DEVELOPMENT OF AN OPTIMIZED METHOD FOR THE CLINICAL ASSESSMENT OF BINAURAL FUNCTION

By

Parnell Davis MacNevin

Submitted in partial fulfillment of the requirements

for the degree of Master of Science

at

Dalhousie University

Halifax, Nova Scotia

March 2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Obscure auditory dysfunction describes auditory deficits despite normal audiometric results. The causes are difficult to elucidate and not adequately assessed by currently available clinical assessments. Interaural phase difference (IPD) detection may be behavioural indicator of peripheral synaptic function because it reflects phase locking to temporal fine structure, sensitivity declines with age without reduced hearing, and it reflects temporal synchronization in the nerve. The objective was to develop three assessments to establish frequency thresholds for IPD detection, and to evaluate their efficiency and reliability; three alternative forced choice (3AFC), continuous adaptive tracking (CAT), and continuous non-adaptive tracking (CNT). Piloting revealed the 3AFC and CAT warranted investigation, and CNT was removed. Thresholds were correlated and demonstrated reliability. The CAT test was the most efficient and reliable. Age was a predictor of IPD thresholds for both tests. Findings indicate the CAT test is a promising approach to assess IPD detection in the clinic.

LIST OF ABBREVIATIONS USED

3AFC- three alternative forced choice

ABR- auditory brainstem response

CAT- continuous adaptive tracking

CNT- continuous non-adaptive tracking

CRM- Coordinate response measure

EcochG- electrocochleography

EEG- electroencephalography

EHF- extended high frequency

FFR- frequency following response

HHL- hidden hearing loss

IHC – inner hair cell

ILD- interaural level difference

IPD- interaural phase difference

IPM-FR- interaural phase modulating following response

ITD- interaural timing difference

NR- no response

OAD- obscure auditory dysfunction

OHC- outer hair cell

QuickSIN test– Quick speech in noise test

SNHL- sensorineural hearing loss

SNR- signal to noise ratio

SR- spontaneous rate

TFS- temporal fine structure

TFSLF- temporal fine structure low frequency

VIF – variance inflation factor

CHAPTER 1: INTRODUCTION

Hearing loss affects approximately 466 million people worldwide (World Health Organization, 2018), and in Canada approximately 8% of children, 40% of adults and 78% of the elderly are deaf or hard of hearing (Statistics Canada, 2016). People with hearing loss are at an increased risk of experiencing a variety of negative consequences including accidents, occupational problems, and, perhaps most debilitating, disrupted language development and reduced communication ability (National Research Council Committee on Disability Determination for Individuals with Hearing Impairments [NRCCDDIHI], 2004). Without remediation, these issues have the potential to significantly impede functioning and reduce quality of life. Communication problems are a particularly salient consequence of hearing loss and have been strongly linked to the development and/or exacerbation of mental health problems, including depression (Li, Zhang, Hoffman, Cotch, Thermann & Wilson, 2014) and recent studies have identified hearing loss a significant risk factor for dementia (Gurgel, 2014).

The typical approach to determine hearing ability in the clinic is to measure hearing sensitivity with an audiogram. This clinical tool is the foundation of the standard audiometric test battery and is a fairly reliable method to assess hearing and decreases in hearing threshold due to middle ear dysfunction or cochlear hair cell loss. Despite its advantages, it has become increasing clear that many forms of auditory dysfunction will be missed if clinicians rely solely on the audiogram, as it is limited capacity to assess many functional hearing problems.

Obscure auditory dysfunction (OAD) is a term used to describe cases where auditory deficits are present, including trouble understanding speech in noise, hyperacusis and intermittent tinnitus, despite normal audiometric test results (Saunders & Haggard, 1989). At this time, the underlying causes of OAD are difficult to elucidate and are not adequately assessed by the currently available clinical assessment battery (Plack, Barker & Prendergast, 2014; Liberman & Kujawa, 2017). In light of these limitations, there is a need to develop tests that go beyond the audiogram and are sensitive to the integrity of various structures and systems along the auditory pathway.

Obscure auditory dysfunction describes a number of conditions with more than one etiology, and the deficits reported could be caused by peripheral damage and/or due to deficits in central auditory processing, but it is likely that some cases of OAD result from age or noise-induced peripheral damage to auditory nerve fibers (Plack et al, 2014; Liberman & Kujawa, 2017). In fact, by definition OAD describes an array of functional hearing problems without a known underlying mechanism (Saunders & Haggard, 1989). Conversely, hidden hearing loss (HHL) refers to the theorized functional hearing deficits that arise following damage to the sensitive synapses between inner hair cells and auditory nerve fibers (i.e., cochlear synaptopathy) (Prell, 2018). As a result, cochlear synaptopathy would not impact hearing sensitivity per se, but would disrupt phase locking to the temporal fine structure (TFS) of sound, the peripheral mechanism that facilitates the processing of key timing information along the auditory pathway (Demany & Semal, 1989; Moore & Sek, 1996).

Unfortunately, the standard audiometric test battery is not sensitive to the integrity of the peripheral temporal coding system that is posited to be disrupted in cochlear synaptopathy (for review see Prell, 2018). There have been several promising approaches studied in the past, but they require extensive psychoacoustic testing and/or electrophysiology, which is often not feasible in the standard clinic. They have not become incorporated into standard clinical practice because they are complicated, time consuming, have not demonstrated that they sensitive enough to peripheral synaptic damage (Plack, Leger, Prendergast, Kluk, Guest & Munro, 2016), and findings are generally inconsistent (e.g., Stamper & Johnson, 2015; Prendergast et al., 2018).

Given the current limitations to the standard audiometric test battery, there is a need to develop a quick, easy, and cost-effective clinical tests that can differentially diagnose peripheral damage at the level of the auditory synapse. One approach that has recently demonstrated promise is binaural phase detection. Research has found that IPD detection reflects phase locking to the temporal fine structure of sound (TFS) (e.g., Füllgrabe, Harland, Sek, & Moore, 2017; Füllgrabe et al., 2018; Hopkins & Moore, 2010), sensitivity to IPD declines with age despite little to no reduction in hearing sensitivity (Ross, Fujioka, Tremblay & Picton, 2007), and IPD detection relies highly on precise temporal synchronization in the nerve that receives input directly from the peripheral system (Lacher-Fougère & Demany, 2005; King et al., 2014). Taken together, IPD detection ability may be a powerful and sensitive behavioural indicator of peripheral synaptic function.

THE PRESENT STUDY

The main objective of the proposed research is to develop a clinically feasible assessment of frequency thresholds for IPD detection. Recently, researchers developed an EEG-based assessment of IPD detection called the interaural phase modulated following response (IPM-FR) (Undurraga, Haywood, Marquardt & McAlpine, 2016). Not only was this approach successful in objectively measuring IPD, but responses were also strongly correlated with the subject's ability to discriminate between phase modulated and static stimuli. These findings suggest that it may be possible to develop a simple behavioural measure of IPD detection, rather than having to rely on complex EEG measures that are not well-suited for routine clinical use.

Research has demonstrated that detecting IPDs is an easy task at low frequencies and becomes more challenging as the frequency increases (Ross et al., 2007), which suggests that measuring frequency thresholds for binaural phase detection could be a sensitive and specific behavioural approach to assess the function of peripheral synapses. Thus, the goal of the present study is to develop a short and simple frequency-based IPD detection test that is appropriate for clinical use so that peripheral damage can be identified earlier, the progression can be tracked, and this information can be used to aid in client consultation, education, prevention efforts, and treatment strategies.

THE ROLE OF BINAURAL CUES IN HEARING

Hearing is a complex system that involves multiple peripheral and central pathways (Gelfand, 2010). The process begins in the peripheral auditory system and auditory input from each ear enters the auditory nerve and is relayed to lower order brain regions where it is compared and integrated by lower order cortical structures (e.g. brainstem nuclei) before it is relayed to higher order regions of the auditory cortex (Gelfand, 2010). This mechanism, referred to as binaural integration, helps listeners localize environmental sounds and is salient to navigating complex acoustic environments (e.g., speech intelligibility in noise) (Gelfand, 2010).

BINAURAL PROCESSING CUES

In general, acoustic perception can be divided into two planes; the horizonal (i.e., azimuth), and the vertical (i.e., medial). While the vertical/medial plane primarily uses monaural spectral cues, the horizontal azimuth utilizes binaural cues for sound localization (Gelfand, 2010). Overall, there are three types of horizontal binaural cues; interaural level difference (ILD), interaural time difference (ITD), and interaural phase difference (IPD) and each cue contributes uniquely to sound localization (Akeroyd, 2006). These cues are also not equally effective across the frequency range, with intensity cues being most effective for high frequency sounds, and timing cues most effective for low frequencies (Gelfand, 2010; Litovsky, 2012).

INTERAURAL LEVEL DIFFERENCE

In the midline (0 degrees) the sound waves reach each ear simultaneously, so there are no differences in timing, intensity and phase. In ILD, when the location of the sound moves across the horizontal plane it triggers the head-shadow effect, an acoustic phenomenon where the sound waves encounter interference from the head that alters the intensity of sound waves arriving to each ear (Akeroyd, 2006). When the sound source is at maximum displacement along the azimuth (i.e., 90 or 270 degrees), the difference elicited by the shadow effect is largest (Akeroyd, 2006). As a result, the ear proximal to the sound source will receive a stronger intensity signal and the signal arriving at the furthest ear is dampened, providing important cues that assist in the localization sounds in the environment (Akeroyd, 2006).

INTERAURAL TIMING AND PHASE DIFFERENCES

The central auditory system is also sensitive to slight differences in the timing of acoustic signals arriving at each ear (Akeroyd, 2006). When a sound is located outside of the midline of the azimuth, waves arrive at each ear at slightly different times and exhibit timing-related phase differences (Gelfand, 2010; Akeroyd, 2006). In general, ITDs are detectable at the onset of a signal and IPDs are best detected when sounds are presented continuously (Akeroyd, 2006; Ross, Termblay, Picton, 2007). Despite the transient nature of these timing cues, a well-functioning peripheral auditory system is able to encode this information and relay it to midbrain regions (e.g., the superior olivary complex), where the binaural timing differences are compared, integrated and relayed bilaterally to higher order structures along the auditory pathway (Gelfand, 2010; Celesia, 2015). Given the rapid and transient nature of these cues, encoding this information requires precise timing in the peripheral auditory system, specifically at the level of the IHC synapse (Schmiedt et al., 1996).

NOISE EXPOSURE AND COCHLEAR SYNAPTOPATHY

While the underlying mechanisms are still being investigated, excessive exposure to noise and age-related sensorineural hearing loss (SNHL), i.e., presbycusis (Statistics Canada, 2016; WHO, 2018) have been linked to functional hearing problems (i.e., hidden hearing loss). Noise is common in everyday life, and exposure to occupational (e.g., industry, construction etc.), recreational (e.g., using earbuds, attending concerts etc.), and environmental noise can have a negative impact on the auditory system (WHO, 2018). Despite advancements in hearing protection, workplace policies to mitigate exposure, and increased public health education about the dangers of noise, noise-induced hearing loss is still a significant global health issue. In particular, hazardous recreational noise is becoming increasingly prevalent in youth populations. So much so, that the World Health Organization recently highlighted the dangers associated with recreational noise and estimated that over a billion young people worldwide are at risk of experiencing hearing loss as a result of this lifestyle (WHO, 2018).

In extreme cases, noise exposure can lead to a permanent decrease in hearing sensitivity that can be easily measured in clinic, but recent animal research has demonstrated that exposure to noise intense enough to trigger a temporary threshold shift permanently damages fragile auditory synaptic connections and disrupts temporal processing (Shi, Chang, Li, Aiken, Liu & Wang, 2016). Research has also indicated that similar synaptic degradation is linked to aging, with animal studies demonstrating a significant reduction in auditory nerve fiber synapses across the lifespan (Sergeyenko, Lall, Liberman & Kujawa, 2013), and these findings have also been reported in human post-mortem studies (Viana et al., 2015). Given this evidence, it has been theorized that exposure to noise can damage auditory nerve fiber synapses (i.e., cochlear synaptopathy).

THE COCHLEAR SYNAPSE

The primary source of information for the central auditory system is via input from the IHC to the cochlear nerve (Spoendlin, 1972; Gelfand, 2010). Cell bodies for afferent cochlear nerve fibers are housed within the spiral ganglion, with axons and dendrites located in the osseous spiral lamina and organ of Corti respectfully (Gelfand, 2010). Afferent cochlear nerve fibers are the main conduits of auditory information and the fiber terminals of dendrites form unique synaptic connections to IHCs (Gelfand, 2010). These connections are characterized by unique ribbon synapses that facilitate the storage and rapid release of neurotransmitter vesicles across the synaptic cleft (Gelfand, 2010). The postsynaptic site contains glutamate receptors that bind with neurotransmitters, activating cochlear neurons that transmit acoustic information along the auditory pathway (Gelfand, 2010).

ANIMAL STUDIES

Animal studies have reliably shown that IHC synaptic connections to cochlear nerve fibers are sensitive to noise damage, even in cases where IHC's are intact and functional (e.g., Shi et al., 2016). Research exploring noise exposure in mice has found that exposure intense enough to cause a temporary threshold shift results in a significant loss of cochlear neuronal synapses even when IHC's are not significantly damaged (Kujawa & Liberman, 2006; 2009). One prominent theory is that noise damage is caused by glutamate excitotoxicity that is restricted to the IHC synapse (Puel, Safieddine, D'Aldin, Eybalin, & Pujol, 1995; Puel, Ruel, D'Aldin & Pujol, 1998), leaving OHC terminals relatively unaffected (Puel, Pujol, Ttibillac, Ladrech & Eybalin, 1994; Kujawa & Liberman, 2009). Noise also appears to selectively target low and medium spontaneous rate (SR) IHC terminals (Furman, Kujawa & Liberman, 2013; Kujawa & Liberman, 2009; Shi et al., 2016; for review see Liberman & Kujawa, 2017), which have large dynamic ranges, high thresholds, and respond selectively to high intensity, suprathreshold stimuli (Schmiedt et al., 1996).

Recent studies have reported that noise damaged synapses can repair over time, but that coding deficits will often persist. In a recent study by Shi and colleagues (2016), guinea pigs exposed to a single session of continuous high intensity broadband noise exhibited significantly reduced post synaptic fiber density and a marked reduction in click-induced compound action potentials (CAP). Despite the recovery of approximately 80% of the synapses one month later and a significant improvement in CAP responses, a significant decrease in the peak latency spike rate of afferent nerve fibers persisted. This finding suggests that the noise-induced processing deficits may not be caused by a loss of synaptic connections per se; rather, they may be due to aberrant function of repaired synapses that disrupt the peripheral auditory systems ability to rapidly and efficiently encode temporal information (e.g., phase locking to temporal fine structure).

COCHLEAR SYNAPTOPATHY IN HUMANS

Animal studies have yielded considerable evidence that noise exposure damages fibers synapsing on IHCs (Furman, Kujawa & Liberman, 2013) and leads to cochlear synaptopathy and temporal coding deficits (Shi et al., 2016), but to date these findings have not been reliably observed in human studies. Researchers have linked synaptic degradation to aging in both animal (Sergeyenko, Lall, Liberman & Kujawa, 2013), and human post-mortem studies (Viana et al., 2015), and experimental research has linked noise exposure history with various indicators of cochlear synaptopathy (e.g., high frequency hearing thresholds, speech in noise testing [word recognition, monaural low redundancy tests], and electrocochleography) (e.g., Liberman, Epstein, Cleveland, Wang, & Maison, 2016)

If cochlear synaptopathy also occurs in humans, it could account for the development and persistence of functional hearing difficulty following noise exposure or across the lifespan, despite threshold recovery and normal audiometric results. This is particularly salient in elucidating the mechanisms underlying OAD because low-SR IHC fibers are less vulnerable to masking (Schmiedt et al., 1996) and play a key role in processing TFS, spatial release from masking, and sensitivity to binaural sound localization cues that are central to speech discrimination in noise (Bharadwaj, Masud, Verhulst, Mehraei, & Shinn-Cunningham, 2015; Schmiedt et al., 1996), which is a common functional deficit associated with OAD and HHL.

METHODS TO ASSESS SYNAPTIC FUNCTION IN HUMANS

OBJECTIVE ELECTROPHYSIOLOGICAL (EEG) ASSESSMENTS

In the past, researchers have investigated various EEG techniques in an effort to establish a reliable and objective clinical method that is sensitive to damage to peripheral temporal encoding in cochlear synaptopathy (Plack et al., 2016).

1. *The Auditory Brainstem Response (ABR)*

Measuring ABR responses are a promising strategy to identify peripheral synaptic damage because they are fairly simple to perform, and the technology is already available in many clinics. Wave 1 of the ABR is generated by the auditory nerve (Plack et al., 2016), and animal studies have demonstrated that noise exposure causing temporary threshold shifts and low-SR fiber synapse damage significantly reduces ABR wave 1 responses (Kujawa & Liberman, 2009). In line with this, studies in humans have found that self-reported noise exposure is significantly associated with reduced ABR wave 1 amplitudes (Stamper & Johnson, 2015).

2. *The Frequency Following Response (FFR)*

Similar to the ABR. this robust steady state response is an indicator of the brainstem neuronal activity that corresponds to a central temporal processing mechanism, phase locking to a stimulus waveform and the TFS of sound (Plack et al., 2016). Research in humans has found that attenuated reduced FFR responses have been documented in cases of noise exposure (Plack et al., 2016). In addition, a reduction in FFR responses to binaural cues consistent with OAD, such as attending to target speech in the presence of competing background speech, as well as frequency or modulation discrimination have been documented (Plack et al., 2016; Bharadwaj et al., 2015).

3. *Electrocochleography (EcochG)*

There is growing evidence that the ratio of summating potential (SP) and action potential (AP) may be a reliable indicator of cochlear synaptic function. Studies have shown that subjects with high self-reported noise exposure exhibited reduced SP/AP ratios compared to controls (e.g., Liberman et al, 2016). This SP/AP reduction is believed to be an early indicator of damage to the peripheral nerve synapse due to noise and, consistent with OAD/HHL, it occurs in conjunction with a marked reduction in speech in noise performance and hyperacusis.

LIMITATIONS OF EEG ASSESSMENTS

While these techniques have yielded fairly promising results, there are a number of noteworthy limitations that both impede their routine use in the clinic and reduce their diagnostic value for cochlear synaptopathy. Beyond the increased complexity, assessment time, and cost associated with EEG testing, responses are easily influenced by a number of factors. For example, ABR amplitudes are often highly variable across trials, even at the individual level, and responses can be influenced by the patients age, sex, stress level, and degree of physiological noise and/or physiologically generated artifacts (Plack et al., 2016). There are also a number of external factors that can influence ABR responses, such as differences in electrode placement and/or impedance across trials, and variability in environmental sources of electrical interference (Plack et al., 2016) that require the generation of site-specific normative data to minimize. The FFR is also subject to these limitations and can be influenced by differences in central auditory processing.

Another significant limitation the use of ABR and the FFR is that response reflect auditory activity in the brainstem and are not particularly sensitive to the integrity of the peripheral auditory system, which considerably limits their value as tool to differentially diagnose cochlear synaptopathy (Plack et al., 2016). Also, despite promising findings using EcochG, the procedure involves placing electrodes close to the tympanic membrane or, in some cases, proximal to the auditory nerve, which is invasive and not normally conducted in clinic (Gibson, 2017). Further, research findings using more feasible clinical methods (ABR, FFR), are generally inconsistent (e.g., Prendergast et al., 2018). Taken together, the lack of consistent findings in the research combined with the complications associated with employing EEG testing in the clinic, the need for site specific normative data (Chalak, Kale, Deshpande & Biswas, 2013), and overall high variability of responses, particularly for wave amplitudes, call into question the value of objective measures as a standalone diagnostic tool for HHL (Plack et al., 2016).

BEHAVIOURAL ASSESSMENTS OF BINAURAL PROCESSING

A number of behavioural tests have been developed to assess temporal processing and the capacity of the auditory systems to use interaural cues effectively:

1. *Temporal fine structure:*

One of the more promising behavioural approaches is assessing sensitivity to the temporal fine structure of sound. In the cochlea, complex sound signals coded as temporal fine structure (Moore, 2008) via neural phase locking in the auditory nerve (McAlpine, 2005). There is a growing body of research that demonstrates reduced temporal fine structure processing with age, which often accompanies trouble understanding speech in noise (Füllgrabe et al., 2015; Grose & Mamo, 2010; Lunner, Hietkamp, Andersen, Hopkins, & Moore, 2012). In addition, experimental research has shown that degrading temporal fine structure in normal hearing subjects has a minimal impact on speech intelligibility in quiet, but significantly reduces speech intelligibility in the presence of background noise (for review see Moon & Hong, 2014).

To date, several temporal fine structure assessments have been developed, but the most promising are the temporal fine structure 1 (TFS1) test, the temporal fine structure low frequency test (TFSLF), and the temporal fine structure adaptive frequency test (TFSAF) (Moore & Sek, 2009; Hopkins & Moore, 2010; Fullgrabe & Moore, 2017). The TFS1 assesses TFS sensitivity monaurally by testing the ability to identify shifts in harmonic complex tones (i.e., inharmonics), while the TFSLF is a binaural test that establishes the minimum detectable IPD in 500 Hz sinusoidal tone bursts (Moore & Sek, 2009; Hopkins & Moore, 2010). Similar to the approach in the current study, the TFSAF test was designed to establish a frequency threshold for phase difference discrimination. Given that temporal fine structure processing is believed to be central to IPD discrimination (e.g., Füllgrabe, Harland, Sek, & Moore, 2017; Füllgrabe et al., 2018; Hopkins & Moore, 2010), the TFSLF and TFSAF are particularly promising approaches to identify temporal processing deficits. These tests are designed as adaptive two interval, two alternative forced choice tasks and the software is readily available for free online (Sek & Moore, 2012).

While the TFS tests have demonstrated value, there are a number of noteworthy limitations. First, the TFSLF only assesses IPD at low frequencies (i.e., 500 Hz), which may limit the assessments sensitivity to the integrity of peripheral temporal processing. Also, in a recent study by Hoover and colleagues (2019) comparing the clinical utility of a variety of suprathreshold behavioural assessments, the TFSLF test produced stable scores but required an average of seven trials to achieve thresholds, which considerably lengthened the assessment time (i.e., median task time = 41 min), (Hoover, Kinney, Bell, Gallun & Eddins, 2019). The TFSAF uses a similar approach as the present study, but at least three trials are required in order to estimate a reliable threshold (Fullgrabe & Moore, 2017). The developers posited that the average assessment duration would take approximately eighteen minutes (Fullgrabe & Moore, 2017), which is still considerably long for use in a standard clinical setting. While there may be ways to improve the efficiency of these tests (Hoover et al., 2019), in their current form these tests would not be efficient for routine clinical use.

2. *Masking level difference (MLD):*

Another well established and widely used binaural interaction test (McFadden, 1966), MLD assesses the auditory systems ability to utilize IPD cues to identify tones in the presence of a masker (Yost, 1974; Green, 1966; Hirsh, 1948; Webster, 1951). By alternating the phase of the masker and the tone (i.e., tone and/or noise presented binaurally in phase or 180 degrees out of phase), the test establishes the degree to which tones out of phase are released from masking. Masking is most effective when the signal and noise are in phase (i.e., homophasic), and least effective (i.e., maximum release from masking) when both the masker and signal are is out of phase (antiphasic) (Yost, 1974; Green, 1966; Hirsh, 1948; Webster, 1951).

This test can be administered at a variety of frequencies, but the largest MLD (i.e., 10-15dB) is observed at 500Hz, making this the most often employed test procedure used in clinic (McPherson, Senderski, Brunham, Fujiki, Harris, Skarzynski & Kochanek, 2011). The MLD decreases significantly to approximately 3dB at higher frequencies (e.g., above 2000 Hz) (Yost, 1974; Green, 1966; Hirsh, 1948; Webster, 1951), so obtaining MLD at frequencies about 500Hz is not routinely performed in clinic (McPherson et al., 2011). There are many positive features of the MLD test; it is a valuable test in determining binaural integration at the level of the brainstem, and not linguistically loaded like other CAPD assessments, but it is not sensitive enough to distinguish peripheral and central temporal processing deficits, so it likely has limited usefulness in the differential diagnosis of cochlear synaptopathy.

ASSESSMENTS OF SPEECH IN NOISE PERFORMANCE

There are a number of speech-in-noise tests that are routinely used to assess functional hearing ability in the clinic and in research. For example, the Quick Speech-in-Noise (QuickSin) test, BKB-Sin, Hearing-in-Noise test (HINT), and the Words-in-Noise test are widely used in the clinic and have well established normative data and are relatively quick and simple to administer. The coordinate response measure (CRM) is a binaural speech test commonly employed in research (Bolia, Nelson, Ericson & Simpson, 1999), and there are also self-report assessments available (e.g., Speech and spatial qualities scale [SSQ]) (Gatehouse & Nobel, 2004) that can provide valuable information about functional hearing ability in real world environments. While these play a vital role in the assessment OAD and HHL, they are unable to distinguish between peripheral and central processing deficits and have limited usefulness with diverse populations because they are linguistically loaded, which reduces their validity in non-native English speakers, young children, and patients who do not use oral communication.

FREQUENCY THRESHOLDS FOR INTERAURAL PHASE DIFFERENCE DETECTION

In a well-known early study, Yost (1974) employed a discrimination task to investigate the role of IPD's in detecting changes in the lateralization of pure tones. He reported that phase discrimination operates best in the midline (0 or 360 degrees in the azimuth) and is frequency dependent. Specifically, he observed that changes in lateralization could only be detected for tones 900 Hz or lower, and subjects could not discriminate at higher frequencies (e.g., above 2000Hz), even when the phase difference was maximized (Yost, 1974). Building on this work, psychoacoustic research has begun to elucidate the value of IPD detection in assessing auditory peripheral damage and the development of binaural processing deficits across the lifespan.

In a seminal study, Ross and colleagues (2007) employed magnetoencephalography (MEG) to assess binaural phase difference evoked responses and behavioural binaural phase detection across carrier frequencies in young, middle aged, and elderly subjects. Cortical responses were elicited in all age groups for stimuli within the low frequency range (500-750 Hz), but responses began to diminish with increasing frequency. In addition, the frequency thresholds were age dependent, with subjects in the younger group eliciting detectable responses over 1250 Hz, while the middle and older age group yielded response thresholds between 1250-1000, and 750-1000 Hz respectfully.

The younger group also outperformed the middle age and older subjects on a binaural discrimination task, and the groups exhibited frequency thresholds that were in line with the previous MEG findings, although the variability of performance on this task was considerably high in the middle age and older groups (Ross et al., 2007). Taken together, it is possible that the findings of Ross and colleagues (2007) are the result of age-related peripheral synaptic damage, that similar damage occurs following noise exposure, and the damage manifests as a marked reduction in the highest frequency at which IPM can be detected.

Recent studies have examined the interaural phase modulation following response [IPM-FR]; a cortical response that can be measured when tones presented binaurally in and out of phase (at a periodic rate) within the auditory systems detectable range (Undurraga, Haywood, Marquardt & McAlpine, 2016). In normal hearing subjects, Undurraga and collogues (2016) demonstrated that the IPM-FR becomes more robust as the phase difference between binaural pure tones increases. In addition, the IPM-FR is present when binaural phase differences are perceivable by listeners, and listeners are able to correctly discriminate between the IPM and control, static stimuli (Undurraga et al., 2016). This finding suggests that it may be possible to develop a simple behavioural measure of IPD detection, rather than having to rely on complex, costly, and time-consuming EEG measures.

DEVELOPING A BEHAVIOURAL THRESHOLD ESTIMATION TEST

BASICS OF THRESHOLD ESTIMATION

Reliable and accurate threshold estimation is the foundation of behavioural audiometric testing and is considered the gold standard approach to assess hearing function. In audiology, behavioural thresholds are defined as the condition (level, signal to noise ratio etc.) where a subject correctly identifies a stimulus (e.g., tone, speech) 50% of the time it is presented (Walker, Cleveland, Davis & Seales, 2013). In psychoacoustics a number of approaches have been used determine behavioural thresholds for the detection of a target stimulus and are based on signal detection theory (Green & Swets, 1966).

BEHAVIOURAL THRESHOLD TEST DESIGN

There are a number of psychoacoustic tests designed to establish behavioural thresholds, including tests where the subject responds only when a signal is identified, as well as forced choice tasks, where the subject is presented with a set of stimuli and must choose the one that contains the target signal (Green & Swets, 1966). When subjects are asked to identify a target stimulus in noise, there are a finite number of possible outcomes; correct responses (e.g., hits, correct rejections) and incorrect responses (misses, false alarms). Alternatively, in forced choice tasks, when the subject is no longer able to identify the target stimulus they are forced to guess, which can result in the selection of the target or not, based on chance (i.e., the probability of correctly guessing). While there are advantages and disadvantages to each test design, both approaches have been found to produce a valid estimate of behavioural thresholds in the research (e.g., Lesmes, Lu, Baek, Tran, Dosher & Albright, 2015; Vancleef et al., 2018).

1. *The Up-Down Method*

In psychoacoustics, the most common procedure to establish behavioural thresholds is the up-down method (Levitt, 1970). In this approach the stimulus presentation will change adaptively depending on the subject's ability to detect the target stimulus such that the task becomes more difficult following a correct response, and easier following an incorrect response until a threshold is reached (Levitt, 1970). According to a seminal paper by Levitt (1970), up-down testing procedures are based on four main assumptions, the most critical of which that the proportion of positive responses must depend on the level of the target stimulus. Additionally, the psychometric function should remain stable throughout testing, have a specific parametric form, and the subject's responses should be independent (e.g., not influenced by the preceding stimuli). While these are not essential to the up-down testing procedure, they do considerably strengthen the value of the approach overall (Levitt, 1970).

The placement of observations throughout testing and the approach to estimating thresholds based on the subject's responses are key elements in the design of a threshold estimation procedure (Levitt, 1970). When the goal is to determine a subject's true threshold, as in the determination of hearing thresholds in clinic, the initial value chosen should be within the subject's detectable range and measurements should converge as close to the threshold as possible.

2. *Step Size*

Another important feature of a well-designed up-down procedure is defining an adequate increment of change for the stimulus (e.g., step size) (Levitt, 1970). In a simple up-down procedure, step sizes must be chosen carefully because if the step size is too large it will reduce the accuracy of the estimation and if it is too small it will reduce the efficiency of the test (Levitt, 1970). Step sizes can remain constant throughout the test or can adapt to decrease when approaching threshold to obtain a more accurate estimation (Chung, 1954).

3. *Reversals*

The number of reversals (i.e., number of times that the adaptive shifts occur) is also key to ensuring that the threshold is reliably established, and according to Levitt (1970), at least six to eight reversals are required.

4. *Adaptive Response Sequence*

Another important consideration when designing an up-down procedure is assigning a response sequence to determine the adaptive nature of the test. In a simple up-down procedure, a positive response triggers an increase in the difficulty of the test, and a negative response triggers a decrease in difficulty, which corresponds to a 50% probability of a positive response at the convergence of the threshold (Levitt, 1970). An alternative approach is to employ a transformed up-down method where the response sequences are grouped as either up responses, or down responses, thereby increasing the probability of a positive response converging at the threshold. For example, when the response sequence increases (i.e., becomes more difficult to detect) following two sequential positive responses, and decreases (i.e., becomes easier to detect), following a negative response or a positive response followed by a negative response, the probability of a positive response at the convergence of the threshold increases to approximately 71% (Levitt, 1970). While altering the response sequences can influence the likelihood of converging on a true threshold, a simple up-down approach has demonstrated reliability at estimating 50% threshold level and the transformed up-down approach takes considerably longer than the traditional approach, which can reduce the efficiency of the test (Letitt, 1970).

5. *Threshold Estimation*

When designing a threshold estimation test, it is also important to consider the technique used to estimate the threshold. One common technique that is often employed with the traditional up-down method is to calculate mid-run estimates (Wetherill, 1963). In this approach an even number of reversals are used in conjunction with decreasing step size, and the range of responses across runs are averaged to generate a threshold estimate. This approach is simple to employ and has demonstrated efficiency, reliability, and low estimation bias in the research (Wetherill, Chen & Vasudeva, 1966).

POTENTIALS FOR BIAS

Up-down procedures are an efficient approach to threshold estimation (Levitt, 1970) that are regularly employed to develop psychoacoustic signal detection tasks; however, there are other factors that must be considered when designing the test that can influence reliability, validity, and efficiency. For example, signal detection tests, where subjects' respond when they detect a target stimulus that is presented sequentially, can be subject to response bias. To minimize this, the temporal presentation of the stimulus must be random to reduce the subject's ability to recognize the pattern of the up-down procedure and anticipate the presentation of the stimulus (Levitt, 1970). In addition, responses to a continuously presented stimulus may be affected by the subject's reaction time (Woods, Wyma, Yund, Herron & Reed, 2015), so a sufficient response time window must be available.

Conversely, alternative forced choice procedures are less subject to response bias, but the number of choices presented to the subject can influence the validity of the test. For example, if only two options are presented, and the subject is unable to detect the stimulus, they have a 50% chance of guessing the correct option. If the number of options is increased to three it will reduce the probability to 33.33%. While increasing the number of choices will increase the likelihood of true positives (Vancleef et al., 2018), it is reasonable to assume that if there are too many options the test may begin to rely too heavily of the subject's working memory which could undermine the validity of the test. In addition, since alternative forced choice procedures rely on the presentation of the number of decoy stimuli in conjunction with a target, threshold estimation may take longer to establish compared to other approaches.

THE CURRENT STUDY

OVERVIEW

The objective of the study was to develop a frequency-based behavioural assessment of IPD discrimination. Research has demonstrated that detecting interaural phase differences is an easy task at low frequencies and becomes more challenging as the frequency increases (Ross et al., 2007). Given this finding, it seemed plausible that establishing a frequency threshold for binaural phase detection could provide insight into the integrity of binaural processing at the level of the brainstem, and may also be sensitive enough to detect the peripheral temporal processing deficits that may underlie cochlear synaptopathy (e.g., at the level of the auditory nerve). In line with previously developed psychometric tests (Levitt, 1970), our goal was to develop a series of assessments that establishes frequency thresholds for the detection of phase modulated binaural stimuli.

The objective is to evaluate the efficiency and reliability of these clinical approaches to determine which is the most efficient (e.g., time sensitive) and reliable for clinical use.

STUDY GOALS

To accomplish this, the goal of the present study was to design and assess the sensitivity and clinical utility of three approaches to determine frequency thresholds for interaural phase difference detection:

1. A three alternative forced choice (3AFC) task designed to establish a specific frequency threshold using the up-down method outlined by Levitt (1970)
2. A continuous adaptive tracking (CAT) task designed to establish a specific frequency threshold in a manner similar to the 3AFC task
3. A continuous non-adaptive tracking (CNT) task that presents the stimulus randomly across set frequencies to establish a percent correct score that can be used to estimate subject's thresholds.

STUDY HYPOTHESES

There were five hypotheses in this study:

1. Subjects would be able to reliably discriminate phase differences at low frequencies and this ability will diminish as frequency increases
2. The thresholds generated across all tests would be significantly correlated
3. The thresholds estimated by each test would demonstrate consistency across trials
4. The continuous assessments (CAT and CNT) will take significantly less time to complete than the 3AFC assessment

5. Binaural phase detection thresholds, high frequency hearing thresholds, and speech in noise performance will decline with age

CHAPTER 2: METHODS

SUBJECTS

SAMPLE SIZE AND ELIGIBILITY CRITERIA

A sample (N=30) of adults (male and female) were recruited to participate in this study. In order to be eligible to participate, subjects had to be at least 17 years of age and have pure tone audiometric average within the normal range. Also, since the study materials (i.e., consent forms) require that the participants read, write and understand English, all participants were required to be fluent English speakers. This sample size chosen for this study (N=30) is in line with other research studies in this area (McAlpine, Haywood, Undurraga & Marquardt, 2016; Ross et al., 2007; Undurraga et al., 2016), and provided enough statistical power to conduct our analysis.

RECRUITMENT

Participants were recruited on Dalhousie campus and in the city of Halifax, and the primary recruitment method for this study was word of mouth, displaying posters on public message boards around campus and in the community, as well as online and on social media. This research project was approved by the Dalhousie Research Ethics board.

EQUIPMENT

All audiometric testing was conducted on a Grason-Stadler (GSI) 61 audiometer with up to date calibration. The binaural phase detection test was designed in LabVIEW (Bitter, Mohiuddin & Nawrocki, 2006) using a standard desktop computer. Participants were seated in a soundproof booth on a comfortable chair in front of a screen that was connected to the laboratory computer located outside the testing booth. Participants provided responses during testing with a desktop mouse. Standard audiometric testing, speech in noise testing, and binaural phase discrimination tests were conducted using insert earphones and ER3-14A disposable foam ear tips (regular, 13mm) (Etymotic). Extended high frequencies were measured using over the ear headphones (Sennheiser HAD 200) specifically calibrated for high frequency audiometry.

STIMULUS

All stimulus programming was conducted in LabVIEW (Bitter, Mohiuddin & Nawrocki, 2006). The stimulus consisted of two sinusoidal amplitude modulated tones that were presented binaurally. In all conditions, the starting frequency was 250Hz with an initial step size of ¼ octave. After three reversals that step size reduced to 1/20th octave. Tones were calibrated to a level of 70dBSPL in an IEC 711 ear simulator at all frequencies.

In the three alternative forced choice condition, the stimulus was presented three times, with each presentation lasting four seconds. The target tones contained a phase shift and the other two tones were static and did not contain a phase shift. At stimulus onset the target tones were identical (i.e., no phase difference was present) and after one second a phase difference corresponding to +90 degrees in the one ear and -90 degrees in the other ear (i.e., a maximal 180-degree interaural phase difference) was presented for two seconds. The program randomly determined which ear would lead the phase shift in each trial. The phase shift was introduced when the modulation of the tone had reached its minimal point (i.e., the null). This technique has been employed in previous studies (e.g., Ross et al., 2007) to ensure that cues related to the introduction of the phase shift (i.e., tone discontinuity, audible clicks) were not perceivable by the participants.

In the continuous adaptive tracking condition, the same stimulus was used, but the phase difference was introduced randomly between four and eight seconds following the stimulus onset and/or an adaptive frequency change. The response allowance time window was two seconds. When participants did not respond to the stimulus within two seconds (i.e., miss) or responded outside of the allowance time window (i.e., false positive), it was counted as an error and resulted in a lowering of the frequency. When participants responded correctly it triggered a two second cool down timer during which responses did not register. The cool down period was introduced to prevent the program from registering double responses (e.g., cases when the participant presses the button twice in response to a single target).

ASSESSMENT PROCEDURE

INFORMED CONSENT

Following initial contact and confirmation that they were interested in acting as a participant in this study, a time was arranged for the participant to come to the lab. Upon arrival they were presented with the study consent form and were given an opportunity to read the form and ask any questions that they had about the study. Once they were satisfied, they were invited to sign the consent form if they chose to participate.

ELIGIBILITY SCREENING

In order to participate in this study, all participants were required to have a pure tone average within the normal range bilaterally. To ensure that this criterion was met, they underwent a pure tone audiometric assessment, identical in procedure to a standard clinical hearing test. The participant entered a soundproof booth and audiometric thresholds were obtained by presenting pure tones at 250, 500, 1000, 2000, 3000, 4000, and 8000Hz at varying levels until a threshold was reached. They indicated when they heard a tone by pressing a button.

EXTENDED HIGH FREQUENCY AUDIOMETRY

Using the aforementioned audiometric testing procedure and high frequency calibrated headphones, extended high frequency (EHF) audiometric thresholds were obtained by presenting pure tone at 10,000, 12,500, 14,000, and 16,000Hz.

SPEECH IN NOISE TESTING

The QuickSIN speech in noise test was conducted in the right and left ear according to the testing procedure outlined by Etymotic Research (2006). Presentation level was set to 70dB and calibration was performed prior to each use. Participants were presented with 6 sentences in each ear (12 sentences in total) and were asked to repeat the sentence to the best of their ability. The sentences began with an advantageous signal to noise ratio (i.e., 25dBSNR) and decreased with each sentence. The signal to noise ratios for each sentence were 25, 20, 15, 10, 5, 0dBSNR.

BINAURAL PHASE DETECTION TESTING PROTOCOL

PRACTICE AND TESTING PROCEDURE

Prior to beginning the first trial of testing, all subjects participated in a practice session. This was intended to ensure that the participants were familiar with the stimulus that they were required to identify in the subsequent trials. The practice session was identical to the three alternative forced choice condition but only presented the phase difference at a low frequency (i.e., 250 Hz) so that the stimuli could be easily identified. An explanation of the stimuli and procedure was provided, and the subjects were told to practice until they felt confident that they could identify a binaural phase change. Once the participant indicated that they were ready to begin, the first test trial was set up by the researcher.

When the subject was ready, they were told to push the start button and begin. To avoid order effects the start condition of the tests was varied across participants. As a result, half of the participants (n=15) began with the three alternative forced choice test and half (n=15) began with the continuous adaptive tracking test. To minimize practice effects, tests were alternated across trials such that when participants began by performing trial 1 of the three alternative forced choice test, they would then go on to perform trial 1 of the continuous adaptive tracking test and continue alternating until all three trials in each test were completed.

THREE ALTERNATIVE FORCED CHOICE (3AFC) TESTING PROCEDURE

This assessment is based on previous research (e.g., Ross et al., 2007) demonstrating that under similar experimental conditions, subjects are more likely to detect phase differences between tones at low frequencies and will begin to find this task challenging as frequency increases. Participants were presented with three buttons on a screen numbered 1-3 with an indicator light located below each button. Three binaural tones were presented in succession, with the corresponding button lighting up with each presentation. The participant was required to wait until all three tones had played before they could respond. Responses were obtained by using a mouse to click the button that corresponded to the tone that contained the phase difference. If they were not able to detect the phase difference, they were instructed to guess. The tone containing the phase change was presented randomly (i.e., 33.33% chance of occurring at any of the three buttons during each trial).

The stimulus frequency was 250Hz at onset, with a correct response triggering a ¼ octave increase in frequency and an incorrect response a ¼ octave decrease in frequency. This up-down method is consistent with Levitt (1970) and is associated with a 50% chance of converging at threshold with a positive response. At onset, step size was ¼ octave, which was decreased to 1/20th octave following three reversals. A total of twelve reversals was needed to establish the participants frequency threshold for phase difference discrimination in this test, which is in line with the requirements outlined by Levitt (1970). Thresholds were estimated using the mid-run estimate procedure outlined by Wetherill (1963). This procedure was replicated three times to assess test-retest reliability.

CONTINUOUS ADAPTIVE TRACKING (CAT) TESTING PROCEDURE

In this test, subjects were asked to track changes in the binaural phase of continuously presented tones. Similar to the three alternative forced choice procedure, frequency at the onset of testing was 250 Hz. At the start of the test there was no phase difference. The first phase difference was presented randomly after a four-eight second duration had passed. The participants had a two second time window to indicate that they perceived the phase difference by clicking a large button located in the middle of the screen. When the participant did not click the button during the time interval following the phase difference, or they clicked the button when there was no phase difference introduced, it was categorized as a miss/false positive and the frequency was reduced. If the participant correctly identified the phase difference the frequency was increased.

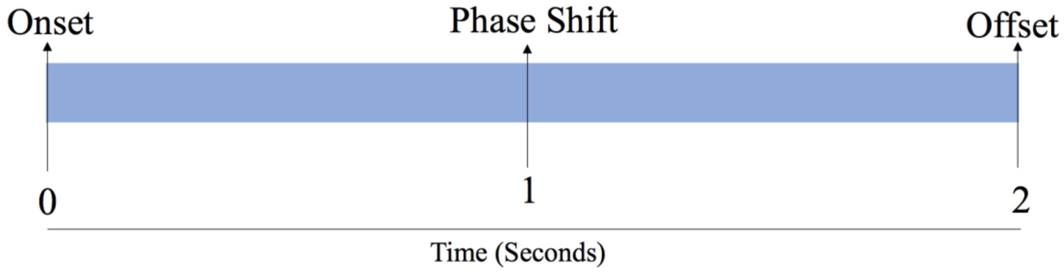
The on-screen button had a light located at each end. The light on the left side flashed red in response to a miss or false positive, and the light on the right flashed green following a correct response. Similar to the three alternative forced choice procedure, step sizes for adaptive frequency changes began with ¼ octave and reduced to 1/20th octave after three reversals had occurred. A total of twelve reversals was needed to establish the participants frequency threshold for phase difference discrimination in this test, which is in line with the requirements outlined by Levitt (1970). Thresholds were estimated using the mid-run estimate procedure outlined by Wetherill (1963). This procedure was replicated three times to assess test-retest reliability.

CONTINUOUS NON-ADAPTIVE TRACKING (CNT) TESTING PROCEDURE

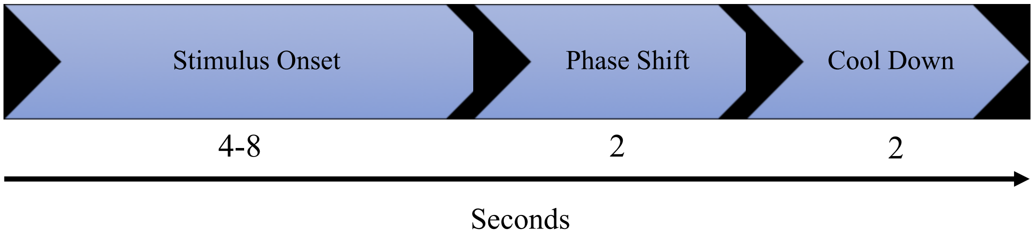
Similar to the previous test, subjects tracked continuously presented tones. The on-screen set-up was identical to the continuous adaptive tracking test. In this test, the phase differences were presented in five blocks, each testing a different frequency (250, 500, 750, 1000, 1250Hz). In each testing block the frequency remained constant and the phase difference was presented 20 times randomly. Similar to the continuous adaptive tracking test, if the participant did not click the button during the time interval following the phase difference, or when the participant clicked the button when there was no phase difference introduced, it was categorized as a miss/false positive and when the participant correctly identified the phase difference it was categorized as a hit/correct response. The button had a light located at each end, with the left light flashing red following a miss or false positive, and the right flashing green following a correct response. This test was designed to generate a percent correct score for each trial (i.e., across each frequency block).

FIGURE 1: STIMULUS TIMING SEQUENCE

Three Alternative Forced Choice:



Continuous Tracking:



DATA ANALYSES

At the onset of the project it was observed that the CNT test procedure was not time efficient. Since the thresholds estimated by this test would not provide the same level of specificity as the 3AFC and CAT test, and the test would take considerably longer to complete, it was removed from the study protocol. As a result, 3 trials of data were collected for the 3AFC and CAT test, yielding a total of 6 trials. All data was analyzed using SPSS statistical software, IBM SPSS Statistics for Mac, version 25 (IBM Corp., 2017).

Pure tone averages at 500, 1000, and 2000 Hz and average EHF thresholds were calculated for each ear. There were a number of cases when high frequency thresholds were outside the limits of the audiometer. To facilitate the calculation of an average EHF threshold, these cases were substituted with the level limit at each frequency +10dB. An average QuickSIN score was calculated by averaging scores in the left and right ear. Binaural phase detection thresholds for each trial were generated using the mid-run estimate approach (Wetherill, 1963; Levitt, 1970). In this approach the range of responses across runs are averaged to generate a threshold estimate.

An intraclass correlation (ICC) and their 95% confidence intervals was performed to assess test-retest reliability and was calculated based on mean-rating (k=3), absolute agreement, using a 2-way mixed-effect model (Fisher, 1954; Koo & Li, 2016). The reliability of each test was interpreted according to guidelines by Koo and Li (2016), with values less than 0.5 indicating poor reliability, between 0.5 and 0.75 indicating moderate reliability, between 0.75 and 0.9 indicating good reliability, and greater than 0.90 indicating excellent reliability (Koo & Li, 2016).

In cases where the difference in thresholds across the first two trials exceeded the mean difference by two standard deviations, the ICC of the two trials of highest agreement were compared (Hoover et al., 2019). To assess differences in reliability between tests, the 95% confidence intervals for ICC estimation were compared (Lu & Shara, 2007). Final thresholds were calculated by taking the mean of the first two trials. Similar to the approach employed by Hoover and colleagues (2019), in cases where one of the first two trials exceeded the mean difference by two standard deviations, the two trials of highest agreement were used to calculate the subject's final threshold.

Repeated measures ANOVA's were employed to examine the mean differences, main and interaction effects across tests for test time, binaural detection threshold estimates, the impact of age group on final thresholds, and impact of start condition (i.e., three alternative forced choice vs. continuous adaptive tracking). Post-hoc testing consisted of paired t-tests with Bonferroni correction. Hierarchical regression was performed to examine the degree to which binaural phase detection thresholds were predicted by age, high frequency thresholds, and speech in noise performance (i.e., binaural QuickSIN). Preliminary data analysis indicated that one participant in the study was a significant outlier. A multiple regression diagnostic analysis (Field, 2005) was performed and the results indicated that the subject was an influential case (e.g., covariance ratio=0.24; Standardized DFbeta absolute value= 1.5; DFfit= 0.99). After concluding that the participant was an outlier that was exerting undue influence on the regression models they were excluded from the analysis.

CHAPTER 3: RESULTS

DESCRIPTIVE STATISTICS

Descriptive statistics are reported in table 1. Ages ranged from 17 to 72, with an average age of 37 years (SD=15.90). The majority of the sample was female (79%). Pure tone averages were all within the normal range bilaterally (right mean=6.34, SD=4.46; left mean=5.41, SD=4.31). Mean QuickSIN scores were within the normal range bilaterally (right mean=1.29, SD=1.24; left mean=1.91, SD=1.64). The mean average EHF threshold was 29dB for the right ear (SD=29dB) and 28dB for the left ear (SD=28dB).

TABLE 1: *DESCRIPTIVE STATISTICS: DEMOGRAPHIC*

	Mean	Standard Deviation	Minimum	Maximum
Age	37.10	15.90	17	72
PTA Right	6.34	4.46	-3	15
PTA Left	5.41	4.30	-2	18
QSIN Right	1.29	1.24	-0.50	3.50
QSIN Left	1.91	1.64	-0.50	6.50
Ext. High Freq Av Right	28.32	28.77	-5.00	96.25
Ext. High Freq Av Left	28.31	28.00	-5.00	100.00

TEST RELIABILITY

THREE ALTERNATIVE FORCED CHOICE

An ICC was employed to examine the absolute agreement of binaural phase detection threshold estimates across 3 trials of the alternative forced choice test. When all three trials were compared, ICC estimates and their 95% confidence intervals indicated moderate reliability (ICC=0.65, 95% CI=0.37-0.82). Cronbach's alpha=0.66. In cases where the difference in thresholds across the first two trials exceeded the mean difference of the two trials (84.43Hz) with the highest agreement by two standard deviations (i.e., 179.58Hz) 500Hz (n=14; 47%), the ICC of the two trials of highest agreement were compared, which yielded ICC estimates and 95% confidence intervals that indicated good reliability (ICC=0.85, 95% CI=0.67-0.93). Cronbach's alpha=0.85.

CONTINUOUS ADAPTIVE TRACKING

An ICC was employed to examine the absolute agreement of binaural phase detection threshold estimates across 3 trials of the continuous adaptive tracking test. When all three trials were compared, ICC estimates and their 95% confidence intervals indicated good reliability (ICC=0.75, 95% CI=0.53-0.87). Cronbach's alpha= 0.75. In cases where the difference in thresholds across the first two trials exceeded the mean difference of the two trials with the highest agreement (59.17 Hz) by two standard deviations (i.e., 129.5 Hz) (n=10; 33%) , the ICC of the two trials of highest agreement were compared, which yielded ICC estimates and 95% confidence intervals that indicated excellent reliability (ICC=0.95, 95% CI=0.89-0.98). Cronbach's alpha=0.95.

INTRACLASS CORRELATION BETWEEN TESTS

To assess differences in reliability between tests, the 95% confidence intervals for ICC estimation were compared (Lu & Shara, 2007). In all cases 95% confidence intervals overlapped, indicating that there were no significant differences in test-retest reliability across all three trials of each test, or across the two trials in highest agreement.

THE INFLUENCE OF START CONDITION ON THRESHOLD

A two-way repeated measures ANOVA was employed with binaural phase detection threshold as the dependent variable, start condition (three alternative forced choice, continuous adaptive tracking) as the between subjects' factor, and test type (2 levels) and trial (3 levels) as the within subjects' factors. Mauchly's test indicated that the assumption of sphericity was not violated ($\chi 2(2) = 5.54$, $p = .06$). The analysis indicated that start condition was not significant ($p=0.87$) but yielded a significant main effect of test type, $F(1,27)=44.95$, $p<0.001$, partial $\eta 2 =0.625$. All other within subjects comparisons were not significant (main effect of trial, $p=0.14$; interactions [test X start condition, $p=0.63$; trial X start condition, $p=0.15$; test X trial, $p=0.31$; test X trial X start condition, $p=0.72$]).

THRESHOLD ESTIMATION

Descriptive statistics for binaural detection thresholds for each trial can be found in table 2. Mean binaural phase detection thresholds for all three trials were 1554Hz (SD=346.48Hz) for the three alternative forced choice test and 1220Hz (SD=305.90Hz) for the continuous adaptive tracking test. Final thresholds were calculated by taking the mean of the first two trials. In cases where one of the first two trials exceeded the mean difference by two standard deviations, the two trials of highest agreement were used to calculate the subject's final threshold. Mean final thresholds were 1479.89Hz (SD=209.57Hz) for the three alternative forced choice test and 1284.81 Hz (SD=240.06Hz) for the continuous adaptive tracking test. A correlation analysis was performed to examine the extent to which thresholds on both tests were in agreement, and the analysis revealed a significant positive correlation (r(29)=0.73, p<0.001).

TABLE 2 *DESCRIPTIVE STATISTICS: BINAURAL PHASE DETECTION THRESHOLDS*

Test	Trial	Mean	Standard Deviation	Minimum	Maximum
Forced Choice	1	1502.17	288.86	1084.00	2228.00
	2	1646.28	432.36	1154.00	2732.00
	3	1540.41	328.47	1062.00	2849.00
	Total	1554.02	346.48	1062.00	2849.00
Adaptive Tracking	1	1174.41	362.28	205.00	1862.00
	2	1228.48	298.76	362.00	1936.00
	3	1258.62	251.26	712.00	1783.00
	Total	1220.51	305.90	205.00	1936.00

TIME EFFICIENCY ACROSS TESTS

Descriptive statistics for the time of each test across trials can be found in table 3. The mean time to complete a three alternative forced choice trial was 4.5 minutes (SD=0.73), and the mean time to complete a continuous adaptive tracking trial was 2.9 minutes (SD=0.69). A two-way repeated measures ANOVA was employed with test time as the dependent variable and test (2 levels) and trial (3 levels) as the within subjects' factors. For both trial and the test by trial interaction the assumption of sphericity was violated (p<0.001), so Greenhouse-Geisser corrections were used. The analysis indicated a significant main effect of test (F(1, 28)=199.31, p<0.001, partial η2 =0.88). The main effect of trial was not significant (p=0.68) and the interaction between test and trial was also not significant (p=0.45).

TABLE 3 *DESCRIPTIVE STATISTICS: TEST TIME*

Test	Trial	Mean	Standard Deviation	Minimum	Maximum
Forced Choice	1	4.56	0.73	3.34	6.12
	2	4.35	0.79	3.13	6.67
	3	4.53	0.68	3.38	6.06
	Total	4.48	0.73	3.13	6.67
Adaptive Tracking	1	2.86	0.58	1.59	4.32
	2	2.91	0.56	2.13	6.90
	3	2.86	0.88	1.59	4.32
	Total	2.88	0.69	1.59	6.90

HIERARCHICAL REGRESSION MODELS: EXAMINING THE EFFECT OF AGE, HIGH FREQUENCY THRESHOLDS, AND SPEECH IN NOISE PERFORMANCE ON BINAURAL PHASE DETECTION

A correlation analysis was employed to examine the relationship between final alternative forced choice and continuous adaptive tracking test thresholds and age, average EHF hearing thresholds, and speech in noise performance (i.e., average QuickSIN scores). Correlation values are reported in table 4. The analysis revealed that final thresholds for both tests were significantly negatively correlated with age (alternative forced choice, r(29)= -0.49, p<0.01; continuous adaptive tracking, r(29)= -0.55, p<0.01). The analysis also revealed a strong positive correlation between age and high frequency hearing thresholds (r(29)=0.9, p<0.001).

TABLE 4 *CORRELATION TABLE: THRESHOLDS, AGE, HIGH FREQUENCY HEARING, QUICKSIN*

		Age	Average QuickSIN	Average EHF	AFC Thresholds
Average QuickSIN	Pearson Correlation	0.23			
	Sig. (2-tailed)	0.23			
Average EHF	Pearson Correlation	0.90**	0.25		
	Sig. (2-tailed)	0.00	0.19		
AFC Thresholds	Pearson Correlation	-0.50**	-0.35	-0.50**	
	Sig. (2-tailed)	0.01	0.06	0.01	
CAT Thresholds	Pearson Correlation	-0.55**	-0.34	-0.52**	0.73**
	Sig. (2-tailed)	0.002	0.08	0.004	0.00

Note: ** Correlation is significant at the 0.01 level (2-tailed).

A hierarchical regression analysis was performed to examine the ex tent to which age, EHF thresholds, and speech in noise performance predicted binaural phase detection thresholds for both tests. The relevant assumptions of hierarchical regression were assessed prior to analysis and all assumptions were met. The sample size (N=29) was appropriate to accommodate the use of three independent variable is the regression model (Tabachnick & Fidell, 2001; Field, 2005). The assumption of singularity was satisfied for all independent variables included in the model (i.e., age, EHF thresholds, average QuickSIN scores). EHF hearing was highly correlated with age (r(29)=0.9, p<0.001), indicating that multicollinearity could be problematic (Field, 2005; Coakes, 2005; Hair et al., 1998). To address this, Variance Inflation Factor (VIF) and tolerance values were assessed for both regression models (AFC, CAT), and were found to be within the acceptable limits (e.g., VIF<10, tolerance value > 0.1) (Coakes, 2005; Hair, Black, Babin

& Anderson, 1998). Finally, P-P plots (Figure 2A and 2B) and scatterplots of standardized residuals and predicted values (Figure 3A and 3B) were generated and indicated that normality and homoscedasticity were satisfied in both models (Hair et al., 1998; Pallant, 2001).

FIGURE 2 P-P PLOTS

A) Alternative Forced Choice Test B) Continuous Adaptive Tracking Test

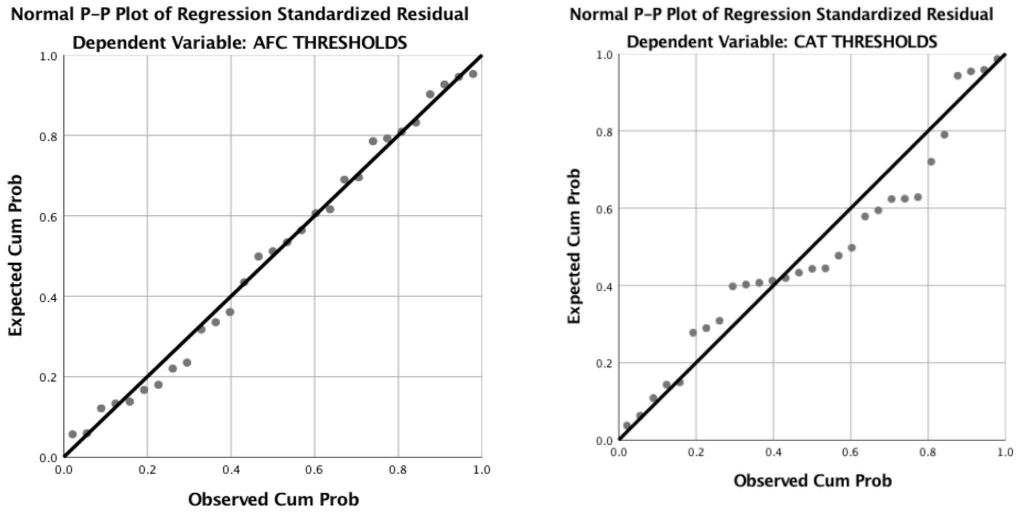
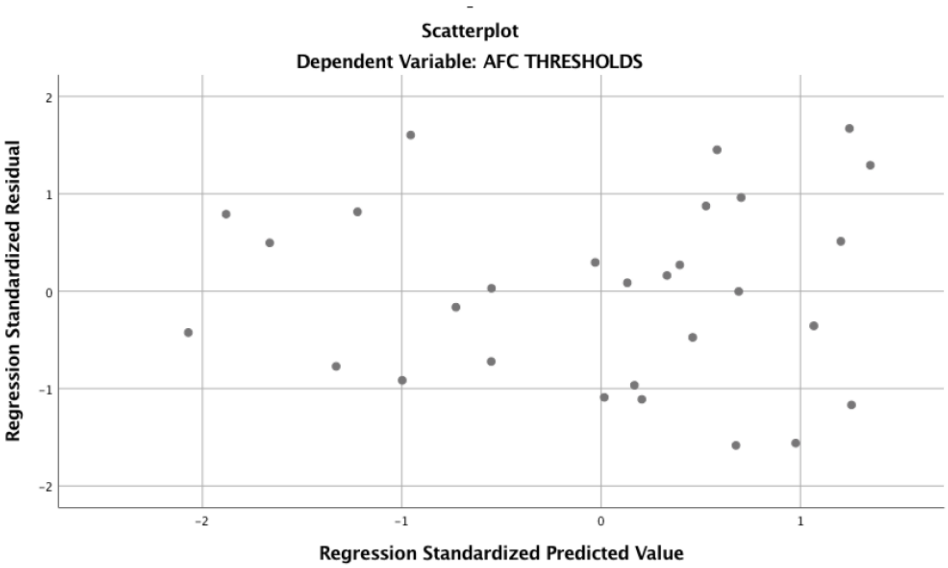
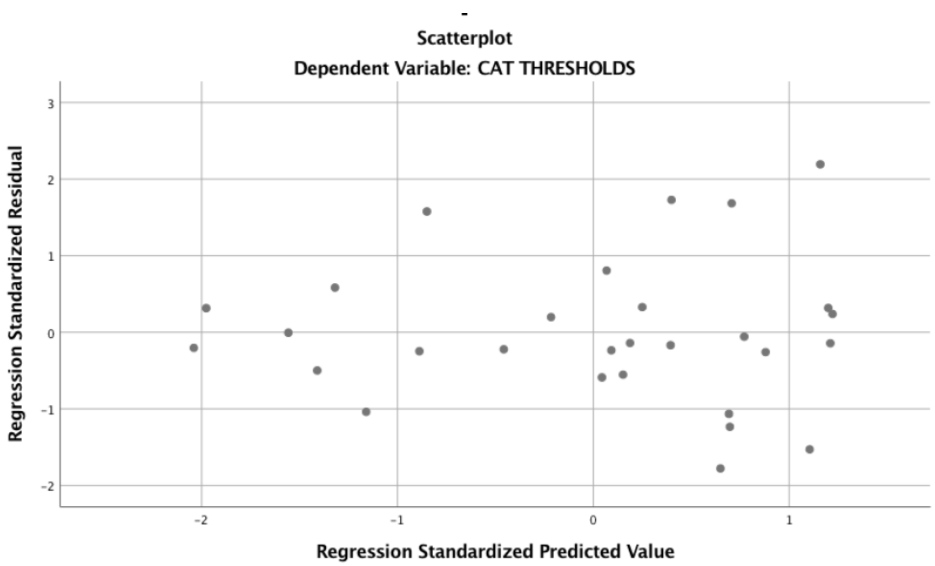


FIGURE 3: SCATTER PLOT OF STANDARDIZED RESIDUALS AND PREDICTED VALUES

A) Alternative Forced Choice



B) Continuous Adaptive Tracking



MODEL 1:THREE ALTERNATIVE FORCED CHOICE

A two-stage hierarchical multiple regression was conducted with binaural phase detection thresholds as the dependent variable. Multiple regression variables are reported in Table 5. The analysis revealed that, at stage one, age contributed significantly to the regression model, F(1,27)=8.37, $p<0.01$, and accounted for 24% of the variability in binaural phase detection thresholds. In stage 2, high frequency hearing thresholds and binaural QuickSIN scores were added to the model. The addition of these variables accounted for an additional 6% of the variance in threshold; however, the R^2 change was not significant, F(1, 25)=1.27, $p>0.05$). The regression equation generated in the first model is AFC Threshold= 1717.82 - 6.41(age) and the regression equation generated in the second model is AFC Threshold= 1704.32 - 2.75(age) -1.83(average EHF hearing threshold) - 43.46(average QuickSIN score).

TABLE 5 *HIERARCHICAL REGRESSION ANALYSIS FOR VARIABLE PREDICTING THRESHOLDS ON THE THREE ALTERNATIVE FORCED CHOICE TEST*

Variable	β	t	Sr²	R	R²	ΔR²
Step 1				0.49	0.24	0.24
Age	-0.49	-2.89**	0.24			
Step 2				0.55	0.31	0.07
Age	-0.21	-0.54	0.01			
EHF	-0.25	-0.63	0.02			
QuickSIN	-0.24	-1.39	0.07			

Note. N = 29; *p < .05, **p < .01, ***p< .001

MODEL 2: CONTINUOUS ADAPTIVE TRACKING

A two-stage hierarchical multiple regression was conducted with binaural phase detection thresholds as the dependent variable. Multiple regression variables are reported in Table 6. The analysis revealed that at stage one, age contributed significantly to the regression model, F(1,27)=11.56, p<0.01, and accounted for 30% of the variability in binaural phase detection thresholds. In stage 2, high frequency hearing thresholds and binaural QuickSIN scores were added to the model, which accounted for an additional 5% of the variance in threshold; however, the R^2 change was not significant, F(1, 25)=0.90, p>0.05). The regression equation generated in the first model is CAT Threshold=1591.60 – 8.27(age) and the regression equation generated in the second model is CAT Threshold=1617.04 – 6.52(age)-0.62(average EHF threshold)-45.221(average QuickSIN score).

TABLE 6 *HIERARCHICAL REGRESSION ANALYSIS FOR VARIABLES PREDICTING THRESHOLDS ON THE CONTINUOUS ADAPTIVE TRACKING TEST*

Variable	β	t	Sr^2	R	R^2	ΔR^2
Step 1				0.55	0.30	0.30
Age	-0.55	-3.40**	0.30			
Step 2				0.59	0.35	0.05
Age	-0.43	-1.14	0.05			
EHF	-0.72	-0.19	0.0014			
QuickSIN	-0.22	-1.30	0.064			

Note. N = 29; *$p < .05$, **$p < .01$, ***$p < .001$

AGE AND TEST TIME

A correlation analysis was performed to examine the relationship between age and test time across each trial of binaural phase detection testing. Correlation values are listed in table 7. The results of this analysis indicated a significant positive correlation between age and test time on the first trial of the three alternative forced choice test ($r(29)=0.39$, $p<0.05$; two tailed). There were no other significant correlations present.

TABLE 7 CORRELATION TABLE: AGE AND TEST TIME

		AFC Trial			CAT Trial		
		1	2	3	1	2	3
Age	Pearson Correlation	0.39*	0.27	0.21	-0.10	0.12	-0.10
	Sig. (2-tailed)	0.04	0.15	0.28	0.62	0.54	0.62

*. Correlation is significant at the 0.05 level (2-tailed).

CHAPTER 4: DISCUSSION

The goal of the present study was to design three clinical tests to measure frequency thresholds for binaural phase difference detection and compare their respective reliability and clinical efficiency. The test designs chosen for this study were 3AFC, CAT, and CNT. During the pilot phase of the study it was observed that the CNT test procedure was not time efficient. Given that the thresholds estimated by this test would not provide the same level of specificity as the 3AFC and CAT test, and the test was not time efficient, it was removed from the study protocol. As a result, three trials of data were collected for the 3AFC and CAT test, yielding a total of six trials.

For the AFC test, the mean binaural phase detection thresholds for all three trials was 1554Hz (SD=346.48Hz) and estimated final thresholds were 1479.89Hz (SD=209.57Hz). For the CAT test, the mean threshold across all trials was 1220Hz (SD=305.90Hz) for the CAT test and the mean final threshold was 1284.81 Hz (SD=240.06Hz). In general, the CAT test demonstrated better reliability and shorter trial duration than the 3AFC test. The findings, implications, and limitations of the present study are discussed in detail below, in addition to considerations for future research.

HYPOTHESES AND MAIN FINDINGS

There were five hypotheses in this study:

1. Subjects will be able to reliably discriminate phase differences at low frequencies and this ability will diminish as frequency increases
2. The thresholds generated across all tests would be significantly correlated
3. The thresholds estimated by each test would demonstrate consistency across trials

4. The continuous assessments (CAT and CNT) will take significantly less time to complete than the 3AFC assessment
5. Binaural phase detection thresholds, high frequency hearing thresholds, and speech in noise performance will decline with age

HYPOTHESES 1: SUBJECTS WILL BE ABLE TO RELIABLY DISCRIMINATE PHASE DIFFERENCES AT LOW FREQUENCIES AND THIS ABILITY WILL DIMINISH AS FREQUENCY INCREASES

In line with previous research examining the relationship between phase difference detection and frequency (e.g., Kohlrausch, 1986; Ross et al., 2007), subjects could reliably identify phase differences at low frequencies (e.g., 250-840Hz) and phase detection became more challenging as the frequency increased, until a ceiling was reached. In addition, all participants were able to detect phase differences up to 1000Hz (AFC=1062Hz; CAT=1001Hz), which is in line with findings reported in similar studies (e.g., Ross, Tremblay & Picton, 2007, Ross, Fujioka, Tremblay & Picton, 2007). Considering that the relationship between frequency and phase detection is well documented (Kohlrausch, 1986; Ross et al., 2007), and posited to reflect the increased difficulty associated with TFS phase locking to rapidly fluctuating signals (Demany & Semal, 1989; Moore & Sek, 1996), replicating these findings was essential to demonstrating the validity of the stimuli design and testing procedure employed in the current study.

HYPOTHESIS 2: THE THRESHOLDS GENERATED ACROSS ALL TESTS WOULD BE SIGNIFICANTLY CORRELATED

Descriptive statistics for binaural detection thresholds for each trial can be found in table 2. Mean binaural phase detection thresholds for all three trials were 1554Hz

(SD=346.48Hz) for the AFC test and 1220Hz (SD=305.90Hz) for the CAT test (see Figure 4). Final thresholds were calculated by taking the mean of the first two trials. In cases where one of the first two trials exceeded the mean difference by two standard deviations, the two trials of highest agreement were used to calculate the subject's final threshold. Mean final thresholds were 1479.89Hz (SD=209.57Hz) for the AFC test and 1284.81 Hz (SD=240.06Hz) for the CAT test (see Figure 5).

FIGURE 4: BAR GRAPH DEPICTING MEAN THRESHOLDS ON EACH TRIAL OF THE ALTERNATIVE FORCED CHOICE AND CONTINUOUS ADAPTIVE TRACKING TESTS

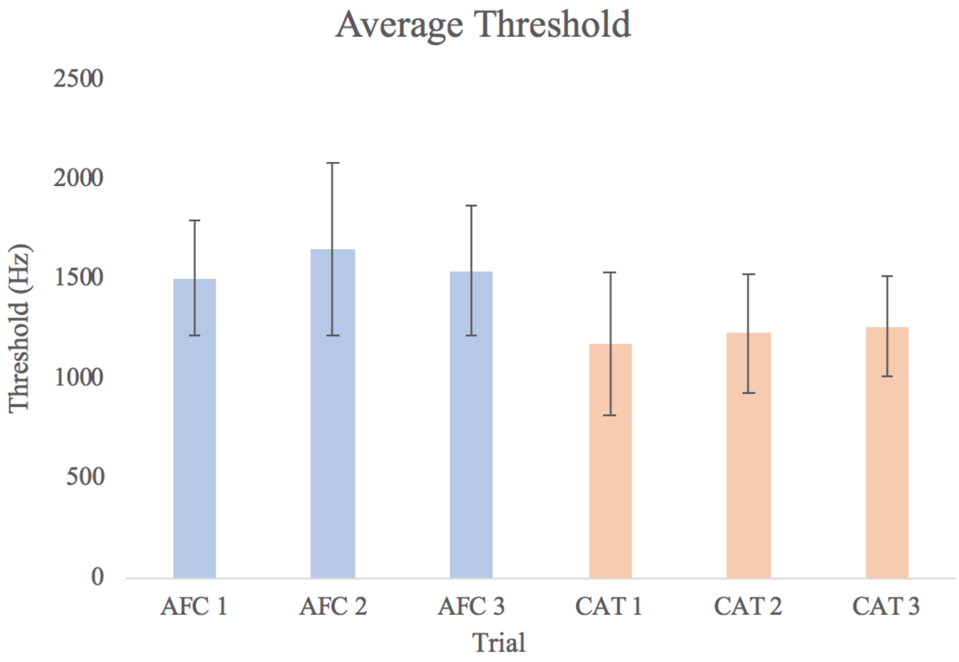
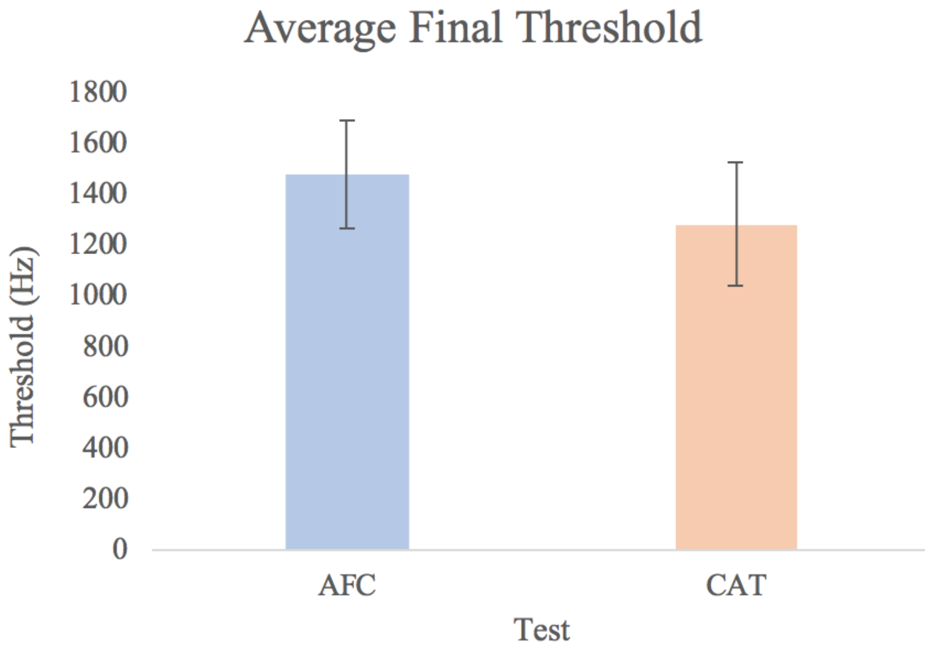


FIGURE 5: BAR GRAPH DEPICTING MEAN FINAL THRESHOLDS FOR THE ALTERNATIVE FORECED CHOICE AND CONTINUOUS ADAPTIVE TRACKING TESTS



Previous studies examining binaural phase difference detection typically employ AFC tests (Ross, Tremblay & Picton, 2007, Ross, Fujioka, Tremblay & Picton, 2007). The present study compared an AFC test with a continuous tracking procedure, which is a novel approach to measuring IPD thresholds. Given that the AFC approach to IPD has been employed successfully in the past, a correlation analysis was conducted to determine the extent to which thresholds estimated by the AFC and CAT tests were in agreement. The analysis indicated that the IPD thresholds obtained in both tests were significantly correlated ($r(29)=0.73$, $p<0.001$), which supports hypothesis 2 of the current study.

In addition, the IPD thresholds obtained in the present study are in line with behavioural thresholds reported in similar studies. For example, Ross and collogues (2007) designed a 2AFC test of binaural phase detection, and reported mean thresholds of 1203Hz, which is in line with the thresholds estimated in the current study

(AFC=1479.89Hz; CAT=1284.81 Hz). Given that the CAT approach has not been tested in past research, the correlation between AFC and CAT thresholds in the current study and similarities to previously reported findings in the literature (e.g., Ross et al., 2007) indicates that the CAT test provides a valid estimation of behavioural IPD thresholds.

While AFC and CAT thresholds were correlated, parametric test results revealed a significant main effect of test, indicating that AFC thresholds were higher overall compared to CAT thresholds. In fact, the CAT thresholds were closer by comparison to those reported in the literature (1284.81 vs. 1203Hz reported by Ross et al., 2007). Also, the maximum threshold on the AFC test was 2849Hz, which is considerably higher than the maximum CAT threshold (1936Hz), and the behavioural thresholds reported in the literature (e.g., Ross et al., 2007: ~1700Hz).

There are a number of factors that could be contributing to this finding. First, both tests were designed using the simple up-down procedure (i.e., a positive response triggers an increase in the difficulty of the test, and a negative response triggers a decrease in difficulty), which corresponds to a 50% probability of a positive response at the convergence of the threshold (Levitt, 1970). Alternatively, Ross and colleagues (2007) and Fullgrabe and Moore (2017) employed a transformed up-down procedure that required two successive positive responses before triggering an increase in frequency, which increases the likelihood of converging on threshold to 70% (Levitt, 1970). On the AFC test when subjects reach their IPD threshold and are no longer able to identify the phase difference they are forced to guess, and in a 3AFC tests, they have a 33.33% chance of guessing correctly.

Given these methodological differences, it is possible that the high thresholds on the AFC test in the currents study can be attributed to a decreased likelihood of converging on threshold and/or the result of guessing the appropriate response correctly. In fact, Fullgrabe and Moore (2017) also observed higher than expected thresholds on the TFSAF test in a number of cases (e.g., 2698 Hz). Given that the TFSAF is an alternative forced choice test, the researchers posited that the threshold inflation resulted from participants making a series of correct guesses. They suggested that threshold inflated could be mitigated by either altering the response procedure from two up, one down to three up, one down, or increasing the number of response options from two to three. The researchers also noted that these changes would improve the accuracy of threshold estimation, but that they would also significantly extend the test duration.

Inflation of responses due to guessing would be less likely to influence the threshold estimated by the CAT test because theoretically subjects will only respond when they actively detect a threshold and are not forced to respond when they are unable to detect the IPD. Considering that both of the tests in the current study used the same adaptive response sequence, and the probability of correctly guessing the response on the 3AFC test is considerably lower than chance (i.e., 33.33%), it is unlikely that the observation that AFC thresholds were higher than both the CAT thresholds and those reported by Ross and colleagues (2007) is due to these methodological factors.

In order to demonstrate that high AFC thresholds were not due to the probability of guessing, response patterns were plotted (see Figure 6 & 7). The plotted responses clearly demonstrate a linear increase at low frequencies, followed by a flattening of the response curve and subsequent fluctuation above and below the threshold.

FIGURE 6: CONTINUOUS ADAPTIVE TRACKING RESPONSE PATTERNS FOR THE PARTICIPANTS WITH THE HIGHEST ESTIMATED THRESHOLDS (P5 & P12)

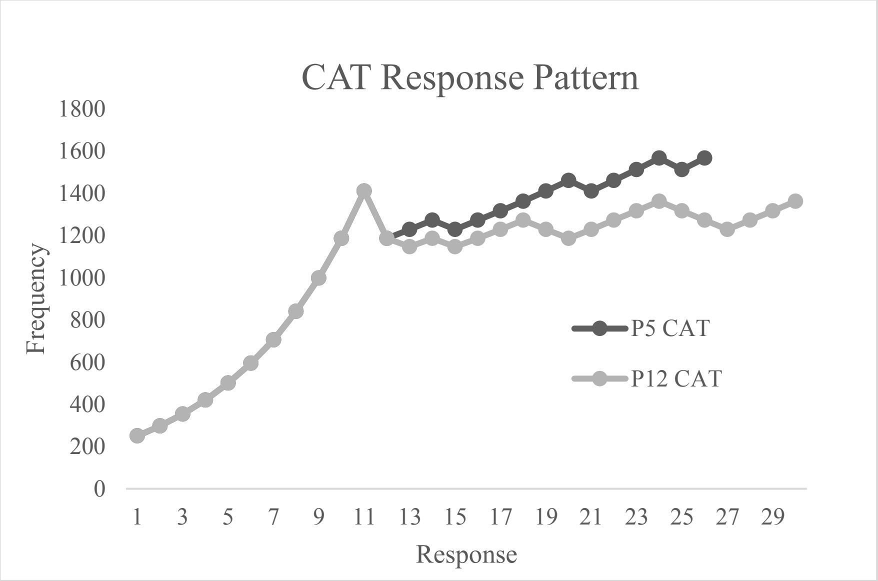
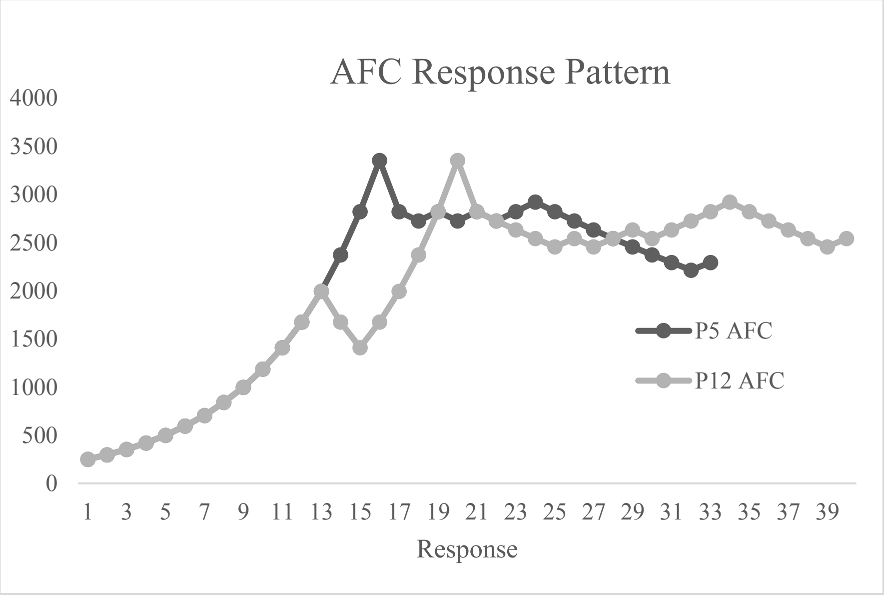


FIGURE 7: ALTERNATIVE FORCED CHOICE RESPONSE PATTERNS FOR THE PARTICIPANTS WITH THE HIGHEST ESTIMATED THRESHOLDS (P5 & P12)



This pattern of responses is consistent with a true threshold estimate, indicating that the higher than expected thresholds on the AFC test were not the result of guessing. We cannot rule out the possibility that there was a subtle acoustic cue present in the AFC test that made it easier for participants to detect the IPD, but given that the test was conducted using insert earphones and the stimuli and equipment was calibrated prior to use, it is unlikely that this was the case. It is also important to note that in all cases where thresholds were higher than expected, they exceeded the mean test difference by two standard deviations and were not used to calculate the subject's final threshold value.

In addition, we cannot rule out the possibility that the thresholds obtained in the AFC test provided a better estimate of IPD threshold, and the CAT thresholds were deflated. If this is the case, one possible explanation is that the CAT test is more challenging than the AFC test. The CAT test requires tracking, which relies heavily on the subjects' attention and reaction time (Woods, Wyma, Yund, Herron & Reed, 2015), and requires that subjects identify phase shifts occurring in both directions, while the AFC test only requires that they identify phase shift in one direction. While we cannot rule out the possibility that CAT thresholds are deflated in the current study, it is unlikely that this is case, given the high agreement of CAT thresholds with those reported in similar studies (Ross, Tremblay & Picton, 2007, Ross, Fujioka, Tremblay & Picton, 2007).

Another important consideration is that normative data for IPD thresholds have not been developed, and thresholds reported in the extant research are using small sample sizes (e.g., Ross et al., n=33). It is possible that the true upper frequency limit for IPD detection has not been established. Finally, it is important to note that, despite the differences in the thresholds estimated by the CAT and AFC tests, they were significantly correlated. Given that there were a number of differences in the testing procedure and approach, it is likely that the threshold estimates generated by each test reflect different points on the psychometric function and are equally meaningful. Future studies may explore this difference by replicating the current study is a large sample of normal hearing subjects, to establish normative data and increase the likelihood of determining a reliable estimate of the upper frequency limit for IPD detection on both AFC and CAT tests.

HYPOTHESIS 3: THRESHOLDS ESTIMATED BY EACH TEST WOULD DEMONSTRATE CONSISTENCY ACROSS TRIALS

One of the main goals of the current study was to demonstrate the reliability of different approaches to establishing IPD frequency thresholds. To accomplish this, an intraclass correlation (ICC) was employed to examine test-retest reliability across all three trials for each test. For the AFC test, ICC estimates and their 95% confidence intervals indicated moderate reliability (ICC=0.65, 95% CI=0.37-0.82, Cronbach's alpha=0.66). For the CAT test, an ICC across 3 trials indicated good reliability (ICC=0.75, 95% CI=0.53-0.87, Cronbach's alpha= 0.75).

For both the AFC and CAT test, there were a number of cases where thresholds were much lower or higher than expected (e.g., greater than +2SD from the mean). The discrepancy across trials could be due to multiple factors, including practice effects, attentional influences, and test fatigue. Given that one of the main objectives of the current study is to assess reliability of the AFC and CAT tests for IPD, we employed a procedure similar to Hoover and colleagues (2019) to calculate final threshold values. Specifically, in cases where one of the first two trials exceeded the mean difference by two standard deviations, the two trials of highest agreement were used to generate a final threshold value.

When final thresholds were calculated, the ICC for the AFC test increased from moderate to good reliability (ICC=0.85, 95% CI=0.67-0.93, Cronbach's alpha=0.85). For the CAT test, when the ICC of the two trials used to generate final thresholds the ICC increased from good to excellent reliability (ICC=0.95, 95% CI=0.89-0.98, Cronbach's alpha=0.95). These findings suggest that the CAT test is a more reliable test overall. In order to explicitly test this, we compared the 95% confidence intervals for ICC estimation (Lu & Shara, 2007). In all cases 95% confidence intervals overlapped, indicating that the difference was not significant. This finding indicates that, while the CAT test may exhibit slightly higher reliability overall, both tests exhibit adequate reliability, particularly when multiple test runs are conducted, and final thresholds are calculated using the two trials in highest agreement.

HYPOTHESIS 4: THE CONTINUOUS ASSESSMENTS (CAT AND CNT) WILL TAKE SIGNIFICANTLY LESS TIME TO COMPLETE THAN THE 3AFC ASSESSMENT

Previous approaches to assess the integrity of the peripheral temporal function (i.e., TFS processing) have relied on complex, time consuming, and relatively insensitive objective and behavioural assessments, one of the main goals of the present study is to determine which approach to measuring behavioural IPD thresholds is the most efficient and suitable for routine clinical use. To establish test efficiency, we compared the mean duration of testing for each trial across all three trials. The mean time to complete an AFC trial was 4.5 minutes (SD=0.73, range: 3.13-6.75 min), and the mean time to complete a CAT trial was 2.9 minutes (SD=0.69, range=1.59-6.90) (see table 3). In line with hypothesis 4, a two-way repeated measures ANOVA revelated that the CAT test duration was significantly shorter than the AFC test duration (See Figure 8 & 9).

FIGURE 8: BOX AND WHISKER PLOT OF TRIAL DURATIONS ON THE ALTERNATIVE FORCED CHOICE AND CONTINUOUS ADAPTIVE TRACKING TESTS

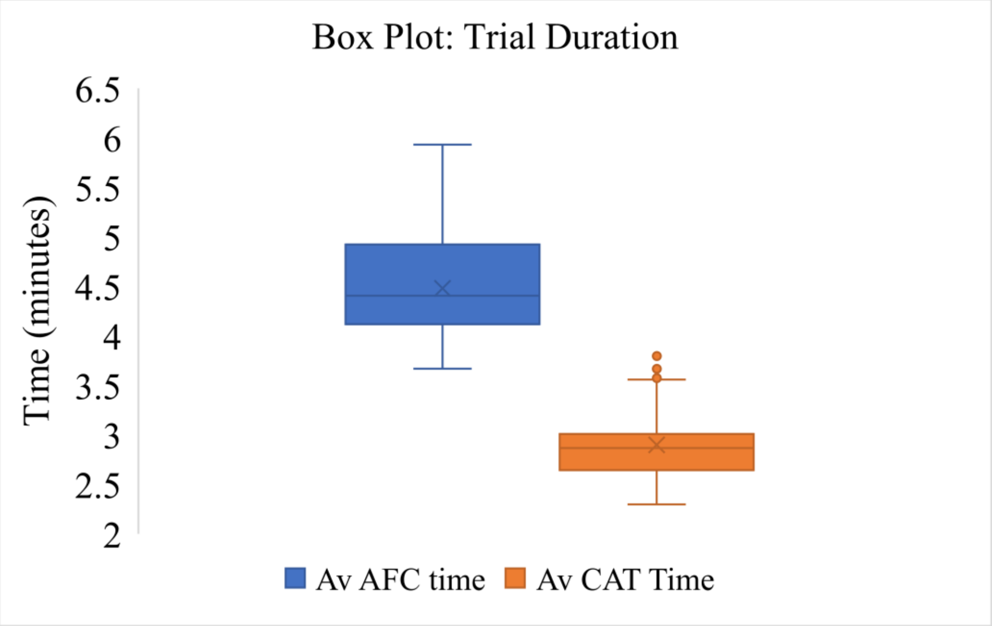
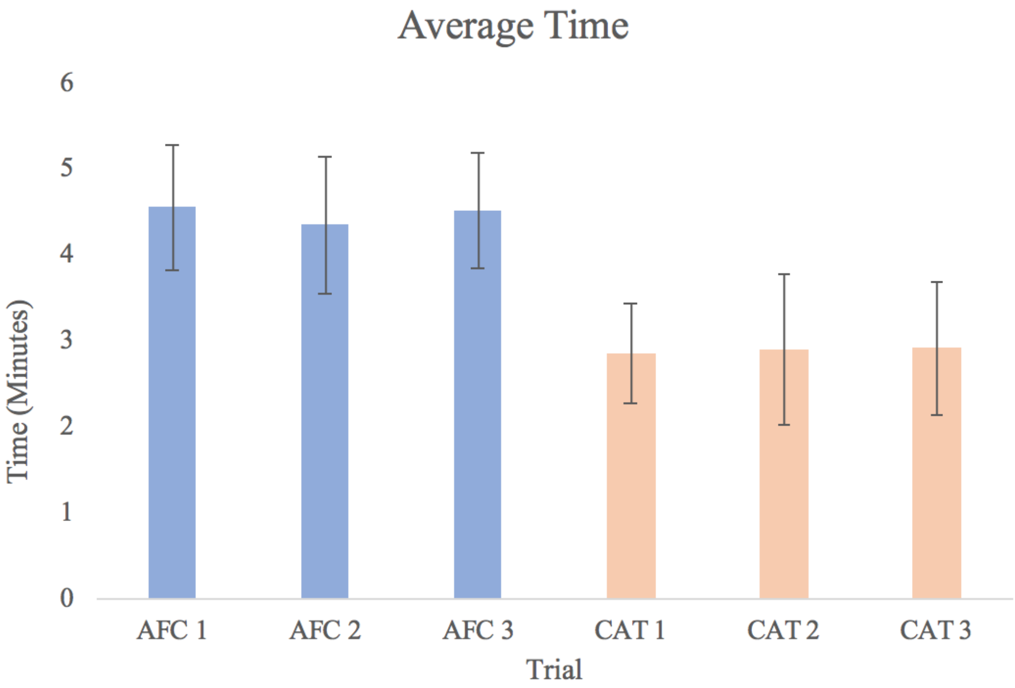


FIGURE 9: BAR GRAPH DEPICTING MEAN TRIAL DURATIONS ON THE ALTERNATIVE FORCED CHOICE AND CONTINUOUS ADAPTIVE TRACKING TESTS



For the AFC test, thresholds were estimated using the first two trials 30% of the time, with 70% requiring a third trial. Similarly, for the CAT test, thresholds were estimated using the first two trials 37% of the time, with 63% requiring a third trial. Given that in the majority of cases more than one trial will be required to establish a reliable threshold, especially when using the AFC test, the maximum test time for both the AFC and CAT test will be approximately 20 minutes, but the average test time for both tests will be much shorter in practice (AFC=13.5 min; CAT=8.7 min).

In order to further evaluate the efficiency of the tests in the current study, the AFC and CAT assessment durations were compared to other tests times reported in the literature. In a recent study, Hoover and colleagues (2019) compared the test duration of various behavioural assessments for TFS processing. The binaural tests of phase difference detection examined in this study were interaural phase difference [TFS–low frequency] and a tone detection in noise test similar to MLD. In order to facilitate the comparison between tests, all assessments were designed as AFC tasks. Of the binaural tests examined, all required at least 5 trials to obtain a valid threshold, which was defined as the number of trials run that produced three thresholds that were within a set number of standard deviations[1] from the mean score. The trial durations reported for these tests ranged from 4.7 minutes (MLD) to 5.5 minutes (TFS-LF), which is comparable to the tests evaluated in the current study; however, they required more trial runs (i.e., 5-7) to arrive at a reliable threshold, which resulted in considerably longer duration estimates than those observed in the present study (e.g., 23.7 [MLD] to 41.0 [TFS-LF] minutes). Additionally, studies show that the TFSAF test requires at least three trials to establish a

---

[1] Hoover and colleagues (2019) did not report the number of standard deviations that were used to determine the consistency of threshold estimates

reliable frequency threshold, which is estimated to take approximately 18 minutes (Fullgrabe & Moore, 2017). While this is comparable to the AFC test duration, (13.5 minutes) the average test time for three trials on the CAT test is considerably shorter (i.e., 8.7 minutes).

It is important to note that the AFC and CAT test were not designed to be directly compared to the tests evaluated by Hoover and colleagues (2019) or the TFSAF test (Fullgrabe & Moore, 2017) and further study is needed to determine if the tests examined in the present study are more efficient. It is also possible that the AFC and CAT test time could be further reduced in the future, perhaps by increasing the start frequency, altering other parameters (e.g., step size), or establishing more reliable age-matched normative data. Taken together, the time constraints of testing in clinic highlight the need to make assessments as efficient as possible. The findings of the current study support the notion that the CAT test is significantly more efficient than the AFC test, which provides further evidence that it may be more appropriate for routine clinical use.

HYPOTHESIS 5: BINAURAL PHASE DETECTION THRESHOLDS, HIGH FREQUENCY HEARING THRESHOLDS, AND SPEECH IN NOISE PERFORMANCE WILL DECLINE WITH AGE

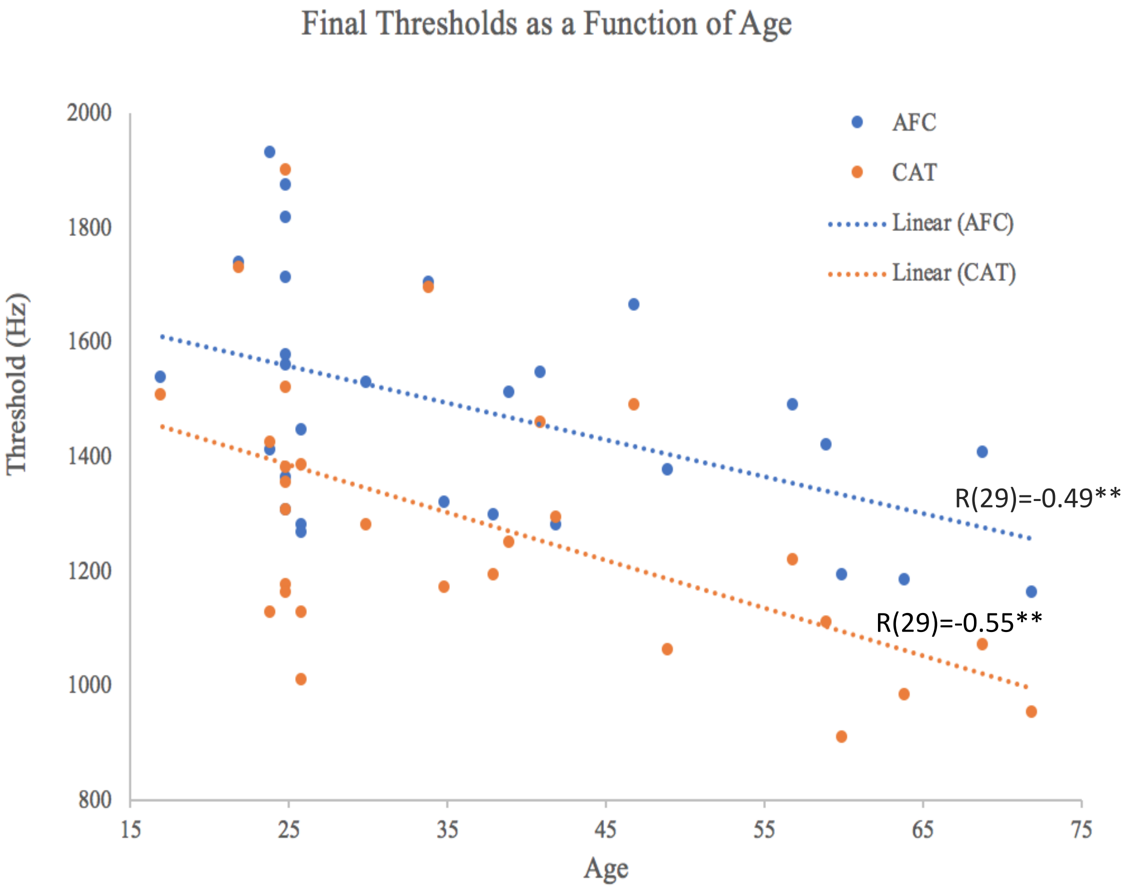
There is a growing body of research demonstrating that binaural hearing deteriorates with age (Füllgrabe et al., 2015; Grose & Mamo, 2010; Lunner, Hietkamp, Andersen, Hopkins, & Moore, 2012). Examining the efficiency and reliability of IPD tests was the primary objective of the present study; however considering the link between age, speech in noise performance and TFS processing, as well as the established relationship between age and decreased EHF thresholds (Hallmo, Sundby & Mair, 1994) a secondary analysis was performed to examine the degree to which these factors predicted estimated AFC and CAT thresholds. The age of participants in this study ranged from 17-72 years, with almost half of the sample (i.e., 48%) younger than 30, 31% between the age of 30 and 49, and the remaining 21% were over 50. The degree to which relationship between age, high frequency hearing, and speech in noise performance (i.e., binaural QuickSIN scores) predicted IPD thresholds was examined by employing hierarchical regression.

AGE AS A PREDICTOR OF IPD THRESHOLDS

In both regression models (AFC and CAT) age was a significant negative predictor of IPD thresholds, such that as age increased IPD thresholds decreased, accounting for 24% of variability in AFC and 30% of CAT thresholds respectfully (see Figure 10). Both high frequency thresholds and QuickSIN scores accounted for a nominal degree of additional variance in the models (5-6%). While these models leave the majority of the variance in IPD thresholds unaccounted for (70-76%), they are in line with similar models reported in the literature.

In a study by Ross and colleagues (2007), magnetoencephalography (MEG) was used to assess binaural phase difference evoked responses and behavioural binaural phase detection across carrier frequencies in young, middle aged, and elderly subjects. Frequency thresholds on an AFC behavioural test similar to the AFC test examined in the current study were age dependent, with subjects in the younger group eliciting detectable responses over 1250 Hz, while the middle and older age group yielded response thresholds between 1250-1000, and 750-1000 Hz respectfully. Ross and colleagues (2007) also employed a regression analysis and reported that age was a significant predictor with thresholds in the range of 1290 Hz at 20, which decreased by 14% every 10 years. Similarly, their model only explained 28% of the variance in IPD thresholds.

FIGURE 10: SCATTERPLOT OF FINAL THRESHOLDS ON THE ALTERNATIVE FORCED CHOICE (AFC) AND CONTINUOUS ADAPTIVE TRACKING (CAT) TEST ACROSS AGE



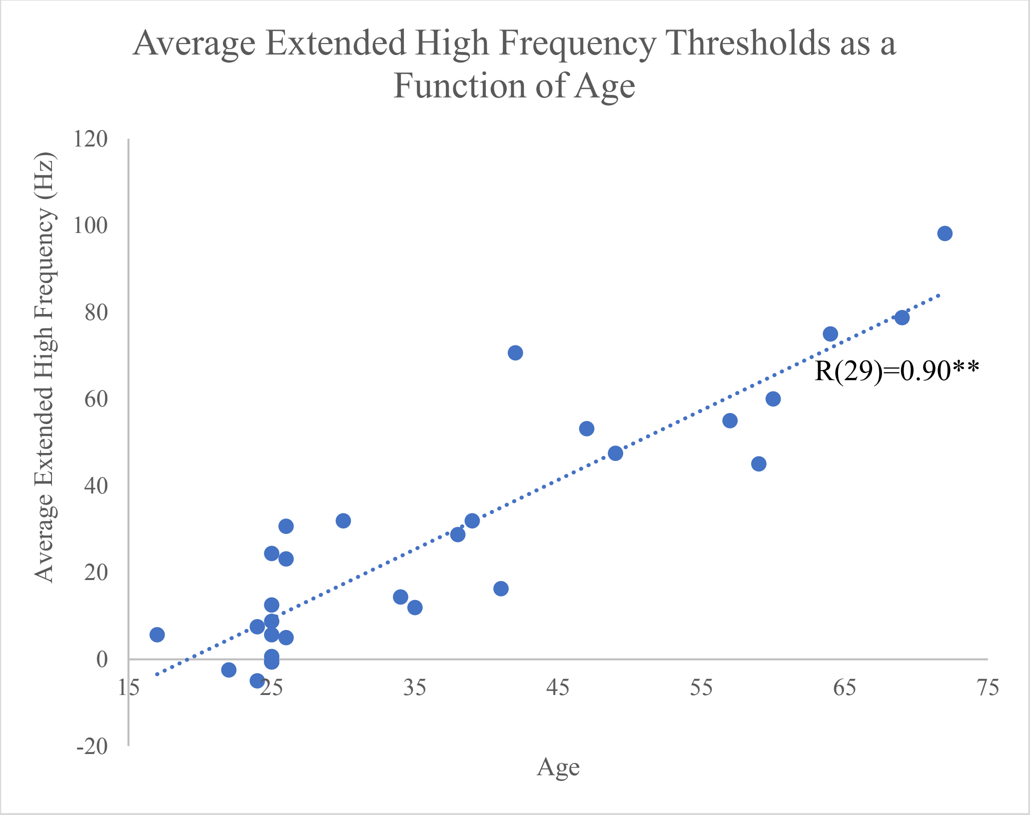
Note: ** $p<0.01$

EXTENDED HIGH FREQUENCY HEARING AS A PREDICTOR OF IPD THRESHOLDS

Age was entered in the first step of the hierarchical regression models, due to the established relationship between age and IPD thresholds in the literature, but it is conceivable that EHF hearing is a significant contributor to this association. Moore, Glasburg, and Stoev (2012) attempted to examine the association between high frequency hearing (i.e., 4000-8000Hz) and temporal fine structure processing using low frequency IPD independent of age. They reported that age and IPD were correlated, but only observed a weak correlation between high frequency hearing and IPD thresholds, which was eliminated when age was accounted for.

Conversely, in a more recent study of young adults (Zadeh, Silbert, Sternasty, Swanepoel, Hunter & Moore, 2019), researchers reported that EHF hearing loss was significantly associated with self-reported difficulty understanding speech in noise. In addition, researchers employed a digits-in-noise test using a broadband and filtered noise masker (e.g., 8000Hz low-pass filtered) and reported that hearing thresholds were significantly worse in the broadband noise condition, which suggests that EHF hearing does contribute to speech in noise performance, independent of age. Ross and colleagues (2007) did not examine EHF hearing thresholds in their study, but they did discuss the possibility elevated high frequency thresholds would contribute to a decline in the binaural cue detection. In the present study, age and EHF thresholds were too highly correlated to tease apart these associations ($r(29)=0.90$, $p<0.001$, see Figure 11).

FIGURE 11: SCATTERPLOT OF HIGH FREQUENCY THRESHOLDS ACROSS AGE



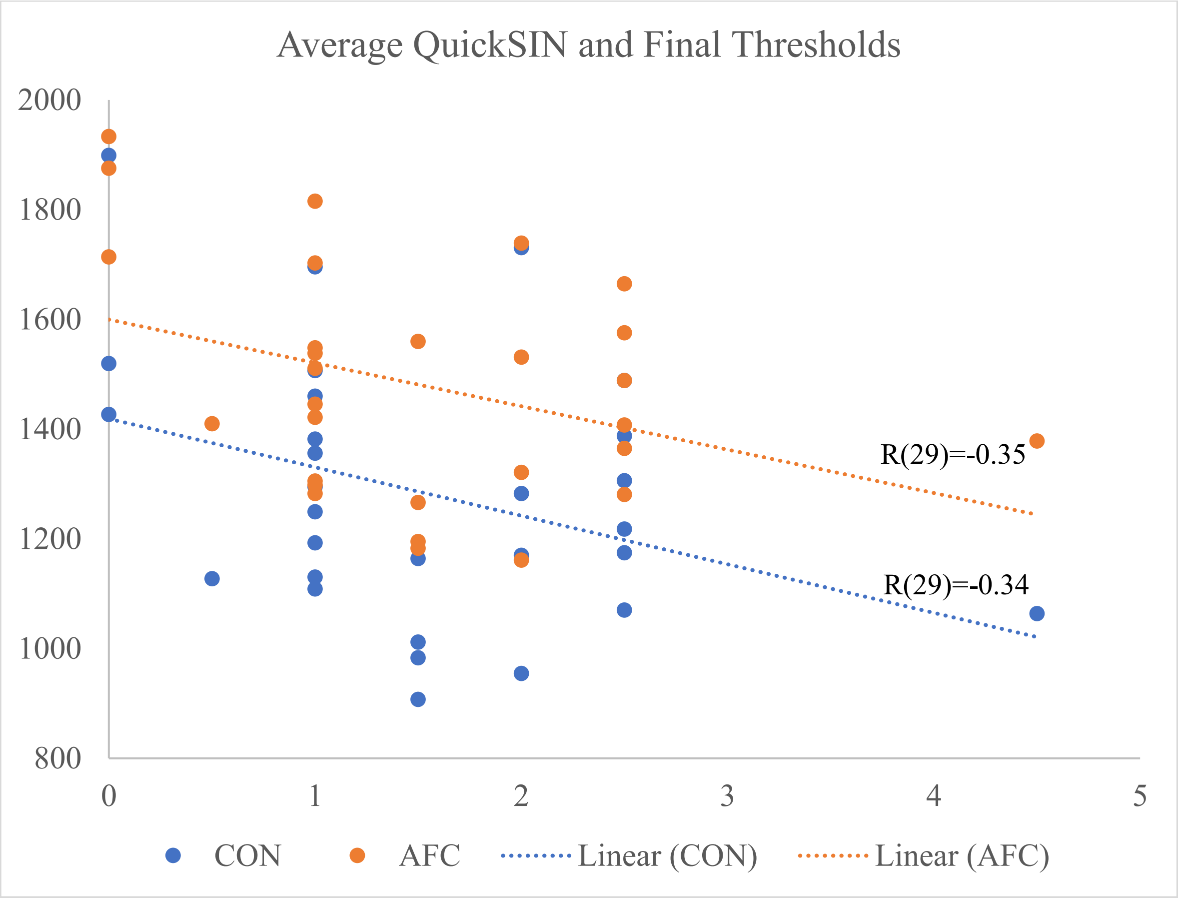
Note: ** $p<0.01$

There is growing evidence that EHF hearing may play a role in the ability to understand speech in noise, but the role of EHF hearing in TFS processing has yet to be elucidated. In the past, EHF's were generally overlooked, due to high variability across subjects, the likelihood of threshold elevation associated with noise exposure and presbycusis (Korres, Balatsouras, Tzagaroulakis, Kandiloros & Ferekidis, 2008), and the assumption that the majority of the speech waveform energy required for speech intelligibility is contained within the 250-8000Hz range (Stevens, 1998). The findings reported in recent studies have highlighted the potential role of EHF in hearing, particularly in challenging acoustic environments (e.g., Zadeh et al., 2019). It has also been suggested that there may be value in measuring EHF upper limits (i.e., the highest frequency where a threshold can be obtained) in the assessment and monitoring of ototoxicity (Rieke, Clavier & Buckey, 2017). In fact, examining frequency thresholds for IPD discrimination, rather than thresholds for phase depth discrimination, can be viewed as a similar approach to assess and monitor TFS processing.

SPEECH IN NOISE PERFORMANCE AS A PREDICTOR OF IPD THRESHOLDS

Speech in noise performance (i.e., average QuickSIN scores) did not significantly contribute to either of the models in the present study. It should be noted that there was very little variability in QucikSIN scores overall and the vast majority of participants had scores within the normal range bilaterally (66%) and only one participant exhibited a mild SNR impairment in both ears. In addition, thresholds on the AFC and CAT test were not significantly correlated with average QuickSIN scores, but the results were approaching significance (AFC, p=0.063; CAT, p=0.08 [see Figure 12]). Given that OAD refers to a population of people with normal audiometric results who experience difficulty understanding speech in noise (Saunders & Haggard, 1989), and that the impaired speech in noise performance seen in HHL is posited to result from impaired temporal fine structure processing at the level of the auditory synapse (Demany & Semal, 1989; Moore & Sek, 1996), the relationship between AFC and CAT thresholds and speech in noise performance requires more rigorous examination before any definitive conclusions can be drawn.

FIGURE 12: SCATTERPLOT OF AVERAGE QUICKSIN SCORES AND FINAL THRESHOLDS ON THE ALTERNATIVE FORCED CHOICE AND CONTINUOUS ADAPTIVE TRACKING TESTS



Note: Thresholds on the AFC and CAT test were not significantly correlated with average QuickSIN scores, but the results were approaching significance (AFC, p=0.063; CAT, p=0.08)

Evaluating the sensitivity of the AFC and CAT tests to speech in noise performance was not the primary objective of the present study, and if this research was replicated in a more variable sample, speech in noise performance should theoretically emerge as a significant contributor. The optimization of clinical IPD assessments conducted in the present study represents a critical first step that will facilitate further examination of the association between speech in noise performance and peripheral temporal processing in future research.

LIMITATIONS

SAMPLE CHARACTERISTICS

While the findings of the current study provide insight into an optimal approach to assess IPD discrimination in the clinic, there are a number of noteworthy limitations. First, the sample size in the present study was relatively small (N=30). While it was larger than those reported in similar studies examining approaches to psychoacoustic testing (e.g., Hoover et al., 2019), there is a need to establish normative data in normal hearing subject before the value of the assessments as a differential diagnostic tool for HHL can be examined. Ideally sample sizes of 50-75 participants are minimally required to establish reliable norms (Bridges & Holler, 2007), which is well above the sample size in the current study.

Another limitation to the current research is that the majority of the sample was young (i.e., 79% we under the age of 50) and female (i.e., 79%). While there is no reason to believe that sex differences would play a significant role in test performance overall, research has reported sex differences in subcortical auditory function (Krizan, Erika & Kraus, 2012), so it cannot be ruled out as a potential confounding factor in the current study. Given that the CAT assessment has emerged as the most efficient and reliable clinical approach to test IPD discrimination, future research would benefit from a large-scale study that establishes age and sex-matched normative data. In addition, there are several potential health related and experiential factors (e.g., ADHD, concussions etc.)

that were not accounted for in the present study but could have influenced performance on the IPD assessments.

SPEECH IN NOISE TESTING

The present study attempted to establish a link between speech in noise performance and IPD thresholds using QuickSIN scores. In addition to the overall lack of variability in the sample with respect to SNR loss, there were several other significant limitations to our approach. First, QuickSIN scores were generated by averaging monaural scores, but, given that our goal was to compare scores with the results of binaural tests, perhaps it would have been more appropriate to administer the lists binaurally. Additionally, only one list was administered to each ear, and research has indicated that administering at least two lists and calculating an average score provides a more accurate assessment of speech in noise performance (Etymotic Research, 2006; Otometrics, 2012). Finally, the QuickSIN is a limited test of speech in noise, as it is designed to be quick and efficient to administer (Etymotic Research, 2006). There are more comprehensive binaural assessments available (e.g., coordinate response measure [CRM]; Bolia, Nelson, Ericson & Simpson, 1999), that would have provided a more accurate representation of speech in noise performance.

TESTING PROCEDURE

There were also a number of limitations related to the testing procedure and design that may have influenced the results of the current study. First, participants engaged in a short stimulus demonstration session prior to beginning their first trial. The session was identical to the AFC test, but only presented phase differences at the 250Hz carrier frequency. While this session was short (e.g., stimulus was presented 2 or 3 times), and only administered for the purposes of demonstrating an easily detectable phase shift, it is possible that it provided an advantage to participants during the subsequent AFC trials.

Given that overall the thresholds on the AFC test were higher than thresholds on the CAT test, we cannot rule out that the demonstration session design contributed to this finding. Ideally, two demonstration sessions should have been used, with one demonstrating the AFC test and another the CAT test, to control for this potential confound. Additional software interface limitations must be noted, including the addition of red and green flashing lights in the CAT test. The AFC assessment did not provide visual cues to participants when they performed a correct or incorrect response, and we cannot rule out the possibility that these cues contributed to the differences in thresholds across tests.

EXTENDED HIGH FREQUENCY THRESHOLDS

There was also a noteworthy limitation associated with the quantification of EHF thresholds. In order to facilitate the statistical analysis of EHF thresholds we were required to assign a numeric value in cases where there was no response (NR) at that frequency. A value 10dB above the limits of the audiometer (e.g., 100dB) was chosen. While this approach was performed out of necessity, it is possible that quantifying NR in this way resulted in an over or under estimation of true EHF thresholds.

CONCLUSIONS AND FUTURE DIRECTIONS

In the future, research should focus on teasing apart the association between age and high frequency hearing thresholds in order to gain a better understanding of their unique contribution to temporal fine structure processing, IPD discrimination and functional hearing. Given that the independent variables examined in the present study only accounted for 24-30% of the variance in IPD thresholds, it is important to note that other factors may play a significant role, including noise exposure history, genetic variability, previous experiences (e.g., musical training), cognitive and/or other comorbid health related factors (e.g., past concussions), and other auditory processing mechanisms (e.g., central processing), and future research should attempt to identify the role that these factors play in IPD discrimination and TFS processing.

Given that the CAT test has emerged as the most efficient and reliable, the next logical step for future studies is to establish normative data. Future normative studies should expand the exclusion criteria to include any potentially confounding disorders that may influence performance on the CAT test (e.g., ADHD, concussion, tinnitus, APD, cognitive impairment). Ultimately, establishing reliable normative data will facilitate future studies examining the efficacy of the CAT test as a differential diagnostic assessment for HHL.

Ideally, future studies will compare the CAT test with MLD, the TFS-LF and TFS-AF tests, and a comprehensive speech in noise assessment, such as the CRM, to examine the extent to which performance on these tests agree. In addition, future studies should administer the CAT test in populations who experience difficulty in challenging acoustic environments. Populations of interest include HHL/OAD, APD, subjects with varying degrees of noise exposure history, and hard of hearing subjects. In an ongoing study conducted by Aiken and colleagues, performance of a variety of psychoacoustic assessments are assessed before and after exposure to damaging noise in a population of military personnel participating in a training session. It would be valuable to test and monitor IPD thresholds in a similar study to examine the effects of acute and controlled noise exposure on IPD thresholds.

In the future, attempts should be made to further optimize the CAT test software. For example, the software could be designed to compare the thresholds of two trials with established normative data to determine if minimum consistency requirements are met. When necessary, the software could automatically run a third trial and use the two highest agreeing thresholds. In addition, many subjects reported that they found the task easier and less fatiguing if they closed their eyes while responding. In the present study subjects completed the task using a mouse, so perhaps future studies should examine the use of a button to facilitate responding in this manner.

It is also possible that other approaches to establishing IPD thresholds that may be equally or more efficient that the CAT test. For example, IPD thresholds could be established using a Bekesy test, where subjects can increase or decrease the frequency of the carrier until they are no longer able to discriminate a phase shift. Future studies should examine the efficacy of other potentially optimal approaches to IPD discrimination testing, to ensure that the most efficient method is made available in clinic.

There are a number of promising avenues to disseminate the CAT assessment for research and clinical use. For example, the Portable Automated Rapid Testing (PART; Gallun et al., 2015, 2018) is an online application available free of charge on the apple platform. This app contains a variety of psychoacoustic tests that can easily be added and modified for use by researchers and clinicians. In the future the feasibility of developing the CAT assessment for use on this platform should be examined.

Finally, it is possible that, similar to APD, a stand-alone differential diagnostic test for HHL is not realistic. In fact, in a recent review of the HHL experimental research, Kobel and colleagues (2017) highlighted the need to develop a HHL test battery the includes a combination of objective and behavioural assessments. This may not be an approach that is feasible for the standard clinic but may be necessary to differentiate HHL from other forms of OAD. Future research should explore the value of administering the CAT test with other promising assessments in an effort to establish a reliable and accurate approach to diagnose HHL.

In conclusion, the goal of the present study was to assess the reliability and efficiency of three approaches to establishing frequency thresholds for IPD detection. Piloting revealed that only two of the tests developed (3AFC and CAT) warranted investigation, so the continuous non-adaptive test was removed from the study. Overall, the CAT test emerged as the most time efficient approach, the thresholds generated with this assessment are in agreement with extant research (e.g., Ross et al., 2007) and exhibit excellent reliability. The findings of the present study indicate that the CAT test is a promising approach to quickly and efficiently assess IPD detection thresholds in a clinic setting. Future research should focus on further optimization of the CAT software, collecting normative data, comparing CAT thresholds with other available objective and behavioural assessments, and testing CAT thresholds in a variety of clinical populations to determine its utility as a differential diagnostic for hidden hearing loss.

REFERENCES

Akeroyd,M. (2006). The psychoacoustics of binaural hearing. *International Journal of Audiology, 45* (suppl 1), S25-S33. DOI: 10.1080/14992020600782626

Bharadwaj, H., Masud, S., Verhulst, G., Mehraei, G. & Shinn-Cunningham, B. (2015). Individual differences reveal correlates of hidden hearing loss. *The Journal of Neuroscience, 35,* 2161-2172. DOI:10.1523/JNEUROSCI.3915-14.2015

Bitter, R., Mohiuddin, T., & Nawrocki, M. (2006). *LabVIEW: Advanced programming techniques*. Crc Press.

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (1999). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America, 107*, 1390–1391. DOI: 10.1121/1.428288

Bridges, A.J. & Holler, K.A. (2007) How many is enough? Determining optimal sample sizes for normative studies in pediatric neuropsychology. *Child Neuropsychology, 13*(6), 528-538. DOI:10.1080/09297040701233875

Celesia, G. (2015) *Hearing disorders in brainstem lesions*. In Aminoff, M., Boller, F & Swaab, D. The Human Auditory System: Fundamental Organization and Clinical Disorders (pp. 509-536). Elsevier.

Chalak, S., Kale, A., Deshpande, V.K. & Biswas, D.A. (2013). Establishment of Normative data for Monaural Recordings of Auditory Brainstem Response and its Application in Screening Patients with Hearing Loss: A Cohort Study. *Journal of Clinical and Diagnostic Research, 7*(12), 2677-2679. DOI: 10.7860/JCDR/2013/6768.3730

Chung, K.L. (1954) On a stochastic approximation method. *The Annals of Mathematical Statistics, 25*, 463-483.

Coakes, S. (2005) *SPSS: Analysis without Anguish*. 1st ed. Milton, Qld.: Wiley Australia

Drennan, W.R., Won, J.H., Dasika, V.K. & Rubinstein, J.T. (2007). Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise. *Journal of the Association for Research in Otolaryngology, 8,* 373-383. DOI:10.1007/s10162-007-0074-y

Etymotic Research Inc. (2006) *QuickSIN Speech-in-Noise Test, Version 1.3*. Retrieved from: https://www.etymotic.com/downloads/dl/file/id/259/product/159/quicksin_user_manual.pdf

Field, A. (2005). Discovering statistics using SPSS (2nd ed.). Sage Publications, Inc.

Fisher RA. *Statistical methods for research workers.* Oliver and Boyd; Edinburgh: 1954.

Füllgrabe, C. & Moore, B. (2017). Evaluation of a method for determining binaural sensitivity to temporal fine structure (TSF-AF Test) for older listeners with normal and impaired low-frequency hearing. *Trends in Hearing, 21*, 1-14. DOI: 10.1177/2331216517737230

Füllgrabe, C. & Moore, B. (2018). The association between the processing of binaural temporal-fine-structure information and audiometric threshold and age: A meta-analysis. *Trends in Hearing, 22,* 1-14. DOI: 10.1177/2331216518797259

Füllgrabe, C., Harland, A., Sek, A. & Moore, B. (2017) Development of a method for determining binaural sensitivity to temporal fine structure. *International Journal of Audiology, 56*(12), 926-935. DOI: 10.1080/14992027.2017.1366078

Furman, A., Kujawa, S. & Liberman, M. (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *Journal of Neurophysiology, 110*, 577-586. DOI: 0.1152/jn.00164.2013

Gallun, F. J., McMillan, G. P., Kampel, S. D., Jakien, K. M., Srinivasan, N. K., Stansell, M. M., & Gordon, S. Y. (2015). Verification of an automated headphone-based test of spatial release from masking. *Proceedings of Meetings on Acoustics, 25*, 050001. DOI: 10.1121/2.0000165

Gallun, F. J., Seitz, A., Eddins, D. A., Molis, M. R., Stavropoulos, T., Jakien, K. M., . . . Srinivasan, N. (2018). Development and validation of Portable Automated Rapid Testing (PART) measures for auditory research. *Proceedings of Meetings on Acoustics, 33,* 050002. DOI: 10.1121/2.0000878

Gatehouse, S. & Nobel, W. (2004) The Speech, Spatial and Qualities of Hearing. *International Journal of Audiology, 43*(2), 85-99. DOI: 10.1080/14992020400050014

Gelfand, S. (2010). *Hearing: An Introduction to Psychological and Physiological Acoustics*. 5th Edition. Informa UK: New York.

Gibson, W. (2017) The clinical uses of electrocochleography. *Frontiers in Neuroscience, 11,* 274-234. DOI: 10.3389/fnins.2017.00274

Green, D. & Swets, J. (1966) *Signal detection theory and psychophysics.* Wiley & Sons, Inc.: New York

Grose, J.H. & Mamo, S.K. (2010). Processing of temporal fine structure as a function of age. *Ear and Hearing, 31,* 755-760. DOI: 10.1097/AUD.0b013e3181e627e7.

Gurgel., R. W. (2014). Relationship of hearing loss and dementia: A prospective, population-based study. *Otology & Neurotology, 35*(5), 775-781.

Hair, J., Black, W., Babin, B. & R. (1998) *Multivariate Data Analysis: A Global Perspective (7th ed.)* Pearson: New York.

Hallmo, P., Sundby, A. & Mair, I. (1994) Extended high-frequency audiometry. Air- and bone-conduction thresholds, age and gender variations. *Scandinavian Audiology, 23*(3), 165-170. DOI:10.3109/01050399409047503

Hirsh, I. J. (1948). Binaural summation and interaural inhibition as a function of the level of masking noise. *The American Journal of Psychology, 61*, 205–213. DOI: 10.2307/1416966

Hoover, E.C., Kinney, B., Bell, K., Gallun, F. & Eddins, D. (2019) A comparison of behavioral methods for indexing the auditory processing of temporal fine structure cues. *Journal of Speech, Language, and Hearing Research, 62,* 2018-2034. DOI: 10.1044/2019_JSLHR-H-18-0217

Hopkins K. & Moore B.C.J. 2010a. Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies. *International Journal of Audiology, 49*, 940–946. DOI: 10.3109/14992027.2010.512613

King, A., Hopkins, K. & Plack, C.J. (2014). The effects of age and hearing loss on interaural phase difference discrimination. *The Journal of the Acoustical Society of America, 135*, 342-351. DOI: 10.1121/1.4838995.

Kobel, M., Prell, c., Liu, J., Hawks, J. & Bao, J. (2017) Noise-induced cochlear synaptopathy: Past findings and future studies. *Hearing Research, 349,* 148-154. DOI: 10.1016/j.heares.2016.12.008

Kohlrausch, A. (1986) The influence of signal duration, signal frequency and masker duration of binaural masking level differences. *Hearing Research, 23*, 267-273.

Koo, T. & Li, M. (2016) A guideline for selecting and reporting intraclass correlation coefficient for reliability research. *Journal of Chiropractic Medicine, 15,* 155-163. DOI: 10.1016/j.jcm.2016.02.012

Korres, G.S., Balatsouras, D.G., Tzagaroulakis, A., Kandiloros, D. & Ferekidis, E. (2008). Extended high-frequency audiometry in subjects exposed to occupational noise. *B-ENT, 4(*3):147-55. PMID:18949961

Krizan, J., Erika, S. & Kraus, N. (2012). Sex differences in subcortical function. *Clinical Neuropsychology, 123*(3), 590-597. DOI: 10.1016/j.clinph.2011.07.037

Kujawa, S. & Liberman, M. (2009). Adding insult to injury: Cochlear nerve degeneration after "tempoarary" noise-induced hearing loss. *Journal of Neuroscience, 29*, 14077-14085. DOI: 10.1523/JNEUROSCI.2845-09.2009.

Kujawa, S. & Liberman, M. (2006). Acceleration of age-related hearing loss by early noise exposure: evidence of a misspent youth. *The Journal of Neuroscience, 26*, 2115-2123. DOI:10.1523/JNEUROSCI.4985-05.2006

Lacher-Fougère, S. & Demany, L. (2005) Consequences of cochlear damage for the detection of interaural phase differences. *The Journal of the Acoustical Society of America, 118*(4), 2519-2526. DOI: 10.1121/1.2032747

Lesmes, L.A., Lu, Z., Baek, J., Tran, N., Dosher, B.A. & Albright, T.D. (2015). Developing Bayesian adaptive methods for estimating sensitivity thresholds (D′) in Yes-No and forced-choice tasks. *Frontiers in Psychology, 6*, 1070, 1-24. DOI: 10.3389/fpsyg.2015.01070

Levitt, H. (1970) Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*, 467-477. DOI: 10.1121/1.1912375

Li, C.M., Zhang, X. Hoffman, H.J., Cotch, M.F., Thermann, C.L. & Wilson, M.R. (2014). Hearing Impairment Associated With Depression in US Adults, National Health and Nutrition Examination Survey 2005-2010. *JAMA Otolaryngology Head & Neck Surgery*, 293-302.

Liberman, C. & Kujawa,S. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hearing Research, 349*, 138-147. DOI: http://dx.doi.org/10.1016/j.heares.2017.01.003

Liberman, M.C., Epstein, M.J., Cleveland, S.S., Wang, H. & Maison, S.F. (2016) Toward a differential diagnosis of hidden hearing loss in humans. *PLOS One, 11*(9), e0162726. DOI: 10.1371/journal.pone.0162726

Litovsky, R. (2012). Spatial release from masking. *Acoustics Today*, April, 18-25.

Lu, L. & Shara, N. (2007) Reliability analysis: Calculate and compare intra-class correlation coefficients (ICC) in SAS. *NorthEast SAS Users Group.* Retrieved from: https://lexjansen.com/nesug/nesug07/sa/sa13.pdf

Lunner, T., Hietkamp, R.K., Andersen, M.R., Hopkins, K. & Moore, B.C. (2012) Effect of speech material on the benefit of temporal fine structure information in speech for young normal-hearing and older hearing-impaired participants. *Ear and Hearing, 33*(3), 377-388. DOI: 10.1097/AUD.0b013e3182387a8c.

McAlpine, D. (2005) Creating a sense of auditory space. *The Journal of Physiology, 566*, 21-28. DOI: 10.1113/jphysiol.2005.083113

McAlpine, D., Haywood, N., Undurraga, J. & Marquardt, T. (2016). Objective measures ofneural processing of interaural time differences. *Advances in Experimental Medicine and Biology, 864,* 197-205. DOI: 10.1007/978-3-319-25474-6_21.

McFadden, D. (1966) Masking-level differences with continuous and with burst masking noise. *Journal of the Acoustical Society of America, 40*(6), 1414-1419. DOI: 10.1121/1.1910241

McPherson, D., Senderski, A., Brunham, M., Fujiki, A., Harris, R., Skarzynski, H. & Kochanek, K. (2011). Masking level difference in an adaptive procedure for clinical investigation, *International Journal of Audiology, 50*(9), 613-620, DOI: 10.3109/14992027.2011.582168

Moon, J. & Hong, S. (2014). What is temporal fine structure and why is it important? *Korean Journal of Audiology, 18*(1), 1-7. http://dx.doi.org/10.7874/kja.2014.18.1.1

Moore, B.C. & Sek, A. (1996) Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *Journal of the Acoustical Society of America, 100*(4), 2120-2031. DOI:10.1121/1.417941

Moore, B. C. J., Sek, A. (2009) Development of a fast method for determining sensitivity to temporal fine structure. *International Journal of Audiology 48*, 161–171. DOI:10.1080/14992020802475235

Moore, B., Glasburg, B. & Stoev, M. (2012) The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies. *The journal of the Acoustical Society of America, 131*(2), 1003- 1006. DOI: 10.1121/1.3672808

National Research Council (US) Committee on Disability Determination for Individuals with Hearing Impairments. (2004). Impact of Hearing Loss on Daily Life and the Workplace. In V. H. Dobie RA, *Hearing Loss: Determining Eligibility for Social Security Benefits.* Washington, DC: National Academies Press (US).

Otometrics (2012). *OTOsuite and QuickSIN Module User Guide*. Retrieved from: https://partners.natus.com/asset/resource/file/otometrics/asset/2019-07/7-50-1200-EN_03.PDF

Pallant, J. (2001). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows version 10.* Buckingham: Open University Press.

Plack, C., Barker, D. & Prendergast, G. (2014). Perceptual consequences of "hidden" hearing loss. *Trends in Hearing, 18*, 1-11. DOI: 10.1177/2331216514550621

Plack, C., Leger, A., Prendergast, G., Kluk, K., Guest, H. & Munro, K. (2016). Toward a diagnostic test for hidden hearing loss. *Trends in Hearing, 20*, 1-9. DOI: 10.1177/2331216516657466

Prell, C. (2018). Hidden versus not-so-hidden hearing loss. *Canadian Audiologist, 5*(6). Retrieved from http://canadianaudiologist.ca/issue/volume-5-issue-2-2018/hidden-versus-not-so-hidden-hearing-loss/

Prendergast, G., Tu, W., Guest, H., Millman, R., Kluk, K., Couth, S., Munro, K. & Plack, C. (2018). Supra-threshold auditory brainstem response amplitudes in humans: Test-retest reliability, electrode montage and noise exposure. *Hearing Research, 364*, 38-47. DOI: 10.1016/j.heares.2018.04.002

Puel, J.L., Pujol, R., Tribillac, F., Ladrech, S. & Eybalin, M. (1994). Excitatory amino acid antagonists protect cochlear auditory neurons from excitotoxicity. *Journal of Comparative Neurology, 341*, 241-256. DOI:10.1002/cne.903410209

Puel, J.L., Ruel, J., d'Aldin, C.G. & Pujol, R. (1998). Excitotoxicity and repair of cochlear synapses after noise-trauma induced hearing loss. *NeuroReport, 9*, 2109-2114.

Puel, J.L., Safieddine, S., D'Aldin, C., Eybalin, M. & Pujol, R. (1995). Synaptic regeneration and functional recovery after excitotoxic injury in the Guinea pig cochlea. *Comptes Rendus de l'Académie des Sciences - Series III, 318*, 67-75.

Rieke, C., Clavier, O. & Buckey, J. (2017). Fixed-level frequency threshold testing for ototoxicity monitoring. *Ear and Hearing, 38*(6), e369-e375. DOI: 10.1097/AUD.0000000000000433

Ross, B., Fujioka, T., Tremblay, K. L., Picton, T. W. (2007) Aging in binaural hearing begins in mid-life: evidence from cortical auditory-evoked responses to changes in interaural phase. *Journal of Neuroscience 27*(42), 11172–11178. DOI:10.1523/JNEUROSCI.1813-07.2007.

Ross, B., Termblay, T. & Picton, T. (2007). Aging in Binaural Hearing Begins in Mid-Life: Evidence from Cortical Auditory-Evoked Responses to Changes in Interaural Phase. *The Journal of Neuroscience, 27*, 11172-11178. DOI:10.1523/JNEUROSCI.1813-07.2007

Saunders, G. & Haggard, M. (1989) The clinical assessment of obscure auditory dysfunction: Auditory and psychological factors. *Ear and Hearing, 10*(3), 200-208. DOI:10.1097/00003446-198906000-00011

Schmiedt, R.A., Mills, J.H. & Boettcher, F.A. (1996). Age-related loss of activity of auditory-nerve fibers. *Journal of Neurophysiology, 76*, 2799-2803. DOI: 10.1152/jn.1996.76.4.2799

Sek, A. & Moore, B. (2012) Implementation of two tests for measuring sensitivity to temporal fine structure. *International Journal of Audiology, 51*, 58–63. DOI: 10.3109/14992027.2011.605808

Sergeyenko, Y., Lall, K., Liberman, M. C., & Kujawa, S. G. (2013). Age-related cochlear synaptopathy: An early-onset contributor to auditory functional decline. *Journal of Neuroscience, 33*, 13686–13694.

Shi, L., Chang, Y., Li, X., Aiken, S., Liu, L. & Wang, J. (2016). Coding deficits in noise-indicued hidden hearing loss may stem from incomplete repair of ribbon synapses in the cochlea. *Frontiers in Neuroscience, 10*, 1-6. DOI: 10.3389/fnins.2016.00231

Spoendlin, H. (1972). Innervation densities of the cochlea. *Acta Otolaryngologica. 73*, 235e248.

SPSS statistical software (2017). IBM SPSS Statistics for Mac, version 25: IBM Corp.

Stamper, G. & Johnson, T. (2015). Auditory function in normal-hearing, noise-exposed human ears. *Ear and Hearing, 36*, 172-84. DOI: 10.1097/AUD.0000000000000107.

Statistics Canada. (2016, October 13). *Hearing Fact Sheets*. Retrieved from Hearing loss of Canadians, 2012 to 2015: https://www150.statcan.gc.ca/n1/pub/82-625-x/2016001/article/14658-eng.htm

Stevens, K.N. (1998) *Acoustic Phonetics.* Cambridge, MA: The MIT Press

Tabachnick, B. & Fidell, Linda (2001). *Using Multivariate Statistics, Fourth Edition*. Needham Heights, MA: Allyn & Bacon.

Undurraga, J., Haywood, N., Marquardt, T. & McAlpine, D. (2016). Neural representation of interaural time differences in humans—an objective measure that matches behavioural performance. *Journal of the Association for Research in Otolaryngology, 17*, 591-607. DOI: 10.1007/s10162-016-0584-6

Vancleef, K., Read, J., Herbert, W., Goodship, N., Woodhouse, M. & Serrano-Pedraza, I. (2018) Two choices good, four choices better: For measuring stereoacuity in children, a four-alternative forced-choice paradigm is more efficient than two. *PLOS, 13*(7): e0201366. DOI: 10.1371/journal.pone.0201366

Viana, L.M., O'Malley, J.T., Burgess, B.J., Jones, D.D., Oliveira, C.A., Santos, F., Merchant, S.N., Liberman, L.D. & Liberman, M.C. (2015). Cochlear neuropathy in human presbycusis: Confocal analysis of hidden hearing loss in post-mortem tissue. *Hearing Research, 327,* 78-88. DOI: 10.1016/j.heares.2015.04.014

Walker, J., Cleveland, L., Davis, J. & Seales, J. (2013). Audiometry screening and interpretation. *American Family Physician, 87*(1), 41-47.

Webster, F. (1951) The influence of interaural phase on masked thresholds: The role of interaural time-deviation. *The Journal of the Acoustical Society of America, 23*(4), 452-462. DOI: 10.1121/1.1906787

Wetherill, G.B. (1963) Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society, B25*, 1-48.

Wetherill, G.B., Chen, H., & Vasudeva, R.B. (1966) Sequential estimation of quantal response curves: A new method of estimation. *Biometrika, 53*, 439-454.

Woods, D., Wyma, J., Yund, W., Herron, T. & Reed, B. (2015) Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience, 9,* 131. DOI: 10.3389/fnhum.2015.00131

World Health Organization. (2018, March 15). *Deafness and hearing loss*. Retrieved from http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

Yost, W. (1974). Discriminations of interaural phase differences. *The Journal of the Acoustical Society of America, 55*, 1299–1303.

Zadeh, L., Silbert, N., Sternasty, K., Swanepoel, D., Hunter, L. & Moore, D. (2019). Extended high-frequency hearing enhances speech perception in noise. *Proceedings of the National Academy of Sciences, 116(*47), 23753-23759. DOI: 10.1073/pnas.1903315116