitude and with phase equal to the EFR phase at the frequency bin of analysis. The vector average of these phasors is computed; the magnitude of the resultant vector (shown as red arrow) yields the PLV. The top panel is an example with high PLV: the phase of the responses varies over a narrow range across trials. The bottom panel is an example with low PLV: response phase relative to stimulus onset is essentially random over the unit circle. B: Relationship between the single-trial SNR of the measurement in the frequency bin of interest and the estimated PLV for a simulated signal in additive noise. At sufficiently high SNR values, the estimated PLV converges to the true PLV (aside from a small sample bias that depends on the number of trials). At lower SNRs, the estimate is biased to be lower than the true value. This is an important consideration when comparing PLVs across sound levels or individuals, since the SNR depends on the magnitude of the true underlying response, the geometry of the generators, and the volume conductor in between. C: Sample PLV spectrum obtained in response to a 100 Hz transposed tone at a carrier frequency of 4 kHz at 65 dB SPL (RMS). Strong peaks are evident in the PLV at multiples of the envelope frequency.



tivity (Purcell et al., 2004; He et al., 2008). Consistent with this, the highest modulation frequency to which EFRs exhibit phase locking decreases with age (Purcell et al., 2004; Leigh-Paffenroth and Fowler, 2006; Grose et al., 2009), supporting the hypothesis that the robustness of supra-threshold modulation coding is reduced with aging. Using measures of both gap detection and word recognition on a sizeable cohort of young and old listeners, Snell and Frisina (2000) concluded that age-related changes in auditory processing occur throughout adulthood. Specifically, they concluded that deficits in temporal acuity may begin decades earlier than age-related changes in word recognition. Though not direct evidence that neuropathy causes these perceptual difficulties, these results are consistent with our hypothesis, especially given animal data suggesting that both aging and noise-exposure degrade ANF responses (especially lower-SR fibers) and degrade supra-threshold temporal coding without affecting thresholds (Schmiedt et al., 1996; Kujawa and Liberman, 2009; Lin et al., 2011a; Furman et al., 2013). If neuropathy underlies deficits in temporal encoding that predict behavioral differences, it may be possible to develop even more sensitive physiological metrics to capture an individual listener's supra-threshold coding fidelity. Section Diagnosing Cochlear Neuropathy is devoted to the discussion of this idea.

### 3.4 Diagnosing cochlear neuropathy

The degree of deafferentation in cochlear neuropathy can be studied directly in animals using invasive methods in combination with histological evaluation, or in humans using post-mortem studies (e.g., Makary et al., 2011). However, assessment in behaving humans must be non-invasive, and therefore must employ indirect methods. Given that neuropathy should impact supra-threshold temporal coding, individual behavioral assessment of envelope and TFS coding of sound at comfortable listening levels may prove useful in assessing neuropathy. In order to expose supra-threshold deficits and individual differences, selective attention tasks in adverse conditions (e.g., in a noise background or in a complex, crowded scene) may be most effective. However, given that aging and noise exposure cause outer

hair cell loss, elevated thresholds, and other (much-studied) effects, assessment of cochlear function is necessary to ensure that supra-threshold deficits are attributable to neuropathy. Measures of brainstem temporal coding, like the ABR and SSSR, may be helpful in assessing neuropathy objectively and passively; exploring these metrics at high sound levels and low modulation depths (which stresses coding of modulations akin to those important when listening in a crowded scene) may be particularly useful (see Section 3.4.2). In order to develop and interpret effective, sensitive tests using these types of non-invasive physiological measures, quantitative models that provide testable predictions will be vital. In this section, we consider some of these points, with a focus on objective measures.

#### **3.4.1 Measuring brainstem coding: ABRs versus SSSRs**

In animal work, the preferential loss of higher-threshold (lower-SR fibers) leads to a decrease in the supra-threshold growth of the amplitude of wave I of the ABR, without a change in ABR threshold (Kujawa and Liberman, 2009; Furman et al., 2013). In both noise-exposed mice and noise-exposed guinea pigs, the proportional decrement in the magnitude of wave I at high levels (i.e., 80 dB SPL) closely corresponds to the percentage of loss of auditory-nerve synapses. However, by limiting the analysis to animals without permanent threshold shifts in the noise-exposed ear, these experiments remove the confound that changes in hearing threshold are likely to affect wave I amplitude; by design, the supra-threshold changes in ABR amplitude found in these experiments cannot be due to differences in threshold sensitivity, but instead reflect differences in the number of fibers responding to supra-threshold sound. Even in populations with normal thresholds, inter-subject variability in ABR amplitudes complicates analysis. One past study showed that in age- and gender-matched mice, the variance in normal ABR amplitude measures is relatively low (Kujawa and Liberman, 2009); however, the mice in this study were genetically identical. In age- and gender-matched guinea pigs, the variance in ABR amplitude is significantly higher. In the genetically heterogeneous guinea pigs, neuropathy-related changes in ABR amplitude are revealed clearly only when data are analyzed within subject, mea-

asuring the effects of noise exposure by normalizing the post-trauma amplitude responses by the responses from the same ear before exposure (Furman et al., 2013). Of course, such a before-and-after approach is unlikely to prove useful for human clinical testing, except in extraordinarily rare circumstances.

The above studies suggest that the ABR may be useful for assessing neuropathy. However, there are a number of reasons why the electrophysiological responses to an AM carrier tone, i.e., the EFR, might be better suited to the assessment of lower-SR neuropathy than the ABR. For one thing, ABR wave I, generated by tone pips, is proportional to the size of the onset responses in the AN. Since, as schematized in Figure 3.2C, the onset responses of lower-SR fibers are small compared to high-SR fiber onset responses (Taberner and Liberman, 2005; Buran et al., 2010), they make a relatively small contribution to the total onset response, rendering the metric fairly insensitive to the integrity of the lower-SR population. In contrast, the steady-state rates of the three SR groups are of more similar magnitude; a loss of lower-SR fibers should thus cause a greater change in steady-state measures like the SSSR or EFR than transient responses like the ABR. Furthermore, as noted above (see Figure 3.2F), lower-SR ANFs synchronize more tightly to the envelope of an AM tone than their high-SR counterparts, especially at moderate and high sound intensities (Johnson, 1980; Joris and Yin, 1992). Synchronization in response to AM-tones can be assessed both by the modulated rate (the amplitude of the peri-stimulus time histogram at the stimulus modulation frequency) and synchronization index (or vector strength; see Joris et al., 2004 for a discussion about different measures of envelope coding). The synchronization index of lower-SR fibers can be larger than that of high-SR fibers of similar best frequency. Indeed, preliminary results suggest that in noise-exposed mice, amplitude decrements in EFR responses to an amplitude-modulated carrier tone presented at the frequency region of maximum cochlear neuropathy are a more sensitive measure of deficit than decrements in ABR wave I amplitude (Shaheen et al., 2013). Perhaps more importantly, a phase-based analysis like the PLV can be used to analyze EFR strength, which can be a more robust and more easily interpreted metric than amplitude measures of these far-field potentials,

which have a weak signal-to-noise ratio (SNR) and depend on factors such as tissue and head geometry.

### **3.4.2 Emphasizing the contribution of lower-SR auditory nerve fibers to the envelope following response**

As previously discussed (Section 3.2.3), one likely consequence of cochlear neuropathy is a reduction in the fidelity of temporal coding in the brainstem. The idea that cochlear neuropathy may preferentially target lower-SR fibers (Schmiedt et al., 1996; Furman et al., 2013) may be exploited to devise EFR measures that are more likely to capture the effects of neuropathy. Focusing on responses to high-frequency envelopes could prove to be an effective way to assess neuropathy, because envelope fluctuations cannot drive saturated high-SR fibers effectively. Even for transposed tones (a modulated high-frequency signal whose envelope mimics the rectified sinusoidal drive of a low-frequency tone operating at low-frequency portions of the cochlea; see van de Par and Kohlrausch, 1997), phase locking of high-SR fibers is reduced at mid to high sound levels (Dreyer and Delgutte, 2006). This effect is likely to be particularly strong for a relatively high-intensity modulated signal with a shallow modulation depth. For such signals, the input intensity of the driving signal will fall within the saturation range of high-SR fibers at all moments; the only fibers that could encode the shallow modulations are the lower-SR fibers. Thus, measures of EFR phase locking to high-frequency, high-intensity, amplitude-modulated signals with shallow modulation may be especially sensitive when assessing lower-SR-fiber status.

Here, we use a simple model of brainstem responses to illustrate why EFRs to shallow amplitude modulations and high sound levels are likely to emphasize the contribution of lower-SR fiber responses to the measurements. Given that EFR responses reflect responses at the level of the brainstem/midbrain, likely the IC, we built a model of IC responses (Figure 3.5A) by combining an established model of the ANF responses (Zilany and Bruce, 2006; Zilany et al., 2009) with previous phenomenological models of amplitude-modulation processing in the IC (Nelson and Carney, 2004). Updated, humanized, ANF model pa-

rameters were used for the simulation (Zilany et al., 2014). This model has been shown to predict ANF single-unit envelope response data quite well (Joris and Yin, 1992). Considering that the simulations included stimuli with high sound levels (as in Dau, 2003; Rønne et al., 2012), a tonotopic array of ANFs (and corresponding IC cells) were included to allow for off-frequency contributions. ANFs with 50 CFs uniformly spaced along the basilar membrane according to a place-frequency map were simulated. For each CF, lower- and high-SR fibers were simulated. In order to obtain a population response at the level of the IC, responses to IC cells driven by lower- and high-SR ANFs were averaged with weights proportional to known population ratios (40% Lower-SR fibers and 60% high-SR fibers, see Liberman, 1978). At the level of the IC, the resulting population response is treated as a proxy for the signal driving the EFR. Responses were simulated for a sinusoidally amplitude modulated (SAM) tone with a carrier frequency of 4 kHz and a modulation frequency of 100 Hz. In order to attenuate the contribution of off-frequency neurons to the population response, a broadband noise masker with a notch centered at 4 kHz and extending 800 Hz on either side was added to the SAM tone, as can be done with real EFR measurements in the laboratory. The SNR for the simulations was fixed at 20 dB (broadband root mean square (RMS)). The IC model parameters were set to the values used in Nelson and Carney (2004), which ensured that the 100 Hz modulation frequency was within the band-pass range of the IC cells. Neuropathy was simulated by progressively attenuating the weights given to the IC population driven by lower-SR ANFs, leaving the high-SR population unchanged.

Figure 3.5 shows the absolute population response magnitude following the 100 Hz modulation in logarithmic units. Results are shown for different amounts of neuropathy, both for different stimulus levels (Figure 3.5B) and for different modulation depths (Figure 3.5C). As seen from the figures, neuropathy has the greatest effect on the population response for stimuli at mid to high sound levels and relatively low modulation depths. This is consistent with the idea that the modulated firing rate of high-SR ANFs is drastically attenuated at moderate to high sound levels and low-modulation depths (Joris and Yin,

1992; Dreyer and Delgutte, 2006). Similar results were obtained (not shown) presenting “transposed” tones to this model as well as when using the Rønne et al. (2012) model, where the EFR is obtained by convolving the ANF population response with a “unitary-response” that is designed to aggregate and approximate all transformations of the ANF population response before being recorded in the EFR. In both model approaches, lower- and high-SR ANF driven IC responses were summed linearly to generate the population response. When the lower- and high-SR ANF responses were mixed non-linearly using a coincidence detection process (i.e., a geometric average instead of an arithmetic average) before being delivered to the IC model, the effects of the lower-SR fiber neuropathy were even larger (not shown). This analysis supports the idea that EFR responses to shallow amplitude modulation at high levels may provide a sensitive, objective correlate of neuropathy. Apart from emphasizing the contribution of lower-SR ANFs, high sound levels are more likely to reveal differences in the number of intact ANFs even if neuropathy is not specific to lower-SR fibers because larger populations of ANFs are recruited overall. These results are also consistent with the report that the ABR wave I amplitude in noise-exposed mice closely corresponds to the amount of neuropathy when the sound level is high (80 dB, Furman et al., 2013) as well as preliminary data from our lab that suggest that individual differences in the EFR are largest at high stimulus levels (Bharadwaj et al., 2013a). In addition, inspection of Figure 3.5B, C suggests that the sizes of the change (i.e., slopes) in the population response with level and with modulation depth both reflect the level of neuropathy. Thus, either of these changes, along with behavioral measures, could be used to assess the ability of the listener to process supra-threshold sound. However, in practice, manipulating modulation depth with the level fixed at a high value may lead to more easily interpreted results than measuring how the EFR changes with overall level (see Section Using Envelope Following Responses to Assess Supra-threshold Coding Fidelity). As explained above, we suggest that individual listeners with normal audiometric thresholds could differ in the number of intact ANFs due to differences in noise exposure, genetic predisposition to hearing damage, and other factors. Given the already-discussed

importance of supra-threshold temporal coding for operating in everyday social settings (understanding speech in noise, directing selective auditory attention, etc.), assessment of neuropathy by measurement of EFRs may have a place in audiological practice, especially because such measures are objective and can be recorded passively (making them suitable for use with special populations in which behavioral assessment is not easy).

### 3.4.3 Isolating cochlear neuropathy

As noted above, in order to assess neuropathy, it is critical to rule out or otherwise account for cochlear dysfunction. One of the most basic characteristics of cochlear function is the frequency selectivity of the basilar membrane (BM). BM frequency selectivity is correlated with cochlear gain at low sound levels (Shera et al., 2002, 2010) and typically decreases with hearing impairment. BM frequency selectivity can be estimated psychophysically (Patterson, 1976; Glasberg and Moore, 1990; Oxenham and Shera, 2003); however, it is possible that such measures may include small contributions from extra-cochlear factors (such as neuropathy). Alternatively, distortion product otoacoustic emissions (DPOAEs) in response to fixed-level primaries (DPgrams; e.g., see Lonsbury-Martin and Martin, 2007) can be used to assess cochlear function. Because OAEs are generated within the cochlea as a consequence of outer-hair-cell activity and do not depend on afferent processing, measuring them may be preferable to measuring psychophysical tuning curve measures. Specifically, normal DPgrams can be used to establish that poor supra-threshold coding arises post transduction (e.g., via cochlear neuropathy) rather than from outer-hair-cell loss or other problems with cochlear amplification (an approach taken in the animal studies of Kujawa and Liberman, 2009; Furman et al., 2013). To test that cochlear compression is intact at the frequencies tested, either stimulus-frequency OAEs (SFOAEs; Schairer et al., 2006) or DPOAE growth functions can be used (Kummer et al., 1998; Neely et al., 2003). DPOAE suppression tuning curves (Gorga et al., 2011; Gruhlke et al., 2012) or SFOAE phase gradients at low stimulus levels (Shera et al., 2002) can provide estimates of cochlear filter tuning. Henry and Heinz (2012) recently demonstrated the importance of considering

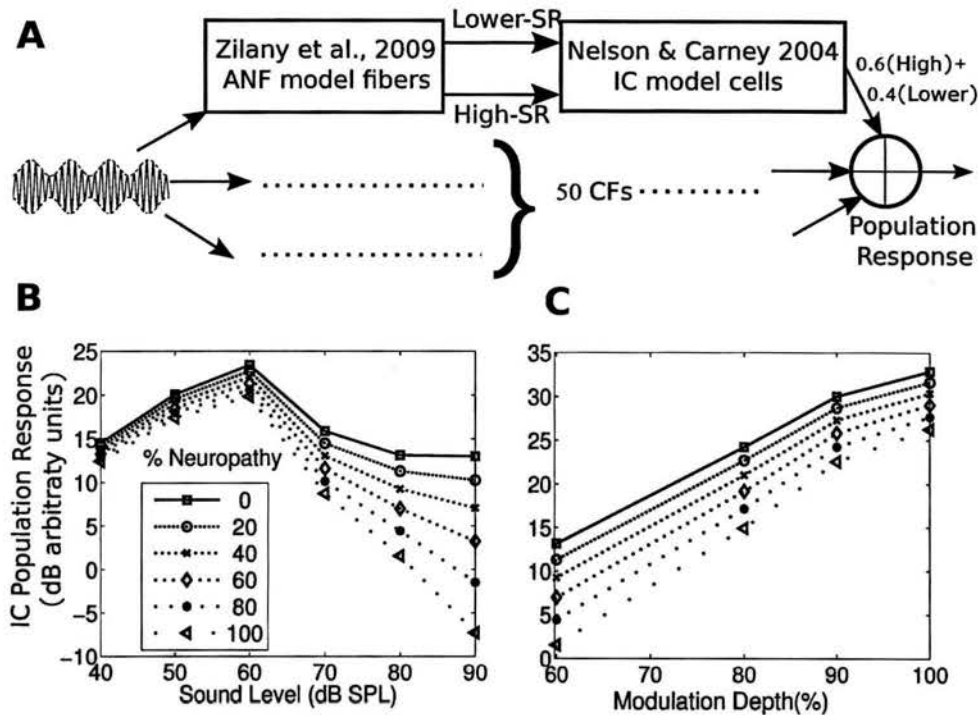


Figure 3.5: A simple model of the population response of inferior colliculus (IC) cells to envelope fluctuations. The model comprised of ANFs (simulated using the Zilany et al., 2009 model) driving the cochlear nucleus (CN), which in turn drives the IC. CN and IC processing of envelope were simulated using the Nelson and Carney (2004) model. A tonotopic array of 50 CFs was used. High-, and lower-SR ANFs were simulated at each CF and the corresponding IC responses were combined with weights equal to the proportion of each group in the population (60% High- and 40% Lower-SR, Liberman, 1978). Neuropathy was simulated by reducing the weight given to the lower-SR driven response. B: Level curves for the population response with different levels of neuropathy for a 100 Hz SAM tone at 4 kHz, with a 60% modulation depth and added broadband noise with a notch centered around 4 kHz and 800 Hz wide on each side. The SNR was fixed at 20 dB (broadband RMS) at all levels. The differences between the levels of neuropathy are most accentuated in the population response at higher stimulus levels. This also suggests that slopes of the level curve at high levels may reflect the level of neuropathy. C: Population response as a function of modulation depth for different levels of neuropathy for an 80 dB SPL SAM tone in notched noise (SNR = 20 dB broadband RMS). The differences between the levels of neuropathy are accentuated better for smaller modulation depths. In addition, this suggests that the slope of the population response strength as a function of modulation depth may be sensitive to the level of neuropathy.



differences in cochlear function in order to interpret differences in measures of temporal coding fidelity properly. As this work shows, establishing that participants have normal cochlear sensitivity by measuring both OAEs and audiometric thresholds is crucial when trying to attribute individual differences in SSSRs and psychoacoustic measures to deficits in supra-threshold coding of sound due to neuropathy.

### 3.5 Future Experiments

A growing body of evidence suggests that (1) NHT listeners vary significantly in how well their auditory systems encode supra-threshold sound; and (2) Noise exposure and aging can lead to considerable amounts of neuropathy without affecting audiometric thresholds. We have argued that cochlear neuropathy in general, and selective neuropathy of lower-SR ANFs in particular, may help explain some of the supra-threshold differences in NHT listeners. Although we believe that the diversity of evidence consistent with this hypothesis is compelling, further experiments are necessary to truly establish these ideas and to understand potential implications for audiological practice. Here, we propose a few key areas that we believe merit future investigation.

#### 3.5.1 Accounting for individual differences in cochlear mechanical function

As discussed in Section Isolating Cochlear Neuropathy, experiments seeking to implicate cochlear neuropathy in human perception must account for individual differences in cochlear processing. There are a number of objective metrics of cochlear health including DPOAE and SFOAE growth functions (Kummer et al., 1998; Schairer et al., 2006), DPOAE suppression tuning curves (Gorga et al., 2011; Gruhlke et al., 2012), and SFOAE group delay measurements (Shera et al., 2002; Shera and Bergevin, 2012). However, there are practical concerns that may limit the utility of many of these methods. For instance, using OAE methods to study neuropathy in patients with elevated hearing thresholds may be difficult, as SFOAE amplitudes critically depend on cochlear gain (Shera and Guinan Jr, 1999).

DPOAE methods depend more on cochlear compression, rather than cochlear gain (Shera and Guinan Jr, 1999), and thus may prove to be a more robust method for assessing contributions of cochlear function to perception in heterogeneous subject populations (Gruhlke et al., 2012). Experiments are needed to determine what tests best quantify cochlear function, enabling such factors to be teased out when appraising cochlear neuropathy, and developing such tests into clinically useful tools.

### 3.5.2 Developing quantitative models of envelope following responses

Because any human measurements of EFRs only indirectly reflect the responses of ANFs, quantitative models of the subcortical generators of the measured response are critical for understanding results and using them to quantify supra-threshold envelope coding. Data suggest that EFRs primarily reflect responses from the mid-brain, and are dominated by responses in the IC (Smith et al., 1975; Sohmer et al., 1977; Dolphin and Mountain, 1992; Kiren et al., 1994; Herdman et al., 2002). However, further experiments are needed to assess if current physiological models capture the behavior of real EFRs. When applied to modulated high-frequency sounds, simple models of IC responses predict a graded loss in the population response with cochlear neuropathy (see Figure 3.5), consistent with the idea that the observed heterogeneity of EFR responses in NHT subjects reflects, in part, differences in ANF survival. Instead of modeling individual neurons, others have modeled brainstem responses (ABRs and FFRs) directly using a kernel method (e.g., Dau, 2003; Rønne et al., 2012). In this approach, all subsequent transformations of the AN responses are modeled by a linear system approximation; model AN responses are used to deconvolve click-ABRs to obtain a unitary response that aggregates all of the transformations occurring from the nerve through to the electrode (including processing within the midbrain nuclei and any summation and filtering influencing what is recorded on the scalp). Despite the obvious simplifying assumptions of such an approach, model predictions capture many of the observed properties of ABRs and FFRs in response to simple stimuli. A slightly more elaborate model of EFRs that combines both these approaches (taking into account single-

unit level phenomena such as in the model in Figure 3.5 as well as scalp- recording properties of the measurements as in Dau, 2003), may be considered. For instance, one recent study explored the consequences of cochlear sensitivity and selective cochlear neuropathy on the latency of simulated ABR responses (Verhulst et al., 2013). Further development, testing, and refinement will ensure that results of EFR experiments are interpreted appropriately in the context of these models. Hence, we identify this as a key area for future efforts devoted to interpreting EFR measures.

### **3.5.3 Using envelope following responses to assess supra-threshold coding fidelity**

A selective loss of lower-SR fibers would likely cause phase locking of the EFR to degrade at high sound levels, in line with the model results presented here (Figure 3.5B). As suggested in Figure 3.5, if neuropathy underlies some supra-threshold deficits, the rate of change of the EFR PLV with sound level (akin to the rate of change of ABR wave I in Furman et al., 2013) would correlate with perceptual abilities on tasks requiring analysis of the envelope of supra-threshold sounds, such as envelope ITD discrimination, spatial selective auditory attention, and related tasks. Preliminary data support this idea (Bharadwaj et al., 2013a). Further experiments are needed to corroborate our hypothesis that neuropathy (especially neuropathy that preferentially affects lower-SR fibers) contributes to individual differences in the ability to analyze complex auditory scenes. The use of narrowband stimuli such as transposed tones (van de Par and Kohlrausch, 1997) with off-frequency maskers may allow for a frequency specific assessment of EFR phase locking at different CFs (i.e., at different frequency channels of the auditory pathway). If the neuropathy hypothesis proves correct, this approach may allow for a frequency-specific diagnosis of cochlear neuropathy from non-invasive physiological measures.

Despite the potential of EFRs (especially the EFR-intensity slope) for assessing cochlear neuropathy, there are some limitations. The EFR is a measure of multi-source population activity and produces scalp potentials that are different mixtures of the source activity at

different scalp locations. These measures depend on the geometry of the generators, properties of the recording electrodes, the volume conductor in between, the level of unrelated electrical activity from cortex and from muscles, and other subject-specific factors (Hubbard et al., 1971; Okada et al., 1997). All of these parameters cause inter-subject variability in the absolute magnitudes of the measured EFRs. This makes interpretation of the raw EFR magnitude difficult. While phase-based metrics such as the PLV are normalized and have a straight-forward interpretation (Zhu et al., 2013), their absolute strength is still influenced by the same factors. Specifically, PLV estimates are biased by the within-band SNR in the raw responses that go into the PLV computation.

This is illustrated in Figure 3.3B, which shows the relationship between estimated and true PLVs for simulated data (signal phase drawn from a von Mises distribution with known concentration and additive noise) as a function of SNR, under the assumptions that the noise phase in any trial is independent of the signal phase (something that can be guaranteed experimentally by jittering the stimulus presentation across trials). In Figure 3.3B, at sufficiently high SNRs, the estimated PLVs converge to the true PLV of the simulated signal, and are insensitive to absolute magnitudes of both signal and noise. However, at intermediate SNR values, the EFR PLV estimates are negatively biased (see Bharadwaj et al., 2014a). This has implications when trying to account for individual differences across subjects, whose raw responses may well have different SNRs. Even in within-subject comparisons, if two experimental manipulations produce responses with very different SNRs, the values of the EFR PLVs will have different biases. This is particularly important when assessing the change in PLV as a function of sound level, since high-level sounds are likely to produce stronger responses (higher SNR measurements) than low-level sounds. While an increase in response power at the stimulus modulation frequency is meaningful in itself, it is not easy to dissociate increases in PLV that result from increases in response synchrony (phase consistency) vs. from increases in response level. Minimally, using recordings in the absence of stimuli might serve to provide estimates of background noise and SNR that can then be used to extract metrics to compare fairly across subjects and conditions. How

important and robust such corrections will prove depends in no small part on where on the SNR curve a particular experimental measurement falls (Figure 3.3B). Additional experiments are needed to characterize these effects in human listeners across different types of stimuli and experimental procedures.

Another limitation is that physiologically, the change in the basilar membrane excitation pattern with sound level also complicates the interpretation of both EFR and psychophysical results. In particular, when seeking to assess cochlear neuropathy within a specific frequency channel using PLV-level growth curves, effects of the spread of excitation are a confounding factor. Use of off-frequency maskers such as notched noise may ameliorate these effects. However, it has also been reported that at least for mid-frequency stimuli (around 1 kHz), the SSSR at the stimulus component frequency can be attenuated by noise even if the peripheral interaction between the signal and the masking noise is expected to be minimal (Gockel et al., 2012).

Alternately, EFRs can be measured in response to narrow-band stimuli with a fixed peak pressure presented at different modulation depths. For deep modulations, high-SR fibers can entrain to the modulation. At shallow modulation depths with a high sound level (carrier level), even the valleys in the signal will have sufficient energy to keep high-SR fibers saturated; thus, the strength of phase locking to shallow modulations may better reflect the contribution of lower-SR ANFs. By computing how the EFR PLV strength changes as the modulation depth is reduced, the spread-of-excitation confounds associated with manipulating the stimulus level may be avoided. Moreover, the approach of fixing the peak sound pressure and progressively decreasing the modulation depth serves to fix the point of operation on the ANF rate-level curve, so that any reduction in PLV with decreasing modulation depth can be interpreted as being related to a drop in synchrony rather than a change in average rate causing a lower SNR. The model results in Figure 3.5C are consistent with this notion. However, as discussed in Section Developing Quantitative Models of EFR Generators, further work is needed to relate EFR results to physiological responses of single neurons. These issues further underscore the importance of combining

electrophysiological, behavioral, and modeling approaches.

### 3.6 Summary and Conclusions

Human listeners with normal audiometric thresholds exhibit large differences in their ability to process supra-threshold sound features. These differences can be exposed in the laboratory by challenging behavioral tasks that necessitate the use of temporal information in supra-threshold sound (e.g., segregating and selecting one auditory object out of a complex scene). While some NHT listeners seek audiological help for difficulties of this sort (a population labeled as having APD), a significant percentage of ordinary, NHT listeners recruited for psychophysical studies in the laboratory, none of whom have known hearing problems, show similar deficits under carefully designed, challenging conditions. These observations hint that perceptual problems with supra-threshold sounds are more widespread than is currently appreciated and that there may be a continuum of abilities across NHT listeners, amongst those who seek audiological help and amongst the general population.

Recent animal work shows that noise exposure and aging can result in a loss of significant proportion of ANFs without any permanent shift in detection thresholds. Moreover, this kind of neuropathy appears to preferentially affect lower-SR ANFs. Both physiological responses to AM stimuli in animals and simplistic computational model simulations suggest that lower-SR fiber loss will degrade temporal coding of sound envelopes at comfortable conversational levels, where high-SR fibers are saturated and therefore unable to entrain robustly to envelopes in input sounds.

A number of studies show that individual differences in the perception of supra-threshold sound are correlated with the strength of brainstem responses measured noninvasively on the scalp (especially SSSRs and EFRs driven by signal modulation). While the absolute strength of EFRs correlates with perceptual abilities, sensitivity of such physiological measures may be improved by using stimuli that mimic conditions akin to adverse listening

conditions, such as high levels and shallow modulations. In addition, differential measures that consider how EFR phase locking changes with stimulus intensity or modulation depth may be especially sensitive when quantifying supra-threshold hearing status, helping to factor out other subject-specific differences unrelated to neuropathy. Interpretation of such measures requires assessment of cochlear function, as well as development of quantitative models of brainstem responses to establish the correspondence between population responses such as EFRs and single-unit physiology.

There are many challenges in trying to relate behavioral and EFR results to underlying physiological changes such as neuropathy, a number of which are due to gaps in current knowledge. However, converging evidence supports the hypothesis that deficits in supra-threshold coding fidelity are relatively common in the population of NHT listeners, and account for at least part of the important differences in how well these listeners can communicate in difficult everyday social settings. Here, we argue that the neuropathy seen in aging and noise-exposed animals may also be occurring in humans and that it may explain observed supra-threshold individual differences. We have also proposed some objective metrics that, based on our hypothesis, should be sensitive measures of the integrity of ANFs, allowing individual assessment of supra-threshold hearing status, and have discussed some of the limitations of the metrics. Still, there remains a large set of questions to be answered, ranging from what mechanisms cause synaptic loss that preferentially affects lower-SR fibers to what physiological or perceptual tests may be most sensitive for assessing neuropathy. We believe these questions should be addressed immediately, given the potential clinical significance of these ideas.

## Chapter 4

# Individual differences in supra-threshold auditory perception are consistent with early effects of noise-exposure and aging

### Abstract

Clinical audiometry involves determining the faintest levels of tones that a listener can detect at different frequencies spanning the normally audible range. Hearing loss is diagnosed only when these hearing thresholds, measured by the audiogram, are elevated. Nonetheless, many listeners with thresholds within the normal hearing range often complain of difficulty hearing in everyday settings and understanding speech in noise. In laboratory settings, listeners with normal thresholds exhibit wide variability in tasks that are engineered to require the use of precise temporal cues. Converging evidence from behavioral and electrophysiological measures point to one potential source of such difficulties: deficits in sound encoding by the early portions of the auditory pathway. Concomitantly, animal studies of the effects of early aging and noise exposure reveal that a significant proportion of the cochlear nerve fiber population could be lost without any discernible changes in the audiogram. Here, using multiple behavioral, otoacoustic and electrophysiological measures in conjunction with computational models of sound processing by the auditory periphery and brainstem, individual differences among listeners with normal hearing thresholds are characterized with a focus on supra-threshold coding of temporal information. Results are consistent with the hypothesis that hidden hearing loss occurring at a very early neu-



ral portion of the auditory pathway contributes significantly to supra-threshold individual differences in auditory perception.

#### 4.1 Introduction

A long-standing tradition in clinical audiology is the characterization of “normal hearing” and hearing loss using threshold audiometry. However, many listeners with normal hearing thresholds (NHTs) complain of difficulty in everyday settings, i.e., in the presence of noise, multiple sound sources, and room reverberation. In the clinic, such listeners are labeled with a range of diagnoses such as obscure auditory dysfunction (Saunders and Haggard, 1992), King-Kopetzky syndrome (Zhao and Stephens, 1996), or more generally as having (central) auditory processing disorders (CAPD/APD, Chermak and Musiek, 1997; Catts et al., 1996; Dawes and Bishop, 2009). The non-specific nature of these labels underscores how little we understand about the mechanisms underlying such supra-threshold deficits. Estimates of the prevalence of such symptoms in adults range from 5-15% of patients seeking audiological help (Cooper Jr and Gates, 1991; Kumar et al., 2007; Hind et al., 2011). In conjunction with the estimated prevalence of hearing loss of one in five individuals in the United States (Lin et al., 2011a), these numbers suggest that supra-threshold hearing deficits may be an important public-health concern that is not currently acknowledged, let alone understood.

Given that threshold audiometry only involves *detection*, it is not surprising that having NHTs does not guarantee good performance in tasks requiring a detailed *analysis* of supra-threshold spectro-temporal sound features. Consistent with this notion, complex, crowded environments such as cocktail parties, noisy restaurants, and busy streets pose special challenges, and are disproportionately frustrating for some NHT listeners as well as for aging listeners and for people with mild hearing loss (Dubno, 1984; Gatehouse and Noble, 2004; Dawes and Bishop, 2009). Several studies have reported that some NHT listeners, particularly older participants, perform poorly on behavioral tasks requiring temporal

acuity, sometimes even on par with hearing impaired listeners (e.g., Grose and Mamo, 2010; Strelcyk and Dau, 2009). Aging is known to degrade auditory temporal perception, seemingly independently of hearing loss (see Fitzgibbons and Gordon-Salant, 2010, for a review). Ruggles and Shinn-Cunningham (2011) showed that NHT listeners exhibit marked differences in how well they can use precise temporal information to direct selective attention, from near chance levels to almost perfect performance. In addition, simple behavioral and electrophysiological measures of coding of temporal fine-structure (TFS) and envelope (ENV) correlated with performance in the complex attention task (Ruggles et al., 2011, 2012). Using measures of both gap detection and word recognition on sizeable cohorts of young and old listeners, Snell and Frisina (2000) concluded that age related changes in auditory processing occur throughout adulthood and that deficits in temporal processing may begin decades earlier than age-related changes in word recognition. Subsequently, they showed a dissociation between threshold sensitivity and temporal processing in adults with NHTs or with mild high-frequency hearing loss (Snell et al., 2002). Thus, a consistent picture appears to be emerging; there are deficits in temporal processing that lead to difficulties in complex everyday environments to which threshold audiometry is insensitive.

Reports of insensitivity of pure-tone thresholds to supra-threshold deficits in processing complex signals date back as early as 1955. Schuknecht and Woellner (1955) reviewed cases of patients who showed abnormally poor speech discrimination scores in relation to their thresholds and hypothesized that cochlear nerve fiber loss may underlie those deficits, consistent with their observation of modest or absent changes in thresholds following cochlear nerve lesions in cats. They reasoned that although only small numbers of cochlear nerve fibers are required to carry the threshold response, more fibers may be required to carry complex signals such as speech. Indeed, one recent study shows that, in animals, thresholds can be normal with only 10-20% of the inner hair cells (IHCs) of the cochlea intact (Lobarinas et al., 2013). This observation is particularly important given recent studies revealing the presence of cochlear-nerve degeneration as a result of noise-exposure and aging. Using groups of unexposed and noise-exposed mice, Kujawa and Liberman (2009) showed

that noise exposure that only leads to reversible/temporary threshold shifts nonetheless leads to an immediate, (seemingly permanent) loss of a considerable proportion (40-50%) of cochlear-nerve terminals innervating the IHCs. While factors associated with cochlear mechanical function and outer hair cell activity such as thresholds for auditory brainstem responses (ABRs), and distortion-product otoacoustic emissions (DPOAEs) recovered in a couple of days, counts of synaptic ribbons remained reduced and were accompanied by supra-threshold reductions in ABR amplitudes. This finding was later replicated in a genetically diverse pool of noise-exposed guinea pigs (Lin et al., 2011b; Furman et al., 2013) using a before-after type within-ear design. Similarly, in animals, aging has also been shown to lead to a reduction in the cochlear-nerve fiber population with an early-onset loss of nerve terminals followed by a delayed degeneration of cell bodies (Schmiedt et al., 1996; Sergeyenko et al., 2013). Recently, using an age-graded series of human temporal bones, a steady degradation of spiral ganglion cell counts with age was also demonstrated in humans (Makary et al., 2011) with about a 30% loss by the age of 70. Given the marked delay between loss of terminals and the loss of cell bodies (Kujawa and Liberman, 2009; Sergeyenko et al., 2013), it is likely that deafferentation and associated functional impairment is more severe than suggested by the modest erosion of spiral ganglion cell density. Furthermore, both in aging and acoustic overexposure, the cochlear neuropathy appears to be selective to fibers with low spontaneous discharge rates ( $SR < 20$  spikes/s, Schmiedt et al., 1996; Furman et al., 2013). Given the higher threshold sound levels associated with these low-SR fibers, this may be an additional factor contributing to the insensitivity of the threshold audiogram to such nerve fiber loss.

Considering the insensitivity of pure-tone thresholds to such cochlear neuropathy, we wondered if portions of the individual differences in supra-threshold auditory ability among NHT listeners could be attributed to such hidden hearing loss (for a detailed articulation of the hypothesis, see Chapter 3). Indeed, given that convergence of, and pooling across multiple fibers underlies an enhancement in the fidelity of temporal coding at higher nuclei along the auditory pathway (Joris et al., 1994; Oertel et al., 2000; Joris and Smith, 2008), it

is likely that a decrease in the size of the nerve population would degrade temporal coding of both TFS and ENV (Lopez-Poveda and Barrios, 2013), and would lead to elevated thresholds and degraded performance in several perceptual tasks under a model of fixed-variance internal noise (Jepsen et al., 2008; MacDonald et al., 2010). Although different perceptual attributes of natural sounds are encoded by different spectro-temporal acoustic cues, many depend on reliable timing information. For instance, perception of speech (Zeng et al., 2005), source location (Blauert, 1997), grouping of acoustic constituents into objects (Elhilali et al., 2009), and release from various kinds of maskers (Moore, 2008; Christiansen et al., 2013), all depend on using TFS and ENV cues. Thus, degraded temporal coding is likely to affect perception in complex acoustic environments. In addition, considering the relative robustness of low-SR fibers to masking (Costalupes et al., 1984; Costalupes, 1985; Young and Barta, 1986) and better synchrony to amplitude-modulations at moderate to high sound levels (Joris and Yin, 1992), a selective low-SR neuropathy would further increase the likelihood of perceptual difficulties in processing supra-threshold sound.

Here, we sought to characterize cochlear mechanical function, supra-threshold temporal coding, and selective attention performance with complex stimuli, in the same group of individual NHT listeners. As described in section 4.2, we characterized supra-threshold temporal coding using behavioral measures of amplitude modulation detection and discrimination of envelope based interaural time differences (ITDs) in addition to objective electrophysiological measures using envelope following responses (EFR). We used psychophysical tuning curves and DPOAEs to characterize cochlear mechanical function. Finally, we use a spatial selective attention task with complex stimuli that were mixtures of speech and noise to evaluate the differential contributions of supra-threshold temporal acuity and cochlear mechanical function to individual differences in perception in a complex scene.

## 4.2 Methods

Given the empirical observation that noise induced temporary and permanent threshold shifts often appear as notches around 4 kHz (Yost, 2007), probably owing to the resonance of the outer ear (Pierson et al., 1994), we designed our methods and stimuli attempting to focus on that tonotopic region of interest. All measures were obtained with the participants seated in an acoustically and electrically shielded booth (single-walled Eckel C-14 booth, Cambridge, MA). For passive measures (DPOAEs and EFRs), participants watched a silent, captioned movie of their choice, ignoring the acoustic stimuli. A personal desktop computer controlled all aspects of the experiment, including triggering sound delivery and storing data. Special-purpose sound-control hardware (System 3 real-time signal processing systems, including D/A conversion and amplification; Tucker Davis Technologies, Gainesville, FL) presented sound through insert phones (ER-1, Etymotic, Elk Grove Village, IL) coupled to foam ear tips. For otoacoustic emission measures sounds were presented and recorded using insert earphones and microphones coupled to foam eartips (ER-10, Etymotic, Elk Grove Village, IL). All sounds were digitized at a sampling rate of 48828 Hz. For all active behavioral experiments, subjects responded by simple button presses and we given feedback after each trial.

### 4.2.1 Participants

Thirty subjects (thirteen female), aged 20 - 40, were recruited from the Boston University community. All subjects had pure tone hearing thresholds better than 15 dB HL in both ears at octave frequencies between 250 Hz and 8 kHz. Subjects provided informed consent in accordance with protocols established at Boston University. Two subjects (aged 27 and 29) with mild high-frequency hearing loss were recruited in order to evaluate the sensitivity of the measures used to characterize cochlear-mechanical function. While the goal was to obtain each of the measures for every subject, some data points were missing. Table 4.2.1 summarizes the measures acquired from each individual participant.

Table 4.1: A summary of which subject performed which test. While some data points could not be acquired because of subject availability constraints, most measures were obtained in a majority of subjects to allow for a correlational analyses and multiple regressions. Key: ENVITD - Envelope ITD discrimination thresholds, AM - Amplitude modulation detection thresholds, ERB - Psychophysical tuning curve measures, PTA - Pure-tone audiometry, OAE - Distortion product otoacoustic emission growth curve measures, ATT - Spatial (ITD based) attention task, EFR - Envelope following responses.

SNo.	ENVITD	AM	ERB	PTA	OAE	ATT	EFR
1				x	x		x
2	x	x		x	x	x	x
3	x	x	x	x	x	x	x
4	x	x		x		x	x
5				x	x		x
6	x	x	x	x	x	x	x
7	x	x		x	x	x	x
8	x	x	x	x	x	x	x
9	x	x	x	x	x	x	x
10	x	x	x	x	x	x	x
11	x	x	x	x	x	x	x
12		x		x		x	x
13	x	x		x		x	x
14	x	x	x	x	x		x
15	x	x	x	x	x	x	x
16	x	x		x			x
17	x	x		x			
18	x	x	x	x	x	x	x
19	x	x		x		x	x
20			x	x	x		
21	x	x	x	x	x	x	x
22	x	x		x		x	x
23				x			x
24	x	x	x	x	x	x	x
25				x			x
26	x	x		x		x	x
27	x	x		x		x	x
28	x	x	x	x	x		x
29	x	x	x	x	x		
30	x	x		x		x	x

## 4.2.2 Characterization of cochlear-mechanical function

### 4.2.2.1 Psychophysical tuning curves

An important aspect of basilar membrane responses is frequency selectivity. Psychophysical tuning curves were measured at a fixed low probe tone sensation level of 10 dB SL using the notched-noise method (Patterson, 1976) in a forward-masking paradigm (Oxenham and Shera, 2003). This approach avoids non-linear cochlear effects that are known to affect tuning measurements obtained using simultaneous masking procedures (Moore and Vickers, 1997) and has been shown to produce measurements that correspond to physiological measures of tuning more closely than other methods (Shera et al., 2002, 2010). The stimulus parameters were identical to those used by Oxenham and Shera (2003), with the one modification that discrete prolate-spheroidal sequences (DPSS, Slepian, 1978) were used as ramping function for the noise and the probe in order to limit spectral leakage (Thomson, 1982). Filter equivalent rectangular bandwidths (ERBs) were estimated by fitting a  $roex(p, w, t)$  function, as in Oxenham and Shera (2003).

### 4.2.2.2 DPOAE input-output curves

To obtain an objective correlate of cochlear-mechanical compression, DPOAE growth functions were measured as a function of level of the  $f_2$  primary tone ( $f_2 = 4$  kHz, Neely et al., 2003). The frequency the level of the  $f_1$  tone were varied according to the formula provided by Johnson et al. (2006) to maximize the level of the DPOAE for each level of the  $f_2$  tone. No artifact rejection was performed; instead, to obtain robust estimates despite artifacts, trials were combined by calculating the sample by sample median instead of the mean. The DPOAE level was calculated using a DPSS-tapered spectral estimate at the distortion frequency of  $2f_1 - f_2$ . The primary ( $f_2$ ) level at which the DPOAE level was 0 dB SPL was quantified as the DPOAE threshold for each subject.

### 4.2.3 Behavioral measures of temporal coding

Based on the hypothesis that cochlear neuropathy underlies supra-threshold individual differences in temporal coding fidelity, we argued that stimuli at relative high levels are more likely to expose differences than quieter stimuli, because higher level stimuli are more likely to recruit larger proportions of the overall population of nerve fibers (Bharadwaj et al., 2014b). In addition, focusing on envelope coding at high levels is more likely to emphasize the contribution of low-SR fibers (see also Joris and Yin, 1992). Given the observed low-SR selective neuropathy in both noise exposure and aging (Figure 4.1), we hypothesized that these stimuli would further increase the likelihood of exposing any supra-threshold temporal coding deficits that may arise as a result of neuropathy (Bharadwaj et al., 2014b).

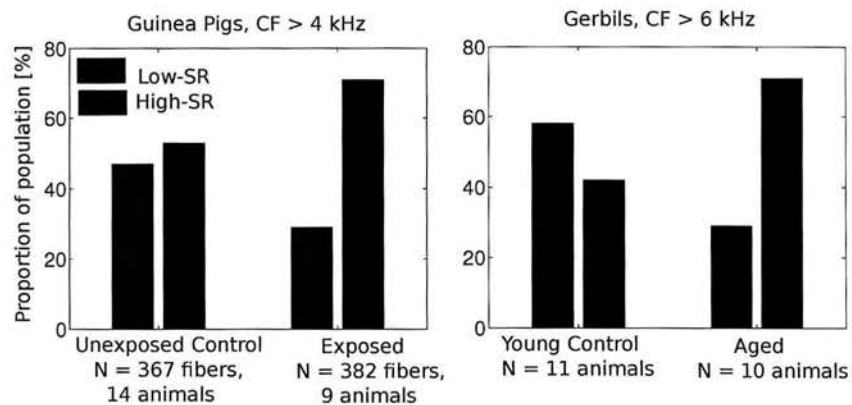


Figure 4.1: Effects on population distribution of cochlear nerve fibers with low- and high-spontaneous discharge rates in acoustic overexposure (left, Furman et al., 2013) and aging (right, Schmiedt et al., 1996). While the split of low- and high-SR fiber counts is roughly 50-50 in both control groups (unexposed guinea pigs and young gerbils respectively), both noise-exposure and aging appear to lead to a selective neuropathy of low-SR fibers. As argued by Bharadwaj et al. (2014b), this selectivity may be used to design stimuli that are more likely to be sensitive to supra-threshold temporal coding deficits that arise as a result of neuropathy.

Thus, detection thresholds for amplitude modulation (AM) were obtained using a broadband noise stimulus at 80 dB SPL (root-mean-square (RMS), Viemeister, 1979).



A 500 Hz wide band centered at 4 kHz was modulated at 19 Hz to create a modulated target. The unmodulated bands of the noise above and below the carrier served as maskers to reduce off-frequency cues. Reference signals were identical but had no modulation. The stimuli were presented diotically in a two alternatives-forced-choice (2AFC) paradigm ((i.e., one interval with the reference signal and another with the target signal) for different modulation depths sampled randomly between 0 and 100%. By using noise stimuli, spectral cues for AM detection were eliminated. Threshold depths for detection were determined using a bayesian approach by fitting the parameters of sigmoidal psychometric function to the responses from 1500 trials and calculating the posterior mean threshold using a monte-carlo markov chain (MCMC) sampling procedure from the posterior density (Kuss et al., 2005). As in Kuss et al., 2005, a normal and log-normal prior were used for the location (threshold) and the slope parameters respectively. The chance level was fixed at 50% and a beta prior was used for the lapse rate parameter.

In order to obtain a binaural measure of temporal coding, we used an envelope ITD discrimination task. A “transposed” tone (van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002) with a carrier at 4 kHz and a modulation frequency of 40 Hz was used. The carrier phase was identical in the two ears and the ITD was only applied to the 40 Hz envelope. The envelope was ramped slowly (and simultaneously in the two ears) over a 100 ms time window to minimise the use of onset-only cues. The stimulus level was set at 75 dB SPL. Off-frequency notched-noise (notch width of 800 Hz) maskers, realized independently in each trial and uncorrelated across the two ears were presented at an SNR of 10 dB (broadband-RMS). The off-frequency masker extended to 20 Hz on the low-frequency side and 20 kHz on the high frequency side and served to attenuate off-frequency cues, including cues from distortion products. Each trial consisted of a sequence of two intervals with the second interval having no ITD. The stimulus in the first interval had an ITD sampled uniformly from the set  $\{50, 100, 200, 400, 800\} \mu s$  with the leading ear randomized across trials. In each trial, the subjects were asked to indicate whether the sound moved from “left to center” or “right to center”. As with the AM threshold

measurement, the threshold was determined by fitting a sigmoidal-psychometric function to responses from 1200 trials and estimating the posterior mean using MCMC sampling.

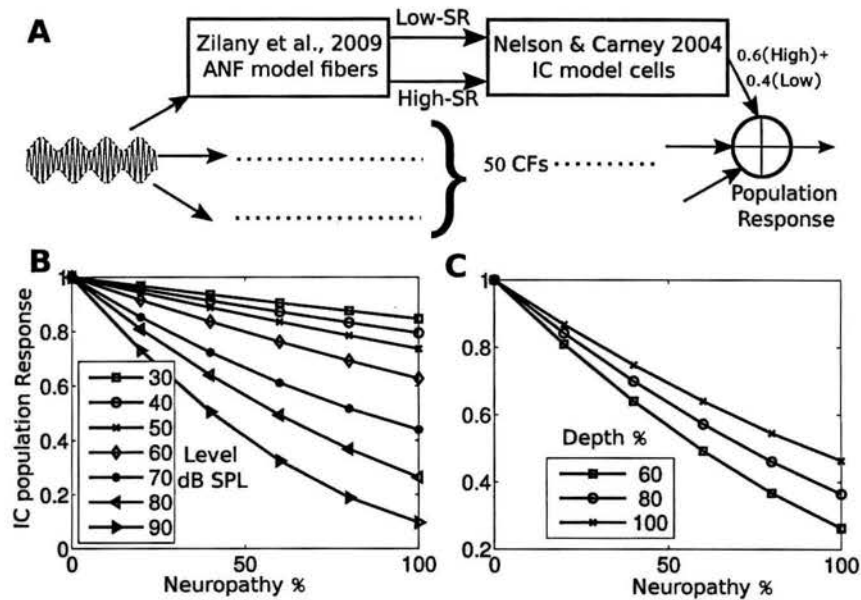


Figure 4.2: A: A simple model of the population response of inferior colliculus (IC) cells to envelope fluctuations. The model comprised of ANFs (simulated using the Zilany et al., 2009 model) driving the cochlear nucleus (CN), which in turn drives the IC. CN and IC processing of envelope were simulated using the Nelson & Carney, 2004 model. A tonotopic array of 50 CFs was used. High-, and lower-SR ANFs were simulated at each CF and the corresponding IC responses were combined with weights equal to the proportion of each group in the population (60% High- and 40% Lower-SR, Liberman 1978). Neuropathy was simulated by reducing the weight given to the lower-SR driven response. B: Degradation of the population response with neuropathy for different levels of a 100 Hz SAM tone at 4 kHz, with a 60% modulation depth and added broadband noise with a notch centered around 4 kHz and 800 Hz wide on each side. The population response degrades more rapidly with neuropathy for higher level stimuli. C: Degradation of the population response with neuropathy for an 80 dB SPL SAM tone for different modulation depths. The population response degrades more rapidly with neuropathy for shallower modulations.

#### 4.2.4 Electrophysiological measure of temporal coding

Envelope following responses (EFR) were measured to 100 Hz transposed tones with a carrier frequency of 4 kHz at a level of 75 dB SPL and for varying modulation depths ( $m$ ). Off-frequency notched-noise maskers were applied at an SNR of 20 dB (broadband-RMS)

with a notch width of 800 Hz to attenuate off-frequency contributions. The noise level was fixed based on pilot experiments which consisted of many more trials than presented to actual test subjects and showed that effect of the noise level on the attenuation of the EFR of the was drastic at initially, i.e., at low noise levels and became a progressively weaker exhibiting a “knee-point” at about an SNR of 20 dB. Attenuations at higher-noise levels were consistent with interactions between noise and signal in measurement space, rather than physiologically, i.e., the differential attenuation of the EFR was proportional to the differential elevation of the noise-floor in the measurements. EFRs were obtained using a 32 channel EEG system (Biosemi Active II system, Amsterdam, Netherlands) using 1000 presentations of each stimulus with half in each polarity. The off-frequency noise was realized independently in each trial. The EFR power was estimated in the frequency domain using a complex-principal component analysis approach to combine across channels. This approach allows to combine measurements from multiple channels with adjustments for phase disparities and has been shown to improve the SNR of the extracted EFR significantly (Bharadwaj and Shinn-Cunningham, 2014).

Under the hypothesis the cochlear neuropathy underlies some of the individual differences in temporal coding, using a simple model of the EFR as a summed population of response of model inferior colliculus cells (Nelson and Carney, 2004), Bharadwaj et al. (2014b) suggested that stimuli at high-levels and shallow modulation depths might be most effective in accentuating the effects of low-SR fiber loss across listeners. In addition, they suggested that changes in EFR amplitude with modulation depth might also be sensitive as a self-normalized measure (Figure 4.2). EFRs were recorded for six different  $m$  values with  $20\log(m)$  varying in steps of  $-4$  dB starting with a value of 0 dB (100% modulation). A multichannel-estimate of phase-locking-value (PLV) was used to infer whether a significant EFR peak above the noise-floor was present. The PLV is convenient for this purpose because the noise-floor distribution depends only on the number of trials that go into the PLV computation and is independent of the frequency bin of analysis and the distribution of background-noise levels (Zhu et al., 2013). Because more than half of the subjects did

not show a significant EFR peak at the modulation frequency of 100 Hz for modulation depths of -12, -16 and -20 dB, only the EFRs in response to stimuli with the largest three  $m$  values was used for further analyses. As suggested by Bharadwaj et al. (2014b), the slope of the EFR amplitude with modulation depth, i.e., the drop in EFR strength for a 4 dB drop in stimulus modulation was computed by fitting a straight line over the three data points corresponding to  $20\log(m) = 0, -4$  and  $-8$  dB respectively.

#### 4.2.5 ITD based attention task

To evaluate the contributions of individual differences in cochlear-mechanical function and supra-threshold temporal coding to differences in listening performance in a complex task, a “spatial” attention task similar to Bharadwaj et al. (2014a) was used. Spoken digits recorded in house in the voice of a female speaker were monotonized to 184 Hz (close to the natural pitch of the voice) using PRAAT (Boersma and Weenink, 2009). To emphasize the use of high-frequency cues, the digits were high-pass filtered at 1500 Hz using FIR filter designed to have 60 dB attenuation at 1000 Hz. Further, uncorrelated low-pass filtered noise (cut-off of 3 kHz) was added to each ear. The stimuli were spatialized using ITDs (broadband). Each trial consisted of two simultaneous sequences of three spoken digits each, with the streams differentiated only with ITDs (Figure 4.3). A visual cue presented 2 seconds before the onset of the sound streams identified the target (left or right) stream. The ITD values in each trial were drawn uniformly from the set 50, 100, 200, 300, 400, 800  $\mu s$ . The target location was randomized on each trial. At the end of each trial a visual response circle cued the subjects to indicate the target sequence by a sequence of button presses and were given visual feedback. The effect of cochlear-mechanical function and supra-threshold temporal coding fidelity on performance in the attention task were determined using a multiple-regression analysis.

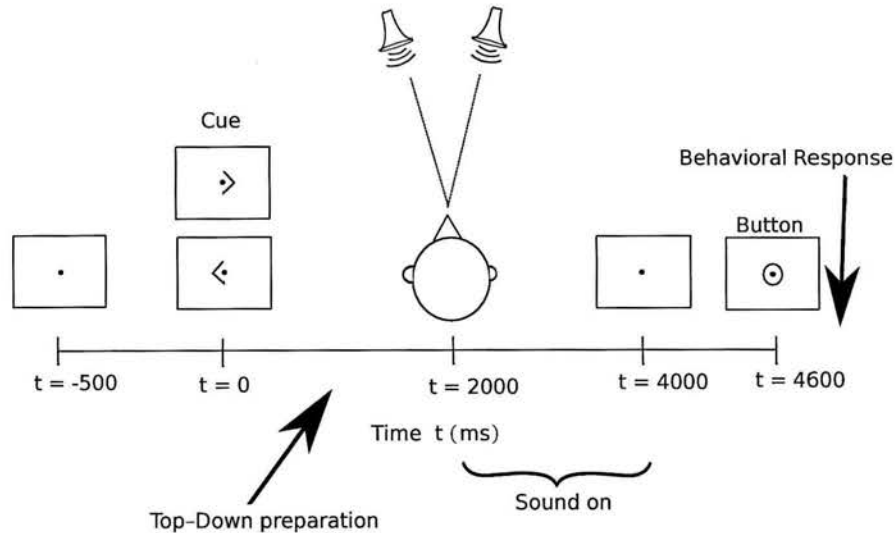


Figure 4.3: A schematic illustration of the ITD based attention task. Each trial began with the subject visually fixated at the center of the screen. A visual cue (left or right arrow) appears two seconds before the onset of the sounds identifying the direction of the target stream (left or right, based on ITDs). Two simultaneous sequences of digits spoken by the same speaker and monotonized to the same pitch were then presented. Then, a visual response circle cued the subject to respond and indicate the three digits in the target sequence by means of button presses. Finally, feedback is given to the subject as follows: a green circle indicating that all three digits were identified correctly (or) a blue circle indicating that two of the three digits were identified correctly (or) a red cross indicating that fewer than two response digits match the correct target sequence.

## 4.3 Results

### 4.3.1 Correlates of peripheral processing

Figure 4.4A shows the DPOAE growth function results for the individual NHT listeners (gray dotted lines) and for the two subjects with hearing loss (solid red lines). The mean and standard error of the mean over the NHT cohort are shown in blue for reference and appear to be in good agreement with published DPOAE input-output curves for NHT listeners (Neely et al., 2003; Johnson et al., 2006). There appear to exist differences across individual listeners in both the shape and overall location of the DPOAE curve leading to differences in thresholds as defined by the primary level at which a DPOAE of 0 dB SPL is obtained. The relatively large attenuation in the DPOAE for the two subjects with hearing

loss testifies to the sensitivity of the measure to cochlear-mechanical function deficits.

The horizontal axis of Figure 4.4B shows the distribution pure tone thresholds at 4 kHz for the cohort NHT listeners. As shown, there are small differences in the thresholds (average across the two ears) across listeners. Psychophysical filter widths also produced results comparable to published literature (Oxenham and Shera, 2003) with a mean ERB of 249 Hz and a standard error of 24 Hz. Importantly, the three measures of cochlear amplifier function correlated with each other (ERB versus pure-tone-thresholds:  $r = 0.51, p \ll 0.01$ , DPOAE versus pure-tone-thresholds:  $r = 0.42, p < 0.01$ ), suggesting that the variance in these measures across subjects is not dominated by measurement noise.

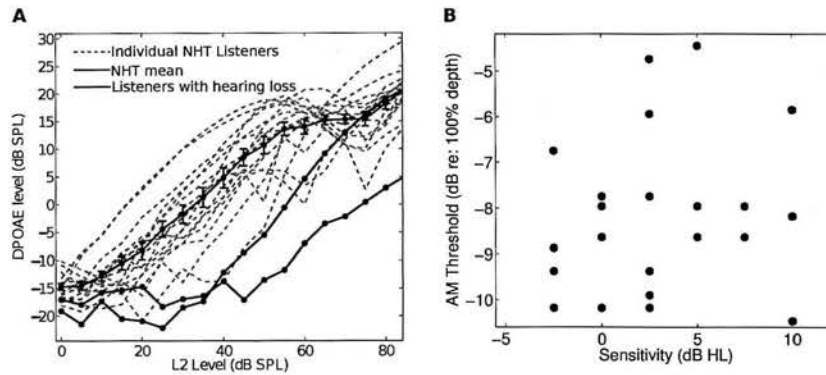


Figure 4.4: A: Results of the DPOAE growth function measurements in the cohort of NHT listeners and two subjects with elevated 4 kHz pure tone thresholds. An examination of the curves corresponding to individual listeners reveals some individual differences in this objective threshold measure (the primary level at which a DPOAE of 0 dB SPL is elicited). The observation that the DPOAE thresholds are much elevated in the subjects with hearing loss suggest that the DPOAE measure is sensitive to deficiencies in cochlear mechanical function. B: Relationship between pure-tone thresholds (averaged across ears) at 4 kHz and threshold for AM detection. While some individual differences exist in the audiogram within the NHT range, they appear to be unrelated to the differences in temporal coding fidelity as measured by AM detection.

### 4.3.2 Correlates of temporal coding fidelity

The vertical axis of Figure 4.4B shows the AM detection thresholds for individual subjects. Similar to previous reports of large individual differences in temporal coding in NHT

listeners (Ruggles et al., 2011, 2012), large individual differences are also seen in the AM thresholds. The thresholds appear to be considerably higher than typical number reported for AM thresholds with noise stimuli (e.g., Viemeister, 1979), and are likely due to narrowband nature of the modulation employed in this study along with the unmodulated off-frequency bands at the same spectral level. Crucially, these differences in AM thresholds appear to be unrelated to the small individual differences in pure-tone thresholds at 4 kHz.

Comcomitantly, large individual differences are also seen in the binaural measures of temporal coding. As shown, in Figure 4.5A, envelope ITD sensitivity for a carrier frequency of 4 kHz varies over a wide range of 46 dB to 59 dB re:  $1\mu s$ , i.e., about  $200\mu s$  to  $900\mu s$ . This range also appears to be larger with values higher than typical reports (e.g., Bernstein and Trahiotis, 2002), but can be reconciled considering the high off-frequency noise masker levels employed in this study (see Bernstein and Trahiotis, 2008). While neither measure of temporal coding fidelity correlated with the audiogram, they exhibited strong correlation with each other ( $r = 0.71, p \ll 0.01$ , Figure 4.5). In addition, psychophysical tuning curve bandwidths (ERBs) did not correlate with AM thresholds ( $r = 0.11, p > 0.05$ ) or envelope-ITD thresholds ( $r = 0.21, p > 0.05$ ) at the 5% false-alarm-rate level.

### 4.3.3 Envelope following response results

As discussed in section 4.2, EFR magnitudes were extracted using frequency-domain principal component analysis for the highest three modulation depths (0 dB, -4 dB and -8 dB). A straight line was fit to calculate the slope of the EFR magnitude as a function of modulation depth. As shown in Figure 4.5B, slopes varied from close to 0 dB per 4 dB drop to 12 dB per 4 dB drop. As predicted by the model (Figure 4.2), the slope of the EFR magnitude as a function of modulation depth was a strong correlate of perceptual AM thresholds. Subjects whose EFR magnitude dropped precipitously with drop in input modulation showed higher AM thresholds. While the EFR magnitude at any one modulation depth also correlated with the perceptual AM thresholds (e.g., EFR at -4 dB versus

AM thresholds:  $r = 0.51, p < 0.01$ ), the slope exhibited a stronger correlation with the perceptual thresholds ( $r = 0.68, p \ll 0.01$ ). To compare the correlations, we used the T2-test (Williams, 1959), which has been shown to perform well when comparing dependent correlations (Steiger, 1980). Based on this test, the slope metric was a significantly stronger correlate of AM thresholds than the EFR at any one modulation depth ( $p < 0.05$ ).

#### 4.3.4 Relationship between attention task performance and correlates of cochlear amplification and supra-threshold temporal coding

Results of the spatial attention task revealed significant individual differences in performance ranging from chance level to ceiling, as in Ruggles and Shinn-Cunningham (2011). An examination of the errors made by subjects suggested the errors were a result of selecting the masker instead of the target, rather than lapses in intelligibility or memory; 91% of the incorrectly reported digits were the digits from the masking stream at the same position in the sequence. This suggested that sensory encoding limitations in being able to select the right digit based on ITD were in play.

In order to evaluate the relative contributions of peripheral processing and neural temporal coding to individual differences in the spatial (ITD based) attention task with broadband speech and noise stimuli, we entered the data into a multiple regression analysis. Because the different measures (audiogram, psychophysical tuning curves and DPOAE thresholds) of cochlear amplifier function were correlated with each other, we summarized all the metrics together by a single factor obtained using principal component analysis (PCA). The same was done for the measures of temporal coding (AM threshold, envelope ITD thresholds and EFR slopes). Before entering the different measures into the PCA, they were each centered (by subtracting across subject average) and normalized (by dividing by the across subject standard deviation) in order to bring them to a common scale. The two factors, one corresponding to peripheral processing and one corresponding to neural temporal coding, were then entered into a multiple-regression analysis. Results revealed that individual differences in temporal coding accounted for a significant fraction of the



across-subject variance in performance in the attention task (49%,  $p \ll 0.01$ ), whereas peripheral processing correlates did not (9%,  $p > 0.05$ ). Thus it appears that for this particular design of the spatial attention task, individual differences in temporal coding dominated the variance in performance. Similar results were obtained when using only the objective measures of peripheral processing and temporal coding (DPOAE thresholds and EFR slopes, respectively).

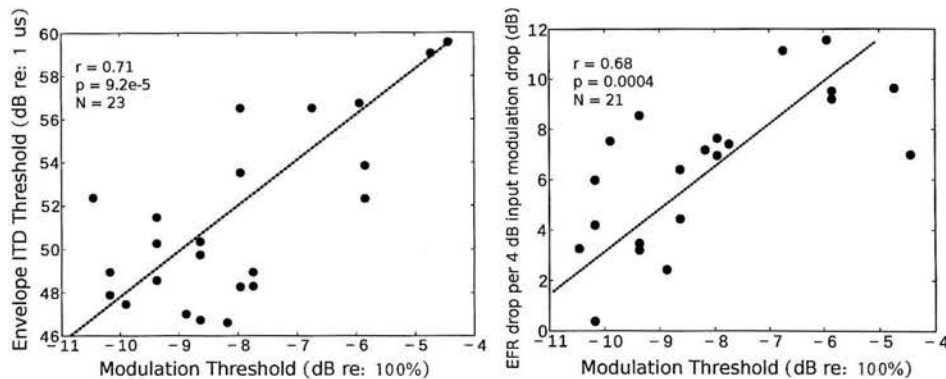


Figure 4.5: A: Results of the behavioral measures of temporal coding. Large individual differences are seen in both AM thresholds and in thresholds for discrimination of envelope ITD. Importantly, there exists a strong correlation between the two measures across individual listeners. Subjects with higher sensitivity to shallow modulations are also able to use envelope timing information better to discriminate interaural temporal differences. B: Relationship between EFR based measure of temporal coding fidelity and behavioral AM thresholds. The slope of the EFR with modulation depth, i.e., a drop in EFR power at the modulation frequency for a 4 dB drop in modulation depth in the stimulus, correlates strongly with behavioral AM sensitivity.

#### 4.4 Discussion

In our high-level task designed to expose individual differences in sensory encoding, large individual differences emerge in performance. Our main finding is that a significant proportion of the variance of which could be accounted for by individual differences in supra-threshold temporal coding of acoustic information. Importantly, measures of cochlear amplifier function, though correlated with each other, accounted for very little of the variabil-

ity in performance in the attention task and were uncorrelated with measures of temporal coding fidelity. In addition, EFR-based metrics that were designed to be more sensitive to supra-threshold temporal coding based on the neuropathy hypothesis, were indeed stronger correlates of behavioral differences. Taken together, these results support the idea that encoding at a very early neural (as opposed to cochlear mechanical) portion of the auditory pathway may be affected in NHT listeners in a manner consistent with early effects of noise-exposure and aging.

For a high-level task such as selective attention, extra-sensory factors, such as differences in language, familiarity of target sounds, memory etc., would likely contribute to performance, and in some cases dominate (e.g., see Surprenant and Watson, 2001; Kidd et al., 2007; Conway et al., 2001). However, in this study, we are concerned with sensory factors that may not be reflected in the audiogram, with a focus on examining “hidden hearing loss” that may exist in the general NHT population. Accordingly, we designed the spatial attention task such that sensory encoding limitations are more likely to be exposed. Errors made by the subjects were consistent with this notion. Clinically identified hearing loss (i.e., a permanent threshold shift) is thought to reflect deficiencies in the function of the cochlear amplifier, reflecting primarily outer hair-cell integrity (e.g., see Stebbins et al., 1979). Recent animal studies, in contrast, identify afferent terminal loss following moderate noise-exposure and early aging even though the hair cells (and presumably cochlear mechanical function) themselves remain intact (Kujawa and Liberman, 2009; Lin et al., 2011a; Makary et al., 2011; Sergeyenko et al., 2013). These findings suggest that traditional definitions of hearing loss are inadequate. Here, under the hypothesis that this cochlear neuropathy may be present in the general NHT population (Bharadwaj et al., 2014b), we employed measures which might be more sensitive to supra-threshold differences in temporal coding than previously used (Ruggles et al., 2011, 2012). To this end, we used stimuli with relatively high sound levels and examined electrophysiological responses with varying modulation depths. In that these modifications indeed resulted in stronger electrophysiological correlates of behavioral measures, the data are consistent

with the neuropathy hypothesis.

While many studies seek to understand the effects of hearing loss and other neurological disorders on auditory temporal processing, unfortunately, not many studies have been conducted that quantify/describe the normal variability that may exist in the NHT population or what the sources of such variability may be. With the aging of world populations in a civilization of incessant noise exposure, the question of how much of the “normal” variability may be due to biophysical processes like neuropathy is one of importance. Indeed, an estimated 17% of the general population, are exposed to noise levels that produce temporary threshold shifts (and lead to considerable neuropathy in animal models) in occupational settings, and experience hearing difficulties that can be attributed to such exposure (Tak and Calvert, 2008; Tak et al., 2009). Recreational noise exposure on an individual basis on the other hand is more difficult to estimate and quantify. Noise exposure levels from recreational, household, hobby and transportation noise have been estimated in several studies (see Clark, 1991; Clark and Bohne, 1999, for reviews) and often exceed 100 dBA. Thus it is plausible that a considerable amount of the inter-individual variability may originate in hidden neural loss owing to noise exposure or its interactions with aging and genetic predisposition to such noise susceptibility (Kujawa and Liberman, 2006; Davis et al., 2001, 2003).

While aging and noise exposure history likely contribute to deficiencies in supra-threshold temporal processing, inherent variability from genetic, epigenetic and experiential factors no doubt play a role in individual outcomes. Indeed, literature exists to suggest that experience and training dependent long term plasticity effects might exist at several levels of the auditory system (Polley et al., 2006; Karmarkar and Dan, 2006). Studies also suggest that such long-term plasticity effects modulate subcortical temporal coding as reflected by the EFR and that this translates to benefits in behavioral performance in auditory tasks (see Chandrasekaran and Kraus, 2010; Anderson et al., 2010; Musacchia et al., 2008; Russo et al., 2005). Further studies are needed to assess the relative contributions of such long-term effects to individual differences in electrophysiological measures such as the EFR

in the general NHT population.

Finally, notwithstanding the findings of strong and meaningful correlations across the multitude of measures employed in this study, the sample size is relatively small for a study seeking to understand the sources of inter-subject variability in supra-threshold listening. While our results suggest that supra-threshold temporal coding differences dominate individual differences in such listening conditions as employed in this study, differences in cochlear mechanics may well be important and account for a significant portion of the individual differences if a larger cohort of listeners is studied.

#### 4.5 Conclusions

Individual differences in supra-threshold auditory perception amongst NHT listeners are revealed in difficult environments where sensory limitations in encoding of precise spectro-temporal features are exposed. In our NHT listeners, these differences appear to be unrelated to functioning of the cochlear amplifier as measured by DPOAEs and psychophysical tuning curves, consistent with the insensitivity of the audiogram to such differences. Instead, results suggest that there are differences in the fidelity with which temporal features are encoded by very early levels of the neural pathway and are consistent with differences in surviving sizes of the cochlear-nerve populations responding to sound. Considering the phenomenology of these individual differences in NHT listeners and the implication in the context of occupational/recreational noise exposure and early aging, these results suggest that considerable effort must be directed towards their study.

## Chapter 5

# Conclusions and Future Work

### 5.1 Summary of findings

#### **Aim 1: Electrophysiological measures of supra-threshold temporal coding**

Chapter 2 demonstrated how to apply a multichannel approach to acquire subcortical steady state responses efficiently. This approach provides a significant improvement in SNR for a given recording session length; equivalently, this approach reduces the data acquisition time by a factor of about three. This improvement makes it feasible to acquire SSSRs in response to several stimulus manipulations within a single session, allowing for a detailed examination of supra-threshold temporal coding in a large cohort of individual listeners.

#### **Aim 2: Model predictions and the cochlear neuropathy hypothesis**

Chapter 3 reviews in detail evidence for cochlear neuropathy, a loss of synapses and afferent nerve terminals that carry information from the inner hair cell to the central nervous system. The consequences of the same for supra-threshold temporal coding are discussed and evidence for such temporal coding deficits in humans is reviewed. Taken together, evidence suggests that individual differences in supra-threshold temporal coding among listeners with normal thresholds may be the first indications of the hidden hearing loss from noise-exposure and aging, thereby explaining the constellation of symptoms affecting NHT listeners seeking audiological help. We present a phenomenological model of population responses that represent the sources of the SSSR (the EFR in particular), built

based on existing models of auditory nerve and inferior colliculus processing of amplitude modulated sounds. The model predicts that examination of envelope coding at high levels and shallow modulations might best expose temporal coding deficits that arise as a result of neuropathy. In addition, the model predicts that the change in the EFR as a function of modulation depth would be a sensitive correlate of supra-threshold temporal coding under the neuropathy hypothesis. These predictions are put to test in chapter 4.

**Aim 3: Relationship between correlates of cochlear mechanical function, supra-threshold temporal coding and spatial attention**

Using an array of objective and behavioral measures on a cohort of 30 listeners, the findings in chapter 4 suggest that supra-threshold temporal coding differences are dominant contributors to individual differences in performance in complex listening tasks where sensory coding limitations are exposed in listeners with normal hearing thresholds. Cochlear mechanical function differences as measured by DPOAEs, thresholds and psychophysical tuning curves account only for a small fraction of supra-threshold differences. On the other hand, as predicted in chapter 3, the change in the EFR for a given drop in input stimulus modulation depth accounts for more than 50% of the individual differences in behavioral measures. In addition, behavioral differences in binaural and complex listening (spatial attention) tasks are partly accounted for by simpler monaural temporal coding measures such as AM detection thresholds. Thus, the data support the hypothesis that suprathreshold sensory deficits arise from a very early neural portion of the auditory pathway, consistent with the cochlear neuropathy effects seen in noise-exposure and early aging.

## 5.2 Significance

Given that the reported research was conducted using a cohort of NHT listeners with no known hearing deficits or complaints, these findings suggest the possibility that supra-threshold effects of noise-exposure and aging may be widespread in the general population,

further implying that “normal-hearing” really spans a continuum of abilities. The establishment of a mechanistic link between measurable supra-threshold coding deficits and heretofore clinically unexplained hearing difficulties is very significant. The identification of fast, objective measures that can assess supra-threshold ability may prove important, clinically. The link between noise-exposure and supra-threshold temporal coding deficits has tremendous implications and provides new considerations for occupational exposure regulations and the use of hearing protection, and points to the importance of educating the general population about the irreversible consequences of noise exposure.

### 5.3 Future research

Our results suggest that noise-exposure and aging may lead to hidden hearing loss that affects communication in social settings. Because these experiments were conducted with a relatively small cohort of NHT listeners recruited from the local community, the ability to control or assess the levels of prior noise-exposure was limited. While occupational noise exposure can be quantified, recreational noise exposure is difficult to assess. However, the clinical significance of these results merit a detailed study of a large sample of listeners spanning the spectrum of occupational and recreational exposures. The statistical power gained from studying such a large and diverse sample may make it possible to tease apart the effects of various kinds of exposures.

As discussed in chapter 3, our knowledge of the relationship between the neurophysiology of single-units and gross population measures such as the EFR are limited. Further experiments and quantitative models are needed in order to help interpret the results of such non-invasive in vivo measure as the EFR.

This thesis focused on sensory coding differences in the periphery and early neural portions of the auditory pathway in NHT listeners. Further research is needed in order to understand the consequences of such sensory differences for higher order processing and vice-versa. A wealth of experiments could be conducted studying individual differences in

sensory coding and its relationship to top-down control that may serve to illuminate the neural mechanisms of selective attention and cocktail-party listening.

Finally, the relationship between such *neural* supra-threshold deficits and cochlear hearing loss needs to be examined in order to be able to design more effective assistive devices and treatments.



## List of Journal Abbreviations

Acta Otolaryngol .....	Acta Otolaryngologica
Acta Otolaryngol Suppl ...	Acta Otolaryngologica Supplementum
Acta Otorhinolaryngol Belg	Acta Acta Otorhinolaryngologica Belgica
Am J Ind Med .....	American Journal of Industrial Medicine
Am J Otolaryngol .....	American Journal of Otolaryngology
Am J Phys Med Rehabil ...	American Journal of Physical Medicine and Rehabilitation
Anesth Analg .....	Anesthesia and Analgesia
Ann Stat .....	Annals of Statistics
Arch Intern Med .....	Archives of Internal Medicine
Audiol Neurotol .....	Audiology and Neurotology
Behav Brain Res .....	Behavioural Brain Research
Bell Syst Tech J .....	Bell System Technical Journal
Br J Audiol .....	British Journal of Audiology
Brain Topogr .....	Brain Topography
Clin Neurophysiol .....	Clinical Neurophysiology
Curr Biol .....	Current Biology
Ear Hear .....	Ear and Hearing
Front Biosci .....	Frontiers in Biosciences
Front Int Neurosci .....	Frontiers in Integrative Neuroscience
Front Neurosci .....	Frontiers in Neuroscience
Front Syst Neurosci .....	Frontiers in Systems Neuroscience

Hear Res .....	Hearing Research
Hum Brain Mapp .....	Human Brain Mapping
Int J Audiol .....	International Journal of Audiology
Int J Bioelectromagn .....	International Journal of Bioelectromagnetism
Int J Lang Commun Disord	International Journal of Language and Communi- cation Disorders
J Acoust Soc Am .....	Journal of the Acoustical Society of America
J Am Acad Audiol .....	Journal of the American Academy of Audiology
J Am Med Assoc .....	Journal of the American Medical Association
J Assoc Res Otolaryngol ...	Journal of the Association for Research in Oto- laryngology
J Clim Appl Meteorol .....	Journal of Climate and Applied Meteorology
J Comp Neurol .....	The Journal of Comparative Neurology
J India Inst Speech Hear ..	Journal of the All India Institute of Speech and Hearing
J Laryngol Otol .....	The Journal of Laryngology and Otology
J Neurophysiol .....	Journal of Neurophysiology
J Neurosci .....	Journal of Neuroscience
J Neurosci Meth .....	Journal of Neuroscience Methods
J Occup Environ Med .....	Journal of Occupational and Environmental Medicine
J R Soc Med .....	Journal of the Royal Society of Medicine
J R Stat Soc .....	Journal of the Royal Statistical Society
J Speech Hear Res .....	Journal of Speech, Language and Hearing Research
J Vis .....	Journal of Vision
Nat Neurosci .....	Nature Neuroscience
Noise Health .....	Noise and Health

Physiol Rev .....	Physiological Reviews
Proc IEEE .....	Proceedings of the IEEE
Proc Meet Acoust .....	Proceeding of Meetings on Acoustics
Proc Natl Acad Am U S A	Proceedings of the National Academy of Sciences of the United States of America
Psychol Bull .....	Psychological Bulletin
Psychon Bull Rev .....	Pyschonomic Bulletin and Review
Psychophysiol .....	Psychophysiology
Sig Proc IEEE .....	IEEE Signal Processing Magazine
Trends Amplif .....	Trends in Amplification
Trend Cogn Sci .....	Trends in Cognitive Sciences
Trends Neurosci .....	Trends in Neurosciences

## Bibliography

- Aiken, S. and Picton, T. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Res*, 245(1-2):35–47.
- Alain, C., Arnott, S. R., et al. (2000). Selectively attending to auditory objects. *Front Biosci*, 5:D202–D212.
- Anderson, S., Skoe, E., Chandrasekaran, B., and Kraus, N. (2010). Neural timing is linked to speech perception in noise. *J Neurosci*, 30(14):4922–4926.
- Arnold, S. and Burkard, R. (2002). Inner hair cell loss and steady-state potentials from the inferior colliculus and auditory cortex of the chinchilla. *J Acoust Soc Am*, 112(2):590–599.
- Baillet, S., Mosher, J. C., and Leahy, R. M. (2001). Electromagnetic brain mapping. *Sig Proc IEEE*, 18(6):14–30.
- Bernstein, L. R. and Trahiotis, C. (2002). Enhancing sensitivity to interaural delays at high frequencies by using “transposed stimuli”. *J Acoust Soc Am*, 112(3):1026–1036.
- Bernstein, L. R. and Trahiotis, C. (2008). Discrimination of interaural temporal disparities conveyed by high-frequency sinusoidally amplitude-modulated tones and high-frequency transposed tones: Effects of spectrally flanking noises. *J Acoust Soc Am*, 124(5):3088–3094.
- Bharadwaj, H. M., Lee, A., and Shinn-Cunningham, B. (2014a). Measuring Auditory Selective Attention using Frequency Tagging. *Front Int Neurosci*, 8(6).
- Bharadwaj, H. M., Masud, S., and Shinn-Cunningham, B. G. (2013a). Bottom-up and top-down contributions to individual differences in auditory spatial attention task performance. Presented at the Abstracts of the Midwinter Meeting of the ARO XXXVI: #887, San Diego, CA.
- Bharadwaj, H. M., Masud, S., and Shinn-Cunningham, B. G. (2013b). The role of high-frequency cues for spatial hearing in rooms. In *Proc Meet Acoust*, volume 19, page 015049. Acoustical Society of America.
- Bharadwaj, H. M. and Shinn-Cunningham, B. G. (2014). Rapid acquisition of auditory subcortical steady-state responses using multichannel recordings. *Clin Neurophysiol*, page in press.
- Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., and Shinn-Cunningham, B. G. (2014b). Cochlear neuropathy and the coding of supra-threshold sound. *Front Syst Neurosci*, 8:26.

- Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann Stat*, pages 1196–1217.
- Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. Cambridge: MIT press.
- Boersma, P. and Weenink, D. (2009). Praat: doing phonetics by computer (version 5.1.05)[computer program]. retrieved may 1, 2009.
- Bohne, B. A., Kenworthy, A., and Carr, C. D. (1982). Density of myelinated nerve fibers in the chinchilla cochlea. *J Acoust Soc Am*, 72(1):102–107.
- Bokil, H., Purpura, K., Schoffelen, J.-M., Thomson, D., and Mitra, P. (2007). Comparing spectra and coherences for groups of unequal size. *J Neurosci Meth*, 159(2):337–345.
- Bregman, A. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: MIT Press.
- Brillinger, D. R. (2001). Principal components in the frequency domain. In *Time series: data analysis and theory*, volume 36, pages 337–366. Siam.
- Buran, B. N., Strenzke, N., Neef, A., Gundelfinger, E. D., Moser, T., and Liberman, M. C. (2010). Onset coding is degraded in auditory nerve fibers from mutant mice lacking synaptic ribbons. *J Neurosci*, 30(22):7587–7597.
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends Cogn Sci*, 8(10):465–471.
- Catts, H. W., Chermak, G. D., Craig, C. H., Johnston, J. R., Keith, R. W., Musiek, F. E., and et al. (1996). Central auditory processing: Current status of research and implications for clinical practice. *American Speech-Language-Hearing Association [Technical Report]*.
- Chandrasekaran, B. and Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, 47(2):236–246.
- Chermak, G. D. and Musiek, F. E. (1997). *Central auditory processing disorders: New perspectives*. San Diego: Singular publishing group.
- Cherry, E. (1953). Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am*, 25:975–979.
- Christiansen, C., MacDonald, E. N., and Dau, T. (2013). Contribution of envelope periodicity to release from speech-on-speech masking. *The Journal of the Acoustical Society of America*, 134(3):2197–2204.
- Clark, W. W. (1991). Noise exposure from leisure activities: a review. *J Acoust Soc Am*, 90(1):175–181.
- Clark, W. W. and Bohne, B. A. (1999). Effects of noise on hearing. *J Am Med Assoc*, 281(17):1658–1659.

- Cohen, L. T., Rickards, F. W., and Clark, G. M. (1991). A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans. *J Acoust Soc Am*, 90(5):2467–2479.
- Colflesh, G. and Conway, A. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychon Bull Rev*, 14(4):699–703.
- Cone-Wesson, B., Parker, J., Swiderski, N., and Rickards, F. (2002). The auditory steady-state response: full-term and premature neonates. *J Am Acad Audiol*, 13(5).
- Conway, A., Cowan, N., and Bunting, M. (2001). The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychon Bull Rev*, 8(2):331–335.
- Cooper Jr, J. and Gates, G. A. (1991). Hearing in the elderly—the Framingham cohort, 1983–1985: part II. Prevalence of central auditory processing disorders. *Ear Hear*, 12(5):304–311.
- Costalupes, J. A. (1985). Representation of tones in noise in the responses of auditory nerve fibers in cats. I. Comparison with detection thresholds. *J Neurosci*, 5(12):3261–3269.
- Costalupes, J. A., Young, E. D., and Gibson, D. J. (1984). Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat. *J Neurophysiol*, 51(6):1326–1344.
- Darwin, C. and Carlyon, R. (1995). Auditory Grouping. In Moore, B. C. J., editor, *Hearing*, pages 387–424. Academic Press, San Diego, CA.
- 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.
- Davis, R., Kozel, P., Erway, L., et al. (2003). Genetic influences in individual susceptibility to noise: a review. *Noise Health*, 5(20):19.
- Davis, R. R., Newlander, J. K., Ling, X.-B., Cortopassi, G. A., Krieg, E. F., and Erway, L. C. (2001). Genetic basis for susceptibility to noise-induced hearing loss in mice. *Hear Res*, 155(1):82–90.
- Dawes, P. and Bishop, D. (2009). Auditory processing disorder in relation to developmental disorders of language, communication and attention: A review and critique. *Int J Lang Comm Dis*, 44(4):440–465.
- Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Meth*, 134(1):9–21.
- Demanez, L., Boniver, V., Dony-Closon, B., Lhonneux-Ledoux, F., and Demanez, J. (2003). Central auditory processing disorders: some cohorts studies. *Acta Otorhinolaryngol Belg*, 57(4):291–299.

- Dietz, M., Marquardt, T., Salminen, N. H., and McAlpine, D. (2013). Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds. *Proc Natl Acad Sci U S A*, 110(37):15151–15156.
- Dobie, R. A. and Wilson, M. J. (1993). Objective response detection in the frequency domain. *Electroencephalogr Clin Neurophysiol - Evoked Potentials Section*, 88(6):516–524.
- Dobie, R. A. and Wilson, M. J. (1994). Objective detection of 40 Hz auditory evoked potentials: phase coherence vs. magnitude-squared coherence. *Electroencephalogr Clin Neurophysiol - Evoked Potentials Section*, 92(5):405–413.
- Dolphin, W. and Mountain, D. (1992). The envelope following response: scalp potentials elicited in the Mongolian gerbil using sinusoidally AM acoustic signals. *Hear Res*, 58(1):70–78.
- Drennan, W. and Watson, C. (2001). Sources of variation in profile analysis. I. Individual differences and extended training. *J Acoust Soc Am*, 110(5):2491–2497.
- Dreyer, A. and Delgutte, B. (2006). Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: implications for sound localization. *J Neurophysiol*, 96(5):2327–2341.
- Dubno, J. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am*, 76(1):87.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2003). Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. *J Acoust Soc Am*, 113(4):2084–2094.
- Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., and Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*, 61(2):317–329.
- Fitzgibbons, P. J. and Gordon-Salant, S. (2010). Behavioral studies with aging humans: Hearing sensitivity and psychoacoustics. In *The aging auditory system*, pages 111–134. New York: Springer.
- Florentine, M., Buus, S., and Mason, C. R. (1987). Level discrimination as a function of level for tones from 0.25 to 16 kHz. *The Journal of the Acoustical Society of America*, 81(5):1528–1541.
- Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (1990). Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. *Hear Res*, 44(2):99–122.
- Furman, A. C., Kujawa, S. G., and Liberman, M. C. (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *J Neurophysiol*, 110(3):577–586.

- Galambos, R., Makeig, S., and Talmachoff, P. J. (1981). A 40-hz auditory potential recorded from the human scalp. *Proc Natl Acad Sci USA*, 78(4):2643–2647.
- Gatehouse, S. and Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Int J Audiol*, 43(2):85–99.
- Gifford, R. H., Bacon, S. P., and Williams, E. J. (2007). An examination of speech recognition in a modulated background and of forward masking in younger and older listeners. *J Speech Lang Hear Res*, 50(4):857–864.
- Gilbert, C. D. and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, 54(5):677–696.
- Glasberg, B. R. and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear Res*, 47(1):103–138.
- Glaser, E., Suter, C., Dasheiff, R., and Goldberg, A. (1976). The human frequency-following response: its behavior during continuous tone and tone burst stimulation. *Electroencephalogr Clin Neurophysiol*, 40(1):25–32.
- Gnanateja, G., Ranjan, R., Sandeep, M., et al. (2012). Physiological bases of the encoding of speech evoked frequency following responses. *J India Inst Speech Hear*, 31:215–219.
- Gockel, H. E., Carlyon, R. P., Mehta, A., and Plack, C. J. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *J Assoc Res Otolaryngol*, 12(6):767–782.
- Gockel, H. E., Farooq, R., Muhammed, L., Plack, C. J., and Carlyon, R. P. (2012). Differences between psychoacoustic and frequency following response measures of distortion tone level and masking. *J Acoust Soc Am*, 132(4):2524–2535.
- Gorga, M. P., Neely, S. T., Kopun, J., and Tan, H. (2011). Distortion-product otoacoustic emission suppression tuning curves in humans. *J Acoust Soc Am*, 129(2):817–827.
- Grandori, F. (1986). Field analysis of auditory evoked brainstem potentials. *Hearing research*, 21(1):51–58.
- Grose, J. and Mamo, S. (2010). Processing of temporal fine structure as a function of age. *Ear Hear*, 31(6):755–760.
- Grose, J., Mamo, S., and Hall, J. (2009). Age effects in temporal envelope processing: Speech unmasking and auditory steady state responses. *Ear Hear*, 30(5):568–575.
- Gruhlke, A., Birkholz, C., Neely, S. T., Kopun, J., Tan, H., Jesteadt, W., Schmid, K., and Gorga, M. P. (2012). Distortion-product otoacoustic emission suppression tuning curves in hearing-impaired humans. *J Acoust Soc Am*, 132(5):3292–3304.
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413.



- He, N.-j., Mills, J. H., Ahlstrom, J. B., and Dubno, J. R. (2008). Age-related differences in the temporal modulation transfer function with pure-tone carriers. *J Acoust Soc Am*, 124(6):3841–3849.
- Henry, K. S. and Heinz, M. G. (2012). Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nat Neurosci*, 15(10):1362–1364.
- Herdman, A. T., Lins, O., Van Roon, P., Stapells, D. R., Scherg, M., and Picton, T. W. (2002). Intracerebral sources of human auditory steady-state responses. *Brain Topogr*, 15(2):69–86.
- Hind, S. E., Haines-Bazrafshan, R., Benton, C. L., Brassington, W., Towle, B., and Moore, D. R. (2011). Prevalence of clinical referrals having hearing thresholds within normal limits. *Int J Audiol*, 50(10):708–716.
- Horel, J. (1984). Complex principal component analysis: Theory and examples. *J Clim Appl Meteorol*, 23(12):1660–1673.
- Horwitz, A. R., Dubno, J. R., and Ahlstrom, J. B. (2002). Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing. *The Journal of the Acoustical Society of America*, 111(1):409–416.
- Hubbard, J., Llinas, R., and Quastel, D. (1971). Electrophysiological Analysis of Synaptic Transmission. *Am J Phys Med Rehabil*, 50(6):303.
- Irimia, A., Matthew Goh, S., Torgerson, C. M., Chambers, M. C., Kikinis, R., and Van Horn, J. D. (2013). Forward and inverse electroencephalographic modeling in health and in acute traumatic brain injury. *Clin Neurophysiol*, 124(11):2129–2145.
- Irimia, A., Van Horn, J. D., and Halgren, E. (2012). Source cancellation profiles of electroencephalography and magnetoencephalography. *NeuroImage*, 59(3):2464–2474.
- Jepsen, M. L., Ewert, S. D., and Dau, T. (2008). A computational model of human auditory signal processing and perception. *J Acoust Soc Am*, 124(1):422–438.
- Jewett, D. L., Romano, M. N., and Williston, J. S. (1970). Human auditory evoked potentials: possible brain stem components detected on the scalp. *Science*, 167(3924):1517–1518.
- Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J Acoust Soc Am*, 68(4):1115–1122.
- Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (2006). Influence of primary-level and primary-frequency ratios on human distortion product otoacoustic emissions. *J Acoust Soc Am*, 119(1):418–428.
- Jørgensen, S. and Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *J Acoust Soc Am*, 130(3):1475–1487.

- Joris, P., Schreiner, C., Rees, A., et al. (2004). Neural processing of amplitude-modulated sounds. *Physiol Rev*, 84(2):541–578.
- Joris, P. and Smith, P. (2008). The volley theory and the spherical cell puzzle. *Neuroscience*, 154(1):65–76.
- Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. (1994). Enhancement of neural synchronization in the anteroventral cochlear nucleus I. Responses to tones at the characteristic frequency. *J Neurophysiol*, 71:1022–1022.
- Joris, P. X. and Yin, T. C. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *J Acoust Soc Am*, 91(1):215–232.
- Karmarkar, U. R. and Dan, Y. (2006). Experience-dependent plasticity in adult visual cortex. *Neuron*, 52(4):577–585.
- Kidd, G., Watson, C., and Gygi, B. (2007). Individual differences in auditory abilities. *J Acoust Soc Am*, 122(1):418–435.
- Kiren, T., Aoyagi, M., Furuse, H., and Koike, Y. (1994). An experimental study on the generator of amplitude-modulation following response. *Acta Otolaryngol Suppl*, 511:28–33.
- Korczak, P., Smart, J., Delgado, R., M Strobel, T., and Bradford, C. (2012). Auditory steady-state responses. *J Am Acad Audiol*, 23(3):146–170.
- Krishna, B. S. and Semple, M. N. (2000). Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol*, 84(1):255–273.
- Krishnan, A. (1999). Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiol Neurotol*, 4(2):95–103.
- Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res*, 166(1):192–201.
- Krishnan, A. (2006). Frequency-following response. In Burkard, R., Don, M., and Eggermont, J., editors, *Auditory evoked potentials: Basic principles and clinical application*, pages 313–333. Lippencott Williams & Wilkins, Baltimore, MD.
- Krishnan, A., Bidelman, G. M., Smalt, C. J., Ananthakrishnan, S., and Gandour, J. T. (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, 50(12):2849–2859.
- Kujawa, S. and Liberman, M. (2009). Adding insult to injury: Cochlear nerve degeneration after “temporary” noise-induced hearing loss. *J Neurosci*, 29(45):14077–14085.
- Kujawa, S. G. and Liberman, M. C. (2006). Acceleration of age-related hearing loss by early noise exposure: evidence of a missed youth. *J Neurosci*, 26(7):2115–2123.

- Kumar, G., Amen, F., and Roy, D. (2007). Normal hearing tests: is a further appointment really necessary? *J R Soc Med*, 100(2):66–66.
- Kummer, P., Janssen, T., and Arnold, W. (1998). The level and growth behavior of the 2 f1- f2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss. *J Acoust Soc Am*, 103(6):3431–3444.
- Kuss, M., Jäkel, F., and Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *J Vis*, 5(8):478–492.
- Kuwada, S., Anderson, J. S., Batra, R., Fitzpatrick, D. C., Teissier, N., and D’Angelo, W. R. (2002). Sources of the Scalp-Recorded Amplitude Modulation Following Response. *J Am Acad Audiol*, 13(4).
- Kuwada, S., Batra, R., and 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.
- Lachaux, J., Rodriguez, E., Martinerie, J., and Varela, F. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208.
- Léger, A. C., Moore, B. C., and Lorenzi, C. (2012). Abnormal speech processing in frequency regions where absolute thresholds are normal for listeners with high-frequency hearing loss. *Hear Res*, 294(1):95–103.
- Leigh-Paffenroth, E. and Fowler, C. G. (2006). Amplitude-modulated auditory steady-state responses in younger and older listeners. *J Am Acad Audiol*, 17(8).
- Liberman, L. D., Wang, H., and Liberman, M. C. (2011). Opposing gradients of ribbon size and AMPA receptor expression underlie sensitivity differences among cochlear-nerve/hair-cell synapses. *J Neurosci*, 31(3):801–808.
- Liberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber. *J Acoust Soc Am*, 63(2):442–455.
- Liberman, M. C. (1980). Morphological differences among radial afferent fibers in the cat cochlea: an electron-microscopic study of serial sections. *Hear Res*, 3(1):45–63.
- Liberman, M. C. (1982). Single-neuron labeling in the cat auditory nerve. *Science*, 216(4551):1239–1241.
- Liberman, M. C. (1991). Central projections of auditory-nerve fibers of differing spontaneous rate. I. Anteroventral cochlear nucleus. *J Comp Neurol*, 313(2):240–258.
- Liberman, M. C., Dodds, L. W., and Pierce, S. (1990). Afferent and efferent innervation of the cat cochlea: quantitative analysis with light and electron microscopy. *J Comp Neurol*, 301(3):443–460.
- Lin, F. R., Niparko, J. K., and Ferrucci, L. (2011a). Hearing loss prevalence in the United States. *Arch Intern Med*, 171(20):1851–1853.

- Lin, H. W., Furman, A. C., Kujawa, S. G., and Liberman, M. C. (2011b). Primary neural degeneration in the Guinea pig cochlea after reversible noise-induced threshold shift. *J Assoc Res Otolaryngol*, 12(5):605–616.
- Lins, O. G., Picton, T. W., Boucher, B. L., Durieux-Smith, A., Champagne, S. C., Moran, L. M., Perez-Abalo, M. C., Martin, V., and Savio, G. (1996). Frequency-specific audiometry using steady-state responses. *Ear Hear*, 17(2):81–96.
- Lobarinas, E., Salvi, R., and Ding, D. (2013). Insensitivity of the audiogram to carboplatin induced inner hair cell loss in chinchillas. *Hearing Res*, 302:113–120.
- Lonsbury-Martin, B. L. and Martin, G. K. (2007). Distortion-product otoacoustic emissions in populations with normal hearing sensitivity. In Robinette, M. S. and Glattke, T. J., editors, *Otoacoustic Emissions Clinical Applications*, pages 107–130. New York: Thieme Medical Publishers.
- Lopez-Poveda, E. A. and Barrios, P. (2013). Perception of stochastically undersampled sound waveforms: a model of auditory deafferentation. *Front Neurosci*, 7.
- Lorenzi, C., Debrulle, L., Garnier, S., Fleuriot, P., and Moore, B. (2009). Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *J Acoust Soc Am*, 125(1):27–30.
- Lorenzi, C. and Moore, B. C. J. (2008). Role of temporal envelope and fine structure cues in speech perception: a review. In Dau, T., Buchholz, J. M., Harte, J. M., and Christiansen, T. U., editors, *Auditory Signal Processing in Hearing-Impaired Listeners, First International Symposium on Auditory and Audiological Research (ISAAR 2007)*. Denmark: Centertryk.
- MacDonald, E., Pichora-Fuller, M., and Schneider, B. (2010). Effects on speech intelligibility of temporal jittering and spectral smearing of the high-frequency components of speech. *Hearing Res*, 261(1-2):63–66.
- Maison, S. F., Usubuchi, H., and Liberman, M. C. (2013). Efferent feedback minimizes cochlear neuropathy from moderate noise exposure. *J Neurosci*, 33(13):5542–5552.
- Makary, C. A., Shin, J., Kujawa, S. G., Liberman, M. C., and Merchant, S. N. (2011). Age-related primary cochlear neuronal degeneration in human temporal bones. *J Assoc Res Otolaryngol*, 12(6):711–717.
- Marsh, J. T., Brown, W. S., and Smith, J. C. (1975). Far-field recorded frequency-following responses: Correlates of low pitch auditory perception in humans. *Electroencephalogr Clin Neurophysiol*, 38(2):113–119.
- Moore, B. C. (2003). *An introduction to the psychology of hearing*. San Deigo: Academic press, 5th edition.
- Moore, B. C. (2007). *Cochlear hearing loss: physiological, psychological and technical issues*. John Wiley & Sons, 2nd edition.

- Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *J Assoc Res Otolaryngol*, 9(4):399–406.
- Moore, B. C. and Vickers, D. A. (1997). The role of spread excitation and suppression in simultaneous masking. *J Acoust Soc Am*, 102(4):2284–2290.
- Musacchia, G., Strait, D., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Res*, 241(1-2):34–42.
- Nadol Jr, J. (1983). Serial section reconstruction of the neural poles of hair cells in the human organ of Corti. I. Inner hair cells. *Laryngoscope*, 93(5):599–614.
- Neely, S. T., Gorga, M. P., and Dorn, P. A. (2003). Cochlear compression estimates from measurements of distortion-product otoacoustic emissions. *J Acoust Soc Am*, 114(3):1499–1507.
- Nelson, P. C. and Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *J Acoust Soc Am*, 116(4):2173–2186.
- Oertel, D., Bal, R., Gardner, S., Smith, P., and Joris, P. (2000). Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proc Natl Acad Sci U S A*, 97(22):11773–11779.
- Okada, Y. C., Wu, J., and Kyuhou, S. (1997). Genesis of MEG signals in a mammalian CNS structure. *Electroencephalography and clinical neurophysiology*, 103(4):474–485.
- Oxenham, A. J. and Shera, C. A. (2003). Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *J Assoc Res Otolaryngol*, 4(4):541–554.
- Parkkonen, L., Fujiki, N., and Mäkelä, J. P. (2009). Sources of auditory brainstem responses revisited: contribution by magnetoencephalography. *Hum Brain Mapp*, 30(6):1772–1782.
- Pascual-Marqui, R. D. (1999). Review of methods for solving the EEG inverse problem. *Int J Bioelectromagn*, 1(1):75–86.
- Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *J Acoust Soc Am*, 59(3):640–654.
- Pauli-Magnus, D., Hoch, G., Strenzke, N., Anderson, S., Jentsch, T., and Moser, T. (2007). Detection and differentiation of sensorineural hearing loss in mice using auditory steady-state responses and transient auditory brainstem responses. *Neuroscience*, 149(3):673–684.
- Pessoa, L., Kastner, S., and Ungerleider, L. G. (2003). Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *The Journal of Neuroscience*, 23(10):3990–3998.

- Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. (2003a). Human auditory steady-state responses. *Int J Audiol*, 42(4):177–219.
- Picton, T. W., John, M. S., Purcell, D. W., and Plourde, G. (2003b). Human auditory steady-state responses: The effects of recording technique and state of arousal. *Anesth Analg*, 97(5):1396–1402.
- Pierson, L. L., Gerhardt, K. J., Rodriguez, G. P., and Yanke, R. B. (1994). Relationship between outer ear resonance and permanent noise-induced hearing loss. *Am J Otolaryngol*, 15(1):37–40.
- Plack, C. J., Oxenham, A. J., and Fay, R. R. (2005). *Pitch: neural coding and perception*, volume 24. Springer.
- Polley, D. B., Steinberg, E. E., and Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci*, 26(18):4970–4982.
- Pujol, R., Puel, J.-L., D’aldin, C. G., and Eybalin, M. (1993). Pathophysiology of the glutamatergic synapses in the cochlea. *Acta Otolaryngol*, 113(3):330–334.
- Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am*, 116(6):3581–3593.
- Rees, A., Green, G., and Kay, R. (1986). Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res*, 23(2):123–133.
- Rickards, F. W., Tan, L. E., Cohen, L. T., Wilson, O. J., Drew, J. H., and Clark, G. M. (1994). Auditory steady-state evoked potential in newborns. *Br J Audiol*, 28(6):327–337.
- Rønne, F. M., Dau, T., Harte, J., and Elberling, C. (2012). Modeling auditory evoked brainstem responses to transient stimuli. *J Acoust Soc Am*, 131(5):3903–3913.
- Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc Natl Acad Sci U S A*, pages 1–6.
- Ruggles, D. and Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: Individual differences in normal-hearing listeners. *J Assoc Res Otolaryngol*, 12(3):395–405.
- Ruggles, D. R., Bharadwaj, H. M., and Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology*, 22(15):1417–1422.
- Russo, N., Nicol, T., Zecker, S., Hayes, E., and Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, 156(1):95–103.

- Ryugo, D. and Rouiller, E. (1988). Central projections of intracellularly labeled auditory nerve fibers in cats: Morphometric correlations with physiological properties. *J Comp Neurol*, 271(1):130–142.
- Saunders, G. and Haggard, M. (1992). The clinical assessment of “Obscure Auditory Dysfunction” (OAD) 2. Case control analysis of determining factors. *Ear Hear*, 13(4):241–254.
- Schairer, K. S., Ellison, J. C., Fitzpatrick, D., and Keefe, D. H. (2006). Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear mechanics in human ears. *J Acoust Soc Am*, 120(2):901–914.
- Scherg, M. and Von Cramon, D. (1985). A new interpretation of the generators of BAEP waves I–V: results of a spatio-temporal dipole model. *Electroencephalogr Clin Neurophysiol - Evoked Potentials Section*, 62(4):290–299.
- Schmiedt, R., Mills, J., and Boettcher, F. (1996). Age-related loss of activity of auditory-nerve fibers. *J Neurophysiol*, 76(4):2799–2803.
- Schuknecht, H. F. and Woellner, R. C. (1955). An experimental and clinical study of deafness from lesions of the cochlear nerve. *J Laryngol Otol*, 69(02):75–97.
- Sergeyenko, Y., Lall, K., Liberman, M. C., and Kujawa, S. G. (2013). Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *J Neurosci*, 33(34):13686–13694.
- Shaheen, L. A., Delgutte, B., and Liberman, M. C. (2013). Using the auditory steady-state response to measure noise-induced auditory nerve degeneration. Presented at the Abstracts of the Midwinter Meeting of the ARO XXXVI: #649, San Diego, CA.
- Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci*, 34(3):114–123.
- Shera, C. A. and Bergevin, C. (2012). Obtaining reliable phase-gradient delays from otoacoustic emission data. *J Acoust Soc Am*, 132(2):927–943.
- Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci U S A*, 99(5):3318–3323.
- Shera, C. A. and Guinan Jr, J. J. (1999). Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs. *J Acoust Soc Am*, 105(2):782–798.
- Shera, C. A., Guinan Jr, J. J., and Oxenham, A. J. (2010). Otoacoustic estimation of cochlear tuning: validation in the chinchilla. *J Assoc Res Otolaryngol*, 11(3):343–365.
- Shinn-Cunningham, B. (2008). Object-based auditory and visual attention. *Trends Cog Sci*, 12(5):182–186.

- Shinn-Cunningham, B. and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, 12(4):283–299.
- Skoe, E. and Kraus, N. (2010). Auditory brain stem response to complex sounds: A tutorial. *Ear Hear*, 31(3):302–324.
- Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430.
- Smith, J., Marsh, J., and Brown, W. (1975). Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. *Electroencephalogr and Clin Neurophysiol*, 39:465–472.
- Snell, K., Mapes, F., Hickman, E., and Frisina, D. (2002). Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity. *J Acoust Soc Am*, 112(2):720–727.
- Snell, K. B. and Frisina, D. R. (2000). Relationships among age-related differences in gap detection and word recognition. *J Acoust Soc Am*, 107(3):1615–1626.
- Sohmer, H., Pratt, H., and Kinarti, R. (1977). Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol*, 42(5):656–664.
- Stapells, D. R., Linden, D., Suffield, J. B., Hamel, G., and Picton, T. W. (1984). Human auditory steady state potentials. *Ear Hear*, 5(2):105–113.
- Starr, A., Picton, T. W., Sininger, Y., Hood, L. J., and Berlin, C. I. (1996). Auditory neuropathy. *Brain*, 119(3):741–753.
- Stebbins, W. C., Hawkins Jr, J. E., Johnsson, L.-G., and Moody, D. B. (1979). Hearing thresholds with outer and inner hair cell loss. *Am J Otolaryngol*, 1(1):15–27.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychol Bull*, 87(2):245.
- Stellmack, M. A., Byrne, A. J., and Viemeister, N. F. (2010). Extracting binaural information from simultaneous targets and distractors: Effects of amplitude modulation and asynchronous envelopes. *J Acoust Soc Am*, 128(3):1235–1244.
- Strelcyk, O. and Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J Acoust Soc Am*, 125(5):3328–3345.
- Surprenant, A. and Watson, C. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *J Acoust Soc Am*, 110(4):2085–2095.
- Szydlowska, K. and Tymianski, M. (2010). Calcium, ischemia and excitotoxicity. *Cell calcium*, 47(2):122–129.



- Taberner, A. M. and Liberman, M. C. (2005). Response properties of single auditory nerve fibers in the mouse. *Journal of neurophysiology*, 93(1):557–569.
- Tak, S. and Calvert, G. M. (2008). Hearing difficulty attributable to employment by industry and occupation: An analysis of the National Health Interview Survey-United States, 1997 to 2003. *J Occup Env Med*, 50(1):46–56.
- Tak, S., Davis, R. R., and Calvert, G. M. (2009). Exposure to hazardous workplace noise and use of hearing protection devices among US workers - NHANES, 1999–2004. *Am J Indust Med*, 52(5):358–371.
- Thomson, D. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096.
- van de Par, S. and Kohlrausch, A. (1997). A new approach to comparing binaural masking level differences at low and high frequencies. *J Acoust Soc Am*, 101(3):1671–1680.
- Verhulst, S., Bharadwaj, H., Mehraei, G., and Shinn-Cunningham, B. (2013). Understanding hearing impairment through model predictions of brainstem responses. In *Proc Meet Acoust*, volume 19, page 050182. Acoustical Society of America.
- Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am*, 66(5):1364–1380.
- Viemeister, N. F. (1988). Intensity coding and the dynamic range problem. *Hear Res*, 34(3):267–274.
- Wightman, F. and Kistler, D. (1992). The dominant role of low-frequency interaural time differences in sound localization. *J Acoust Soc Am*, 91(3):1648–1661.
- Williams, E. J. (1959). The comparison of regression variables. *J R Stat Soc. Series B (Methodological)*, pages 396–399.
- Yost, W. (2007). The Abnormal Auditory System. In *Fundamentals of Hearing: An Introduction*. Academic Press, 5th edition.
- Young, E. D. and Barta, P. E. (1986). Rate responses of auditory nerve fibers to tones in noise near masked threshold. *J Acoust Soc Am*, 79(2):426–442.
- Zeng, F., Nie, K., Stickney, G., Kong, Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proc Nat Acad Sci U S A*, 102(7):2293–2298.
- Zhao, F. and Stephens, D. (1996). Hearing complaints of patients with King-Kopetzky Syndrome (obscure auditory dysfunction). *Brit J Audiol*, 30(6):397–402.
- Zhu, L., Bharadwaj, H., Xia, J., and Shinn-Cunningham, B. (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.

- Zilany, M. S. and Bruce, I. C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *J Acoust Soc Am*, 120(3):1446–1466.
- Zilany, M. S., Bruce, I. C., and Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *J Acoust Soc Am*, 135(1):283–286.
- Zilany, M. S., Bruce, I. C., Nelson, P. C., and Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *J Acoust Soc Am*, 126(5):2390–2412.

# Curriculum Vitae

**Hari M. Bharadwaj**



---

## Education

- **Boston University**, Boston, MA  
Ph.D., Biomedical Engineering, May 2014
- **University of Michigan**, Ann Arbor, MI  
M.S., Electrical Engineering: Systems, Dec 2008  
M.S., Biomedical Engineering, Apr 2008
- **Indian Institute of Technology (IIT) Madras**, India  
B.Tech., Electrical Engineering, May 2006

## Experience

- **Auditory Neuroscience Laboratory, Boston University, Boston, MA**  
*Research Assistant* Sep 2010 - present
- **Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA**  
*Research Assistant* Feb 2009 - Aug 2010
- **Boston Scientific Cardiac Rhythm Management Research, St. Paul, MN**  
*Research Intern* May - July 2008
- **Functional Magnetic Resonance Imaging Laboratory, University of Michigan, Ann Arbor, MI**  
*Research Assistant* Mar 2007 - Dec 2008

## Teaching

- **Department of Biomedical Engineering, Boston University**  
*Graduate Teaching Fellow* Sep 2011 - Dec 2012  
 Led lectures, discussions, laboratory sessions, mentored design projects
  - BE 401 - **Signals and Systems**
  - BE 491 - **Biomedical Measurements and Instrumentation**
- **Department of Biomedical Engineering, University of Michigan**  
*Graduate Student Instructor* Aug 2007 - May 2008  
 Led discussions, laboratory sessions, mentored design projects for both graduate and undergraduate courses
  - BME 510 - **Medical Imaging Laboratory**
  - EECS 458 - **Biomedical Instrumentation**

## Publications

- Choi, I., Wang, L., **Bharadwaj, H. M.**, & Shinn-Cunningham, B. G. (*under revision*). Individual differences in attentional modulation of cortical responses correlate with selective attention performance.
- Bharadwaj, H. M.**, Masud, S., Verhulst, S., Mehraei, G., & Shinn-Cunningham, B. G. (*In Preparation*). Individual Differences in Suprathreshold Auditory Perception among Normal-Hearing Humans are consistent with Cochlear Neuropathy.
- Bharadwaj, H. M.**, Verhulst, S., Shaheen, L., Liberman, M. C., & Shinn-Cunningham, B. G. (2014). Cochlear Neuropathy and the Coding of Supra-threshold Sound. *Frontiers in Neuroscience*. doi:10.3389/fnsys.2014.00026.
- Bharadwaj, H. M.**, & Shinn-Cunningham, B. G. (2014). Rapid acquisition of auditory subcortical steady-state responses using multichannel recordings. *Clinical Neurophysiology*. doi:10.1016/j.clinph.2014.01.011.
- Bharadwaj, H. M.**, Lee, A. K., & Shinn-Cunningham, B. G. (2014). Measuring Auditory Selective Attention using Frequency Tagging. *Frontiers in Neuroscience*, 8, 6. doi:10.3389/fnint.2014.00006.
- Zhu, L., **Bharadwaj, H. M.**, Xia, J., & Shinn-Cunningham, B. (2013). A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. *The Journal of the Acoustical Society of America*, 134(1), 384-395.
- Shinn-Cunningham, B., Ruggles, D. R., & **Bharadwaj, H. M.** (2013). How Early Aging and Environment Interact in Everyday Listening: From Brainstem to Behavior Through Modeling. *Basic Aspects of Hearing* (pp. 501-510). Springer New York.
- Ruggles, D., **Bharadwaj, H. M.**, & Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology*, 22(15), 1417-1422.

Kenet, T.<sup>†</sup>, Orekhova, E. V.<sup>†</sup>, **Bharadwaj, H. M.**<sup>†</sup>, Shetty, N. R., Israeli, E., Lee, A. K. C. et al. (2012). Disconnectivity of the cortical ocular motor control network in autism spectrum disorders. *NeuroImage*, 61(4), 1226-1234. (<sup>†</sup>equal contribution)

Lee, A. K., Rajaram, S., Xia, J., **Bharadwaj, H. M.**, Larson, E., Hamalainen, M. S., & Shinn-Cunningham, B. G. (2012). Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Frontiers in Neuroscience*, 6: 190.

Ruggles, D., **Bharadwaj, H. M.**, & Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences*, 108(37), 15516-15521.

### Conference Proceedings

**Bharadwaj, H. M.**, Masud, S., & Shinn-Cunningham, B. G. (2013). The role of high-frequency cues for spatial hearing in rooms. Proceedings of Meetings on Acoustics 19:015049.

Verhulst, S., **Bharadwaj, H. M.**, Mehraei, G., & Shinn-Cunningham, B. G. (2013). Understanding hearing impairment through model predictions of brainstem responses. Proceedings of Meetings on Acoustics 19:050182.

Choi, I., Bressler, S., **Bharadwaj, H. M.**, & Shinn-Cunningham, B. (2013). Subcortical and cortical neural correlates of individual differences in temporal auditory acuity. Proceedings of Meetings on Acoustics 19:050125.

Rajaram, S., **Bharadwaj, H. M.**, Shinn-Cunningham, B. G., & Lee, A. K. C. (2011). Cortical functional connectivity inference using MEG. Noninvasive Functional Source Imaging of the Brain and Heart & 8<sup>th</sup> International Conference on Bioelectromagnetism (NFSI & ICBEM) 77-80. IEEE.

**Bharadwaj, H. M.**, Peltier, S., & Noll, D. C. (2008). Simultaneous EEG-fMRI: Effect of choice of MRI pulse sequence and adaptive denoising strategies. Proceedings of the 16<sup>th</sup> Annual Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM), Toronto.

### Invited Talks

*Individual differences revealed by the challenges of listening in a complex, crowded scene.* Special session: Comparative perspectives on the Cocktail Party Problem, 167<sup>th</sup> Meeting of the Acoustical Society of America (ASA), Providence, RI (May, 2014).

*Bottom-Up and Top-Down Contributions to Individual Differences in Auditory Spatial Attention Task Performance.* Young Investigator Symposium, 36<sup>th</sup> MidWinter Meeting of the Association for Research in Otolaryngology (ARO), Baltimore, MD (Feb, 2013).