ASSESSING HUMAN MEDIAL OLIVOCOCHLEAR REFLEX FUNCTION WITH COMPLEMENTARY PRE-NEURAL AND NEURAL ASSAYS

by

Spencer Benjamin Smith

Copyright © Spencer Smith 2017

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF SPEECH, LANGUAGE, AND HEARING SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2017
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Spencer Benjamin Smith, titled Assessing Human Medial Olivocochlear Reflex Function with Complementary Pre-Neural and Neural Assays and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: (05/5/2017)
Barbara K. Cone

Date: (05/5/2017)
David S. Velenovsky

Date: (05/5/2017)
Frank E. Musiek

Date: (05/5/2017)
Elena Plante

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: (05/5/2017)
Dissertation Director: Barbara K. Cone
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interest of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Spencer Benjamin Smith
ACKNOWLEDGEMENTS

Over nearly seven years as a student at the University of Arizona, I have received unwavering support. The following acknowledgements will certainly fall short of articulating the immense gratitude that I feel for everyone who has offered me mentorship and friendship during my time here. Nevertheless, I will try...

Barbara Cone: I am forever grateful that I had you in my corner as a mentor and advocate as we navigated uncharted territory together with the Au.D./Ph.D. program. Thank you for all of your support both in my professional and personal life. Your enthusiasm for teaching, learning, and tinkering will be a model for me in my career as I begin to mentor students of my own. Thank you for introducing me to the wonderful world of electrophysiology and for affording me the flexibility to explore my interests and ideas.

David Velenovsky: I look back at our long afternoon conversations and “Science Fridays” with fondness. We are kindred spirits in curiosity, and I always appreciated your willingness to dive into rabbit holes with me. Thank you for teaching me to be resourceful with lab equipment and that, if all else fails, take it apart!

Elena Plante: Thank you for teaching me to sit down, roll up my sleeves, and write the next grant. Your grant writing class and feedback on my F30 were some of the most valuable experiences of my doctoral training program. I have always valued your honest feedback on all of my work.

Frank Musiek: Thank you for keeping your door open to discuss all brain-related matters. I appreciated your career advice and willingness to connect me with others in the field. Your depth of knowledge regarding the field of audiology is a model for me in my career.

Lastly, I would like to acknowledge friends and family who have offered me encouragement, support, and the occasional question “So, when do you graduate, again?” over the years. My most devout supporter has been my colleague and wife, Rosemary, who traversed this long trail ahead of me and has always looked back to offer love and reassurance as I find my way.

This research was funded by the National Institutes of Health, National Institute on Deafness and other Communication Disorders (F30 DC01418), The American Speech-Language Hearing Association, The American Academy of Audiology Foundation, and the Graduate and Professional Student Council at the University of Arizona.
DEDICATION

This dissertation is dedicated to family, friends, mentors, and colleagues. Thank you all for your support over the years.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... 7

LIST OF TABLES ............................................................................................................... 8

ABSTRACT ....................................................................................................................... 9

CHAPTER 1: DISSERTATION SPECIFIC AIMS AND ORGANIZATION ..................................... 12

CHAPTER 2: ANATOMY AND PHYSIOLOGY OFafferENT AND EfferENT AUDITORY PATHWAYS: AN OVERVIEW ................................................................................. 14

CHAPTER 3: PRE-NEURAL AND NEURAL ASSAYS OF AUDITORY SYSTEM FUNCTION ......................... 39

CHAPTER 4: OLIVOCOCHLEAR SYSTEM FUNCTION – A FOCUSED LITERATURE REVIEW .......................................................... 59

CHAPTER 5: CONTRALATERAL INHIBITION OF CLICK- AND CHIRP-EVOKED HUMAN COMPOUND ACTION POTENTIALS .......................................................... 71

CHAPTER 6: EfferENT MODULATION OF COMPLEMENTARY PRE-NEURAL AND NEURAL DISTORTION PRODUCTS ........................................................................... 98

CHAPTER 7: EXAMINING HUMAN MOC REFLEX “UNMASKING” IN SUBCORTICAL SPEECH-IN-NOISE ENCODING ........................................................................ 124

CHAPTER 8: SUMMARY AND FUTURE DIRECTIONS ................................................................ 146

REFERENCES .................................................................................................................... 148
LIST OF FIGURES

2.1. Anatomy of Afferent and Efferent Auditory Pathways..........................15
2.2. Cochlear Micrographs........................................................................17
2.3. Organ of Corti Micro-Optical Coherence Tomography.........................20
3.1. Otoacoustic Emissions Coherent Reflection and Distortion Sources......41
3.2. Electrocochleography Waveform Components.....................................47
3.3. Convolution Model of Compound Action Potential Waveform..............51
3.4. Frequency Following Response Morphology........................................53
5.1. Chirp-to-Click CAP Amplitude Ratios................................................83
5.2. Average Chirp- and Click-Evoked CAP Amplitudes...............................84
5.3. Chirp and Click CAP Waveforms with and without CAS.......................85
5.4. Average Level-Series Functions for Chirps and Clicks..........................86
5.5. CAP Normalized Amplitude Reduction...............................................87
5.6. Effective Attenuation of CAPs..............................................................89
6.1. Pre-neural and Neural Correlates of the Two Source Model..................101
6.2. Effects of CAS on DPOAE Fine Structure and FFRs.............................115
6.3. DPOAE Fine Structure Peak Inhibition...............................................116
6.4. FFR Peak Amplitudes in Quiet and with CAS.......................................117
6.5. Relationships Between DPOAE and DPFFR Inhibition..........................118
7.1. Three Types of CEOAE Inhibition Measurement....................................129
7.2. sABR Stimulus and Response.............................................................132
7.3. CEOAE Inhibition by Level and Sex......................................................137
7.4. Average sABR Waveforms for the Four Experimental Conditions.........138
7.5. CEOAE Inhibition and sABR Latency Unmasking Relationships............140
7.6. CEOAE Inhibition and sABR Onset Latency Unmasking Relationships....141
LIST OF TABLES

6.1. DPFFR Stimulus Characteristics and Response Prevalence……………..114
7.1. sABR Condition Parameters and Rationales………………………………134
ABSTRACT

The auditory nervous system contains an extensive and distributed network of efferent pathways connecting auditory cortices to cochleae. At the most caudal level of the efferent auditory system, cochlear outer hair cells (OHCs) receive direct innervation from the auditory brainstem via the medial olivocochlear (MOC) bundle. Through the MOC bundle, the brainstem modulates cochlear amplifier gain – an effect termed the MOC reflex. One putative role of the MOC reflex is improving the signal-to-noise ratio by reducing cochlear gain for noise (i.e., “unmasking”).

The human MOC reflex has been studied using pre-neural assays of OHC function, such as otoacoustic emissions. A limitation of this approach is that it is insensitive to subsequent “downstream” MOC reflex effects on the neural ensembles that mediate hearing. To elucidate the functional role of the MOC reflex, it is imperative to understand relationships between the pre-neural OAE assays of MOC reflex function and their downstream neural complements: compound nerve action potentials and auditory brainstem responses.

The specific aims of this dissertation were to 1) examine predictive relationships between complementary pre-neural and neural assays of MOC reflex function, and 2) test the hypothesis that the human MOC reflex is advantageous in speech-in-noise processing. Three experiments were undertaken to address these aims.
In the first experiment, click-evoked otoacoustic emissions and click- and chirp-evoked auditory nerve compound action potentials were measured with and without activation of the MOC reflex using contralateral noise. We hypothesized that MOC reflex amplitude inhibition of compound action potentials would be larger than otoacoustic emission amplitude inhibition and that compound action potential inhibition would be predicted by otoacoustic emissions inhibition. In the second experiment, distortion product otoacoustic emissions and distortion product frequency following responses were measured with and without activation of the MOC reflex using contralateral noise. We hypothesized that MOC reflex inhibition of distortion product frequency following responses would be larger than distortion product otoacoustic emissions and that distortion product frequency following response inhibition would be predicted by distortion product otoacoustic emission inhibition. In the third experiment, we measured MOC reflex strength using otoacoustic emissions as well as brainstem speech-in-noise processing with and without activation of the MOC reflex. We hypothesized that otoacoustic emission inhibition would predict brainstem speech-in-noise unmasking.

The results of Experiment 1 suggested that compound action potential amplitude inhibition was larger than otoacoustic emission amplitude inhibition when results were reported on the same scale. Further, chirp-evoked compound action potential inhibition was larger than click-evoked compound action potential inhibition, suggesting that chirps may be a better tool for measuring MOC reflex
inhibition of auditory nerve responses. The results of Experiment 2 revealed that distortion product frequency following response inhibition was largest for the component measured at 2f1-f2 than for f1 or f2. Further, distortion product otoacoustic emission inhibition was mildly predictive of distortion product frequency following response inhibition at 2f1-f2 and f2. The results of Experiment 3 revealed that otoacoustic emission inhibition was not predictive of speech-in-noise “unmasking” at the level of the brainstem. Taken together, the experiments suggest that pre-neural inhibition measurements likely underestimate MOC reflex strength and that neural assays may be more beneficial in understanding the functional significance of the MOC reflex in humans.
CHAPTER 1: DISSERTATION SPECIFIC AIMS AND ORGANIZATION

1.1. Specific Aims

Much of what is understood about the neurophysiology of auditory processing has been derived from studying and modeling the *afferent* auditory nervous system, a complex neural network that conveys information from the inner ears to the brainstem and cortex. However, a less extensive body of research in both experimental animals and humans suggests that the *efferent* auditory nervous system, a network descending from auditory cortices to the cochlea, may serve an equally important role in auditory processing.

One specific efferent circuit believed to be involved in the early stages of auditory processing is the olivocochlear bundle in the caudal auditory brainstem. Neurons from the *medial* olivocochlear (MOC) bundle project bilaterally from the brainstem to cochlear outer hair cells (OHCs). Through the MOC bundle, the brainstem modulates cochlear amplifier gain – an effect termed the MOC reflex. One putative role of the MOC reflex based on animal model data is “unmasking” of signals in noise.

The human MOC reflex has been studied using *pre-neural* assays of OHC function, such as otoacoustic emissions. A limitation of this approach is that it is insensitive to subsequent “downstream” MOC reflex effects on the neural ensembles that mediate hearing. To elucidate the functional role of the MOC
reflex, it is imperative to understand relationships between commonly used pre-neural assays of MOC reflex function and their neural complements.

The specific aims of this dissertation were to:

1) Test predictive relationships between complementary pre-neural and neural assays of MOC reflex function

2) Test the hypothesis that the human MOC reflex is advantageous in speech-in-noise processing.

1.2. Organization

The experiments presented in this dissertation were motivated by a large body of literature and enlisted a variety of data collection techniques. Chapters 2-4 provide prerequisite theoretical and methodological context for these experiments. Specifically, Chapter 2 provides an overview of the anatomy and physiology of afferent and efferent auditory systems, Chapter 3 provides an overview of data collection techniques including otoacoustic emissions, electrocochleography, and frequency following responses, and Chapter 4 provides a focused literature review on the proposed functions of the olivocochlear system. Chapters 5-7 correspond to three experiments and are formatted as publication manuscripts each with introduction, methods, results, and discussion sections, respectively. Chapter 8 is a brief synthesis of experimental findings and includes a discussion of future work.
CHAPTER 2: ANATOMY AND PHYSIOLOGY OF AFFERENT AND EFFERENT AUDITORY PATHWAYS: AN OVERVIEW

2.1. Introduction

Making sense of sound is arguably one of the most computationally demanding endeavors of the human brain. In the auditory periphery, sound energy is acoustically and mechanically filtered by the outer and middle ears before it is transduced into a complex spectro-temporal code in the auditory nerve fibers innervating the cochlea. The signals represented in the auditory nerve subsequently initiate cascades of neural activity through parallel and overlapping pathways in the brainstem, midbrain, and cortex. Contrary to classical views of auditory processing, these pathways feed both forward and backward via afferent and efferent tracts, respectively (Figure 2.1). The purpose of this chapter is to provide an overview of the anatomy and physiology of afferent and efferent auditory pathways as a foundation for following chapters. It begins with a peripheral-to-central review of the afferent auditory system and ends with a central-to-peripheral review of the efferent system.
2.2. Afferent Auditory System Anatomy and Physiology

2.2.1. Outer and Middle Ears

The input impedance of the fluid-filled cochlea is approximately 200 times greater than air, meaning that without a mechanism to overcome this mismatch, 98% of sound energy in the environment would reflect away from the inner ear (Nakajima
et al., 2009). The outer and middle ears improve transmission of sound energy from air to fluid by two major mechanisms: passive resonance and impedance transformation.

The convexities and concavities of the pinna as well as the length and diameter of the ear canal have evolved to passively enhance sound energy within certain frequency ranges. Specifically, the combined outer ear resonances increase sound pressure at the level of the eardrum by approximately 15-20 dB for frequencies 2000-8000 Hz (Pickles, 2008). When sound vibrates the eardrum, this vibration is conducted through three bones in the middle ear - the malleus, incus, and stapes - the last of which articulates with the cochlear oval window. The function of the middle ear is to perform impedance transformation, and it does so in two ways. First, the area ratio of the tympanic membrane relative to the stapes footplate is approximately 18/1, resulting in a 25 dB boost in sound energy at the oval window. Second, the arm of the malleus is two times longer than the arm of the stapes, resulting in a lever action that accounts for approximately 6 dB of gain at the oval window. The combined effects of the middle ear transformer are thus approximately 30 dB (Rosowski, 1991). It is interesting to note that the energy sensitivity of the cochlea is relatively constant from 450-10,000 Hz (Pickles, 2008). Therefore, the U-shape of the minimum audible pressure curve (i.e., threshold sensitivity in dB SPL as a function of frequency) reflects the transmission efficiency of the middle ear.
2.2.2. Cochlea

The cochlea is a coiled osseous tube embedded within the petrous portion of the temporal bone. It is comprised of 2.5 turns in humans, with an overall height of 5 mm and width of 1 cm (Pickles, 2008). Within the cochlea are three membranous, fluid-filled scalae that spiral around the modiolus (i.e., “wheel nave”), which carries auditory nerve fibers medially to the brainstem. The scala vestibuli is the superior chamber, which begins at the oval window and ends at the helicotrema. The scala tympani is the inferior chamber which begins at the round window and ends at the helicotrema. The scala vestibuli and tympani are thus conjoined via the helicotrema and are both filled with perilymph. The middle chamber is the scala media, which is filled with potassium-rich endolymph. The membrane separating the scala media from the scala tympani - the basilar membrane – is the floor of the organ of Corti, which houses the receptor cells of hearing (Figure 2.2).

Figure 2.2. Micrographs of (guinea pig) cochlea. The cross-sectional view (a) displays the modiolus (M), scala vestibuli (SV), scala media (SM), and scala
tympani (ST) as well as Reissner's and basilar membranes (RM and BM, respectively). Zoomed-in views show organ of Corti macrostructure (b) and microstructure (c) demonstrating various primary and supporting cell types including auditory nerve fibers (light blue arrow), IHCs (red arrow), and OHCs (blue arrows). *Image from Iyer et al. (2016) reproduced with permission under a Creative Commons CC-BY license.*

The primary role of the organ of Corti is to convert vibrations conveyed to the cochlea by the middle ear into neural impulses. Thus, it must act both as a mechanical frequency analyzer that decomposes sounds into their constituent frequencies and a dynamic compressor that accurately represents an input range of approximately 120 dB. Passive mechanical frequency analysis is achieved by the basilar membrane, which is thick, narrow, and stiff at the base and thin, wide, and flaccid at the apex. Because of this stiffness gradient, sound frequency is mapped onto tonotopic place in the form of traveling waves such that high frequencies are represented in the base and low frequencies in the apex. The frequency response of the human cochlea is approximately 20-20,000 Hz (Ruggero & Temchin, 2002).

Traveling wave peaks are larger and sharper in a living cochlea relative to a postmortem model, which is presumably due to the cycle-by-cycle amplification characteristics of outer hair cells (OHCs). Specifically, when OHC receptor potentials change cyclically during basilar membrane vibration, they contract and elongate (Brownell et al., 1985) due to a motor protein, prestin, that is densely represented in the lateral walls of OHCs (Zheng et al., 2000). OHC electromotility
also allows a large dynamic range to be compressed, as low input stimuli are amplified and this response decreases with increased input level.

Inner hair cells (IHCs) are the sensory receptors of hearing. Like OHCs, IHC receptor currents fluctuate cyclically based on the rate of basilar membrane vibration. Depolarization of IHCs results in the release of glutamate onto postsynaptic auditory nerve fiber dendrite terminals (Pickles, 2008). Unlike OHCs, IHCs express specialized ribbon synapses, which anchor vesicles to the active zone and support more rapid and synchronized neurotransmitter release. Abnormal expression of otoferlin, a protein associated with ribbon synapses, has been implicated in both auditory neuropathy and deafness (Roux et al., 2006; Santarelli et al., 2009). Importantly, without OHC amplification, IHCs would require an input stimulus intensity greater than approximately 50-60 dB to be driven to depolarization. While the mechanisms of IHC stimulation are still poorly understood (Guinan, 2012), the relationship between OHC damage and sensorineural hearing loss is evidence that OHCs are crucial intermediaries for low- to moderate-level input sounds.

2.2.3. Auditory Nerve

The auditory nerve is comprised of approximately 30,000 bipolar nerve fibers, 95% of which synapse with IHCs in the organ of Corti (Musiek & Baran, 2007). The other 5%, known as type II fibers, diffusely innervate multiple OHCs, and their
function is poorly understood.¹ A single auditory nerve fiber synapses at the base of one IHC; however, one IHC may synapse with up to 20 fibers (Spoendlin & Schrott, 1989). The dendrite of each auditory nerve fiber courses medially from its synapse through a small cribiform opening in the osseous spiral lamina (habenula perforata), and the cell bodies of many fibers cluster in Rosenthal's canal to form the spiral ganglia (Figure 2.3). Auditory nerve fibers from the entire length of the organ of Corti coalesce in the cochlear modiolus and exit the temporal bone via the internal auditory meatus as the auditory nerve. The frequency tuning of each auditory nerve fiber is dependent upon the tonotopic location of its presynaptic IHC, with basal and apical fibers representing high and low frequencies, respectively.

¹ New research indicates that Type II fibers may be involved in a non-canonical “pain” pathway signaling noise-induced tissue damage in the cochlea (Flores et al., 2015). This has interesting implications for hyperacusis, as a common complaint is sound-induced otalgia.
Auditory nerve fibers encode frequency in both rate-place and temporal codes (Møller, 2000; Pickles, 1988). The rate-place code conveys information about where motion is happening along the basilar membrane, whereas the temporal code describes how often (i.e., the periodicity of motion). Thus, both mechanisms provide information about the frequency content of sound. Because the temporal code is phase-locked to depolarizing cycles of IHCs, it provides precise periodicity information up to a certain rate. Most auditory nerve fibers can phase-lock to every cycle of the stimulus up to ~0.8 kHz (Pickles, 1988), and a smaller population can discharge up to 5 kHz (Møller, 2000). At higher frequencies, temporal discharge patterns from groups of neurons pool together to collectively code stimulus or envelope periodicity (Hanekom & Kruger, 2001). When stimuli surpass 4-5 kHz, IHC receptor potentials are dominated by a DC component, and essentially remain depolarized (with minor fluctuations in the AC component) throughout the stimulus. Thus, auditory nerve fibers fire in a sustained fashion and therefore only provide frequency information in a rate-place code within this range. Importantly, phase-locking of auditory nerve fibers to steady state, time-variant speech sounds has also been reported (Miller & Sachs, 1983, 1984; Sachs & Young, 1983; Young & Sachs, 1979).
2.2.4 Cochlear Nucleus

The cochlear nucleus (CN) is the caudal-most nucleus complex in the auditory brainstem pathway. It is located anteriorly to the cerebellum in the posterolateral aspect of the pontomedullary junction of the brainstem. From a transverse view of the brainstem, the CN is cashew-shaped with its concave arc facing anteromedially and its dorsomedial wall bordering the fourth ventricle. Three subdivisions, each with complete tonotopic maps, comprise the CN: anterior ventral (AVCN), posterior ventral (PVCV) and dorsal (DCN). At least eight distinct cell types have been reported in the CN, and each type appears to cluster within particular functional subdivisions of the complex. Spherical bushy, globular bushy, and stellate (multipolar) cells are primarily located in the AVCN. Located in the PVCN are stellate, granular, giant, and octopus cells. Pyramidal, giant, small, and granular cells are predominant in the DCN. As auditory nerve fibers enter the CN via the root entry zone at the junction of the AVCN and PVCN, they bifurcate ventrally and dorsally and synapse with all major cell types in each subdivision. Importantly, differences in CN cell structures facilitate the encoding and augmentation of features represented in the auditory nerve.

A description of the response characteristics of each CN cell type is beyond the scope of this review. However, it is pertinent to mention the response patterns of certain cells that preserve and enhance the phase-locking characteristics of the auditory nerve given the techniques used in Chapters 6 and 7. As a general rule,
cells within the DCN respond better to spectral information (i.e., they exhibit sharp tuning curves with inhibitory sidebands), whereas the cells of the AVCN and PVCN respond better to temporal information. Near-field recordings from the AVCN and PVCN indicate that many neurons within these subdivisions can phase-lock up to 3-4 kHz (Moller, 2000). The cell types most likely responsible for precise phase-locking in the CN are bushy, octopus, and stellate (Moller, 2000; Romand & Avan, 1997). Like the auditory nerve, the PSTHs of bushy cells are primary-like; therefore, at low frequencies, they demonstrate phase-locked discharges clustered at intervals equal to the stimulus period (Winter, 2006). Bushy cells also track fundamental frequency of vowel sounds (Recio & Rhode, 2000). Octopus cells exhibit an “on” response (i.e., an abrupt discharge peak in the PSTH at the onset of a stimulus), which is highly sensitive to the onset timing of the input stimulus. They have elongated, extensive dendrites (like the arms of an octopus) that synapse with and integrate inputs from many auditory nerve fibers. Because of their large size and low input resistance, octopus cells work as coincidence detectors (Oertel et al., 2000) that are highly sensitive to “pooled” periodicity; that is, only synchronized discharges across many auditory nerve fibers can initiate an action potential in octopus cells. Stellate cells demonstrate “chopper” responses to periodic sounds; however, these cells can only respond to frequencies of ~ 1 kHz or below.
Information from the AVCN, PVCN, and DCN exits through the routes of the ventral, intermediate, and dorsal acoustic stria, respectively. The ventral acoustic stria travels along the ventrolateral border of the brainstem and projects to ipsilateral (~25% of fibers) and contralateral (75% of fibers) superior olivary complexes (SOCs). The intermediate and dorsal acoustic stria project mainly to the contralateral lateral lemniscus and inferior colliculus.

### 2.2.5. Superior Olivary Complex

The three major auditory structures of the superior olivary complex are the medial superior olive (MSO), lateral superior olive (LSO), and medial nucleus of the trapezoid body (MNTB). These structures lie ventromedially to the cochlear nuclei, and importantly, are the first sites of binaural processing in the auditory system. In humans, the MNTB is described as a fiber tract carrying axons from the contralateral AVCN to the ipsilateral SOC. The LSO is an “S” shaped nucleus (at least in cats) lying dorsolaterally to the elongated nucleus of the MSO; both nuclei contain their own tonotopic maps. Multipolar neurons in the MSO receive inputs from bilateral CN and have complex activation patterns; approximately 65% of MSO neurons are excited by stimuli in either ear, 25% are excited by stimuli in one ear and inhibited by stimuli in the opposite ear, and the remaining 10% are only excited by monaural sound (Musiek & Baran, 2007). The PSTHs of MSO fibers are primary-like and primary-like with notch. These complex activation patterns and firing characteristics are presumably why the MSO is highly sensitive
to interaural time differences to low frequency stimuli (Galambos et al., 1959). Animal studies indicate that neurons within the MSO can phase lock up to 2-3 kHz and likely are the source of the binaural masking level difference (Rouiller, 1997). Nearly all neurons in the LSO are excited by broadband stimuli presented to one ear and inhibited by the other. The PSTHs of LSO neurons are chopper- and primary-like. It is likely that the neurons of the LSO form the basis for interaural level differences (Musiek & Baran, 2007). Importantly, phase-locking is sharpened by the excitatory/inhibitory patterns of the SOC. Most fibers projecting from the MSO innervate the ipsilateral central nucleus of the inferior colliculus. Low frequency fibers from the LSO project ipsilaterally to the inferior colliculus where they are inhibitory. High frequency fibers project contralaterally to the inferior colliculus where they are excitatory.

The LSO and MSO disproportionally represent high and low frequencies, respectively (Cant & Casseday, 1986). Interestingly, the morphology of each structure across species is indicative of auditory sensitivity; for example, MSO size is related to animals’ abilities to localize low and mid-frequency signals (Masterton et al., 1975).

2.2.6. Lateral Lemniscus & Inferior Colliculus

The lateral lemniscus is an ascending fiber tract comprised of projections from the contralateral CN that combine with ipsilateral ascending SOC fibers (Musiek &
Baran, 2007). This fiber tract surrounds two nucleus groups, the ventral nucleus of the lateral lemniscus (VNLL) and dorsal nucleus of the lateral lemniscus (DNLL). The VNLL is elongated, receives inputs from the contralateral AVCN, and is located more caudally than the DNLL. The DNLL receives bilateral inputs from the LSO, ipsilateral inputs from the MSO, and is immediately caudal to the inferior colliculus (IC). The DNLL of each side of the brainstem is connected via the commissure of Probst. The multipolar and globular cells of the VNLL have excellent phase-locking ability and appear to be sensitive to interaural time differences, whereas cells in the DNLL are more sensitive to interaural intensity differences.

The IC is a pearl-shaped structure on the dorsal aspect of the midbrain, immediately rostral to the DNLL. This nucleus is an obligatory stop for all ascending auditory information, and thus receives inputs from contralateral CN, SOC, and DNLL and ipsilateral LSO, MSO, DNLL, and VNLL. There are three main divisions of the IC: central nucleus (ICC), dorsal cortex (DCIC), and lateral zone (LZIC). The ICC is a spherical nucleus that is covered dorsally by the DCIC and laterally by the LZIC. The disc-shaped and stellate cells of the IC are organized such that they form tonotopic isofrequency sheets. Near field recordings from neurons within the isofrequency sheets of the IC indicate strong phase-locking up to 0.6 kHz and high sensitivity to amplitude modulation (Joris et al., 2004). Neurons in the ICC have a strong preference for frequency modulated
and binaurally phase-shifting stimuli, respectively, suggesting that they may be involved in detecting movement (Escabi & Schreiner, 2002; Pollak, 2012). Notably, the electrical fields of LL and IC neurons form rostral-caudal dipoles, making their summed contributions to EEG recordings more robust at the vertex.

2.2.7. Commentary on Functional “Streams” in the Subcortical Afferent Auditory System

As described in the sections above, there are highly specialized neuron types in each of the nuclei comprising the auditory brainstem. Based on similarities between neuron response properties and their anatomical configurations within the brainstem, some anatomists have referred to two subcortical auditory processing “streams” (Pickles, 2008). The ventral auditory stream consists of cell types and nuclei that primarily analyze temporal information, whereas the dorsal auditory stream extracts spectral information. Both streams are likely initially combined in the IC and are further integrated in the thalamus and auditory cortex.

2.2.8. Medial Geniculate Body

The medial geniculate body (MGB) is the primary thalamic auditory hub, receiving input from the IC and sending outputs to auditory cortex. Based on the complex response properties of MGB neurons, it has been suggested that spatial, spectral, and temporal information from separate auditory streams is combined into “auditory objects” (i.e., feature groupings providing the perceptual basis of a
sound’s origin) in the MGB (Pickles, 2008). The MGB is comprised of ventral, dorsal, and medial divisions. The ventral MGB receives primary input from the ICC and projects to the ipsilateral auditory cortex. The dorsal MGB receives inputs from the DCIC and somatosensory system and projects to auditory cortex belt areas. The medial division likely plays a role in multisensory integration, as it receives inputs not only from the ICC, but also from vestibular and visual systems, and the spinal cord. The medial MGB also projects diffusely to the auditory cortex and to the amygdala, which serves a critical role in emotional responses and associations to sound.

The ventral MGB is organized in a laminar structure with low frequencies represented in the dorsal area and high frequencies represented in the ventral area (Cetas et al., 2001). The laminae are inter-digitated, forming local feedback loops between excitatory and inhibitory cell types. Unlike the clear tonotopy expressed in earlier auditory nuclei, frequency tuning maps are discontinuous in the MGB. Within each lamina are functional “slabs” that represent common frequencies, and within each slab, cells are locally organized based on their input preferences (e.g., binaural, ipsilateral, contralateral). Reaching within and orthogonally across isofrequency planes are “patches” of neurons that respond preferentially to complex combinations of stimulus features (binaurality and frequency, for example); MGB patchiness is fed-forward to the functional organization of auditory cortex (Velenovsky et al., 2003). The ventral MGB and
auditory cortex are bi-directionally coupled, suggesting that they behave more as a functional unit than two distinct entities (Suga et al., 1997).

The stimulus-specific adaptation properties of neurons found in the medial and dorsal MGB suggest that these “extra-lemniscal” centers respond to auditory novelty (Bauerle et al., 2011) and may be part of a pre-attentive alerting pathway. The medial MGB is part of an associative learning network that includes the amygdala, as fear conditioned responses to sound are abolished by ablation of this division of the MGB (Duvel et al., 2008; Halverson, et al., 2008). Thus, medial and dorsal MGB divisions are components in a multisensory pathway that is highly modifiable by environmental factors.

2.2.9. Auditory Cortex

The auditory cortices are situated on the superior temporal planes within the Sylvian fissures of left and right cortical hemispheres and are reciprocally connected via the corpus callosum. Based on histologic and physiologic data, the auditory cortex can be broadly divided into three anatomical areas: core, belt, and parabelt. Pyramidal neurons in core areas primarily receive inputs from the ventral MGB in layer IV, whereas the surrounding belt and parabelt areas receive inputs from the extralemniscal divisions of the MGB. While the belt also receives inputs from the core and vice-versa, the parabelt has few, if any, connections with the core (Hackett et al., 1998). The parabelt is diffusely and reciprocally connected to
the frontal lobe, including the frontal eye fields, suggesting that information regarding auditory and visual perception of movement is integrated here.

Detailed descriptions of core, belt, and parabelt subdivisions are beyond the scope and intent of this review; however, a few features of the auditory cortices are pertinent to the broader theme of this dissertation (i.e., that the auditory system is bi-directional and dynamic). Firstly, there are multiple tonotopic maps in the primate auditory cortex core and belt subdivisions, which are organized caudo-rostrally in “frequency-band strips”. This organization is more diffuse in the belt (Barton et al., 2012). Neurons within these frequency-band strips show large variation from very sharp tuning with side band inhibition to broad frequency selectivity over many adjacent frequencies. Additionally, some core and belt neurons demonstrate sensitivity to binaurality/monaurality, frequency and amplitude modulation, and changes in interaural timing and level differences (Stecker et al., 2003). Neurons in the belt respond to more complex acoustic stimuli than the core and may be organized in coincidence detection networks that combine multiple features of sound (Winer, 2006). Importantly, response properties and connectivity in the auditory cortex are continuously adjusted by behavioral demands and experience (Chait et al., 2012; Headley & Weinberger, 2013); this also suggests that corticofugal effects mediated by the efferent system (discussed in Section 2.3) are also modifiable. Secondly, hemispheric specialization of right and left auditory cortices has been reported, although the
basis of these differences continues to be a topic of debate. Based on a large body of fMRI, electrophysiologic, and neurological lesion research, it has been proposed that the left auditory cortex is specialized for temporal resolution, whereas the right cortex is specialized for spectral resolution (see Zatorre et al., 2002 for a comprehensive review). A closely related hypothesis is that the effective sampling rates of left and right auditory cortices differ, with the left sampling at a much higher rate (25-50 Hz) than the right (4-8 Hz), allowing for fine structure and global aspects of sound to be analyzed individually (Hickok & Poeppel, 2007). This too has implications for how the corticofugal system modifies subcortical information, and hemispheric specialization may even be reflected in the inner ear (discussed in Chapter 4).

2.3. Efferent Auditory System Anatomy and Physiology

2.3.1. Corticofugal Pathways

Descending neural tracts from auditory cortex layers V and VI to MGB, IC, SOC, and CN have been identified using tracer, microstimulation, and cryoloop cooling techniques in experimental animal models (e.g., Mellott et al., 2015; Suga & Ma, 2003; Terreros & Delano, 2015). Out of all subcortical auditory nuclei, the MGB receives the most descending projections from the auditory cortex. These projections form ipsilateral tonotopic feedback loops between the cortex and thalamus (Winer & Lee, 2007). Descending cortical fibers innervating the IC are mostly ipsilateral; however, contralateral fibers have also been observed (Bajo et
As with descending pathways from cortex to MGB, the cortico-collicular pathways are also tonotopic. IC response modulation to sound location, duration, and intensity by the cortex have been described (Ma & Suga, 2001; Yan & Ehret, 2002; Zhou & Jen, 2005). In addition to the MGB and IC, the auditory cortex also connects to the SOC and CN both through direct projections and indirect projections via the IC (Mulders & Robertson, 2000). The SOC and CN are components of a well-studied reflex loop, the olivocochlear bundle, that modulates OHC function; therefore, it has been suggested that the auditory cortex may exert control over cochlear function by commandeering the olivocochlear system (Winer, 2006).

2.3.2. The Olivocochlear System

*Medial Olivocochlear Bundle*

Medial olivocochlear (MOC) neuron perikarya are relatively large (7-26 μm), multipolar, and have extensive dendritic branches (Sahley et al., 1997; Warr, 1992). Thick MOC axons (0.5-2.8 μm) comprise the terminal effector pathway of the MOC reflex, which modulates OHC electromotility. Uncrossed MOC fibers originate in the pre- and periolivary nuclei of the medial superior olivary complex ipsilateral to the cochlea that they innervate. Crossed MOC fibers originate in the pre- and periolivary nuclei of the medial superior olivary complex contralateral to the cochlea that they innervate. The axons of the crossed MOC fibers travel dorsomedially from the superior olive and traverse the midline of the brainstem at
the ventral border of the fourth ventricle before curving laterally. The axons of the uncrossed MOC fibers travel dorsally and coalesce with the crossed fibers to form what is generally called the olivocochlear bundle (OCB) proper. The OCB proper travels along (and sends interneurons into) the medial perimeter of the cochlear nucleus (CN) before exiting anterolaterally as part of the inferior vestibular branch of the eighth nerve. The terminal ends of MOC neurons diverge from the inferior vestibular branch in the internal auditory meatus and enter the cochlear modiolus with the auditory branch. They then travel laterally through the habenula perforata and tunnel of Corti to synapse directly onto OHCs.

Cell bodies of the crossed and uncrossed MOC pathways are respectively innervated by the CN (specifically, the AVCN and PVCN) contralateral to them via axons traveling through the ventral acoustic stria. Therefore, the MOC reflex is a double-crossed circuit; the crossed MOC fibers mediate the ipsilateral reflex (i.e. MOC activation and OHC effects are in the same ear), whereas the uncrossed MOC fibers mediate the contralateral reflex (i.e., MOC activation is in the opposite ear from OHC effect). The ventral acoustic stria projects to MOC fibers in a tonotopic fashion, therefore MOC fibers have tuning curves almost as sharp as auditory nerve fibers (Liberman & Brown, 1986). The terminal ends of MOC fibers project to areas of the organ of Corti within 0.1-1 octave of their own center frequencies (Guinan, 2006). Despite this precise tonotopicity, MOC fibers appear to integrate inputs over large frequency ranges, as narrowband elicitor noise...
placed up to two octaves above and below a probe tone can inhibit neural responses to the tone (Lilaonitkul et al., 2002).

In small animals, the ratio of crossed to uncrossed MOC fibers is 3:1 (Venicia et al. 2005; Warr, 1992). Consequently, the ipsilateral MOC reflex is up to three times stronger than the contralateral MOC reflex (Liberman, 1989). The preponderance of crossed and uncrossed fibers in humans is unknown but is approximately 1:1 in primate models (Guinan, 2006).

Most of what is understood about MOC physiology has been gleaned from invasive electrophysiologic recordings and pharmacologic studies in animal models. In the former method, recordings from OHCs, IHCs, and auditory nerve fibers are made with and without acoustic (Liberman, 1988a, 1988b; Liberman & Brown, 1986; Robertson & Gummer, 1985) or electric (Galambos, 1956; Gifford and Guinan, 1983, 1987) MOC activation. One way that single MOC neurons are identified in a physiologic preparation is that their discharge rates in response to steady state noise are extremely periodic, whereas afferent fibers discharge at random (Brown, 2014). In the latter method, the efferent system is activated or deactivated by perilymphatic chemical perfusion (e.g., Bobbin & Konishi, 1971, 1974; Comis & Leng, 1979; Gisselsson, 1960; Konish, 1972; Kujawa et al., 1992, 1993, 1994; Roberston & Johnstone, 1978).
Biochemistry of the MOC Bundle

When MOC neurons are stimulated electrically or acoustically, they release acetylcholine (ACh) into the synaptic cleft (Elgoyhen et al., 2001; Sridhar et al., 1997). ACh binds to α-9 nicotinic receptors on OHC membranes, which results in an influx of Ca$^{2+}$ into the intracellular space. The fast effect of Ca$^{2+}$ influx is that $K_{Ca}$ channels are opened and $K^+$ travels down its concentration gradient into the extracellular space. Consequently, membrane conductance quickly increases (on the order of ~ 100 ms) and the OHC becomes hyperpolarized. This fast effect adapts over time and is quickly reversible with the cessation of MOC activation. With sustained MOC activation, ACh remains in the synaptic cleft, which keeps the α-9 nicotinic receptors open and allows Ca$^{2+}$ to remain in the intracellular space. The lingering Ca$^{2+}$ binds to ryanodine receptors in the subsurface cisternae of OHCs and releases more Ca$^{2+}$ into the intracellular space. This additional Ca$^{2+}$ activates maxi-$K_{Ca}$ channels embedded in the lateral wall and allows even more $K^+$ to leave the cell. Because this second process requires sustained MOC activation on the order of tens of seconds, it is termed the slow effect. The slow effect is reversible on the order of tens of seconds after cessation of MOC activation activation. Based on differences in basilar membrane phase-lag during fast and slow effects, it is assumed that each alters the OHC in a different way: the fast effect mainly hyperpolarizes OHCs, while the slow effect actually reduces their stiffness (Cooper & Guinan, 2003). Most assays of the MOC reflex use either sustained noise or shocks, which engages fast and slow effects...
together (Guinan, 2006). Interestingly, drugs that block the re-uptake of Ca\(^{2+}\) into the subsurface cisternae of OHCs prolong the slow effect because Ca\(^{2+}\) remains in the intracellular space and continues to bind to maxi-K\(_{\text{Ca}}\) channels; thus, intratympanically delivered calcium re-uptake inhibitors may serve as a protective agent against acoustic trauma by prolonging the slow effects of MOC fibers (Sridhar et al., 1997).

*Lateral Olivocochlear Bundle*

LOC neuron perikarya are small (5-25 μm) and either round or fusiform in shape with two to five dendrites (Sahley et al., 1997; Warr, 1992). Their axons are thin (0.3-0.7 μm) and unmyelinated. Unlike MOC fibers, the functional role of LOC fibers in hearing has not been clearly delineated (Guinan, 2014; Liberman et al., 2014). Uncrossed LOC fibers originate in the marginal nuclei (dorsolateral hilus, dorsal and dorsolateral periolivary nuclei, and lateral nucleus of the trapezoid body) of the lateral SOC ipsilateral to the cochlea that they innervate. Crossed LOC fibers originate from the marginal nuclei of the lateral superior olivary complex contralateral to the cochlea that they innervate. The axons of the crossed and uncrossed LOC fibers follow the same trajectory in the brainstem as crossed and uncrossed MOC fibers, respectively, and join the OCB proper. Unlike MOC fibers, LOC terminals do not synapse directly onto hair cells after entering the habenula perforata; rather, they synapse onto type I afferent nerve fibers beneath IHCs (Guinan, 2006).
The innervation pattern of LOC cell bodies differs from that of the MOC system. Cell bodies of uncrossed LOC fibers are innervated by the (relatively) ipsilateral AVCN and PVCN, whereas crossed LOC fibers receive innervation from the (relatively) contralateral AVCN and PVCN via the ventral acoustic stria. Like MOC fibers, the LOC fibers are tonotopically organized; however, they appear to innervate the organ of Corti more diffusely than MOC fibers (Musiek & Baran, 2007). Nearly all LOC fibers (~95%) are uncrossed in small mammals and primates, meaning they innervate the ipsilateral ear (Warr, 1980). Low spontaneous rate auditory nerve fibers, which aid in supra-threshold hearing, receive double the LOC innervation as high spontaneous rate, low threshold fibers; this arrangement suggests that tonic LOC activity regulates auditory nerve fiber spontaneous activity (Liberman, 1988).

LOC neurons are difficult to directly assay given that they are small in diameter and lack myelin (Guinan, 2006; Liberman et al., 2014). It has been demonstrated that the fibers can be activated by sound (Adams, 1995), but there is no direct evidence of a sound-evoked LOC reflex per se (Guinan, 2014). Electrically stimulating LOC fibers can increase or decrease the firing rate of type I neurons, and the time course of activation is very slow relative to MOC fibers - on the order of minutes (Comis, 1970; Groff & Liberman, 2003). There is evidence that the dominant neurotransmitters employed by the LOC system are endogenous opioid
peptides (enkephalins and dynorphins) and dopamine, both of which can cause excitatory and inhibitory effects on type I fibers (Darrow et al., 2006; Garret et al., 2010). Because the active neurotransmitter differs between MOC and LOC systems, chemical manipulation of the LOC system may be the most specific way to study this division of the caudal efferent system.

2.4. Conclusion

In the auditory periphery and subcortex, complex sounds are decomposed, and their constituent features (e.g., frequency composition, loudness) are represented in specific neuron types within functional neural “streams”. These streams are combined first in the midbrain and further in the cortex, where multiple tonotopic maps and hemispheric laterality allow for optimized processing of temporal and spectral features. Contrary to classical views of the auditory nervous system, this process is bi-directional, as afferent and efferent auditory pathways interact at virtually every level. Pertinent to the experiments carried out in this dissertation, the olivocochlear pathway from the brainstem to cochleae modulates OHC and auditory nerve function. It may therefore be a potent mechanism by which input into the auditory nervous system is shaped either reflexively or under corticofugal control.
CHAPTER 3: PRE-NEURAL AND NEURAL ASSAYS OF AUDITORY SYSTEM FUNCTION

3.1 Introduction

The purpose of Chapter 4 is to provide a broad theoretical and methodological overview of the objective techniques used in the experiments presented in Chapters 5-7. Two categories of assessment, pre-neural and neural, are considered. Pre-neural assays indirectly probe OHC function, whereas the neural assays discussed in this chapter probe auditory nerve and brainstem function. Because the pre-neural and neural assays presented here are both influenced by OHC activity, they may provide complementary and non-invasive measurements for studying MOC reflex effects in humans. This chapter does not discuss efferent effects on each type of measure, as that is the focus of the experiments reported in Chapters 5-7.

3.2. Pre-neural Assays of Auditory System Function

3.2.1. Otoacoustic Emissions

OAEs are sound waves generated by the inner ear and recorded in the ear canal (Kemp, 1978). It has long been hypothesized that mammalian OAEs are byproducts of active mechanical processes involved in traveling wave amplification. This hypothesis is supported by the disappearance of OAEs after
asphyxiation or death, the absence of OAEs in preston knockout mice\(^2\), and the absence of OAEs in animals and humans with moderate or greater sensorineural hearing loss (Bergevin, 2007; Brownell, 1990; Cheatham et al., 2004; Dallos et al., 2008; Glattke & Robinette, 2002; Liberman et al., 2002).

3.2.1.1. Human Model of OAE Generation

*Basilar Membrane Tonotopy and the Traveling Wave*

As discussed in Chapter 2, basilar membrane tonotopy arises from a basal-apical stiffness gradient such that higher frequencies are represented in the cochlear base and lower frequencies in the apex (Pickles, 2008). The basilar membrane stiffness gradient has two important consequences on traveling wave patterns and therefore in the origination of OAEs. The first is that traveling waves occurring at the basal end of the cochlea have a much higher velocity (~ 100 m/s) than those occurring at the apex (~ 2 m/s; Dallos, 1996; Zwislocki, 2002). The second is that frequency scaling of the mammalian basilar membrane is exponential, meaning that as octaves increase from apex to base, the membrane area representing each octave remains relatively constant (Shera & Guinan, 1999, 2007). Consequently, the number of wavelengths accumulated by a traveling wave is nearly independent of center frequency. This is demonstrated by the fact that, although a high frequency tone reaches its traveling wave peak before a low frequency tone (due

---

\(^2\) Prestin is a motor protein that is densely expressed in OHC lateral walls. Although preston knockout mice have intact OHCs, they lack the requisite protein supporting electromotility.
to basal-apical differences in velocity), both traveling waves will have accumulated the same number of wavelengths at their peaks. The accumulated number of wavelengths at the center frequency is typically expressed as “phase-lag” relative to the input stimulus. The near independence of phase-lag from center frequency has important consequences for how OAE “coherent reflection-” and “distortion-” sources (discussed below) are generated.

Figure 3.1. Coherent reflections (left) and distortions (right) are the two main sources of OAEs. Coherent reflections arise from place-fixed perturbations along the basilar membrane, whereas distortions are wave-fixed and move with the traveling wave. In this example, coherent-reflections and distortions are considered for Traveling Wave 1 (black) and Traveling Wave 2 (gray). See text for more detail.

Coherent Reflection Sources

In von Bekesy’s box model of the basilar membrane (1947), the passive traveling wave built up smoothly from the base, peaked at its center frequency place, and quickly dissipated. In contrast, traveling waves in the living mammalian cochlea
encounter “pre-existing perturbations” as they build up to the center frequency place (Shera & Guinan, 1999). These perturbations are place-fixed entities that coherently reflect (or “scatter”) traveling wave energy in both directions, creating additional reverse and forward traveling waves that were not present in the incoming stimulus. Pre-existing perturbations are theoretically conceptualized as passive or active micromechanical irregularities in the cytoarchitecture of the organ of Corti, which may arise from OHC number, geometric position, and the relative forces generated by differing densities of prestin molecules within a given region (Lonsbury-Martin et al., 1988; Shera & Guinan, 2007).

As mentioned above, the phase-lag pattern of a traveling wave is nearly frequency independent, with similar phase accumulation occurring at the traveling wave’s center frequency, regardless of its position along the basilar membrane. However, shifting the center frequency of a traveling wave around a single place-fixed perturbation will significantly affect the phase-lag of the coherent reflection (Figure 3.1, right). For example, assume that the tail of “Traveling Wave 1” passes through a place-fixed perturbation “P”, and at this point in the tail, the (arbitrary) phase-lag of the traveling wave is 2 cycles. The coherent reflection of the traveling wave caused by “P” will send a reverse-propagated wave toward the stapes that will eventually be recorded in the ear canal as an OAE. Now assume that the peak of “Traveling Wave 2” is centered over “P”, and at this point the phase-lag of the traveling wave is 6 cycles. Again, the coherent reflection of the traveling wave at
“P” will send a reverse-propagated wave to the stapes and eventually will become an emission. The relative phase-lag difference between the two emissions would be 4 cycles, and this phase difference would continue to increase as sequential traveling waves (3, 4, 5, etc.) moved basally away from “P”. Coherent reflection-sources, therefore, have a rapidly rotating phase as a function of stimulus frequency (Shera & Guinan, 1999, 2007). In the real mammalian cochlea, there are several “Ps” within spatially broad regions of the traveling wave. Collectively, these sources generate many wavelets that constructively and destructively interfere to form OAEs, but the principle of coherent reflection described above similarly applies.

**Distortion Sources**

While some perturbations are pre-existing and place-fixed along the basilar membrane (as was “P” in the example above), others are induced by the traveling waves themselves (Shera & Guinan, 1999). In the mammalian cochlea, mechanical nonlinearities near the peak of a single traveling wave can scatter stimulus wave energy in both directions along the basilar membrane. Because these distortion-sources (“D” in Figure 3.1) move with the traveling wave, their phase-lag gradient changes very little with shifts in stimulus frequency; again, this is attributable to the scaling symmetry of the basilar membrane described above. Two traveling waves can also overlap to generate areas of intermodulation distortion, which induces “new” waves that radiate in both directions. When the
spacing between the two traveling waves remains at a fixed ratio, the amount of overlap will also remain constant. As is discussed in the section on distortion product OAEs below, an optimal spacing ratio between two tones has been determined empirically in the mammalian cochlea, at least for primaries above 1000 Hz (Harris et al., 1989; Knight & Kemp, 2001).

3.2.1.2. Measurement-based Taxonomy of Human OAEs:

The two OAE source types described above (coherent reflection and distortion) do not necessarily occur in a mutually exclusive manner. The following section provides a description of how coherent reflection- and distortion-sources respectively contribute to each “type” of OAE used in the experiments reported in Chapters 5-7.

Transient Evoked OAEs

Click evoked OAEs (CEOAEs) are measured in a post-stimulus time window (~20 milliseconds) immediately following the presentation of a brief (100 μs) click (Kemp, 1986). The CEOAE waveform that emerges in the ear canal within this post-stimulus epoch represents the reverse-propagated impulse response of the basilar membrane with the basal end responding first, followed by more apical regions. Interestingly, the spectrum contour of the CEOAE is highly regular (i.e.,

---

3 Shera & Guinan (1999) recommended using a mechanism-based taxonomy to explain how different source combinations contribute to different types of OAEs; however, for this review, a measurement-based taxonomy will be used for clarity.
it is smooth), even though the spacing of perturbations along the basilar membrane at different frequencies is presumably irregular. This “filtering effect” is thought to occur because the size of the traveling wave peaks relative to the spacing of perturbations is quite large (Shera & Guinan, 2007). CEOAE phase-lag characteristics as a function of frequency are consistent with coherent reflection-sources in that they rotate rapidly with changes in stimulus frequency.

**Distortion Product OAEs**

Distortion product OAEs (DPOAEs) are obtained from healthy ears during simultaneous presentation of two primary tones (f1 and f2, respectively). Empirical experiments in humans have demonstrated that the optimal ratio of f2/f1 is ~ 1.2 (Harris et al., 1989; Knight & Kemp, 2001), and the optimum presentation levels of f1 and f2 are 65 dB SPL and 55 dB SPL, respectively. In the cochlea, the f1 and f2 traveling wave envelopes overlap and interact to create an intermodulation distortion-source slightly basal to the f2 place at the frequency 2f1-f2. This distortion-source generates both forward- and reverse-propagating waves. The reverse-propagating wave vibrates the stapes and is emitted into the ear canal. The forward-propagating wave travels apically to the 2f1-f2 place where it is coherently reflected toward the stapes and is eventually also emitted.

---

4 While there are other distortion products created by intermodulation distortion, this summary will focus on the cubic distortion product at 2f1-f2.
The recorded DPOAE is a mixture of coherent reflection- and distortion-sources (Shera & Guinan, 1999). When the primary tones (and consequently, 2f1-f2) are frequency swept, it becomes apparent in the DPOAE “fine structure” spectrum that the coherent reflection- and distortion-sources of 2f1-f2 have different rates of phase change that constructively and destructively interfere to form peaks and troughs. The distortion-source component of the DPOAE fine structure can be isolated by placing a suppressor tone at the 2f1-f2 place, thus removing the reflection component. When this is done, frequency-shifting the primary tones has little effect on the phase-lag of the distortion-source emission due to frequency scaling symmetry, and the fine structure flattens out (Shera & Guinan, 2007).

3.2.1.3. OAE Stimulus-Response Characteristics

Transient-evoked OAEs (a term including clicks, chirps, and tone bursts) and DPOAEs can be reliably measured using stimuli presented from ~ 40-80 dB SPL (Robinette & Glattke, 2007); below 40 dB SPL, responses are too small to be measured above the noise floor and above 80 dB SPL are contaminated by stimulus artifact or residual “ringing”. Growth functions of transient-evoked OAEs are compressive, indicating greater OHC amplification at lower stimulus levels. DPOAE growth functions are more complex, and can be compressive, monotonic, or diphasic (Elsisy & Krishnan, 2008; Nelson & Kimberley, 1992); this is likely due to the relative contributions of coherent reflection- and distortion-sources as a function of stimulus level in DPOAE recordings. Commercially
available OAE analyzers allow for OAEs to be measured between ~500-8000 Hz, and some have modules extending high-frequency capability to 10,000 Hz or above. Some researchers, however, have recorded OAEs to stimuli below 300 Hz (Christensen et al., 2016) and up to 20,000 Hz (Dewey & Dhar, 2016).

3.3. Neural Assays of Auditory System Function

3.3.1 Electrocochleography

Electrocochleography is a technique used to measure electrical potentials of the inner ear and auditory nerve. Recording electrodes can be placed on the scalp, earlobe, ear canal wall, tympanic membrane (TM), cochlear promontory or round-window (Ferarro, 2010). Generally, larger electrical potentials are recorded as electrodes are placed closer to the cochlear promontory (Schoonhoven et al., 1995). Three classic potentials are recorded electrocochleographically from the ear in response to sound: the cochlear microphonic, summating potential, and auditory nerve compound action potential (CAP).

![Figure 3.2. Components of the electrocochleographic waveform. In panel A, separate responses to condensation (dotted line) and rarefaction (solid line) click](image)
polarities are shown, and a large compound action potential can be seen in both waveforms. Panel B shows the summating and compound action potentials in a waveform calculated by averaging responses to condensation and rarefaction polarities. In Panel C, compound action and summating potentials have been removed from the response by subtracting condensation and rarefaction waveforms, and only the cochlear microphonic remains.

*Cochlear Microphonic*

OHCs and IHCs generate receptor potentials that phase lock to an input stimulus on cycle-by-cycle basis (Dallos et al., 1982; Russel & Sellick, 1978). These cyclical changes produce an alternating current that is detectible by electrocochleography electrodes (Figure 3.2). The cochlear microphonic response is dominated by OHC receptor currents, as IHC currents are approximately 30-40 dB lower in intensity (Dallos & Wang, 1974). Further, when electrodes are oriented toward the cochlear promontory, as is almost exclusively the case in humans, OHCs from the cochlear base are more represented in the cochlear microphonic than those from the apex (Eggermont, 2017). While healthy skepticism regarding the clinical utility of cochlear microphonic measurements has been expressed since the early use of electrocochleography in humans (e.g., Eggermont, 1976), a renewed interest in using this potential to assess OHC survival is apparent in the recent work of several laboratories (Charaziak et al., 2017; Chertoff et al., 2012, 2014; Kamerer et al., 2016).

The cochlear microphonic has an onset of < 1 ms, and, unlike neurally-generated potentials, it does not increase in latency as stimulus presentation level decreases.
Further, the cochlear microphonic does not fatigue with rate and waveform polarity inverts with a 180-degree phase shift of the stimulus (i.e., condensation vs. rarefaction). Consequently, the cochlear microphonic can be essentially cancelled out by using an alternating polarity stimulus presentation.

**Summating Potential**

The summating potential is a negative or positive direct-current shift observable in electrocochleographically recorded waveforms (Tasaki et al., 1954). Like the cochlear microphonic, the summating potential begins almost immediately after the onset of a stimulus and is sustained throughout its duration (Figure 3.2); it is therefore a non-adapting direct-current shift that corresponds to the stimulus envelope (Dallos, 1973). The summating potential arises mainly from OHCs, and is thought to reflect a “baseline shift” in the operating state of the organ of Corti during acoustic stimulation. This potential has most notably been studied in individuals with Meniere’s disease/endolymphatic hydrops, which increases endolymphatic pressure and distends both the basilar and Reissner’s membranes (Ferraro, 2010). Consequently, the organ of Corti’s natural operating state is distorted and auditory stimulation causes a larger than normal summating potential in these patients.

**Compound Action Potential**
The auditory nerve CAP represents the summed response of synchronously firing auditory nerve fibers (Figure 2.5). Goldstein and Kiang (1958) were the first to model an auditory nerve CAP waveform as a convolution integral:

\[ \text{CAP}(t) = N \int_0^t s(\tau) a(t - \tau) \, d\tau \]

where \( N \) is the estimated number of nerve fibers, \( s(\tau) \) is the latency distribution function, and \( a(\tau) \) is the unit response. Essentially, this equation uses a diphasic action potential response from a single auditory nerve fiber as a template, and applies that template over the temporal response characteristics of the basilar membrane while accounting for the number of nerve fibers and nerve fiber density along the length of the cochlea (Figure 3.3). Two important aspects of CAP generation are revealed by this model. The first is that auditory nerve fibers tuned to the cochlear base are more readily stimulated than those in the apex due to tonotopy. The second is that neural responses from the cochlear base are highly synchronized, whereas those from the apex are minimally represented in the CAP waveform due to temporal cancellation. This means that when a click stimulus is used, the CAP waveform is dominated by basal auditory nerve fibers around 4 kHz and above. It also means that the latency of CAPs evoked by tone bursts or clicks with high-pass masking will increase as tone burst or high-pass masking cutoff frequency decreases (Eggermont, 1976). One way to correct for delays related to tonotopy and thus synchronize auditory nerve fiber activity along the length of the cochlea is to construct a rising frequency chirp (Chertoff
et al., 2010; Fobel & Dau, 2004). While chirps are still being validated for clinical use, they hold promise for clinical applications, such as improving newborn hearing screenings. A chirp optimized for evoking CAP responses is described in Chapter 5.

Figure 3.3. Basilar membrane tonotopy results in displacement delays between the cochlear base and apex (A). Based on these delays, action potentials originating from auditory nerve fibers innervating the cochlear base are more synchronized than those initiated by apical fibers (B). The electrocochleographically-recorded CAP reflects the summated responses of auditory nerve fibers and is dominated by synchronized discharges from the cochlear base at the recording electrode.
3.3.1.2. Electrocochleography Stimulus-Response Characteristics

CAPs are generally larger in females than in males and in right ears versus left ears (Chatrian et al., 1985). As stimulus intensity increases, CAP amplitude increases and latency decreases. A plateau region is typically evident in CAP level-series functions at mid-intensities, which is related to the compressive growth function of OHC amplification (Picton, 2011). As stimulus rate increases, CAP amplitudes decrease and latencies increase until a point at which the CAP is effectively adapted out of the recording and only cochlear microphonic and summating potential components remain. Broadband stimuli such as clicks and chirps evoke larger CAPs than tone bursts, and tone burst frequencies above 2000 Hz evoke larger and sharper CAPs than tone burst frequencies below 2000 Hz.

3.3.2 Frequency Following Response

The frequency following response (FFR) is a sustained evoked potential generated by the pooled phase-locking of neural ensembles within the auditory brainstem (Krishnan, 2007). The morphology of the FFR waveform mimics the cycle-to-cycle or envelope periodicity of the evoking sound (Figure 3.4) and its spectrum reveals discernable peaks at constituent stimulus frequencies and/or the envelope frequency. The FFR was named and first described by Worden and Marsh (1968), who recorded neural phase-locking to tone bursts (0.8-5 kHz) from various auditory brainstem nuclei in cats. In their pioneering work, these authors demonstrated that the FFR was neurogenic and distinct from the pre-neural
cochlear microphonic, which had received much attention since its (inadvertent) discovery by Wever and Bray in 1930. Moushegian et al. (1973) subsequently were first to record FFRs from the vertex of the human scalp.

4.3.2.1. FFR Neural Generators

Figure 3.4. Pure tones (500 Hz) in rarefaction and condensation polarities (Panel A, top) are overlaid. Each pure tone evokes a frequency following response (Panel B, bottom) that follows the period of the stimulus. Note that there is a half-period shift between the frequency following response waveforms to each polarity due to rectification from the mechano-electrical transduction process. Amplitude modulated pure tones (500 Hz, AM @ 80 Hz) in rarefaction and condensation polarities are overlaid (Panel B, top). Note that, while the fine structure polarity flips between condensation and rarefaction polarities, the envelopes of each stimulus are the same. Likewise, the envelope following responses (Panel B, bottom) to 80 Hz are essentially the same between fine structure polarities.
At stimulus levels less than 50 dB SPL, FFRs to low frequency tones are likely initiated by the phase-locked activity of auditory nerve fibers innervating the most apical 1/3 of the cochlea around the stimulus center frequency (Gardi et al., 1979; Moushegian et al. 1978; Pfieffer & Kim, 1974). When the intensity of a low frequency stimulus is increased to approximately 70 dB SPL, the cochlear origin of the FFR becomes more diffuse and is biased toward basal regions due to “upward spread” of the traveling wave envelope (Dau, 2002). These findings indicate that the place specificity of the FFR is poor at high-intensities; however, more place-specific information can be obtained with high-pass masking techniques (e.g., Gardi et al., 1979).

Animal models have revealed much about the brainstem neural generators of the FFR. Worden and Marsh (1968) presented several key observations from cats in their seminal paper. Perhaps most importantly, they noted that all auditory brainstem nuclei contained phase-locking neurons based on near-field recordings. As they repositioned their recording electrodes from caudal to rostral nuclei, the peak latencies of the phase-locked responses became increasingly delayed relative to stimulus onset, which is consistent with accruing synaptic delays in higher brainstem areas. They also noted that the upper frequency limit of phase-locking decreased systematically from cochlear nucleus to inferior colliculus, and no phase-locking responses were recorded beyond the inferior colliculus. Smith et al. (1975) observed that the vertex-recorded FFR in cats disappeared with cryogenic cooling of the contralateral inferior colliculus and
concluded that the inferior colliculus is the major generator of the vertex FFR. In a similar experiment, Gardi et al. (1979) sequentially aspirated auditory brainstem nuclei beginning with the inferior colliculus and noted a 50% amplitude reduction of the FFR when the contralateral lateral lemniscus was ablated.

Brainstem generators of the human FFR have been inferred by measuring response latencies with various recording montages (Galbraith et al., 2001) and by studying individuals with focal brainstem lesions (Sohmer et al., 1977a,b). Galbraith et al. (2001) recorded FFRs to tone bursts simultaneously from vertical and horizontal montages. The vertical montage accentuated peaks with latencies of 6-8 ms, which is consistent with rostral generators. The horizontal montages accentuated peaks with latencies of ~3.5 ms, which is consistent with activity from the auditory nerve, cochlear nucleus, or superior olivary complex (Moller et al., 1983, 1988). Similarly, Sohmer et al. (1977) recorded FFRs from individuals with known rostral brainstem lesions and missing waves IV and V from the auditory brainstem response. Their results demonstrated that “short-latency” FFRs (~3.5 ms) were present, whereas “long-latency” FFRs (~6-8 ms) were not. More recently, FFR group delay data has suggested that the rostral brainstem dominates FFR responses to frequencies below ~500 Hz, above which more caudal sources contribute (King et al., 2016; Shaheen et al., 2015; Tichko & Skoe, 2017). Beyond ~1000-1500 Hz, neural phase locking becomes poor and pre-neural potentials such as the cochlear microphonic and summating potential may dominate the response.
4.3.2.2. Human FFR Stimulus-Response Characteristics:

Signal Frequency

FFR response amplitudes increase from 100 Hz to 250 Hz, where they reach a peak, and begin to decrease at greater frequencies (Hoorman et al., 1992). By 600 to 800 Hz, response amplitudes are quite small and acquiring FFRs to higher frequencies may require much more averaging. The reduction in FFR amplitude with increasing frequency reflects a decrease in phase-locking ability at higher frequencies. Krishnan (2007) recommends using a horizontal montage (i.e., mastoid-to-mastoid) to record FFRs to stimuli > 1 kHz based on the observation that more caudal structures can phase-lock at higher rates. As the evoking stimulus becomes more spectrally complex (e.g., harmonic sound; speech), the amplitude of the FFR decreases relative to those evoked by tone bursts (Krishnan et al., 2000, 2004).

Signal duration and Intensity Effects

Slowly ramped continuous tones evoke FFRs without an overlapping onset response; however, longer tones result in significant adaptation of amplitude over time (Krishnan, 2007). Tone bursts have wider spectral splatter and therefore typically evoke an onset response in addition to the FFR. To reduce the overlap of an onset response to tone bursts, it is recommended that a plateau of at least 4 signal cycles be used (Hoormann et al., 1992).
FFR stimuli must be presented at suprathreshold levels (≥ 45-50 dB SPL) for stable responses that are sufficiently above the background EEG noise to be recorded (Pandya et al., 2004). Increasing stimulus levels beyond the "electrophysiologic threshold" results in peak amplitude growth up to ~ 75 dB SPL. Above this level, the amplitude growth function saturates or even begins to decline; thus the amplitude growth function has a small dynamic range (~ 20-40 dB). The growth in FFR amplitude with stimulus intensity likely represents greater synchronization of auditory nerve fibers innervating regions that fall within the traveling wave envelope. As mentioned above, high intensities can lead to neural recruitment from more basal regions and thus desynchronize the response and/or decrease its latency.

Two-tone Complexes and Speech

As discussed above, appropriately spaced tone pairs can create nonlinear intermodulation distortion products along the basilar membrane. While these products propagate into the ear canal as acoustic waveforms (i.e., DPOAEs), they can also evoke phase-locked neural activity at the quadratic (f2-f1) and cubic (2f1-f2) distortion frequencies (Bhagat & Champlin, 2004; Elsisy & Krishnan, 2008; Goldstein & Kiang, 1968; Rickman et al., 1990). Thus, the distortion product FFR (DPFFR) is a neurologic measure that, unlike the DPOAE, represents “feed-forward” effects of nonlinear cochlear processing (see Chapter 6 for more detail).
When the FFR is evoked by speech sounds, it is also called the “speech evoked ABR” or “complex ABR” (Skoe & Kraus, 2010). The FFR preserves many components of the speech signal including voice pitch, spectral characteristics (e.g., harmonics and formants), and time-variant characteristics (e.g., diphthongs, tones, prosody). When the complex waveform of speech is broken into component frequencies along the basilar membrane, the relative energy of the harmonics (i.e., formants) is preserved in traveling wave envelopes (Krishnan, 1999). Spectral analysis of the FFR reveals that formants are robustly encoded (Chandrasekaran & Kraus, 2010; Krishnan & Agrawal, 2010; Skoe & Kraus, 2010). Additionally, harmonics falling outside of formants may be suppressed, indicating that cochlear two-tone suppression may accentuate formant peaks (Krishnan, 2007). Krishnan et al. (2004) recorded FFRs to Mandarin Chinese tones and demonstrated with autocorrelation analysis that they tracked changes in pitch contour; the tracking was more precise in native Mandarin speakers, which suggested that linguistic experience shapes subcortical temporal encoding. Similarly, the FFR has been demonstrated to track the pitch and timbre of musical notes, and this phenomenon is more robust in musicians (Chandrasekaran & Kraus, 2010).
4.1. Introduction

Chapter 2 provided an overview of the anatomy and physiology of afferent and efferent auditory nervous systems and Chapter 3 discussed some pertinent tools used to assess these systems. The aim of Chapter 4 is to review the extant literature examining olivocochlear system *function* and to provide specific context and motivation for the experiments described in Chapters 5-7. The review is organized in two sections focused on animal and human experiments, respectively. Within each section are subsections describing evidence for one putative role of the olivocochlear system or summarizing a specific research topic from the literature. The chapter concludes with a discussion highlighting limitations of olivocochlear assays and some outstanding questions regarding olivocochlear function in humans.

4.2. Olivocochlear Function in Animal Models

Germinal experiments on the olivocochlear system were performed using animal models. Rasmussen (1946) was the first to anatomically describe the olivocochlear efferents and Galambos (1956) first demonstrated that electrical stimulation of olivocochlear fibers resulted in a reduction of auditory nerve compound action potentials. Although replicated many times (Desmedt and Monaco, 1961; Fex, 1967; Desmedt, 1962; Wiederhold, 1970), this was initially a paradoxical finding, as the evolutionary benefit of such a reflex was not
immediately clear. In the decades that followed, a more sophisticated understanding of efferent system function was developed from animal data, and several putative roles have been gleaned from this work.

4.2.1 Auditory Nerve “Unmasking”
In the presence of masking noise, the dynamic range of auditory nerve fibers is reduced, as signal thresholds increase and saturation levels decrease due to synaptic depletion. Electrically or acoustically activating the caudal efferents, specifically the MOC system, desensitizes the auditory nerve to masking noise by reducing the cochlear excitation response to continuous broadband sounds. Thus, the “unmasking” effect of the MOC reflex restores the dynamic range of auditory nerve fibers and aids in the detection of transient signals in noise such as tones and speech (Dolan & Nuttall, 1988; Kawase et al., 1993; Kawase & Liberman, 1993; Robles & Delano, 2008; Winslow & Sachs, 1988). Importantly, this unmasking effect is preserved in the behavior of inferior colliculus neurons, suggesting potent downstream MOC effects (Seluakumarnan et al., 2008).

4.2.2 Protection from Noise-Induced and “Hidden” Hearing Loss
Animals raised in persistent, low-level acoustic stimulation are less likely to suffer cochlear damage when exposed to high levels of noise than animals living in quiet environments (Cody & Johnstone, 1982; Rajan & Johnstone, 1988). Presumably this is due to “strengthening” of the caudal efferent system with low-level noise exposure (Brown et al., 1998), and there appears to be a positive relationship
between MOC reflex strength and resilience to noise induced hearing loss (Maison & Liberman, 2000). This protective effect was abolished when the MOC system was either pharmacologically disabled or severed (Kujawa & Liberman, 1997). One criticism of the hypothesis that the MOC reflex protects from acoustic overexposure is that there are very few natural environments in which noise is sufficiently intense to cause temporary or permanent threshold shifts; therefore, intense sound exposure in the natural environment is likely too rare a phenomenon to have caused selective evolutionary pressure for the emergence of the olivocochlear system (Kirk & Smith, 2003).

More recently, it was demonstrated that “hidden hearing loss” (i.e., normal audiometric thresholds with suprathreshold hearing impairment due to auditory nerve fiber synaptopathy) was exacerbated in both noise exposed (Maison et al., 2013) and aging ears (Liberman et al., 2014) when the olivocochlear system was selectively lesioned. While this body of literature supports the role of the MOC system in protecting the ear from acoustic trauma, the LOC system may also reduce noise-induced excitotoxicity at the level of the inner hair cell/auditory nerve fiber synapse (Darrow et al., 2007; Ruel et al., 2001). Far less is known about the behavior and function of the LOC system since the fibers are unmyelinated and in the minority among all olivocochlear neurons. Exciting work is underway examining the extent to which the caudal efferent system can be pharmacologically modulated in animals (e.g., Le Prell et al., 2014), with the hope
that prophylactic treatments against noise induced hearing impairments may be developed for humans.

4.2.3. Top-down Control of Cochlear Processes During Attention

As described in Chapter 2, the efferent auditory system is extensive, beginning in the cortex and terminating at the level of the cochlea. This organization suggests that top-down influences, such as attention, may modulate cochlear function by “commandeering” the caudal efferent system\(^5\) (e.g., Smith & Cone, 2015). Delano and colleagues (2007) measured compound action potentials and cochlear microphonics from chinchillas during auditory and visual attention tasks. They found that compound action potential amplitude was significantly reduced while the cochlear microphonic was significantly increased during visual attention tasks compared with auditory attention tasks. Their explanation for these findings was that the MOC system increases the conductance of OHCs, and therefore simultaneously increases cochlear microphonic and decreases compound action potential amplitudes. Some, but not all, experiments have shown a decrease in auditory detection or discrimination after the olivocochlear system had been severed (e.g., May & McQuone, 1995).

\(^5\) Note that this organization, wherein a reflex is modulated by higher order brain centers, is ubiquitous in sensory systems. For example, the pupillary reflex is passively activated when photons are absorbed by the retina but can also be modulated during attention. Similarly, spinal reflexes can be “overridden” with attention.
Perhaps some of the most compelling data on the top-down effects of the efferent auditory system come from experiments in which cortical or rostral brainstem neurons are deactivated (either pharmacologically or via cryoloop cooling) or electrically stimulated. For example, OHCs and auditory nerve fibers can be modulated with activation or de-activation of the efferent system “upstream” in the rostral auditory brainstem or cortex (Dragicevic et al., 2015; Leon et al., 2012; Mulders & Robertson, 2000; Rajan, 1990; Ota et al., 2004; Zhang & Dolan, 2006). Additionally, the characteristic frequency of stimulated neurons in the auditory cortex (Xiao & Suga, 2002) and inferior colliculus (Mulders & Robertson, 2000) corresponds with the frequency of maximum OHC and auditory nerve fiber inhibition. Groff and Liberman (2003) reported long lasting neural inhibition and enhancement with inferior colliculus stimulation, even in instances where OHC responses were unaltered. Furthermore, these findings persisted after selectively sectioning the MOC system, leading them to conclude that the LOC system may be involved in slow potentiation of auditory nerve responses.

4.3. Olivocochlear Function in Humans

With the advent of commercial otoacoustic emission (OAE) analyzer systems, it became possible to non-invasively study olivocochlear function in the human cochlea, and most human MOC reflex studies have been performed using this technique. Several putative roles of the human caudal efferent system have been proposed based on empirical evidence.
4.3.1. Improving the Neural Encoding of Speech-in-Noise

There has only been one study describing auditory nerve “unmasking” of simple stimuli with activation of the MOC reflex in humans (Kawase & Takasaka, 1995). More commonly, correlations between the magnitude of contralateral OAE inhibition (described in Chapters 5-7) and signal- and speech-in-noise perception ability have been reported (Abdala et al., 2014; Giraud et al., 1997; Kumar & Vanaja, 2004; Micheyl et al., 1995; Yilmaz et al., 2007). These studies suggest that the olivocochlear system is involved in speech-in-noise processing or, at the very least, is indicative of the overall strength of the efferent system. Importantly, some studies have shown no correlation between the two measurements (Scharf et al., 1997; Wagner et al., 2008). Abdala and colleagues (2014) reported that the strength of the MOC reflex begins to decrease in middle age, which may be related to poorer speech perception abilities in the elderly.

4.3.2. Top-down Control of Cochlear Processes During Active Listening

As with animal models, attention effects on cochlear function have also been studied in humans. Because the efferent system enters the inner ear as part of the vestibular branch of the eighth nerve, patients with normal hearing who have undergone vestibular neurectomy for intractable vertigo provide unique opportunities to study the proposed top-down effects of the efferent system on various auditory abilities. In their studies on vestibular neurectomy patients, Scharf and colleagues (1994, 1997) tested many auditory abilities pre- and post-
surgery including frequency and intensity discrimination of tones, detection of masked tones, and loudness adaptation. None of these measures significantly differed between pre- and post-surgery, calling into question the functional benefit of the olivocochlear system, at least for perception of simple stimuli. They did find, however, that unexpected off-frequency tones were more easily detected post-neurectomy, indicating that the efferent system may actively inhibit non-relevant frequency channels during active listening. Khalfa and colleagues (2001) reported decreased MOC function in patients undergoing temporal lobe resection for intractable epilepsy in the ear opposite the surgery, providing further evidence of corticofugal cochlear control in humans.

Other studies have indicated that OHC function is altered in normal hearing listeners during attention (Avan & Bonfils, 1992; de Boer & Thornton, 2007; Ferber-Mart et al., 1995; Froehlich et al., 1990; Garinis et al., 2011; Harkrider & Bowers, 2009; Maison et al., 2001; Perrot et al., 2006; Puel et al., 1988; Smith & Cone, 2015; Wittekind et al., 2014); however the size and direction of the modulation effect is not consistent between studies. The largest effects of attention were observed in studies in which the participants were cued prior to the task and OAE measurements were time-locked to the cue (e.g., Wittekind et al., 2014). This method likely avoids “washing out” attention modulation effects over long averaging blocks in which subjects are inattentive during interstimulus interval periods. Interestingly, there may be a developmental component to MOC reflex modulation during attention, as Smith and Cone (2015) found smaller top-down
effects in a group of 8-14 year-old children.

4.3.3. Predicting Noise-Induced Hearing Loss and Auditory Training Benefit

In two large longitudinal studies, Lapsley-Miller and colleagues (2006, 2009) reported an average reduction in OAE amplitudes in military servicemen exposed to noise, even though group hearing thresholds remained stable on average. Individuals with lower baseline OAE amplitudes were six- to nine-times more likely to develop permanent threshold shifts, indicating that OAEs may be a predictor of noise-induced hearing loss susceptibility. In the future, individuals with lower baseline OAE measurements and weaker MOC reflexes may be good candidates for prophylactic pharmaceutical treatments to prevent noise induced hearing loss.

The neuroplastic effects of auditory training may, in part, be facilitated by the olivocochlear system. For example, it is reasonable to assume that speech-in-noise training bolsters the anti-masking effect of the caudal efferent system and thus improves listening in noise. de Boer and Thornton (2008) reported that weaker baseline MOC reflex strength was associated with greater training-induced improvement on a speech-in-noise perception task. Further, MOC reflex strength increased in individuals demonstrating a significant improvement on the speech-in-noise task after training. These measurements may be useful in identifying which patients would be most likely to benefit from auditory training and would provide an objective means of quantifying training-based neuroplasticity.

4.3.4. MOC Reflex Function in Disordered Populations
Activation of the MOC reflex is dependent upon proper function of afferent (inner hair cells, auditory nerve fibers, cochlear and superior olivary nuclei) and efferent (MOC fibers and OHCs) components of the reflex loop. Patients with auditory neuropathy and multiple sclerosis affecting the auditory brainstem do not demonstrate OAE inhibition due to poor neural synchrony in this circuit (Berlin et al., 1993b; Coelho et al., 2007). Patients with myasthenia gravis, a muscular disorder affecting the synaptic release of acetylcholine, also show reduced OAE inhibition, as acetylcholine is the efferent neurotransmitter released by MOC fibers onto the base of OHCs; notably, OAE inhibition strength increases with treatment of myasthenia gravis (Paludetti et al., 2001). Ototoxicity and diabetic hyperglycemia, both of which alter the metabolism of outer hair cells, can also be monitored with OAE inhibition (Jacobs et al., 2012; Riga et al., 2007).

Individuals with more global deficits also show aberrant OAE inhibition. For example, children with selective mutism, learning disabilities, and auditory processing disorder demonstrate reduced OAE inhibition versus normal controls (Bar-Haim et al., 2004; Boothalingam et al., 2015; Garinis et al., 2008; Muchnik et al., 2004). Tinnitus patients with hyperacusis but otherwise normal hearing demonstrate "hyperactive" MOC activity (Knudson et al., 2014), and those with autism have reportedly shown both a lack of OAE inhibition (Danesh & Kaf, 2011) as well as hyperactivity (Wilson et al., 2015). The discrepancy in these results is perhaps because some studies have failed to sort autistic participants into
hyperacusic and non-hyperacusic groups. It should be noted that methodology differs widely in this body of literature, making it difficult to compare results across studies (Mishra, 2014). Additionally, while the results of these studies are academically interesting, they presently have no clinical diagnostic value for the abovementioned disorders.

4.3.5. Ear and Sex Differences in Human MOC Reflex Function

OAE and OAE inhibition asymmetries have been reported in the literature. For example, spontaneous OAEs are more often observed in right versus left ears, and this appears to be related to slightly better right ear hearing sensitivity (Bilger et al., 1990; McFadden, 1993). CEOAEs are larger in right ears for both adults and infants (Glattke et al., 1995; Newmark et al., 1997). Interestingly, Sininger and Cone-Wesson (2004) reported larger CEOAEs in newborn right ears than left, but this pattern was flipped for DPOAEs. This suggests that well known cortical functional asymmetries may exert corticofugal influences on cochlear processing via the olivocochlear system.

Binaural presentation of an inhibitor sound causes the largest reduction of human OAEs (Berlin et al., 1993). The effects of ipsilateral and contralateral inhibitor presentation (re: probe) are dependent on elicitor bandwidth. For broadband noise, ipsilateral and contralateral inhibition are similar; however, as bandwidth decreases, ipsilateral inhibitors are twice as potent as contralateral inhibitors.
(Lilaonitkul & Guinan, 2009). This suggests that MOC reflex laterality effects are influenced by central processes that respond in some way to elicitor features. Garinis et al. (2011) reported that, when CEOAE inhibition was reported as a fractional change in emission amplitude, contralateral OAE inhibition was significantly larger for OAEs recorded in the left ear versus right.

In general, females have larger OAEs than males (Bilger et al., 1990; Talmadge et al., 1993; McFadden, 1998). Because this difference is present at birth, some researchers have suggested that OAEs are sensitive to androgenic hormones (see McFadden et al., 2006 for review). Based on a review of his own work, McFadden proposed the MOC reflex strength is greater for the left ear than right ears and less potent in females than in males. However, when this hypothesis was tested by Ismail and Thornton (2003), they reported that the MOC reflex is more potent in the right ear than left with no significant differences in this pattern between the sexes. This finding suggests that, although OAE magnitude is influenced by hormones, MOC reflex strength and laterality may be more influenced by cortical organization than sex per se.

### 4.4 Outstanding Questions Regarding Human Olivocochlear Function

As demonstrated in the preceding sections, human MOC reflex assays have mainly consisted of OAE measurements. While much has been learned with this technique, it remains unclear how MOC reflex-induced changes in OHC motility
influence downstream neural processing, which is the foundation for auditory perception. It is thus imperative to develop neural assessments of MOC reflex function and to understand the relationships between pre-neural and neural assays.

Stimuli used to evoke OAEs also evoke neural responses from the auditory nerve or brainstem. For example, clicks or chirps evoke both transient OAEs and auditory nerve compound action potentials. Appropriately-spaced tone pairs evoke both distortion-product OAEs and auditory brainstem frequency following responses. Thus, with careful consideration of stimulus parameters, traditional OAE paradigms can be modified to also study neural correlates. Using complementary methods assessing pre-neural and neural MOC reflex effects provides promise in understanding the functional role of this mechanism in hearing.
5.1. Introduction

Cochlear outer hair cells (OHC) receive direct efferent feedback from the caudal auditory brainstem via the medial olivocochlear (MOC) nerve bundle. The MOC bundle inhibits OHC motility and indirectly modulates basilar membrane motion and inner hair cell (IHC) sensitivity – an effect termed the MOC reflex (Mountain, 1980; Siegel & Kim, 1982; Murugasu & Russell, 1996; Cooper & Guinan, 2003, 2006). Experiments in animal models have revealed that excitation of the MOC reflex “unmasks” signal representation in the auditory nerve by reducing mechano-electrical transduction of noise within the cochlea and therefore may play an active role in hearing in noise (Kawase & Liberman, 1993; Kawase et al., 1993). The functional importance of the MOC reflex in human hearing, however, remains unclear.

Because otoacoustic emissions (OAEs) likely originate from mechanics associated with OHC motility (Liberman et al., 2002; Cheatham et al., 2004; Dallos et al., 2008), they are sensitive to MOC reflex-induced changes in OHC function and provide a non-invasive, albeit indirect, method to study efferent effects in humans. In the classic contralateral inhibition of OAEs paradigm, OAEs are measured without and with presentation of a contralateral acoustic stimulus (CAS; e.g., broadband noise, BBN), which activates the uncrossed MOC fibers of the reflex circuit. Magnitude and/or phase differences between OAEs recorded without and with CAS are then used to quantify MOC reflex-induced shifts in OHC function (Guinan, 2006). Such studies have quantified characteristics of human
MOC reflex strength (e.g., Backus & Guinan, 2007; Marshall et al., 2014), tuning (e.g., Veuillet et al. 1991; Chéry-Croze et al. 1993; Lilaonitkul and Guinan 2007; Zhao & Dhar, 2012), and laterality (e.g., Frances & Guinan, 2010; Garinis et al., 2011). However, OAEs are pre-neural measurements and are therefore less informative about the “downstream” MOC effects on inner hair cell excitation and the subsequent neural ensembles that mediate hearing.

Few experiments have reported MOC reflex effects on evoked compound action potentials (CAPs) from the human auditory nerve (Folsom & Owsley, 1987; Kawase & Takasaka, 1995; Chabert et al., 2002; Lichtenhan et al., 2015; Najem et al., 2016). Both the dearth of research in this area and the wide range of reported inhibition with CAS (2-20 dB) may stem from technical issues related to CAP inhibition measurements. For example, OAE experiments have shown that the effect of MOC reflex inhibition on OHC activity is more potent at lower stimulus levels (e.g., Hood et al., 1996); however, clicks and tone bursts presented at these levels evoke less synchronized neural responses from a smaller population of auditory nerve fibers and therefore produce CAP waveforms with poorer morphology than higher stimulus levels. Without adequate response averaging, CAP waveforms evoked by low- to moderate-level clicks or tone bursts are highly variable with poor signal-to-noise ratios and “true” physiologic changes attributable to the MOC reflex (i.e., reduction in CAP amplitude) are difficult to separate from measurement variation.
Stimuli evoking more robust CAP responses than clicks or tone bursts, such as rising frequency chirps, may circumvent some of the technical issues related to neural MOC reflex measurements. Unlike a click, which initiates synchronized responses predominately from more basal auditory nerve fibers (Kiang, 1975; Abdala & Folsom, 1995), chirps synchronize auditory nerve fiber excitation along the length of the cochlear spiral by correcting for temporal delays associated with tonotopicity (Shore & Nutall, 1985; Fobel & Dau, 2004). Recently, Chertoff and colleagues (2010) demonstrated that chirps optimized for eliciting human CAPs produced significantly larger amplitudes than those evoked by clicks in young, normal-hearing adults at moderate to high stimulus levels (75-125 pSPL). The improved signal-to-noise ratio of chirp-evoked CAPs, compared to those from clicks, may thus provide a higher fidelity response to assay CAS-induced MOC reflex effects on the auditory nerve. Additionally, MOC fibers innervate the length of the cochlear spiral with tuning similar to afferent auditory nerve fibers (Warr, 1992). Chirp-evoked CAPs may therefore be more sensitive to the summed CAS-induced MOC reflex effects along the entire length of the cochlea and thus show greater inhibition than click-evoked CAPs.

In this experiment, we tested two hypotheses: 1) That chirps evoke larger CAP amplitudes than clicks using low to moderate stimulus levels, which engage the cochlear amplifier and are thus more sensitive to MOC effects, and 2) That MOC reflex inhibition of chirp-evoked CAPs is larger than for click-evoked CAPs due to the broader basilar membrane area represented in chirp responses. To relate our findings to more commonly used MOC
reflex assays, we also compared average chirp- and click-evoked CAP inhibition to click evoked OAE (CEOAE) inhibition measured in the same subjects.

5.2. Method

Participants

The University of Arizona Human Subjects Protection Program approved the following methods. Eighteen adult participants without history of neurologic or otologic disease were enrolled in the study; however, due to attrition, 14 subjects (average age 22.25 years; 10 females) completed all six testing sessions. Otoscopy examinations found that all ear canals were free of excess cerumen and that tympanic membranes (TMs) appeared healthy in all subjects. Participants had normal tympanograms bilaterally, defined as ear canal volume of 0.6-1.5 cc and peak-compensated static admittance between 0.3-1.4 mL (Margolis & Heller, 1987), and contralateral acoustic reflex thresholds to 1-10 kHz BBN ≥ 70 dB SPL, measured using conventional admittance methods (Sun, 2008). The latter requirement was to mitigate the possible involvement of middle ear muscle contractions during MOC inhibition measurements, although others have shown that acoustic reflex thresholds can be lower when measured using more sensitive techniques (e.g., Zhao & Dhar, 2010; Lichtenhan et al. 2016). Air conduction hearing thresholds from 0.25-8 kHz were within normal limits (≤ 25 dB HL) bilaterally for all subjects.

Equipment and Procedures
A 100-μs click and 10-ms chirp were used to evoke CAPs. The click was created using the Intelligent Hearing Systems Smart-EP stimulus generator (Intelligent Hearing Systems, Miami, FL). The chirp was created in WAV file format in MATLAB (The Mathworks, Inc., Natick, MA, USA) using a modified “O-Chirp” from Fobel and Dau (2004), as implemented by Chertoff et al. (2010). The O-Chirp is a flat-spectrum stimulus relating frequency to basilar membrane delay using parameters from stimulus frequency OAEs. To optimize the O-Chirp for evoking CAPs, forward traveling wave delays were estimated from Eggermont’s (1979) derived-band CAP latencies as opposed to stimulus frequency OAEs. The relationship between basilar membrane delay in milliseconds and frequency was expressed as:

\[ \tau_{BM} = c^* f^\alpha \]

where 0.45 kHz \( \leq f \leq 10 \) kHz and \( c \) (0.69) and \( \alpha \) (-77) are constants. The chirp WAV file was converted into a stimulus file suitable for presentation by the Intelligent Hearing Systems Smart-EP program.

The click and chirp were presented through ER-3A insert earphones (Etymotic Research, Elk Grove Village, IL) to a 2-cc coupler and calibrated in units of \( \text{dB peak-to-peak equivalent sound pressure level} \) (ppeSPL) using a 1000 Hz tone as a reference (Burkard, 2006). Click and chirp spectra were comparable with the exception that the chirp had 3-5 dB less energy below ~3.5-4 kHz (see Chertoff et al., 2010, Figure 1).
Behavioral thresholds for clicks and chirps were obtained from the right ears of 18 subjects using a modified Hughson-Westlake procedure. Stimuli were presented at a starting presentation level of 50 dB ppeSPL. Presentation level was decreased by 4 dB after every positive response and increased by 2 dB after each failure to respond. Threshold was defined as the lowest presentation level at which three positive responses occurred. These measurements were made without electrodes in the ear canal, as our previous work demonstrated that TM electrode contact with the eardrum can influence audiometric thresholds, particularly to low frequencies (Smith et al., 2016). Average behavioral thresholds were 32 dB ppeSPL and 30 dB ppeSPL for clicks and chirps, respectively. While we express stimulus levels in units of dB ppeSPL throughout this paper, behavioral thresholds to clicks and chirps can be subtracted from these values to convert from dB ppeSPL to normalized hearing level (nHL).

_Tympanic Membrane Electrodes_

Using a modified protocol by Ferraro and Durrant (2002), we assembled TM electrodes in our laboratory that were suitable for our evoked potentials recording system. The electrodes were constructed from 11.43-cm long sections of PFA-insulated silver wire (0.1 mm gauge) encased in 10.16-cm long pieces of flexible silastic medical tubing. The PFA-insulation was removed from the last 0.635 cm of each end of the wire. One uninsulated end was crimped with a female machine pin that was connected to an electrode cable interfacing with the bio-amplifier. The other uninsulated end was bent to form a hook around a 0.25 gram wisp of cotton, and the end of the hook was tucked back
into the opening of the silastic medical tubing to ensure that it did not directly make contact with the eardrum when it was inserted. Prior to each recording session, the cotton-tipped end of a TM electrode was saturated with 1-cc of Synapse electrode cream (Kustomer Kinetics, Arcadia, CA) using a 27-gauge needle. TM electrodes were inserted into the right ear canal of each subject and advanced until the TM was contacted, which was verified by subject report of the occlusion effect and by monitoring electrode impedance changes until they were consistently $\leq 7$ kΩ on the Intelligent Hearing Systems bio-amplifier (Ferraro, 2010). Further confirmation of electrode contact with the TM was indicated by areas of acute redness and accumulation of electrode gel observed otoscopically after TM electrodes were removed at the end of each testing session (see Smith et al., 2016, Figure 1). Each electrode was held in place throughout the session by a 13 mm ER3-14A foam ear tip coupled to the ER-3A insert earphone.

**CAP Measurements and Amplitude Calculations**

Each subject participated in six two-hour CAP recording sessions – three in which clicks were used to evoke CAPs and three in which chirps were used. The order in which subjects participated in click or chirp sessions was randomized. In every session, subjects were comfortably reclined in a lounge chair in an electromagnetically shielded sound booth and remained awake and alert throughout recordings. CAPs were acquired using a single-channel electrode montage: right TM electrode (+), left earlobe (-), and forehead (⩽). Waveforms were sampled at a rate of 40 kHz over a 25.6 ms epoch, filtered from 0.1-3 kHz, and amplified by 150,000. Stimulus presentation rate was
Gaussian distributed from 9.1/s to 13.1/s with a mean rate of 11.1/s. This relatively slow range of presentation rates was selected to ensure that the stimuli did not temporally summate to activate the MOC reflex, which has been shown to affect OAE measurements at stimulus presentation rates as low as 30/s to 50/s (Veuillet et al., 1991; Frances & Guinan, 2010; Boothalingam & Purcell, 2015). A Gaussian-distributed (i.e., “temporally jittered”) presentation rate was selected to facilitate subject alertness, as this may influence MOC reflex strength (Aedo et al., 2015).

CAP level-series measurements without and with CAS (1-10 kHz flat spectrum BBN at 60 dB SPL, delivered to left ears through an ER-2 earphone) were interleaved throughout the duration of each two-hour session with the exception that the first 20 minutes of the sixth session was devoted to CEOAE measurements (described in Section 2.2.4.). A 60 dB SPL CAS presentation level is commonly used for MOC reflex experiments, as it is the highest BBN level, on average, that elicits MOC reflex activity without triggering the middle ear muscle reflex (Guinan, 2006). CAP level-series were obtained using a chained stimulus paradigm (Hamil et al., 1991), which randomized stimulus levels from 50-80 dB ppeSPL using 10 dB steps. Each of the interleaved recording blocks automatically stopped after 2,048 averages were collected at each of the four stimulus levels and a 120 second break was inserted between each interleaved trial to allow subjects to reposition, etc. Advantages of using the chained paradigm in this context were that complete level-series functions could be obtained relatively quickly (~12 minutes) in a single testing block and that the effects of electrophysiologic or myogenic noise were randomly distributed
across responses to all stimulus levels as opposed to one. In a typical recording session, three to four pairs of level-series functions without and with noise were obtained and averaged at the end of the session. Each recording session thus resulted in eight grand average waveforms (2 conditions x 4 stimulus levels) with each grand average waveform being comprised of approximately 6,144-8,192 sweeps. At the end of six recording sessions, there were 48 waveforms (2 conditions x 4 stimulus levels x 2 stimulus types x 3 sessions) for each subject.

The 48 CAP waveforms for each subject were saved as ASCII files and analyzed offline in MATLAB. CAP waveforms were grouped based on stimulus type (click or chirp), level (50-80 dB ppeSPL), and whether they were obtained without or with CAS. CAP amplitudes for each waveform were expressed in two different ways: 1) **Raw amplitude** was calculated as the $\mu$V difference between the pre-stimulus baseline average amplitude and the N1 peak, which was automatically selected as the largest waveform minimum within a restricted time epoch at each level based on normative click and chirp latency data from our laboratory. Responses were “not present” if the raw amplitude of a peak was less than one standard deviation of the pre-stimulus baseline amplitude. 2) **Normalized amplitude** expressed each CAP peak magnitude as a percentage of the maximum raw amplitude (either without or with CAS) in the level-series in which it was acquired:

$$\text{Normalized Amplitude} = \left( \frac{\text{Raw CAP Amplitude (uV)}}{\text{Single Session Level Series Maximum Amplitude (uV)}} \right) \times 100$$
Treating the data in this manner produced normalized level-series functions for each subject at the end of each recording session. We hypothesized two advantages to this approach. First, normalizing data obtained in each recording session would be expected to minimize differences in raw CAP amplitudes within subjects that were due to changes in electrode placement or orientation in the ear canal between visits, which can significantly influence raw amplitudes (e.g., Alhanada, 2012). Second, a normalized scale would be expected to make level-series functions between subjects more similar; because we analyzed group data in this experiment, it was imperative to reduce the effects of inter-subject differences in raw amplitude on our results.

**CEOAE Measurements**

Three pairs of CEOAE level-series functions (60-80 ppeSPL\(^6\)) without and with CAS were obtained using a Mimosa Acoustics HearID System (Mimosa Acoustics, Inc. Champaign, IL). Responses were collected using “linear” clicks (i.e., consistent stimulus polarity and level across all presentations) presented at 11/sec for 250 sweeps in each trial. CEOAEs were considered present if they were ≥ 6 dB above the noise floor and if emission waveform sub-averages from response bins A and B were ≥ 80% correlated. CEOAE files were saved and offline analyzed in MATLAB, which extracted the composite values representing total emission amplitude and noise floors for each level and CAS condition.

---

\(^6\) CEOAE responses to clicks at 50 dB ppeSPL were absent in most subjects based on our criteria; therefore, CEOAE level-series measurements were made from 60-80 dB ppeSPL.
All response amplitudes were converted from dB to a pressure scale in order express CEOAE level-series on a linear ordinate scale, as was done with CAPs.

Analyses

Chirp-to-Click CAP Amplitude Ratios and Amplitude Comparisons

Chirp-to-click CAP amplitude ratios were calculated for each stimulus level using grand-averaged chirp and click raw amplitudes obtained without CAS for each subject. The purpose of this analysis was to determine the relative amplitude advantage of the chirp at each stimulus level. Paired t-test comparisons between chirp and click raw amplitudes without CAS at each level were also conducted.

CAP Inhibition Measurements: Amplitude Reduction and Effective Attenuation

The first step in testing the hypothesis that chirp-evoked CAPs were more sensitive to MOC reflex inhibition than clicks was to determine whether level-series functions were less variable when expressed either in units of raw amplitude or normalized amplitude. While we expected that normalizing amplitudes for each recording session would decrease between-subject CAP amplitude variability and provide a better scale on which to analyze group data, this was tested empirically. Coefficients of variation, which allow for variability comparisons between data sets with different units (e.g., μV versus %), were calculated at each level and compared for raw and normalized level-series functions for each stimulus type. The amplitude scale producing the smallest coefficients of variation at each stimulus level and across all stimulus levels was used in subsequent analyses of
CAP inhibition under the assumption that the less variable scale would be more sensitive to “true” physiologic changes induced by the MOC reflex.

Group CAP inhibition for chirps and clicks was quantified using two measures reported in the literature: 1) **Amplitude reduction** was calculated as the average “vertical” (ordinate) difference in CAP amplitudes without and with CAS at each level of the level-series function. This method of quantifying MOC reflex strength is most commonly used in the OAE inhibition literature; 2) **Effective attenuation** of chirp and click CAPs was calculated as the “horizontal” (abscissa) difference between linear regression fits to level-series without and with CAS using all subject data. Effective attenuation expresses the amount of dB that the stimulus would need to be increased to overcome the effects of MOC reflex inhibition; it is therefore useful in quantifying inhibition in terms of input level, which allows for gross comparisons of pre-neural and neural responses on the same scale (e.g., Puria et al., 1996).

5.3. **Results**

**Amplitude differences between Chirp and Click Evoked CAPs**

With few exceptions, chirps produced larger raw peak amplitudes than clicks in individual ears, as evidenced by chirp-to-click CAP amplitude ratios (Figure 5.1). The size of the chirp/click amplitude ratio differed between subjects and showed a range of 0.76-4.22 across all stimulus levels. For most participants, the amplitude ratios decreased slightly as level was increased. Note that click-evoked responses at 50 dB ppeSPL were
separable from the noise floor in all three test sessions in only 9 of the 14 participants; thus, amplitude ratios were calculated for only 9 participants at this level.

![Figure 5.1](image.png)

Figure 5.1. Chirp-to-click CAP amplitude ratios for waveforms obtained without CAS. Each symbol represents a single subject. Symbols falling above the dotted line indicate larger chirp responses than clicks.

The mean raw amplitudes of chirp-evoked CAPs without CAS were larger than those for clicks at each level tested (Figure 5.2). Paired t-tests with Bonferroni corrections for multiple comparisons ($\alpha = 0.0125$) revealed that these differences were significant at 50 ($t(8) = -2.85, p=0.008$), 60 ($t(13) = -7.19, p=0.0009$), 70 ($t(13) = -4.28, p=0.001$), and 80 dB ppeSPL ($t(13) = -2.57, p=0.007$).
Figure 5.2. The average amplitudes for CAPs in response to chirps (▲) and clicks (●) without CAS are shown as a function of level. Chirp raw amplitudes were significantly larger than click raw amplitudes at every level using a corrected alpha level for multiple comparisons ($\alpha = 0.0125$). Error bars = SEM; NF = Noise Floor.

**CAP Inhibition**

Representative chirp- and click-evoked CAP waveforms without and with CAS from a randomly selected participant are plotted in Figure 5.3. This figure demonstrates three pertinent observations that were noted in most subjects including: 1) the overall amplitude advantage of chirps, especially at lower stimulus levels, 2) the small reductions in chirp- and click-evoked CAP amplitudes with CAS, and 3) the stability of pre-stimulus baselines prior to the N1 peak of the CAP.
Figure 5.3. Average CAP waveforms evoked by chirps (top) and clicks (bottom) for a representative subject. Small reductions in N1 peak amplitudes with CAS can be seen in the chirp-evoked CAP waveforms at lower stimulus levels, but not for the click at any level. Note that at 50 dB ppeSPL, a CAP was not identified in this subject (Black = without CAS; Red = with CAS).

Figure 5.4 displays chirp and click average level-series functions across all subjects and sessions without and with CAS. Level-series functions are expressed in both normalized and raw amplitudes for each stimulus type. For chirps, the average coefficient of variation across four stimulus levels and two noise conditions was 45% when expressed in raw amplitude and 20% when expressed in normalized amplitude; this mean difference was significant ($t(14) = 3.46, p = 0.0038$). For clicks, the average coefficient of variation was 47% when expressed in raw amplitude and 29% when expressed in normalized...
amplitude, which was also a significant mean difference ($t(14) = 2.86$, $p = 0.013$). Thus, we used the less-variable measurements expressed in normalized amplitude for subsequent MOC reflex inhibition of CAP analyses.

Figure 5.4. Average level-series functions for chirps (top) and clicks (bottom) expressed in raw (left) and normalized (right) amplitudes. Response variability was smaller across all subjects when amplitudes were expressed on a normalized scale. Error Bars = SEM; NF = Noise Floor.
Normalized CAP Amplitude Reductions

Average normalized amplitude inhibitions were largest for stimulus levels below 80 dB ppeSPL for both chirps and clicks (Fig 5.5). Normalized amplitude reduction with CAS was statistically significant only for chirp-evoked responses at 50 ($t(30)=3.55, p=0.006$) and 60 dB ppeSPL ($t(38)=4.18, p<0.0001$), respectively, using an alpha level ($\alpha = 0.0125$) to account for multiple comparisons.

Figure 5.5. Average normalized chirp and click amplitude reductions for all subjects. Asterisks indicate significant reductions ($p < 0.01$). Note that at every stimulus level, chirp inhibition was less variable than clicks, as indicated by the 95% confidence interval bars.
CAP Effective Attenuation

Separate linear regression models were fit to the normalized group level-series data obtained without and with CAS for chirps ($y=1.33x-11.96$, $R^2 = 0.47$; $y=1.59x-33.22$, $R^2=0.57$) and clicks ($y=1.77x-50.13$, $R^2=0.51$; $y=1.86x-60.75$, $R^2=0.54$), respectively (Figure 5.6 A-B). For both stimulus types, the models fit to CAP amplitudes without and with CAS diverged at low stimulus input levels and converged at higher stimulus input levels, indicating a greater effect of CAS on CAP amplitudes at low input levels. Regression coefficients as a function of condition (without or with CAS) were not significantly different for chirps ($t=1.63$, $p=0.103$) or clicks ($t=0.45$, $p=0.66$). Effective attenuation for each stimulus type was calculated as the difference in the abscissa between without and with CAS linear regression lines for equivalent ordinate values (Figure 5.6 D). At the lowest stimulus level, effective attenuation was 5.07 dB for chirps and 3.02 dB for clicks (Figure 5.6 D).

---

Note that while normalizing CAP amplitudes to the maximum value in a subject's level-series function reduced amplitude variation across all levels, it also introduced heteroscedasticity; therefore, robust standard errors were used for each regression model, which allowed for the presence of heteroskedastic data by relaxing the assumptions that errors were independent and identically distributed (Hayes & Cai, 2007).
Comparison of CAP and CEOAE Effective Attenuation

Based on our findings that CAP amplitudes were less variable when expressed on a normalized scale (see Section 3.2), we only report CEOAE inhibition in terms of effective attenuation of normalized responses in the present experiment for comparison. CEOAE normalized level-series data from all subjects obtained without ($y=2.84x-131.42, R^2=0.78$)
and with CAS \( (y=2.92x-141.87, R^2=0.80) \) were also fit with separate linear regression models (Fig 6C). The CEOAE models were better fit than CAP data, as normalized CEOAE amplitudes were less variable across subjects. The largest differences in without and with CAS models occurred at the lowest input level, as was observed in the CAP data. Regression coefficients as a function of condition (without or with CAS) were not significantly different \( (t=0.43, p=0.67) \). Effective attenuation was calculated in the same manner as the CAP data. A comparison of chirp-evoked CAP, click-evoked CAP, and CEOAE effective attenuations at 60 dB ppeSPL revealed that inhibition was largest for chirps (3.42 dB), followed by click CAPs (2.49 dB) and CEOAEs (1.93 dB).

5.4. General Discussion

The findings of this study were that: 1) Chirps evoked larger CAP amplitudes than clicks at low to moderate stimulus levels; 2) Normalized CAP amplitude reductions with CAS were largest at the group level using chirps at 50 and 60 dB ppeSPL (5.89 and 7.75%, respectively). These were the only statistically significant amplitude reductions observed; 3) Effective attenuation measurements were largest at the group level for chirp-evoked CAPs followed by click-evoked CAPs and CEOAEs, respectively, at the lowest stimulus levels where all three could be measured (i.e., 60 dB ppeSPL).

The Chirp Advantage

The chirp generated larger CAP amplitudes at each stimulus level in most subjects; however, the size of this advantage varied considerably across subjects. This finding is
consistent with the observations of Chertoff and colleagues (2010; see Figure 5.3) who used higher presentation levels than the present study. Inter-subject differences in the chirp advantage may be related to multiple factors. First, the chirp used in the present study and by Chertoff and colleagues (2010) related frequency to basilar membrane delay using derived band CAP latencies from 15 normally hearing subjects reported by Eggermont (1979). Subject characteristics, such as sex, were not reported in that study, but it has been inferred that sex differences in cochlear length may affect basilar membrane delays and therefore the degree to which neural responses are synchronized (e.g., Don et al., 1993, 1994). With the current participant pool of 10 females and 4 males, it is possible that the chirp was not optimized for individual ears based on these differences. One way to quickly construct a CAP chirp that is more optimized for an individual ear than a click may be to use basilar membrane delay estimates from otoacoustic emissions, as derived-band CAP masking procedures are time consuming.

Secondly, some authors have encouraged the use of chirps that are optimized for different presentation levels (Elberling et al., 2010; Elberling & Don, 2010; Kristensen et al., 2012), suggesting that cochlear frequency place maps do not scale simply with level. Our use of a chained stimulus paradigm, which allows for random level presentation of a single stimulus file, did not provide the flexibility to use multiple chirps optimized for different levels in this investigation.

The chirp advantage reported here and by Chertoff and colleagues (2010) suggests that chirps may also be a useful tool in studying animal and human synaptopathy – a
pathology in which noise exposure predominately insults high threshold auditory nerve fibers but spares low threshold fibers and hair cells (Kujawa & Liberman, 2009). Synaptopathy has been postulated as the basis of severe hearing difficulties in patients with normal audiograms (i.e., “hidden hearing loss”) and may also be involved in the generation of tinnitus (e.g., Schaette & McAlpine, 2011). The synaptopathy “phenotype” in animal models presents as significantly reduced CAP amplitudes evoked by suprathreshold sounds in the presence of normal (electrophysiologic) audiometric thresholds and OAE responses. Because chirp-evoked CAPs are larger and represent the summed activity from auditory nerve fibers along the length of the basilar membrane, they may provide a more sensitive measure of synaptopathy. Further, narrowband chirps tailored to evoke CAPs may be even more sensitive to synaptopathy in distinct cochlear regions.

**MOC reflex effects on CAPs**

Our findings suggest that chirps may be more suitable than clicks in studying the neural consequences of MOC reflex inhibition for a few reasons. First, chirp-evoked CAPs were larger than clicks even at the lowest stimulus level, which allowed for more accurate N1 peak identification in quiet and with CAS conditions (e.g., compare 60 dB ppeSPL waveforms for chirps and clicks in Figure 5.3). Since OHCs are more potently inhibited by the MOC reflex at low input levels, using a chirp may allow for more accurate estimates of MOC effects in this range. Because chirp-evoked CAPs reflect the summed activity over broader cochlear regions, they may also be more sensitive to the summated effects
of MOC fibers than click-evoked CAPs, which mainly reflect neural synchrony from fibers innervating the cochlear base (Don & Eggermont, 1978). Second, the variability of CAP inhibition for chirps was smaller relative to clicks on both amplitude reduction and effective attenuation measurement scales (Figures 5.5 & 5.6A-B). This finding suggests that chirps may be more sensitive to “true” physiologic changes attributable to MOC reflex activation than clicks. It is important to note, however, that chirp and click CAP effective attenuation was calculated from relatively weak linear regression fits to group data, which may have been caused by individual differences in both level-series function contours and magnitude of inhibition. An analysis of individual data using the same method resulted in even poorer linear fits due to the fewer data points in the models. Thus, a limitation of our work is that we were unable to reliably resolve efferent inhibition of CAPs at the single-subject level, which is of interest in studying individual variation in MOC function and in understanding the predictive relationships between pre-neural and neural efferent assays. This issue may have been resolved by focusing recording time on obtaining more response averages to fewer low-intensity input levels (Lichtenhan et al., 2016); however, an advantage of acquiring a level-series function spanning 40 dB was that CAS effects on CAPs evoked by different stimulus levels could be evaluated.

Involvement of the middle ear muscle reflex is always a consideration in MOC reflex experiments, as CAS can activate both mechanisms. The observed CAP amplitude reductions with CAS were unlikely to be the consequence of “sub-threshold” middle ear compliance changes from activation of the middle ear acoustic reflex because such a
change would be expected to reduce responses to all input levels of the level-series function. In contrast, the CAS-induced changes in our data were primarily at low input levels, which is suggestive of changes in OHC function. Nevertheless, the possibility of middle ear muscle involvement cannot be fully ruled out, as some reports indicate that standard measures of acoustic reflex threshold, like the one used in our screening protocol, may overestimate the level at which the stapedius muscle is activated by CAS (Feeney & Keefe, 2001; Zhao & Dhar, 2010).

**CEOAE and CAP Effective Attenuation Comparisons**

OAE measurements are used far more often as an indirect assay of MOC reflex effects than CAP amplitudes, as they require less time to collect and are less inherently variable than electrophysiologic techniques (see Figure 5.6). This difference is presumably because far-field CAP recordings are influenced by more sources of noise (e.g., background EEG, myogenic and electrical noise, high electrode impedance due to small surface area) than OAEs. Based on these technical differences, a compelling argument can be made for using OAE based assays of the MOC reflex in a clinical setting, for example. It is, however, of great importance to understand the relationships between pre-neural and neural inhibition because the latter reflects modulation of the auditory nerve signal mediating hearing, which cannot be assessed with OAEs. Pre-neural and neural inhibition comparisons must be made in light of evidence that there is not a one-to-one correspondence between changes in OHC function and modulation of IHC neurotransmitter release, which is the basis for auditory nerve fiber depolarization.
(Guinan, 2012). However, by expressing CEOAE and CAP inhibition in terms of effective attenuation, direct comparisons can be made between MOC reflex effects on each type of response.

Our observation that CEOAE effective attenuation underestimated chirp- and click-evoked CAP effective attenuation by up to approximately 1.5 dB at low stimulus levels was consistent with previous reports in animals and humans (e.g., Puria et al., 1996; Lichtenhan et al., 2016). The source of this consistently reported discrepancy is not clear. OHCs are postsynaptic only to MOC fibers, whereas the auditory nerve is post synaptic to both MOC and lateral olivocochlear (LOC) fibers, which directly contact type I auditory nerve fibers (Warr & Guinan, 1979). This anatomical configuration suggests that CAP inhibition reflects the summation of MOC and LOC inhibition, whereas OAEs only reflect MOC inhibition. However, several lines of evidence appear to refute this suggestion. Gifford and Guinan (1987) measured CAP inhibition from cats while electrically stimulating different regions of the caudal brainstem. They observed that stimulating the floor of the fourth ventricle (which diffusely activates the OCB proper) is comparable to the combined inhibitory effects of directly stimulating MOC neurons. When LOC neurons were directly stimulated, no inhibitory effects on the CAP were observed. The investigators also documented that increases in cochlear microphonic amplitude were related to decreases in CAP amplitude during OCB stimulation, indicating that the same process (i.e., direct modulation of OHCs) likely mediates each effect. Brown and colleagues (1983) measured IHC receptor potential tuning curves (from the AC
component) with and without fourth ventricle electrical stimulation and observed 9-24 dB of inhibition at the tuning curve “tips” (i.e., center frequency) with no change away from center frequencies. Basilar membrane displacement tuning curves show similar effects (Cooper & Guinan, 2003; Murugasu & Russell, 1996). While these measurements are pre-neural, they are remarkably similar to auditory nerve tuning curves using the same paradigm (Bonfils et al., 1986; Wiederhold & Kiang, 1970). Thus, there is strong evidence that the MOC system is the main effector of inhibition in both pre-neural and neural assays. In contrast, there is no evidence that the LOC system can be excited with acoustic stimulation and its role in hearing remains poorly understood. The best available evidence suggests that the LOC system’s influence on hearing is likely through slow “top-down” potentiation of auditory nerve activity (Sahley & Nodar, 1994; Le Prell et al., 2003; Groff & Liberman, 2003), which may protect auditory nerve fibers from acoustic trauma (e.g., Darrow et al., 2007).

If the MOC reflex accounts for inhibition measured from both OHCs and the auditory nerve, it may be expected that effective attenuation slopes of CEOAEs and CAPs would be parallel. We observed at the group level that the slopes of click CAP and CEOAE effective attenuation were similar to each other and quite different than chirp CAP effective attenuation (Figure 5.6D). Because we did not measure chirp-evoked OAEs, it is unclear if this difference is stimulus related or explained by some other mechanism. The temporal differences between clicks and chirps make the chirp a better stimulus for evoking synchronized neural responses, but these differences would not be expected to
produce significantly dissimilar composite emission amplitudes evoked by each stimulus. Previous work has indicated that stimulus frequency OAEs (SFOAEs) and CEOAEs are generated in a nearly equivalent manner through coherent reflection when the spectral power within a bandwidth on the basilar membrane is equal (Neumann et al., 1994; Kalluri & Shera, 2007); if a chirp is conceptualized as a swept SFOAE, the effect of MOC reflex inhibition on chirp-evoked OAEs and CEOAEs would be expected to be similar. To our knowledge, there have been no experiments comparing MOC reflex inhibition of click- and chirp-evoked OAEs; therefore, the origin of the differences in effective attenuation slopes between chirp-evoked CAPs and preneural and neural measurements evoked with clicks in the group data is not clear.

5.5. Conclusions

The present study is the first in which a chirp was used to evoke CAPs from the human auditory nerve with and without MOC reflex activation. Our findings indicate that, at least at the group level, the chirp may be a more sensitive stimulus for evaluating neural efferent effects than a click because it evokes a larger response at lower stimulus intensities and may be more sensitive to summed efferent activity along the cochlear spiral. Additionally, our findings are consistent with previous work indicating that OAE assays of the MOC reflex underestimate neural inhibition (i.e., Puria et al., 1996; Lichtenhan et al., 2016). Future experiments which optimize chirp parameters for individual ears and allow for reliable within-subject neural measurements of MOC reflex inhibition are warranted.
CHAPTER 6: EFFECTIVE MODULATION OF COMPLEMENTARY PRE-NEURAL AND NEURAL DISTORTION PRODUCTS

6.1. Introduction

The medial olivocochlear (MOC) bundle is an inhibitory neural circuit originating in the mammalian auditory brainstem and terminating directly onto cochlear outer hair cells (OHCs). When activated, the MOC bundle alters OHC motility and indirectly influences basilar membrane motion and inner hair cell (IHC) sensitivity – an effect termed the MOC reflex (Mountain, 1980; Siegel & Kim, 1982; Murugasu & Russell, 1996; Cooper & Guinan, 2003, 2006). Experimental work in animal models has demonstrated that although the MOC reflex modulates pre-neural signal processing via direct control of the cochlear amplifier, this mechanism may potently influence “downstream” neural encoding of sound. For example, MOC reflex activation results in “unmasking” of signals-in-noise represented in the auditory nerve (Dolan & Nuttall, 1988; Kawase & Liberman, 1993; Ferry & Meddis, 2007), and this effect appears to be preserved in the behavior of inferior colliculus neurons (Seluakumaran et al., 2008). One proposed function of the MOC reflex based on this data is to peripherally reduce mechano-electrical transduction of noise, which may enhance listening in complex acoustic environments.

The human MOC reflex has been studied almost exclusively using pre-neural assays of OHC function such as otoacoustic emissions (OAEs). OAE-based measurements have revealed characteristics of MOC reflex tuning (e.g., Veuillet et al. 1991; Chéry-Croze et al. 1993; Lilaonitkul and Guinan 2007; Zhao & Dhar, 2012), strength (e.g., Backus &
Guinan, 2007; Marshall et al., 2014), and the mechanism’s possible involvement in
directed auditory attention (e.g., Froehlich et al., 1993; de Boer & Thornton, 2007; Garinis
et al., 2011) or listening in noise tasks (e.g., Giraud et al., 1997; Kumar & Vanaja, 2004;
de Boer & Thornton, 2008; Smith & Cone, 2015; de Boer et al., 2012). However, a major
limitation of this approach is that OAEs are insensitive to MOC reflex effects on the neural
ensembles that mediate human hearing. To directly assess whether efferent-induced
changes in inner ear mechanics are maintained downstream, it is necessary to
understand the relationship between complementary pre-neural and neural
measurements of the MOC reflex (Lichtenhan et al., 2015; Smith et al., 2016).

**DPOAE Inhibition – A pre-neural MOC reflex assay**

Distortion product OAEs (DPOAEs) are sounds putatively generated by cochlear OHCs
and recorded in the ear canal. They are evoked by two simultaneously presented pure
tones, f1 and f2 with an f2/f1 frequency ratio of 1.22 generating the most robust distortion
products (Kemp, 1978; Probst et al., 1991). The 2f1-f2 distortion product is most
commonly obtained for clinical and research purposes due to its robustness, although
others have been reported in the human literature (Wittekindt et al., 2009). Distortion
products other than 2f1-f2, such as the quadratic distortion product at f2-f1, may prove to
be more informative regarding changes in the in vivo transducer function; however, these
measurements have not been optimized in humans (see Bian & Chen, 2008; Abel et al.,
2009).
The contemporary “two source” model of DPOAE generation posits that the ear canal-recorded emission is a mixture of components arising from at least two cochlear initiation sites (Kemp, 1986; Zweig & Shera, 1995; Mauermann et al., 1999 a,b; Shera & Guinan, 1999; Talmadge et al., 1999). The first is the “distortion source” generated by intermodulation distortion of overlapping f1 and f2 traveling waves (Figure 6.1a), which arises due to the nonlinear growth characteristics of OHC amplification within this region (see Bian et al., 2002, Appendix A). The second is the “reflection source”, which arises from coherent scattering of distortion source energy as it reaches its center frequency place on the basilar membrane and travels backward. The recorded DPOAE represents the vector summation of distortion and reflection source magnitudes and phases as they propagate peripherally from the cochlea and into the ear canal (Kalluri & Shera, 2001). When f1 and f2 are continuously swept or adjusted in small frequency steps, the rate of phase rotation of each source differs significantly and a quasi-periodic “fine structure” of maxima and minima emerges in the DPOAE magnitude-frequency function (Figure 6.1b) due to constructive and destructive interference of energy from both sources (Brown et al., 1996; Heitmann et al., 1998; Talmadge et al., 1999; Knight & Kemp, 2000, 2001; Dhar & Shaffer, 2004).
Figure 6.1. Pre-neural and neural correlates of the two source model.  

(a) F1 and f2 stimuli (blue and red waveforms, respectively) mix in the ear canal to form a two-tone complex waveform (purple). The complex waveform is decomposed into two cochlear traveling waves representing f1 (blue arrow) and f2 (red arrow) frequencies. Energy from the distortion source, where f1 and f2 traveling waves overlap, propagates backward toward the ear canal and forward to the CDP place. The reflection source arises from coherent scattering of energy at the CDP place, which also propagates toward the ear canal and mixes with distortion source energy. Auditory nerve fibers tuned to f1, f2, and CDP center frequencies feed each component forward into the neural code.  

(b) The DPOAE fine structure represents peaks at which distortion and reflection sources constructively (black arrow) and destructively (subsequent trough) interfere.  

(c) Phase locking to f1, f2, and CDP components initiated by auditory nerve fibers is represented in the ensemble behavior of auditory brainstem nuclei and recorded from the scalp as the FFR. (Note: The f2-f1 (or ASSR) potential corresponding to the amplitude modulated envelope of the two tone stimulus (purple) is not shown.

Because of the likely relationship between DPOAEs and OHC motility (Liberman et al., 2002; Cheatham et al., 2004), DPOAEs are sensitive to changes in OHC activity induced by the MOC reflex and provide a non-invasive method to study efferent effects in humans.
Most commonly, DPOAEs are recorded first in quiet and then during presentation of a contralateral acoustic stimulus (CAS; e.g., broadband noise), which activates the uncrossed fibers of the MOC bundle (Puel & Rebillard, 1990; Chery-Croze et al., 1993; Williams and Brown, 1995, 1997; Guinan, 2006). Amplitude and phase differences between DPOAEs recorded in quiet and with CAS can then be used to quantify attenuation of the cochlear amplifier mediated by the MOC reflex.

An important caveat to consider when measuring DPOAE inhibition is that distortion and reflection sources are differentially affected by CAS (Sun, 2008a; Abdala et al., 2009; Deeter et al., 2009). This occurs because OHCs at distortion and reflection source sites are respectively operating at different “points” on the cochlear amplifier input-output function and the MOC reflex more potently inhibits responses at lower intensities on this function. CAS-induced changes in distortion and reflection source amplitudes and phases can thus combine to cause “artefactual” differences in DPOAE fine structure resulting in gross over- or underestimation of MOC reflex strength (Muller et al., 2005; Wagner et al., 2007; Sun, 2008a; Abdala et al., 2009; Deeter et al., 2009). This issue is minimized and a “true” estimate of total inhibition is achieved when analyses are conducted at spectral fine structure peaks where reflection and distortion source energy are in phase and constructively interfering. At DPOAE fine structure peaks, average measured inhibition in young, normally hearing adults is ~ 0.5-2.5 dB, depending on frequency and probe tone levels (Lisowska et al., 2002; Zhang et al., 2007; Sun, 2008a; Deeter et al., 2009; Abdala et al., 2009).
Neural representation of auditory distortion products – A potential assay of “downstream” MOC reflex effects

The neural correlates of DPOAEs and their evoking tone pairs can be recorded from electrodes placed on the human scalp (Krishnan, 2007). These “frequency following responses” (FFRs) represent cycle-to-cycle phase locking of auditory nerve and brainstem ensembles to primary tones (F1-FFR and F2-FFR) and multiple distortion product frequencies (Figure 6.1c). Observations in animals and humans suggest that the 2f1-f2 cubic distortion product FFR (CDP-FFR\(^8\)) is initiated by or tightly coupled to the same nonlinear cochlear processes that generate the DPOAE reflection source component (Elsisy & Krishnan, 2008). For example, the presence of CDP-FFRs is contingent upon monotic presentation of probe tones with f2/f1 ratios consistent with those used to evoke DPOAEs (Rickman & Chertoff, 1991; Chertoff et al., 1992; Bhagat & Champlin, 2004). Auditory nerve fibers tuned to the 2f1-f2 center frequency place phase-lock at a rate equivalent to the 2f1-f2 period, and the amplitude and phase characteristics of these responses are nearly indistinguishable from those evoked by acoustic pure tones of the same frequency (Goldstein & Kiang, 1968; Kim, 1980; Kim et al., 1980). These findings suggest that the energy evoking CDP-FFRs is present in the gross motion of the 2f1-f2 basilar membrane place, as is the case with the reflection source of the DPOAE. While the distortion source largely contributes to the overall DPOAE amplitude, it is

\(8\) We use the notation “CDP-FFR” to refer specifically to the neural 2f1-f2 distortion product. We acknowledge that other cubic (and quadratic) distortion products can sometimes exist in the scalp recorded response.
unlikely that the 2f1-f2 period is represented in discharge patterns of auditory nerve fibers tuned to the f1 and f2 traveling wave overlap area (Goldstein & Kiang, 1968; Kim et al., 1980). Thus, 2f1-f2 DPOAEs appear to arise from a combination of at least two sources, whereas CDP-FFRs may only represent phase locking initiated by the reflection source place (Elsisy & Krishnan, 2008).

F1-FFRs, F2-FFRs, and CDP-FFRs initiated in the auditory nerve are maintained in the phase locking behavior of ventral cochlear nucleus and inferior colliculus neurons (Smoorenburg et al., 1976; Arnold & Burkard, 1998; Faulstich and Kossi, 1999). Given this redundancy, sources of these scalp-recorded potentials are difficult to disentangle and likely represent overlapping responses from ensembles along the caudo-rostral neuraxis (Stillman et al. 1978; Gardi et al. 1979; Batra et al. 1986; Galbraith et al. 2000, 2001; Bidelman 2015; Shaheen et al., 2015). Group delay and latency data suggest that stimulus frequency can be manipulated to bias FFR generators: the rostral brainstem dominates FFR responses to frequencies below ~500 Hz, above which more caudal sources contribute (Batra et al., 1986; King et al., 2016). Beyond ~1000-1500 Hz, neural phase locking becomes poor and pre-neural potentials such as the cochlear microphonic and summating potential can dominate the response. The relationship between stimulus frequency and the dominant neural generator of the scalp recorded response poses important technical considerations in studies comparing DPOAEs and their FFR correlates: F1 and f2 frequencies must not be so high that robust neural responses cannot be obtained; conversely, they must not be so low that DPOAEs have poor signal
to noise ratios (SNRs) and CDP-FFRs are dominated by rostral generators further away from their presumed cochlear source.

Few studies have directly compared DPOAEs and CDP-FFRs and fewer have done so for the purpose of evaluating pre-neural and neural effects of the MOC reflex. Elsisy and Krishnan (2008) simultaneously measured input-output functions of DPOAEs and CDP-FFRs evoked by low frequency tone pairs. They reported that average DPOAE growth functions were biphasic, whereas average CDP-FFR growth functions were compressive. The authors speculated that the biphasic DPOAE input-output functions may have been influenced by complex interactions between distortion and reflection source components, whereas the CDP-FFR growth functions represented activity only related to the reflection source place. This issue likely also influenced the results of an experiment by Elsisy and Krishnan (2005), wherein CAS during simultaneous DPOAE and CDP-FFR recordings produced DPOAE enhancements in some subjects who also showed CDP-FFR inhibition. Measuring DPOAE inhibition only at fine structure peaks may have provided a more straightforward comparison between the two measurements. In a similar study, Bhagat and Champlin (2004) reported CDP-FFR inhibition with CAS in subjects who also demonstrated DPOAE inhibition. However, the authors differentially optimized stimulus parameters for each type of recording by using different f2/f1 ratios, probe frequencies and levels and did not measure inhibition at fine structure peaks, making it difficult to directly assess pre-neural and neural effects of the MOC reflex.
Objective

We hypothesized that pre-neural and neural distortion products would each be inhibited with activation of the efferent system and that inhibition magnitudes would be related given the apparent coupling of both responses. By first measuring CAS-induced DPOAE inhibition at fine structure peaks, we were able to avoid responses influenced by vector summation of out of phase source components and thus measure the “true” MOC reflex pre-neurally. The f1 and f2 combination producing the largest DPOAE fine structure peak for each subject was then used to evoke FFRs with and without CAS to provide a complementary neural measure of efferent inhibition.

6.2. Methods

Subjects

The University of Arizona Human Subjects Protection Program approved the following methods. Thirty-one participants, 18 females and 13 males, were recruited for the study. Participant age ranged from 19 – 27 years (Median = 21.4 years). Screening inclusion criteria for enrollment in the study were: (1) No history of neurologic or otologic disease (2) normal otoscopy with no evidence of outer or middle ear disease or cerumen occlusion (3) bilateral air conduction pure tone sensitivity at 20 dB HL or better from 250-8000 Hz (4) type-A tympanograms, defined as peak pressure between ±50 daPa and peak-compensated static acoustic admittance between 0.4-1.5 mmhos, and (5) a reproducible acoustic reflex threshold to broadband noise ≥ 70 dB SPL. Criterion five is based on the consideration that middle ear muscle activation can attenuate DPOAE amplitudes as well
as probe tones and therefore confound cochlear efferent measurements (Sun, 2008b; but also see Zhao & Dhar, 2010).

**DPOAE Fine Structure Measurements**

During the first visit, DPOAEs were recorded from the right ears of subjects seated comfortably in a sound booth with an ILO 88 OAE System (Otodynamics, Ltd, Hatfield, UK). Proper probe placement was achieved by ensuring that an 80 dB peSPL click stimulus was within +/- 1 dB of target level and that the ear canal transfer function spectrum was visibly flat. DPOAEs were evoked by pure tone pairs (f1<f2; f2/f1=1.22) each calibrated in situ to a level of 70 dB SPL. More intense primary tone levels were chosen for this portion of the experiment to ensure that robust neural responses could be recorded using the same stimuli in the second portion of the experiment. Three “microstructure” acquisition windows (193-195 Hz wide) centered at f2= 800, 1000, and 1200 Hz, respectively, were used to obtain DPOAEs with 12.2 Hz spectral resolution and therefore reveal emission fine structure within the restricted frequency range of 708 Hz ≤ f2 ≤ 1294 Hz when the windows were combined. This frequency range is lower than the typical “DP-gram” and generally produces recordings with poorer SNRs relative to higher frequency tone pairs (Probst et al., 1990; Gorga et al., 1997). However, restricting f2 below 1500 Hz ensured that the DPOAE stimulus pair producing the largest fine structure peak could also be used to generate sufficiently robust neural responses in the FFR portion of the study.
The three DPOAE microstructure windows collectively spanning 708 Hz < f2 < 1294 Hz were obtained in ascending order in four interleaved triads of quiet and with CAS trials. The CAS was a 60 dB SPL flat spectrum broadband noise (0-10,000 Hz) presented continuously throughout the recording to the left ear through an ER-2 insert earphone (Etymotic Research, Elk Grove Village, IL). DPOAEs were monitored online, and each point was averaged for approximately 3-4 seconds. Up to five sweeps were obtained for each microstructure window to ensure that noise floors were minimized. Subject data was only considered for further analysis if three consecutive spectral points were ≥ 6 dB above the average noise floor in at least one microstructure window. Failure to meet this criterion resulted in the exclusion of the subject’s data from further analyses.

DPOAE microstructure windows were saved as ASCII files and analyzed offline in MATLAB (The Mathworks, Inc., Natick, MA, USA). First, the three microstructure windows were combined and expressed as a continuous fine structure spectrum from 708 Hz < f2 < 1294 Hz for each quiet and with CAS trial. Second, the fine structure spectra for were averaged and estimates of error at each point were also calculated. This was done differentially for the trials in quiet and those with CAS. Third, the largest fine structure peaks in the quiet and with CAS average spectra were identified using MATLAB’s findpeaks function and labeled on the average spectral plots. If the largest amplitude value was at the upper (f2 = 1294) or lower (f2 = 708) margin of the fine structure plot, the second largest peak was selected. F1, f2, and 2f1-f2 frequencies, fine structure peak amplitudes, and inhibition (i.e., fine structure peak amplitude differences
between quiet and with CAS) were used in statistical analyses described below. Note that MOC reflex-induced frequency shifts in fine structure peaks have been reported in the literature (e.g., Abdala et al., 2009); however, these shifts are generally smaller (~5-10 Hz) than the spectral resolution of our recordings and were therefore not assessed.

**FFR Measurements**

During the second visit, FFRs were recorded from each subject using the Intelligent Hearing Systems Smart-EP advanced research module (Intelligent Hearing Systems, Miami, FL) with a single-channel electrode montage: hairline (+), C7 vertebra (−), and forehead (±). FFR stimuli were selected for each subject based on the f1 and f2 pair producing the largest quiet DPOAE fine structure peak in the previous visit. The stimuli were 92 ms pure tones shaped by a trapezoidal envelope with a rise and fall time of 15 ms and presented at 8.1/sec using condensation and rarefaction polarities. Both tones were delivered via separate shielded ER-2 insert earphones coupled to an ER 10-B+ probe microphone assembly and acoustically mixed in the ear canal (Etymotic Research, Elk Grove Village, IL). Prior to each recording, the tones were calibrated in-situ to 70 dB SPL using the probe microphone.

Throughout the FFR recordings, subjects remained awake and relaxed in a reclining chair in an electrically-shielded sound booth and electrode impedances were maintained at ≤ 3 kΩ. The EEG was monitored online and recording was paused during obvious subject movements. Individual sweeps with EEG peaks exceeding +/- 20 µV were rejected online.
as artifact. Response waveforms to condensation and rarefaction stimulus polarities were recorded in interleaved quiet and with CAS trials. Individual response waveforms consisted of 2048 artifact-free sweeps amplified 200,000 times and sampled at 40,000 Hz over a 102.4 ms epoch. Responses were online filtered from 10-5000 Hz and more conservatively filtered from 70-2000 Hz offline. Quiet and with CAS trials were repeated up to six times over the two-hour testing block and replicated waveforms were saved as ASCII files for offline analysis in MATLAB.

The MATLAB program first sorted waveforms into quiet and with CAS conditions and subtracted paired condensation from rarefaction responses within each condition to derive difference waveforms; this technique reduces envelope following responses at f2-f1 and accentuates F1-FFR, F2-FFR, and CDP-FFR “fine structure” components (Greenberg et al., 1987; Rickman et al., 1991; Krishnan, 1999; Pandya & Krishnan, 2004). Next, the difference waveforms for quiet and with CAS conditions were averaged, weighted with a Hanning window, and a 4096-point fast Fourier transform (FFT) analysis was performed, resulting in a spectral bin width of 9.77 Hz. The spectral peaks corresponding to f1, f2, and 2f1-f2 were automatically identified by the MATLAB routine based on each subject’s stimulus tone pair which was manually entered by the user. Amplitudes were expressed in dB re: 1 nV. F1-FFR, F2-FFR and the CDP-FFR were considered “present” if they were each ≥ 6 dB above the noise floor, which was calculated as the average amplitude of the five frequency bins above and below each component.
These FFR component amplitudes for quiet and with CAS conditions were used in statistical analyses described below.

**Evaluation of Equipment Distortion and Electromagnetic Stimulus Artifact**

OAE and EEG analyzers can themselves generate nonlinear distortion or artifacts mimicking biological responses and thus contaminate experimental data (Chertoff & Hecox, 1990). We assessed non-biological distortion and stimulus artifact contamination of OAE and EEG equipment in pilot experiments conducted prior to recruiting subjects for this study. Non-biological distortion of the ILO 88 system was assessed by sealing the OAE probe tip into a 2-cc coupler and measuring 2f1-f2 spectral amplitude for each of the 51 frequency pairs in the 708 Hz < f2 < 1294 Hz range presented at 70 dB SPL. The average amplitude for 2f1-f2 across all tone pairs presented in the coupler was -10.3 dB (± 2.7), indicating that non-biologic equipment distortion was minimal. To assess distortion related to stimulus presentation from the EEG equipment, we selected the highest (f1=1060 Hz; f2=1294 Hz) and lowest (f1=580 Hz; f2=708 Hz) possible tone pair combinations that could be used for the FFR portion of this experiment. The probe assembly was inserted into a 2 cc coupler, each tone pair was presented, and waveforms recorded by the ER-10B+ microphone were saved for offline spectral analysis in a custom MATLAB script (The Mathworks, Inc., Natick, MA, USA). Average spectral amplitude at 2f1-f2 for the two tone pairs was 5.2 dB (± 2.4 SD), which was inseparable from the spectral noise floor. Lastly, the possibility of stimulus-related electromagnetic artifact contamination of EEG responses was examined. Using the parameters described in
Section 2.3., FFRs were collected from one subject in rarefaction and condensation polarities with the highest and lowest possible tone pair combinations. Responses were recorded with earphone tubes patent and clamped and saved for offline analysis. With patent earphone tubes, all F1-FFR, F2-FFR, and CDP-FFR components were identifiable at ≥ 6 dB above the spectral noise floor except for F2-FFR = 1294 Hz. With clamped tubes, no responses were identifiable above the noise floor, suggesting that the EEG was not contaminated by non-biologic electromagnetic artifact.

Statistical Analyses

DPOAE, F1-FFR, F2-FFR, and CDP-FFR peak amplitudes identified at ≥ 6 dB above the noise floor were counted to determine the prevalence of these responses under our experimental conditions. A paired t-test was used to assess the effect of noise (quiet vs. with CAS) on DPOAE fine structure amplitude. F1- and F2-FFRs were analyzed separately from the CDP-FFR component due to differences in how the responses are initiated. For example, F1- and F2-FFRs correspond to pure tones with a known input level, whereas the CDP-FFR is initiated by a distortion product of an unknown quantity of energy on the basilar membrane. A two-factor repeated measures analysis of variance (ANOVA) was used to evaluate the effects of peak (F1-FFR and F2-FFR) and Noise (quiet vs. with CAS) on FFR amplitude. A paired t-test was used to assess the effect of noise (quiet vs. with CAS) on CDP-FFR amplitude.
DPOAE and FFR inhibition were calculated by subtracting response amplitudes with CAS from response amplitudes in quiet. Univariate regression analyses between DPOAE inhibition and each FFR peak inhibition measurement were conducted both with and without identified outliers, as identified by variable box plots.

6.3. Results

*DPOAE Fine Structure Peak Prevalence and Effects of CAS*

DPOAE fine structures from 22 (12 female) out of 31 recruited subjects (71%) met the 6 dB SNR acceptance criteria. Table 1 reports the frequency pairs producing the largest DPOAE fine structure peak for each subject, which mostly fell within the f2 region corresponding to the middle and high frequency microstructure measurement windows (i.e., $903 < f2 < 1294$). Average DPOAE fine structure amplitudes were 12.44 dB ($\pm$ 4.0) in quiet and 10.79 dB ($\pm$ 4.27) with CAS, which was a statistically significant difference ($t_{21} = 9.90, p < 0.005$). The average noise floors corresponding to quiet (-6.51 dB, $\pm$ 2.02) and with CAS (-6.02 dB, $\pm$ 2.40) fine structure peaks did not significantly differ.
Table 6.1. F1, f2, and 2f1-f2 frequency corresponding to the largest DPOAE fine structure peak in each subject. The bottom row represents the prevalence of each FFR peak (2f1-f2 = CDP-FFR).

<table>
<thead>
<tr>
<th>Subject</th>
<th>f1</th>
<th>f2</th>
<th>2f1-f2</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>818</td>
<td>1001</td>
<td>635</td>
</tr>
<tr>
<td>505</td>
<td>830</td>
<td>1013</td>
<td>647</td>
</tr>
<tr>
<td>507</td>
<td>806</td>
<td>977</td>
<td>635</td>
</tr>
<tr>
<td>511</td>
<td>916</td>
<td>1111</td>
<td>721</td>
</tr>
<tr>
<td>512</td>
<td>903</td>
<td>1099</td>
<td>707</td>
</tr>
<tr>
<td>513</td>
<td>940</td>
<td>1147</td>
<td>733</td>
</tr>
<tr>
<td>514</td>
<td>830</td>
<td>1013</td>
<td>647</td>
</tr>
<tr>
<td>515</td>
<td>806</td>
<td>989</td>
<td>623</td>
</tr>
<tr>
<td>516</td>
<td>842</td>
<td>1025</td>
<td>659</td>
</tr>
<tr>
<td>518</td>
<td>916</td>
<td>1123</td>
<td>709</td>
</tr>
<tr>
<td>519</td>
<td>806</td>
<td>989</td>
<td>623</td>
</tr>
<tr>
<td>521</td>
<td>806</td>
<td>977</td>
<td>635</td>
</tr>
<tr>
<td>522</td>
<td>977</td>
<td>1196</td>
<td>758</td>
</tr>
<tr>
<td>523</td>
<td>696</td>
<td>854</td>
<td>538</td>
</tr>
<tr>
<td>524</td>
<td>720</td>
<td>879</td>
<td>561</td>
</tr>
<tr>
<td>526</td>
<td>977</td>
<td>1184</td>
<td>770</td>
</tr>
<tr>
<td>528</td>
<td>977</td>
<td>1196</td>
<td>758</td>
</tr>
<tr>
<td>532</td>
<td>916</td>
<td>1111</td>
<td>721</td>
</tr>
<tr>
<td>533</td>
<td>671</td>
<td>818</td>
<td>524</td>
</tr>
<tr>
<td>534</td>
<td>842</td>
<td>1025</td>
<td>659</td>
</tr>
<tr>
<td>540</td>
<td>1025</td>
<td>1245</td>
<td>805</td>
</tr>
<tr>
<td>541</td>
<td>867</td>
<td>1050</td>
<td>684</td>
</tr>
</tbody>
</table>

| FFR Prevalence | 100% | 100% | 91% |

Table 6.1. F1, f2, and 2f1-f2 frequency corresponding to the largest DPOAE fine structure peak in each subject. The bottom row represents the prevalence of each FFR peak (2f1-f2 = CDP-FFR).
Figure 6.2. Effects of CAS on DPOAE fine structure and complementary FFRs in one subject.  a) DPOAE fine structure is plotted as a function of $f_2$ on a linear pressure scale. Fine structure is shown in quiet (black) and with CAS (red). The fine structure peak was identified at $f_2 = 1111$ Hz. b) The probe tones corresponding to the DPOAE fine structure peak were used to evoke FFRs. Spectra are plotted for quiet (black line) and with CAS (red line) conditions. $CDP= 721$ Hz, $F1-FFR = 916$ Hz, $F2-FFR = 1111$ Hz. Note also that an additional FFR corresponding to $3f_1-2f_2$ at 526 Hz is apparent.

Figure 6.2a plots DPOAE fine structures in a single subject for quiet (black) and with CAS (red) conditions. Note that the y-axis of each panel is expressed in linear units to visually accentuate response peaks and differences between quiet and with CAS contours. The largest fine structure peak corresponds to $f_2 = 1111$ Hz, which was inhibited by 1.4 dB with CAS. Figure 6.3 plots DPOAE fine structure inhibition as a function of $f_2$ for all subjects. In all cases, peaks were inhibited with CAS, although the magnitude of inhibition ranged from 0.5-3.5 dB across subjects. The three largest instances of DPOAE inhibition were found in the region of $F2 = 1000$ Hz.
Figure 6.3. DPOAE fine structure peak inhibition for each subject plotted as a function of f2 frequency. (Error bars = SEM).

**FFR Peak Prevalence and Effects of CAS**

FFRs were detected in response to F1 and F2 in all subjects. Yet, 2 subjects did not have a detectible CDP-FFR. Figure 6.2b plots FFR spectra for a single subject for quiet (black) and with CAS (red) conditions using the probe tone pair associated with the DPOAE fine structure peak in 6.2a. F1-FFR and CDP-FFR inhibition are 2.77 dB and 8.80 dB, respectively. In contrast, F2-FFR exhibited *enhancement* of 3.43 dB in this participant. Instances of both inhibition and enhancement were observed in FFR peaks during CAS presentation. Inhibition was most prevalent in CDP-FFRs (95%), followed by F2-FFRs (80%) and F1-FFRs (64%).

Figure 6.4 plots the average amplitudes of each FFR peak in quiet and with CAS. It can be seen that F1-FFR had the largest amplitudes in quiet (55.16 dB) and with CAS (53.92 dB). Further, the F2-FFR and CDP-FFR components had comparable amplitudes in quiet (45.73 dB and 46.71 dB, respectively), but CAS had a much larger effect on the CDP-
FFR than for the F2-FFR, with CAS inhibitory effect of 8.0 dB for the CDP-FFR, whereas it was only 1.42 dB for FFR-2. A two-factor repeated measures ANOVA was conducted on FFR amplitudes to evaluate the effects of response (F1-FFR vs F2-FFR) and noise (quiet vs with CAS). The results demonstrated a significant main effect of peak, (F [1,21] = 17.69, p < 0.0001), with F1-FFR amplitudes larger than those for F2-FFR, and a non-significant effect of noise (F [1,21] = 0.25, p=0.62) or Peak/Noise interaction (F [1,21] = 0.06, p=0.81). As shown in Figure 6.4, the CAS had virtually no effect on F1-FFR amplitude, and a small effect (less than 2 dB) on F2-FFR. However, the decrement in CDP-FFR with CAS statistically significant (t(19) = 5.95, p < 0.005).

Figure 6.4. FFR peak amplitudes in quiet and with CAS. Amplitude is expressed as dB (re: 1nV). (Error Bars = SEM).

Relationship Between DPOAE and FFR Inhibition Measures of Inhibition
DPOAE, F1-FFR, F2-FFR, and CDP-FFR inhibition values were calculated by subtracting response amplitudes obtained with CAS from those obtained in quiet. Shapiro-Wilk tests of normality on each measurement revealed that DPOAE inhibition was non-normally distributed, and this was corrected using a square-root transformation. Inhibition values for FFR peaks were found to be uncorrelated with each other, thus separate univariate analyses between DPOAE inhibition and each FFR peak inhibition measurement were conducted to test the hypothesis that pre-neural inhibition was predictive of neural inhibition (Figure 6.5). There was a statistically significant relationship between DPOAE inhibition and F2-FFR inhibition with all data points ($R^2 = 0.36$, $p= 0.006$) and there was also significance with an outlier removed ($R^2 = 0.24$, $p= 0.04$). Similarly, there was a significant relationship between DPOAE inhibition and CDP-FFR inhibition with all data points ($R^2 = 0.39$, $p= 0.004$) and also with one outlier removed ($R^2 = 0.33$, $p= 0.01$).

Figure 6.5. Scatter plots of (transformed) DPOAE inhibition and FFR peak inhibition. Black trend lines represent regressions with all data included; dashed lines represent fits with outliers (shaded) removed. Regression equations for data with outliers removed were: F1-FFR ($y = -0.9473x + 2.95$), F2-FFR ($y = 5.1641x - 4.252$), and CDP-FFR ($y = 8.3733x - 3.4955$). See text for more detail.
6.4. Discussion

The present study is the first published report using complementary DPOAE fine structure and FFR measurements to explore the relationship between pre-neural and neural MOC reflex effects. The novel findings of this study were that: 1) CAS differentially inhibited FFR peaks associated with stimulus (F1-FFR and F2-FFR) and distortion product (CDP-FFR) components such that CDP-FFR was the most inhibited, and 2) F2-FFR and CDP-FFR inhibition were significantly related to DPOAE fine structure peak inhibition. Our observation that CAS inhibited DPOAE fine structure peaks by 1.65 dB on average is generally consistent with previous reports (Sun, 2008a; Deeter et al., 2009; Abdala et al., 2009). The average magnitude of CDP-FFR inhibition reported in the present study (8 dB) falls within the range of previous reports using wideband noise as CAS (Bhagat & Champin 2004; Elsisy & Krishnan 2005); however, it is imperative to note that none of these studies are parametrically identical.

The Relationship Between DPOAE Sources and FFR Peaks

The observed relationships between pre-neural and neural MOC reflex inhibition are best understood within the context of the two source model of DPOAE generation. According to the model, backward-propagated energy from two cochlear sources contributes to the response measured in the ear canal (Zweig & Shera, 1995; Talmadge et al., 1999). The distortion source arises near or basal to the f2 place where f1 and f2 traveling waves maximally overlap, whereas the reflection source arises from the 2f1-f2 basilar membrane
place (Shera & Guinan, 1999; Shera, 2004). Assessing DPOAE inhibition at fine structure peaks minimizes artifacts related to interactions between DPOAE sources and efferent inhibition and thus represents a “true” measure of the MOC reflex (Abdala et al., 2009). It is important to note that DPOAE measurements obtained in the present study represented a composite inhibition of in-phase distortion and reflection sources combined. More sophisticated DPOAE inhibition measurements in which distortion and reflection sources were “unmixed” and appraised individually have revealed that the distortion source component is larger in amplitude than the reflection source component but that the MOC reflex inhibits the reflection component more potently than the distortion component (e.g., Abdala et al., 2009; Deeter et al., 2009). This pattern likely emerges because the energy present on the basilar membrane at the reflection source is more influenced by cochlear amplifier changes induced by the MOC reflex.

FFR peaks represent forward-fed neural correlates of the DPOAE stimulus pair (F1-FFR and F2-FFR) and, based on data from experimental animals (e.g., Goldstein & Kiang, 1968; Kim, 1980; Kim et al., 1980), the reflection source (CDP-FFR). In this way, the F2-FFR is physiologically coupled to the same basilar membrane place giving rise to the DPOAE distortion source and the CDP-FFR is coupled to the same basilar membrane place giving rise to the reflection source. Thus, an MOC reflex induced change in one of the DPOAE sources may be expected to coincide with a change in its FFR counterpart. Because we did not unmix DPOAEs and assess inhibition of each source separately, we were only able to indirectly assess this relationship in the present study. The observation
that F2-FFRs were less inhibited by CAS than CDP-FFRs is consistent with DPOAE reports demonstrating lesser distortion source than reflection source inhibition. Further, the significant relationships between composite DPOAE inhibition and F2- and CDP-FFRs are suggestive that the measures are coupled and similarly influenced by MOC reflex activation. The lack of relationship between DPOAE inhibition and F1-FFR inhibition may be related to the fact that the f1 basilar membrane place is apical to the distortion source and basal to the reflection source; thus, MOC reflex effects at the f1 place may not be represented in the DPOAE inhibition composite measure.

F2-FFR and CDP-FFR inhibition were both larger than DPOAE inhibition on a dB scale; however, the relationship between 1 dB of DPOAE versus FFR inhibition is not straightforward. In order to directly compare the functional effects of each type of inhibition, it would be necessary to derive a level series function and quantify MOC reflex inhibition in terms of “effective attenuation” (i.e., the amount of dB that the stimulus would need to be increased to overcome the effects of inhibition; see Puria et al., 1996; Lichtenhan et al., 2015, and Smith et al., 2016). Because one presentation level was used in this study, we cannot calculate effective attenuation from our data.

**Experiment Limitations and Future Work**

While the present data provide promise for understanding the relationships between pre-neural and neural efferent effects in humans, our observations should be interpreted with the following limitations and considerations in mind. Even at the pre-neural cochlear level,
the MOC effects imparted at the base of OHCs are unlikely to proceed “unfiltered” by the many complex drives which activate IHCs (Guinan, 2012, 2014). While it has been demonstrated that the MOC reflex modulates basilar membrane motion (Cooper & Guinan, 2006), IHCs (and consequently auditory nerve fibers) are driven by motion at the top of the organ of Corti, not the basilar membrane. Additionally, the effect of CAS on auditory brainstem nuclei is likely much more complex than those imparted by the cochlear efferent system. For example, MOC axons branch collaterally on both OHCs and cochlear nucleus neurons in cat and rodent models (Brown et al., 1988), suggesting that some neural inhibition, while related to MOC fiber activity, is likely not a direct consequence of OHC modulation. It is important to note, however, that collateral MOC fibers have not been identified in humans (Moore & Osen, 1979).

It is also clear that auditory processing of diotic or binaural stimuli becomes more complex at each ascending step of the auditory brainstem (Moore, 1991). We chose the stimulus frequency range in the present study (708 < f2 < 1294 Hz) such that any possible DPOAE fine structure peak would also evoke FFR responses biased towards more caudal brainstem generators (Batra et al., 1986; King et al., 2016) and thus limit the number of processing steps between pre-neural and neural responses. However, the exact sources of our FFRs could not be estimated using neural group delays. Neural group delay measurements are based on the rate of FFR phase change as a function of fine frequency steps. Regions in which the rate of phase change remains constant are suggestive of a stable generator or ensemble of generators (Kuwada et al., 2002; Bharadwadj et al.,
2014; Shaheen et al., 2015). The spacing of F1-FFR, F2-FFR, and CDP-FFR components from each subject in the present study was too large for phase data to be unambiguously unwrapped. Thus, the possibility that more rostral neural generators contributed to our recordings cannot be discounted, especially since we used a single vertical electrode montage (King et al., 2016). Given the frequencies of our probe stimuli, it is also possible that our FFRs included contributions from cochlear microphonics, especially at higher frequencies. Our few observations of FFR enhancement are consistent with reports that the cochlear microphonic amplitude grows with CAS (e.g., Fex, 1959).

In future experiments, we plan to simultaneously record DPOAEs and FFRs (e.g., Elsysy & Krishnan, 2005, 2008) using small frequency steps or swept tones to account for DPOAE fine structure and to obtain FFR spectra that are fine enough for unambiguous phase delay estimates. Such an approach will clarify FFR sources and thus provide a more complete picture regarding the relationship between pre-neural and neural measurements of cochlear distortion and MOC reflex inhibition. This technique could also be used to study corticofugal effects of attention on pre-neural and neural auditory processing (e.g., Wittekind et al., 2014).
7.1. Introduction

The dynamic acoustic characteristics of speech are precisely represented by a combination of temporal and spatial auditory nerve fiber codes (Delgutte, 1980; Ghitza, 1986; Palmer, 1990; Sachs & Young, 1979; Shamma, 1985; Young & Sachs, 1979). Each code's accuracy is contingent upon the fidelity with which cochlear mechanical vibrations are transduced into auditory nerve impulses with minimal information loss. This process is encumbered by the presence of masking noise, which interferes with cochlear traveling waves and depletes neurotransmitter at the inner hair cell-auditory nerve fiber junction (Delgutte, 1990). At the cellular level, synaptic depletion increases auditory nerve fiber thresholds and decreases the input level at which fibers saturate, resulting in a net reduction of nerve fiber dynamic range (Costalupes, Young, and Gibson, 1984; Gibson, Young, and Costalupes, 1985; Kawase, Delgutte, & Liberman, 1993; Winslow & Sachs, 1987). At the behavioral level, masking noise increases detection thresholds and decreases discrimination of speech sounds (Culling & Stone, 2017).

One mechanism hypothesized to reduce the deleterious effects of masking noise on mechano-electrical transduction of speech is the medial olivocochlear (MOC) bundle. The MOC bundle is an efferent neural tract that begins in the brainstem peri-olivary region and terminates at the base of outer hair cells, forming a feedback loop from the central auditory nervous system to the inner ear (Winer, 2005). When activated, the MOC bundle inhibits cochlear amplification—an effect termed the MOC reflex. Experiments in animal
models indicate that the MOC reflex desensitizes the auditory nerve to masking noise by reducing the cochlear excitation response to continuous broadband sounds. This “unmasking” effect partially restores the dynamic range of auditory nerve fibers and aids in the detection of signals in noise (Kawase et al., 1993; Kawase and Liberman, 1993; Winslow & Sachs, 1987).

Studies of human MOC reflex unmasking have used a combination of otoacoustic emissions (OAE) and speech-in-noise behavioral tests. Because OAEs are related to outer hair cell motility, OAE measurements in quiet and with activation of the MOC reflex by a contralateral noise (CN) can be analyzed to elucidate MOC reflex strength for an individual. MOC reflex strength is typically defined as the dB difference in OAE amplitude between without and with CN conditions. In many studies, a positive correlation has been reported between MOC reflex strength and speech- or signals-in-noise perceptual performance (de Boer et al., 2012; de Boer & Thornton, 2008; Giraud et al., 1997; Kumar & Vanaja, 2004; Micheyl & Collet, 1996; Micheyl et al., 1997). Other studies, however, have reported no or negative correlations (Garinis et al., 2011; Micheyl & Collet, 1996; Micheyl, 1995). The cause of these discrepancies is unclear, but task and OAE recording parameter differences likely account for some variation in the results.

One way to discern if MOC reflex strength is correlated with speech-in-noise unmasking is to examine the relationship between OAE inhibition and neural processing of speech-in-noise, as measured using speech-evoked auditory brainstem responses (sABRs).
sABRs are scalp recorded potentials generated by ensembles of subcortical auditory nuclei which preserve the spectro-temporal characteristics of the input speech stimulus with precision (Johnson et al., 2005; Skoe & Kraus, 2010). Masking studies on sABRs demonstrate that in the presence of noise, sABR peak amplitudes are reduced and latencies are increased (Anderson et al., 2010a, 2010b, 2013; Song et al., 2011, 2012). Further, the extent to which sABRs are impacted by masking noise has been correlated with speech-in-noise perceptual performance (e.g., Song et al., 2011) supporting the hypothesis that poor subcortical representation of speech compromises perceptual performance. If individual variation in speech-in-noise processing is related to MOC reflex strength, an expected finding would be that OAE inhibition is positively correlated with sABR unmasking when the MOC reflex is activated. This may be evident as a “recovery” of sABR latency and/or amplitude shifts caused by noise during MOC reflex activation. In a recent experiment, de Boer and colleagues (2012) reported that sABR latency shifts induced by an ipsilaterally presented masking noise were negatively correlated with OAE inhibition strength; that is, subjects with stronger MOC reflex inhibition unexpectedly showed greater sABR masking effects in the presence of noise. The authors also reported that OAE inhibition was positively correlated with speech-in-noise perceptual performance. They concluded that the positive correlation between OAE inhibition and speech-in-noise perceptual performance indicated that the MOC reflex is helpful when commandeered corticofugally by attention; however, the negative correlation between OAE inhibition and sABR masking effects suggested that “passive” activation of the MOC reflex can be detrimental.
The present experiment reexamines the human MOC reflex unmasking hypothesis. A limitation of previous work is that OAE inhibition has commonly been measured using CN, whereas speech stimuli have been presented monaurally in the presence of ipsilateral masking when measuring neural or behavioral responses. This method thus compares objective measures of contralateral MOC reflex strength with the combined ipsilateral MOC reflex and masking effects on the neural or perceptual representation of a signal (Guinan, 2014). In this study, we first measured OAE inhibition with CN to quantify contralateral MOC reflex strength. We then measured sABRs in four conditions designed to collectively extract potential unmasking effects conferred by the contralateral MOC reflex. We hypothesized that, consistent with previous work, ipsilaterally presented masking noise would cause characteristic increases in sABR peak latencies and decreases in peak amplitudes (de Boer et al., 2012; Parbery-Clark et al., 2009). Further, activating the contralateral MOC reflex would result in unmasking of the neural signal. The amount of neural unmasking was hypothesized to positively correlate with the strength of OAE inhibition.

7.2. Method

Participants

The University of Arizona Human Subjects Protection Program approved the following methods. Eighteen adult participants (average age 21.3 years, SD = 3.1; 8 female) without history of otologic or neurologic disease were enrolled in the study. All ear canals
were free of excess cerumen and tympanic membranes (TMs) appeared healthy by otoscopy in all subjects. Participants had normal tympanograms bilaterally (Margolis & Heller, 1987) and contralateral acoustic reflex thresholds to 1-10 kHz broadband noise ≥ 70 dB SPL, measured using conventional admittance methods (Sun, 2008). Acoustic reflex threshold measurements were made to document that middle ear muscle contractions during MOC inhibition measurements were not confounding our results, although others have shown that acoustic reflex thresholds can be lower when measured using more sensitive techniques (e.g., Zhao & Dhar, 2010; Lichtenhan et al. 2016). Air conduction hearing thresholds from 0.25-8 kHz were within normal limits (< 25 dB HL) bilaterally for all subjects. All procedures were conducted in an electromagnetically shielded sound booth.

**CEOAE Inhibition Measurements**

CEOAEs were obtained with a Mimosa Acoustics HearID System (Mimosa Acoustics, Inc. Champaign, IL). Prior to testing, the probe assembly was calibrated in-situ to verify proper fit and flat broadband frequency response in the ear canal. Responses were collected using 100 µs “linear” clicks (i.e., consistent stimulus polarity and level across all presentations) presented at 11/sec for 250 sweeps in each trial. We chose this relatively slow presentation rate based on reports that clicks themselves can activate the ipsilateral MOC reflex at rates as low as 30-50/sec (Veuillet et al., 1991; Frances & Guinan, 2010; Boothalingam & Purcell, 2015). CEOAEs were considered present if they were ≥ 6 dB above the noise floor and if buffers A and B were ≥ 80% correlated. CEOAE files were
saved and offline analyzed in MATLAB (The Mathworks, Inc., Natick, MA), which extracted the composite values representing total emission amplitude and noise floors for each level and condition.

CEOAEs were collected at 60 and 70 ppeSPL in quiet and with CN to assess MOC reflex effects on OHC function. Each condition was repeated three times, resulting in 12 total measurements. The starting presentation level and order of quiet and with CN trials were randomized for each participant. Sixty seconds were interleaved between each collection trial to allow the MOC reflex to "reset" prior to the next run. Trials in which background noise exceeded 60 dB SPL were discontinued and restarted.

![Figure 7.1. CEOAE amplitudes evoked by 60 and 70 dB ppeSPL clicks without (black dots) and with CN (grey dots) for one subject. The three ways of quantifying MOC reflex inhibition are shown in the figure: Amplitude reduction is indicated by the downward-pointing arrow; Effective attenuation at 65 dB is indicated by the horizontal arrow; I/O](image)

Figure 7.1. CEOAE amplitudes evoked by 60 and 70 dB ppeSPL clicks without (black dots) and with CN (grey dots) for one subject. The three ways of quantifying MOC reflex inhibition are shown in the figure: Amplitude reduction is indicated by the downward-pointing arrow; Effective attenuation at 65 dB is indicated by the horizontal arrow; I/O.
inhibition is the difference between the linear regression slopes for quiet and with CN conditions.

CEOAE inhibition was quantified in three ways, which are each represented in the extant literature (Figure 7.1). *Amplitude reduction* was calculated as the difference in dB between CEOAEs collected in quiet and with CN at each presentation level (60 and 70 dB ppeSPL). *Effective attenuation* was calculated as the “horizontal” (abscissa) difference between linear regression fits to CEOAE amplitudes in quiet and with CN (e.g., Puria et al., 1996). For simplicity, a single effective attenuation value was calculated for each subject using the following equation:

\[
EA_{65} = \frac{(m_Q(65) + b_Q) - b_{CN})}{m_{CN}} - 65
\]

where \(m_Q\) and \(m_{CN}\) are slopes and \(b_Q\) and \(b_{CN}\) are intercepts of linear regression fits to quiet and with CN data, respectively. Thus, our effective attenuation value expressed the amount to which a 65 dB ppeSPL input stimulus would need to increase to overcome the attenuation effects of MOC reflex inhibition. *Input/output (I/O) inhibition* was calculated as the difference in linear regression coefficients (i.e., slopes) obtained by fitting separate models to quiet and with CN CEOAE level-series functions. The rationale behind this measurement is that cochlear gain decreases with activation of the MOC reflex, and the
CEOAE I/O slope steepens as a result. Large I/O inhibition (i.e., greater steepening of I/O slope with CN) therefore serves as a proxy measurement indicating greater MOC reflex strength (de Boer et al., 2012).

sABR Measurements

Stimulus

The evoking speech stimulus (Figure 7.2) was a five-formant 110 ms synthetic /da/ generated at a 20,000 Hz sampling rate using a Klatt synthesizer (Klatt, 1980) and based on modified parameters reported by Song and colleagues (2011). As demonstrated in Figure 7.2, the stimulus had a 100 Hz fundamental frequency and consisted of three parts: onset frication burst (0-10 ms), voice onset/formant transition (10-55), and steady state vowel (55-110). The initial frication burst consisted of spectral energy centered at 3300-3750 Hz. Between 10-55 ms, the formants transitioned linearly as follows: F1: 400-720 Hz; F2: 1700-1240 Hz; F3: 2580-2500 Hz; F4: 3300-3300; F5: 3750-3750 Hz. From 55-110 ms, the formants remained stable for five cycles of the vowel. The stimulus was saved as a WAV file and converted into a format suitable for Intelligent Hearing Systems SmartEP software (Intelligent Hearing Systems, Miami, FL) using the Stimulus Generation tool.
Figure 7.2. The sABR evoking stimulus (top) was a 110 ms /da/ created with a Klatt synthesizer. The stimulus consisted of onset (0-10 ms), formant transition (10-55 ms), and steady-state (55-110 ms) components. sABRs (bottom) encode each component of the stimulus with precision. Note that the grey error bars represent the standard error of the mean waveform for 18 subjects.

**Electrophysiologic Recordings**

sABRs were acquired with Intelligent Hearing System’s SmartEP Advanced Research Module using a single-channel three-electrode montage: vertex (+), right mastoid (-), and forehead (±)\(^9\). Responses (Figure 7.2) were artifact rejected at ± 20 µV, online filtered

---

\(^9\) In pilot work for this experiment, we explored the feasibility of recording speech evoked responses electrocochleographically in a second simultaneous channel. The purpose of these recordings was to discern whether speech evoked compound action potentials from the auditory nerve could be reliably measured and whether evidence of MOC reflex unmasking of speech was
from 70-3000 Hz, and sampled at a rate of 20,000 Hz over a 153.6 ms recording epoch, which included a 13.6 ms pre-stimulus baseline period. Participants were comfortably seated in a reclining chair and asked to remain awake during testing. In each condition, the synthetic /da/ stimulus was presented at 70 dB ppeSPL to each subject’s right ear through an electrically shielded ER-3A insert earphone (Etymotic Research, Elk Grove Village, IL) at a rate of 7.1/s. sABRs were recorded in four experimental conditions: Quiet (Q), with CN, with Ipsilateral Noise (IN), and with IN+CN. Table 7.1 lists specific stimulus parameters and rationales for each condition. A total of 2,048 sweeps were acquired in condensation and rarefaction stimulus polarities for each condition. Adding sABR waveforms evoked by condensation and rarefaction stimuli accentuates the envelope of the neural response, while subtracting accentuates high-frequency temporal fine structure components. Analyses were only performed on sABR “envelope” responses since IN and IN+CN conditions were found to obliterate temporal fine structure components.

---

present at the level of the auditory nerve. While there was some evidence of speech evoked potentials from the auditory nerve, the waveforms were variable and only seen at high intensity levels. Thus, electrocochleographic recordings were not included in this experiment to save time and ensure participant comfort throughout the recording process.
Table 7.1. sABR conditions and rationales.

Time Domain Analyses

Absolute Peak Latencies and Latency Shifts Relative to Quiet

sABR absolute peak latencies were calculated for onset and offset components and an average peak latency value was calculated for transition (10-55 ms) and steady-state (55-110 ms) components for each condition. Peak latency shifts relative to Quiet were also calculated for CN, IN, and IN+CN waveforms using a cross-correlation analysis (Skoe & Kraus, 2010). This analysis temporally shifted CN, IN, and IN+CN waveforms until they were maximally correlated with the Quiet waveform. The purpose of quantifying latency shifts relative to Quiet was to understand the effects of noise on sABR responses in the other three test conditions.
Peak Amplitude and Average Amplitude Differences

sABR peak amplitudes were calculated for onset and offset components and an average peak amplitude value was calculated for transition (10-55 ms) and steady-state (55-110 ms) components for each condition.

Statistical Analyses

CEOAE amplitudes were evaluated using a repeated measures analysis of variance (RM ANOVA) with condition (Q and CN) and level (60 and 70 dB ppeSPL) as within-subject factors and sex as a between-subject factor. Sex was included based on previous reports demonstrating male and female differences in both OAE magnitude (McFadden et al., 2009) and inhibition strength (Khalfa & Collet, 1996). sABR latencies and amplitudes were also evaluated using separate RM ANOVAs with condition (Q, CN, IN, IN + CN) as the within-subject factor and sex as a between-subject factor. Sex was also included as a factor in this analysis based on previous reports that males and females demonstrate ABR differences evoked by simple (e.g., Jerger & Hall, 1980) and complex stimuli (Krizman et al., 2012). Whenever the ANOVA assumption of sphericity was violated, F-tests were Greenhouse-Geisser corrected. Bonferroni corrections were applied to post-hoc tests to account for multiple comparisons.

The primary aim of this study was to test the hypothesis that sABR unmasking, if observable, is predicted by MOC reflex strength. sABR unmasking was hypothesized to
manifest as latency and/or amplitude “improvement” in the IN + CN condition relative to the IN condition; this is because the masking effect imposed on the sABR in the IN condition should be reduced when the contralateral MOC reflex is activated in the IN + CN condition. sABR unmasking was quantified as the difference in latency shift or amplitude between IN + CN and IN conditions. Positive unmasking values therefore indicated improvement in latency or amplitude in the IN + CN condition relative to the IN condition. Separate linear regression analyses were performed with MOC reflex strength (i.e., CEOAE amplitude reduction, effective attenuation, and I/O inhibition, respectively) as the independent variable and sABR unmasking (i.e., latency shifts and amplitude changes, respectively) as the dependent variable.

7.3. Results

CEOAE Inhibition

CEOAE amplitudes were reduced with CN compared to Q (Figure 7.3) at each presentation level (60 and 70 dB SPL). Mean CEOAE inhibition with CN was 1.41 dB, which was significant [F(1, 16) =53.27, p<0.001] based on a RM ANOVA, where noise (Q and CN) and level (60 and 70 dB SPL) were within-subject factors and sex was a between-subject factor. There was a marginally significant interaction between click level and noise condition [F(1,16) =7.77, p=0.0592], as inhibition was greater at 60 dB (1.65 dB) versus 70 dB (1.16 dB). The RM ANOVA also revealed a significant main effect of sex on CEOAE amplitude [F(1,16) = 10.24, p = 0.0056], which was due to the fact that females had larger (7.45 dB) CEOAEs than males (3.52 dB); amplitude reduction,
however, did not significantly differ between males (1.20 dB) and females (1.60 dB). When amplitude reduction was expressed as a ratio, males demonstrated larger proportional inhibition (1.2/3.52 dB) than females (1.6/7.45 dB).

Figure 7.3. CEOAEs were significantly inhibited by CN at both 60 and 70 dB ppeSPL presentation levels (left). There was a marginally significant interaction between stimulus level and condition suggesting that inhibition was larger for 60 than 70 dB ppeSPL clicks. Inhibition did not significantly differ between sexes (right), although CEOAEs were generally larger in females than males ( * = p < 0.025).

**Effects of Noise on sABR**

**Latency**

In general, sABR latencies were prolonged and amplitudes were reduced in conditions in which the stimulus and noise were presented to the same ear (Figure 7.4), which was consistent with previous work (e.g., Parbery-Clark et al., 2009; de Boer et al., 2012). Latency and amplitude data were analyzed in separate RM ANOVAs with condition (Q, CN, IN, IN+CN) and peak label (Onset, Transition, Steady-State, Offset) as within-subject factors and sex as a between-subjects factor.
Figure 7.4. Average sABR waveforms for CN (red), IN (green), and IN+CN (blue) conditions are overlaid onto Q (black) waveforms for comparison (left). Average latencies (right) and amplitudes (bottom) of onset, transition, steady-state, and offset components are plotted for comparison. Latencies were significantly prolonged in IN and IN+CN conditions with onset components being most affected by masking. ( * = p < 0.01; ** = p < 0.001). Amplitudes were significantly reduced in IN and IN+CN conditions relative to Q and CN.

There was a significant main effect of condition on peak latencies [F(3,48) = 93.80, p < 0.001] due to the large shifts observed in IN and IN+CN relative to Q and CN conditions.
Post hoc analyses revealed that IN and IN+CN latencies were significantly longer than Q and CN latencies (p < 0.001). There was a marginally significant interaction between condition and peak on latency \([F(9,142) = 1.83, p = 0.0671]\) due to larger masking effects on the onset component in IN and IN+CN conditions relative to the transition and steady-state components (Figure 7.4). There was no significant effect of sex on sABR latency in our sample.

Amplitude

Consistent with previous literature (e.g., Burkard & Hecox, 1980), variation in peak amplitude measurements was larger than latency measurements. There was a significant main effect of condition on peak amplitudes \([F(3,48) = 24.15, p < 0.01]\), which was driven by the large amplitude decreases observed in IN and IN+CN conditions. Amplitude differences between IN and IN+CN conditions were not significant. No significant interactions between condition and peak or between sex and amplitude were observed.

Relationships Between CEOAE Inhibition and sABR Unmasking

Latency

13/18 subjects demonstrated sABR latency unmasking according to our operational definition (i.e., earlier peak latencies in the IN+CN relative to the IN condition). The average size of the latency unmasking effect was quite small (0.3 ± 0.21 ms), however, and as mentioned above, sABR latency shifts were not significantly different between
IN+CN and IN conditions for any component. CEOAE inhibition, regardless of how it was quantified, was not predictive of sABR latency unmasking (Figure 7.5).

![Graphs showing unmasking in different conditions](image)

Figure 7.5. CEOAE inhibition was not predictive of sABR latency unmasking. $R^2$ and $p$-values for each model are presenting in the bottom right-hand corner of each frame.

The marginally significant interaction between peak and condition on sABR latency revealed by the RM ANOVA suggested that responses to transient aspects of speech are more susceptible to noise than steady-state components. Because we initially quantified CN, IN, and IN+CN latency shifts relative to Q using a waveform cross-correlation technique, it is possible that sABR component-specific unmasking effects were washed out in this analysis. This is because waveform cross-correlation is biased by larger amplitude periodic components than the onset component. To address this issue, unmasking values were also calculated separately for the onset component only. An additional linear regression analysis was conducted with CEOAE inhibition as the independent variable and sABR onset unmasking as the dependent variable. CEOAE inhibition did not predict onset unmasking (Figure 7.6).
Figure 7.6. CEOAE inhibition was not predictive of sABR onset component latency unmasking. $R^2$ and p-values for each model are presenting in the bottom right-hand corner of each frame.

We also explored whether CEOAE inhibition was predictive of “resilience to speech masking” – the magnitude of latency shift induced by the IN condition relative to the Q condition. Although this method compares contralateral MOC reflex strength with combined ipsilateral MOC and masking effects, we wanted to explore whether our dataset supported previous work demonstrating a relationship between contralateral CEOAE inhibition and neural representation/perception of signals in ipsilateral masking. As with our measures of unmasking, CEOAE inhibition was not predictive of resilience to speech masking.

**Amplitude**

9/18 subjects demonstrated amplitude unmasking when amplitudes were averaged across all peaks for IN and IN+CN conditions. As with the latency data, there were no
significant relationships between CEOAE inhibition and amplitude unmasking. Further, sABR amplitude resilience to speech masking was not related to CEOAE inhibition.

7.4. Discussion

The purpose of this study was to reassess the hypothesis that the human MOC reflex is beneficial in the neural encoding of speech in noise. This hypothesis is based on a large body of experiments in animal models demonstrating MOC reflex unmasking at the level of the auditory nerve (see Chapter 4 for review) and fewer human studies linking MOC reflex strength to perceptual performance in noise (e.g., Giraud et al. 1997; Kumar & Vinaja, 2004; de Boer & Thornton, 2008). Our results did not demonstrate a relationship between MOC reflex strength and speech unmasking at the level of the auditory brainstem. Additionally, MOC reflex strength was not related to the resilience to speech masking. The latter observation conflicts with the findings of a similar study, which demonstrated that individuals with greater MOC reflex inhibition were susceptible to larger sABR masking effects (de Boer et al., 2012). The cause of the discrepancy between our results and those of de Boer and colleagues is not clear, as the stimulus paradigms used in each study were similar. One possible reason may be that they controlled for hearing sensitivity when exploring the relationship between CEOAE inhibition and sABR resilience to masking after finding that those with better hearing demonstrated larger CEOAE inhibition. Because we only screened our participants for hearing loss, we were unable to adjust for hearing sensitivity in our dataset. An additional finding of de Boer and colleagues was that individuals with larger CEOAE inhibition performed better on a
speech-in-noise perceptual task. Because a behavioral paradigm was not included in the present study, we were unable to evaluate this relationship and the potential effects of attention on MOC reflex function (Maison et al., 2001; Veuillet et al., 2007; de Boer & Thornton, 2007, 2008).

Our method for evaluating sABR unmasking focused on latency and amplitude differences in the neural onset, transition, and phase locking responses to the 100 Hz stimulus envelope between conditions. Temporal fine structure was obliterated in conditions that included ipsilateral masking noise and therefore could not be evaluated. Temporal fine structure is important to consider because MOC reflex-induced changes in cochlear gain appear to be frequency specific (Guinan & Gifford, 1988; Vinay & Moore, 2008; Lilaonitkul & Guinan, 2012); therefore, unmasking may have occurred in frequency channels not assessed in our analyses. In addition, psychophysical MOC reflex studies indicate that pure tone unmasking occurs when masker energy falls within the same frequency band as the target, whereas the effect of off-frequency maskers is enhanced by the MOC reflex (Strickland 2001, 2008; Krull & Strickland, 2008; Jennings et al., 2009; Roverud & Strickland, 2010; Jennings & Strickland, 2012; Verschooten et al., in press).

With a speech stimulus, the interactions between energy within frequency bands over time and MOC reflex unmasking are likely complex. It is possible that a larger signal-to-noise ratio than 10 dB would have preserved temporal fine structure components and allowed for fine-grained analyses of the sABR. It is also possible that using an OAE elicitor that was more like the speech stimulus parameters (e.g., stimulus frequency OAEs
at frequencies corresponding to speech harmonics and formants) would have revealed a relationship between OAE inhibition and sABR unmasking.

At the level of the auditory nerve, unmasking is evident as a recovery of action potential amplitude typically without a significant effect on latency (e.g., Desmedt, 1962; Fex, 1967; Lichtenhan et al., 2016; Smith et al., 2017). Amplitude unmasking in the auditory nerve may or may not be preserved in the behavior of subsequent auditory brainstem ensembles, as input gain is centrally adjusted as the signal ascends to the rostral brainstem (e.g., Polley et al., 2015; Schaette & McAlpine, 2011). sABR amplitudes were highly variable in the present experiment; therefore, it is possible that “true” sABR amplitude unmasking effects were present in our recordings but were hidden by residual electrophysiologic noise. Another complicating factor is that sABRs arise from the activity of multiple overlapping generators in the auditory brainstem; the envelope following response in particular is dominated by potentials from the rostral brainstem (Bharadwaj et al., 2015). It is possible that many factors other than the MOC reflex, such as binaural integration, influenced sABR recordings in various conditions. However, to prevent “binaural unmasking effects” (i.e., improvement of signal encoding due to computation of interaural timing differences in the masking noise), we used non-correlated noise the IN+CN condition.

7.5. Conclusion
Previous studies have suggested a link between MOC reflex strength and signals- or speech in noise processing, although consensus on this topic is lacking. In the present study, we did not find a relationship between CEOAE inhibition, measured in a variety of ways reported in the literature, and sABR unmasking.
CHAPTER 8: SUMMARY AND FUTURE DIRECTIONS

8.1. Summary
The results of Experiment 1 suggested that compound action potential amplitude inhibition was larger than otoacoustic emission amplitude inhibition when results were reported on the same scale. Further, chirp-evoked compound action potential inhibition was larger than click-evoked compound action potential inhibition, suggesting that chirps may be a better tool for measuring MOC reflex inhibition of auditory nerve responses. The results of Experiment 2 revealed that distortion product frequency following response inhibition was largest for the component measured at 2f1-f2 than for f1 or f2. Further, distortion product otoacoustic emission inhibition was mildly predictive of distortion product frequency following response inhibition at 2f1-f2 and f2. The results of Experiment 3 revealed that otoacoustic emission inhibition was not predictive of speech-in-noise “unmasking” at the level of the brainstem. Taken together, the experiments suggest that pre-neural inhibition measurements likely underestimate MOC reflex strength and that neural assays may be more beneficial in understanding the functional significance of the MOC reflex in humans.

8.2. Future Work
Future studies should be conducted to expand upon the findings of this dissertation and to attempt to resolve questions that have arisen from our data. The following is a brief list of planned studies following the work reported here:
1. While neural MOC reflex inhibition tended to be larger than pre-neural inhibition, the variation of neural recordings was higher and the data acquisition needs (i.e., electrodes, etc.) were more involved. Therefore, more work needs to be done to justify (or condemn) the use of neural over pre-neural measurements in human MOC reflex evaluation. As discussed in Chapter 5, optimizing the chirp stimulus based on basilar membrane delays for an individual may be a way to obtain more robust and stable CAP responses. “Online” stimulus development that accounts for subject/patient characteristics is an underexplored area both with research and clinical diagnostic applications.

2. A current limitation of DPOAE measurements is that they are poor below ~500-1000 Hz. This is because the physiologic noise floor for acoustic recordings is largest below ~ 500 Hz. DPFFRs may be a more effective way to measure non-linearity - and efferent effects on non-linearity - in the cochlear apex for clinical and research purposes. Further, this method may be a better way to explore the “breaking” of scaling symmetry in the cochlear apex (Dhar et al., 2011). To this end, more work needs to be done to understand the specific cochlear and neural generators of DPFFRs. This can be explored using simultaneous acquisition of DPOAEs and DPFFRs in both humans and animal models in which the cochlear apex can be precisely manipulated (e.g., Lichtenhan et al., 2016).
REFERENCES


180. References:


efferent effects on ipsilateral and contralateral cochlear potentials in the guinea pig.
*Brain research*, *1081*(1), 138-149.

spontaneous otoacoustic emissions. *Journal of the Association for Research in
Otolaryngology*, *11*(1), 53-67.

221. Zhao, W., & Dhar, S. (2012). Frequency tuning of the contralateral medial

hearing loss: Contralateral inhibition of otoacoustic emissions in mice. *Journal of
Comparative Neurology*, *503*(5), 593-604.

2047.