ABSTRACT

Title of Document:	EFFECTS OF AGING ON VOICE-PITCH PROCESSING: THE ROLE OF SPECTRAL AND TEMPORAL CUES
	Kara Chantal Schvartz Doctor of Philosophy, 2010
Directed By:	Associate Professor Monita Chatterjee Department of Hearing and Speech Sciences University of Maryland, College Park

Declines in auditory temporal processing are a common consequence of natural aging. Interactions between aging and spectro-temporal pitch processing have yet to be thoroughly investigated in humans, though recent neurophysiologic and electrophysiologic data lend support to the notion that periodicity coding using only unresolved harmonics (i.e., those available via the temporal envelope) is negatively affected as a consequence of age. Individuals with cochlear implants (CIs) must rely on the temporal envelope of speech to glean information about voice pitch [coded through the fundamental frequency (*f0*)], as spectral *f0* cues are not available. While cochlear implants have been shown to be efficacious in older adults, it is hypothesized that they would experience difficulty perceiving spectrally-degraded voice-pitch information. The current experiments were aimed at quantifying the ability of younger and older listeners to utilize spectro-temporal cues to obtain voice pitch information when performing simple and complex auditory tasks. Experiment 1 measured the ability of younger and older NH listeners to perceive a difference in the frequency of amplitude modulated broad-band noise, thereby exploiting only temporal envelope cues to perform the task. Experiment 2 measured age-related differences in *f0* difference limens as the degree of spectral degradation was manipulated to approximate CI processing. Results from Experiments 1 and 2 demonstrated that spectro-temporal processing of f0 information in non-speech stimuli is affected in older adults. Experiment 3 showed that age-related performances observed in Experiments 1 and 2 translated to voice gender identification using a natural speech stimulus. Experiment 4 attempted to estimate how younger and older NH listeners are able to utilize differences in voice pitch information in everyday listening environments (i.e., speech in noise) and how such abilities are affected by spectral degradation. Comprehensive results provide further insight on pitch coding in both normal and impaired auditory systems, and demonstrate that spectro-temporal pitch processing is dependent upon the age of the listener. Results could have important implications for elderly cochlear implant recipients.

EFFECTS OF AGING ON VOICE PITCH PROCESSING: THE ROLE OF SPECTRAL AND TEMPORAL CUES .

By

Kara Chantal Schvartz

Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park, in partial fulfillment of the requirements for the degree of Doctor of Philosophy
2010

Advisory Committee:
Associate Professor Monita Chatterjee, Chair
Professor Sandra Gordon-Salant
Professor Peter Fitzgibbons
Associate Professor Rochelle Newman
Assistant Professor Tracy Fitzgerald
Professor Robert Dooling

Dedication

This dissertation is dedicated to my parents, Lars and Gennell, for their unconditional love, guidance, and encouragement. I am also grateful to my dearest friends and classmates who have provided me with immense support through the years. I am truly blessed.

Acknowledgements

First and foremost, I would like to express my gratitude to my mentor Dr. Monita Chatterjee, who has provided me with incredible support and friendship during my time as a graduate student. I am also grateful to Dr. Sandra Gordon-Salant who has also served as an invaluable advisor and mentor. Additional thanks to the members of my dissertation committee for their input and guidance in directing my dissertation research: Dr. Tracy Fitzgerald, Dr. Peter Fitzgibbons, Dr. Rochelle Newman, and Dr. Robert Dooling.

Many thanks to Dr. Qian-Jie Fu at House Ear Institute for allowing use of his experimental software, and to John J. Galvin III (also of House Ear Institute) for his advice and expertise when I was planning these experiments. I am also grateful to Dharma Teja for software support. Additional thanks to Dr. Frederick Gallun at the National Center for Rehabilitative Auditory Research for his advice during data analysis. Special thanks to the participants for their time and willingness to participate in these studies.

Last but certainly not least, many thanks to my fellow lab mates in the Cochlear Implants and Psychoacoustics lab for their help with this dissertation and for providing wonderful camaraderie...and a plentiful amount of gummy bears.

This project was funded by NIH/NIDCD grant number R01DC004786 to Monita Chatterjee, and a fellowship from training grant T32DC000046 (PI: Dr. Arthur Popper) to the author.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: GENERAL INTRODUCTION

Several cues aid in one's ability to listen to a target message in the presence of extraneous distracters, but adequate processing of fundamental frequency *(f0)* information is critical to communicating in such environments. Speech is a harmonic complex signal, and the pitch of listeners' voices is roughly equal to the *f0* of the signal. Individuals use voice-pitch information not only to determine talker-gender (Childers & Wu, 1991; Linke, 1973; Titze, 1989), but also to separate competing sound sources (Brokx & Nootebohm, 1982; Brungart, 2001; Brungart, Simpson, Ericson, & Scott, 2001; Summers & Leek, 1998), both of which are essential in most communicative situations. Although not a focus of the current discussion, *f0* also helps to convey information about prosody or intonation, which aids in the lexical segmentation of speech discourse (Cooper, 1983; Rosen, 1992). The range of human voice-pitch is approximately 100-300 Hz, with males exhibiting lower *f0*s compared to females (Peterson & Barney, 1952).

While the importance of the *f0* cue to everyday communication is indisputable, the mechanism for coding such information is still debated. In general, two primary mechanisms are hypothesized to code pitch in a healthy auditory system. The first pitch coding mechanism, commonly referred to as *'spectral' coding (place coding, or rate-place coding)*, exploits the tonotopic organization of the cochlea. Briefly, each point along the cochlea is tuned to optimally process information of a certain frequency; nerve fibers associated with that location in the cochlea respond to basilar membrane excitation thereby causing a pitch percept. High frequencies and low frequencies are best coded in the base and apex of the cochlea, respectively. The second theory of pitch coding is referred to as *'temporal' coding (or time coding).* As the name would suggest, auditory fibers respond to an acoustic stimulus by "time-locking", or "phase-locking" to a given phase in the stimulus. Despite the fact that these two disparate and general theories exist, pitch of lower frequencies (<5 kHz) is likely coded using both place and time cues (commonly referred to as *spectro-temporal coding*) (see de Cheveigné, 2005, for review).

In a seminal paper, Rosen (1992) characterized temporal cues based on the rate or frequency (cycles/second) of the acoustic signal. The term *envelope* was used to describe the slowest modulation rates (<50 Hz), *periodicity* was used to describe slightly faster rates (50-500 Hz), and *temporal fine-structure* (TFS) was used to describe the fastest modulation rates (>600 Hz). Each of these frequency ranges helps to convey particular aspects of the speech signal (e.g., voicing, manner, stress, intonation). Relevant to the current discussion, periodicity information aids in the perception of voice-pitch and/or talker-gender since the average *f0* values of most male and female talkers are roughly 100 and 200 Hz, respectively (Peterson & Barney, 1952).

Figure 1 shows the time waveform of a harmonic complex with an *f0* of 100 Hz. As shown in this figure, the TFS information consists of the fast modulations caused by the summing of the harmonics, or partials. The periodicity information is conveyed in the 'pitch period' of the envelope and occurs at a rate equal to that of the *f0*. It should be noted that while Rosen (1992) defined only those temporal frequencies <50 Hz as "envelope cues", the same term is commonly used to refer to the slower temporal fluctuations (those not belonging to the TFS) of any complex sound. For example, the envelope of the harmonic complex in Figure 1 repeats at a frequency of 100 Hz but would still be referred to as the envelope of the amplitude waveform. Such cues will be referred to as envelope periodicity cues for the remainder of the discussion. Although it is

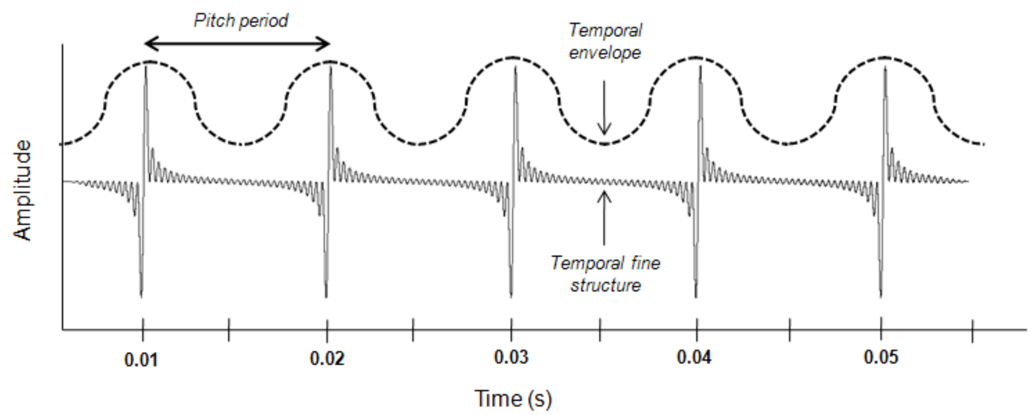

Figure 1. Schematic of a time waveform showing a harmonic complex with a fundamental frequency of 100 Hz. The pitch period (1/*f*, where *f*=frequency), temporal envelope and temporal fine structure are labeled.

widely believed that these temporal cues are essential in pitch coding, evidence suggests that adequate spectral resolution is also needed for optimal coding.

In recent years, for example, cochlear implant (CI) research has demonstrated the importance of spectral resolution in pitch coding. Current technology affords CI users with a crude representation of the signal, and in particular, CI processing results in a severe reduction of spectral (‘place’) cues (Chatterjee & Shannon, 1998; Shannon, 1990). Although speech processing strategies differ slightly depending on the particular CI device, the most common strategy is referred to as Continuous Interleaved Sampling (CIS). For this process, the incoming speech signal is divided into different frequency bands (or channels) using a number of band-pass filters, and the temporal envelope within each band is extracted and preserved. In pulsatile stimulation strategies (more commonly used), TFS of the signal is discarded and is replaced with biphasic electric pulse trains (presented at a static frequency or rate across all electrodes). The temporal envelopes associated with each spectral band are used to modulate a train of current pulses. The rate of the carrier (current pulses) is also known as the stimulation rate. The modulated current pulse train from each band-pass filter is sent to the electrodes which correspond to the specific band of frequencies. That is, if the amplitude envelope originated in the frequency band centered at 1000 Hz, then the associated modulated pulse train would be sent to the electrode that corresponds best to that frequency according to the tonotopic organization of the cochlea (for review, see Loizou, 1998).

Although the design of the CI attempts to take advantage of the tonotopically organized cochlea, electrical current overlap (“cross-talk” between electrodes, or channels) limits spectral resolution. Studies have demonstrated that CI users listen with eight spectral channels, at best (Fishman, Shannon & Slattery, 1997; Friesen, Shannon, Baskent & Wang, 2001), which means that if pitch is coded using spectral mechanisms, then CI users would not be able to derive *f0* information. Recall though, that the original temporal envelope and periodicity information is retained and unaltered in CI processing; therefore, CI users could theoretically gather *f0* information using temporal envelope periodicity cues. Psychoacoustic studies in normal-hearing (NH) listeners demonstrate the ability to hear some representation of 'pitch' via the temporal envelope or frequency alone (Burns & Viemeister, 1976; Formby, 1985; Grant, Summers, & Leek, 1998; Hanna, 1992; Miller & Taylor, 1948; Mowbray, Gebhard, & Byham, 1956; Patterson, Johnson-Davies & Milroy, 1978; Pollack, 1952).

Two general methods have been used to estimate performance outcomes in CI listeners. The first method is to simply evaluate performance in actual CI listeners when stimuli are presented via their everyday speech processors. The second method, noise-band vocoding (NBV), is used to simulate CI processing for NH individuals. Cochlear-implant simulations are accomplished by dividing the speech signal into a specified number of frequency bands, and extracting the temporal envelope from each band. The TFS within in each band is replaced by a noise-band of the same frequency range, which is then modulated by the corresponding temporal envelope information. This type of simulation mimics the current, popular speech coding strategy previously mentioned (i.e., CIS). The result is a stimulus with unaltered temporal cues, but limited spectral cues (e.g., Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

Cochlear-implant simulations allow for increased control of the experimental design, as several variables can be manipulated in NBV processing. The *number of* *channels* can be manipulated to simulate more or less channel interaction. For example, four channels represents more channel interaction whereas 32 channels represents less channel interaction. The *filter slope* of each channel can be manipulated and is an additional way to represent more or less channel interaction between filter bands. For example, a steeper filter slope (96 dB/octave) would represent almost complete independence of channels, whereas a filter slope of 6 dB/octave represents broad overlap of filters and decreased spectral resolution. Investigations show that decreases in speech recognition are present with a filter slope of less than 12 dB/octave (Fu & Shannon, 2002), therefore most investigations using CI simulation maintain a filter slope of at least 24 dB/octave. As described before, the time-varying envelope is extracted from each of these channels; this is commonly accomplished by using half-wave rectification followed by low-pass filtering or implementing a Hilbert transform. The cut-off frequency of the low-pass filter is used to determine the range of temporal envelope frequencies that will be extracted from each frequency band. The cut-off frequency should be greater than 16 Hz, as studies show that lowering the frequency of the low-pass filter to this point (16 Hz) causes decrements in performance (Shannon et al., 1995).

There are several variables that cannot be easily replicated when using CI simulations. First, electrical stimulation of the auditory nerve can sometimes result in pronounced temporal coding as CI listeners sometimes exhibit excellent performance on tasks measuring modulation sensitivity or gap detection (Fu, 2002; Shannon, 1989). Results from Shannon (1993) and Snyder, Vollmer, Moore, Rebscher, Leake and Beitel (2000), however, suggest that temporal coding is quite comparable across acoustic and electric stimulation once differences in loudness and dynamic range are taken into account. Regardless, it is likely that CI simulations do not precisely represent the temporal cues in actual CI processing. Furthermore, other factors such as duration of deafness (and consequential neural degeneration) also contribute to the success of cochlear implantation (Leung, Wang, Yeagle, Chinnici, Bowditch, Francis, & Niparko, 2005; Shea, Domico, & Orchik, 1990; Shepherd, Roberts, & Paolini, 2004). As there is no simple way to simulate the effects of duration of deafness, results obtained using CI simulations represent the outcomes of a CI listener with excellent neuron survival.

As previously mentioned, both methods have been employed to determine how well CIs convey *f0* information. Overall, evidence among CI users and NH individuals listening to CI simulations demonstrates that most CI users are able to use differences in the temporal envelope pitch (TEP) pattern to discriminate between *f0s* (when stimuli are presented in isolation) (Chatterjee & Peng, 2008; Green, Faulkner, & Rosen, 2002; Green, Faulkner, Rosen, & Macherey, 2005). Collective findings, however, suggest that auditory tasks involving *f0* processing (e.g., gender identification, talker identification, sound source segregation based on *f0* cues) remain difficult for most CI users (Fu, Chinchilla, & Galvin, 2004; Fu, Chinchilla, Nogaki, & Galvin, 2005; Qin & Oxenham, 2003; Stickney, Zeng, Litovsky, & Assman, 2004; Stickney, Assman, Chang, & Zeng, 2007; Vongphoe & Zeng, 2005). Generally speaking, CIs convey weak *f0* cues, and they are far less salient compared to those cues available to a listener with NH.

Although the ability to discriminate voice pitch is well-documented in CI users, it is reasonable to conjecture that chronological age may influence one's ability to perceive TEP cues. To date, the interaction between cochlear implantation and the aging auditory system is not well understood. While there is some evidence that older individuals might experience more difficulty understanding spectrally-degraded phonemes compared to younger individuals (Schvartz, Chatterjee, & Gordon-Salant, 2008), several studies demonstrate equal performance across younger and older experienced CI users when measured on simple tasks such as understanding speech in quiet (Chatelin, Kim, Driscoll, Larky, Polite, Price, & Lalwani, 2004; Leung et al., 2005; Pasanisi, Bacciu, Vincenti, Guida, Barbot, Berghenti, & Bacciu, 2003). Declines in auditory temporal processing, however, are commonly cited as a consequence of aging among adults (Abel, Krever, & Alberti, 1990; Fitzgibbons & Gordon-Salant, 1994; 1998; Helfer & Wilber, 1990; Phillips, Gordon-Salant, Fitzgibbons, & Yeni-Komshian, 1994; Schneider, Pichora-Fuller, Kowalchuk, & Lamb, 1994; Snell, 1997; Snell & Frisina, 2000) and it is therefore highly plausible that older listeners would experience difficulty when forced to hear pitch using primarily temporal cues (as required in CI processing). This wealth of literature collectively demonstrates that temporal processing of both simple and complex stimuli is negatively affected as a consequence of advanced age, and recent literature shows temporal processing deficits begin to manifest as early as middle-age (Snell & Frisina, 2000; Grose, Hall, & Buss, 2006; Lister, Besing, & Koehnke, 2002; Ostroff, McDonald, Schneider, & Alain, 2003).

Despite such findings, one aspect of temporal coding that has yet to be measured in older listeners is the ability to code pitch using primarily temporal cues. A few studies have demonstrated that the ability to *detect* amplitude modulation of broad-band noise does not change as a function of age (e.g., Takahashi & Bacon, 1992), but no studies have measured the ability of younger and older listeners to *discriminate* between amplitude modulation frequencies of broad-band noise. Neurophysiologic evidence obtained from animals suggests that older listeners might require larger differences to reliably discriminate between modulation frequencies in the temporal envelope (Shadduck-Palombi, Backoff & Caspary, 2001; Walton, Simon & Frisina, 2002), while electrophysiologic data obtained in humans further support age-related declines in temporal envelope processing (Grose, Mamo, & Hall, 2009; Leigh-Paffenroth & Fowler, 2006; Purcell, John, Schneider, & Picton, 2004).

Based on these data, it is expected that older individuals listening to spectrally degraded signals would experience difficulty perceiving voice pitch cues via the temporal envelope. World census data show that the number of elderly individuals grows by 795,000 each month, and hearing impairment is known to be a common consequence of natural aging (Kinsella & Velkoff, 2001; Pearson, Morrell, Gordon-Salant, Brant, Metter, Klein, & Fozard, 1995). Consequently, the number of potential elderly CI candidates grows each year as cochlear implantation is considered a viable option for elderly hearing-impaired (HI) listeners. It is crucial to understand how aging might affect *f0* processing in these listeners. *The purpose of these experiments is to determine how aging influences listeners' abilities to code voice pitch when spectral cues are degraded, and individuals are therefore required to rely on temporal envelope cues for periodicity information.* The overall aim is to form a comprehensive understanding of such processes using several methods and different types of stimuli (non-speech and speech).

All experiments were performed in a group of younger (ages 21-30) and older (ages 60 and older) NH listeners, and NBV was used to simulate CI processing. Experiment 1 measured the ability of listeners to discriminate between the frequencies of sinusoidally amplitude modulated (SAM) broadband noise bursts (TEP coding using non-speech stimuli). Experiment 2 estimated the ability of listeners to perform a *f0* discrimination task when stimuli were spectrally degraded harmonic complexes (spectro-temporal pitchcoding using non-speech stimuli). Experiment 3 quantified the ability of listeners to correctly identify the gender of a talker, when spectral and temporal cues were manipulated (spectro-temporal processing using simple speech stimuli). Lastly, Experiment 4 measured the ability of listeners to hear a target sentence in the presence of a competing sentence uttered by a different speaker or multi-talker babble when spectral cues are degraded (spectro-temporal processing using complex speech stimuli). The same listener was not allowed to participate in more than two of the four experiments. Further, each participant was only a permitted to participate in one of the two experiments that used non-speech stimuli (i.e., Experiment 1 *or* Experiment 2) and in one of the two experiments that used speech stimuli (i.e., Experiment 3 *or* Experiment 4). From a clinical perspective, it is hoped that results of these experiments will aid in understanding potential interactions between aging and cochlear implantation. From a scientific perspective, it is anticipated that the results of these studies will further elucidate the consequences of aging on temporal coding in the auditory system.

CHAPTER 2: PERCEPTION OF SPECTRALLY DEGRADED PERIODICITY PITCH IN ADULTS USING NON-SPEECH STIMULI

Literature Review

The ability to properly perceive voice-pitch information is essential to everyday communication. While the precise mechanisms of pitch coding have served as a topic of debate among scientists for many decades, there is unequivocal evidence that adequate spectral coding is required to hear voice-pitch information with high accuracy. For example, individuals who utilize CIs exhibit poor spectral resolution and consequently demonstrate a reduced ability to discriminate between voice pitches. While CI users are generally not able to use spectral cues to gather voice-pitch information, several studies have demonstrated that adult listeners are able to perceive a rough estimate of voice pitch using periodicity information contained in the temporal envelope.

Although several studies have measured the ability of younger adults to discern voice-pitch information using purely temporal pitch, evidence from studies using neurophysiologic and electrophysiologic measures provide reason to suspect that older CI users might have greater difficulty using such cues when compared to younger CI users. To date, the interactions between aging and periodicity processing are not well understood. This relationship certainly warrants further investigation since a significant number of adult CI recipients are, in fact, older HI individuals. Therefore, any interactions between aging and voice-pitch processing have important implications for numerous adult CI recipients.

In order to better understand why aging might affect spectrally degraded voice-pitch processing, the following section reviews how voice pitch is coded in both healthy and impaired auditory systems and also discusses the ability of NH listeners to perceive voice-pitch information using spectral and temporal cues. Further, literature supporting senescent declines in spectro-temporal *f0* processing will be discussed. The current experiments focus on the ability of younger and older NH listeners to perceive pitch using spectral and/or temporal cues. More explicitly, the following experiments aim to determine if younger and older listeners differ in their ability to code *f0* information when spectral cues are limited or absent altogether, as in CI processing.

Voice-pitch Perception in Younger Listeners

Complex pitch perception in a healthy auditory system

The pitch of a harmonic series roughly approximates that of its lowest component, or the *f0* of the harmonic series (Moore, Glasberg, & Peters, 1985; Plomp 1967; Ritsma, 1967). Generally speaking, there are three theories used to explain how complex pitch is coded in the auditory system: 1) Spectral, 2) Temporal, and 3) Spectro-temporal theories. Those pitch theories subscribing to spectral and spectro-temporal coding mechanisms (Goldstein, 1973; Licklider, 1951; Meddis & Hewitt, 1991a; 1991b; Terhardt, 1972; von Hemholtz, 1877; Wightman, 1973), propose that the *f0* and its first 5 to 10 harmonics fall within the tonotopically organized auditory filter bands within the cochlea and, consequently, these partials are referred to as '*resolved harmonics'* (Bernstein and Oxenham, 2003; Moore & Ohgushi, 1993; Plomp, 1964; Plomp and Mimpen, 1968). In this case, the frequency-tuned cochlear filters are used to resolve the *f0* and associated harmonics, thereby promoting coding of voice-pitch information. Temporal theories, however, propose that pitch information is coded using only temporal cues (Schouten, 1940). In this case, multiple harmonics fall within the same auditory filter band and are therefore referred to as *unresolved harmonics*. The pitch is coded based on the summed waveform of adjacent harmonics, which results in a complex waveform modulated at a rate equal to that of the *f0* ('pitch period'). This type of temporal pitch has been referred to as "residue" or "virtual" pitch since it does not require that the *f0* or any harmonics be resolved within the cochlea.

Figure 2 is a schematic demonstrating the supposed processing of resolved and unresolved harmonics in the peripheral auditory system. This figure shows how a harmonic complex with an *f0* of 100 Hz (such as the one shown in Figure 1) is processed in a normal auditory system. As shown in this figure, the first few harmonics of a harmonic complex are each resolved in the narrow, frequency tuned auditory filters found in the apex of the cochlea. The higher harmonics, however, become unresolved since the auditory filter width begins to widen in more medial and basal portions of the cochlea. The unresolved harmonics sum together to create a complex waveform with an envelope that repeats at the rate of the *f0*. While results from studies suggest that neither spectral nor temporal theories can singly account for pitch perception in human listeners (see de Cheveigné, 2005 for review) resolved harmonics have been shown to provide a more salient pitch percept (Bernstein & Oxenham, 2003; Houtsma and Smurzynski, 1990; Plomp, 1964; Plomp & Mimpen, 1968; Qin & Oxenham, 2005). While it is possible to gather some pitch information using TEP cues, optimum pitch coding requires resolution of individual harmonics. Results have shown that the perception and discrimination of *f0* information by younger NH listeners is quite exceptional using resolved harmonics, as *f0* difference limens (*f0*DLs) are roughly 1-2%, on average, when either speech or non-

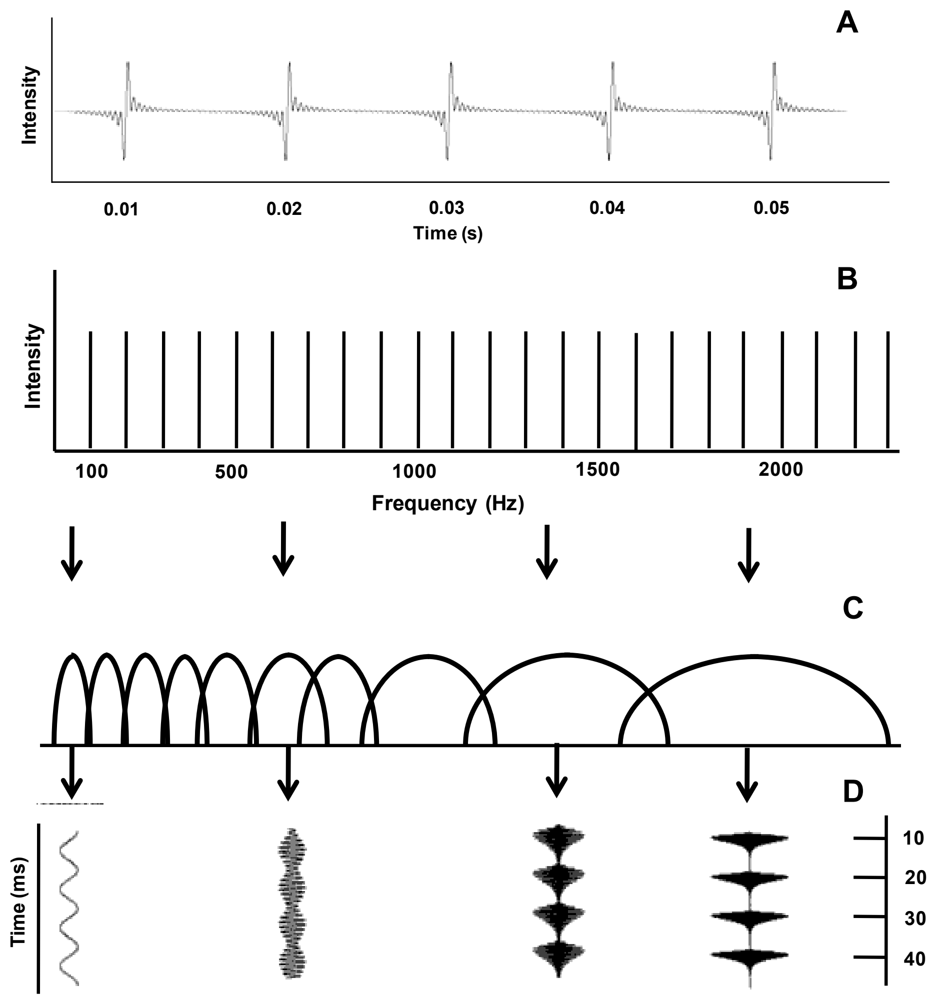

Figure 2. Schematic representing the processing of a complex harmonic series in the peripheral auditory system. The first panel (A) shows the input waveform (100 Hz harmonic series). The second panel (B) shows the spectrum of the input waveform. The third panel (C) shows the auditory filterbank through which the input waveform is processed (wider filterwidth with increasing center frequency). The fourth panel (D) shows the time-averaged excitation pattern for some of the filters. The lower harmonics are resolved, whereas the multiple higher harmonics fall within a single auditory filter and are therefore unresolved. Note that not all figures are drawn to scale. Adapted from Plack and Oxenham (2005).

speech stimuli are used (Arehart, 1994; Moore & Peters, 1992; Qin & Oxenham, 2005; Summers & Leek, 1998; Vongpaisal & Pichora-Fuller, 2007).

Complex pitch perception in an impaired auditory system

Individuals with auditory impairment typically exhibit difficulty processing *f0* information. In HI listeners, for example, it is hypothesized that changes in TFS processing and/or broadened auditory filters cause a decline in pitch processing (Moore, Glasberg, Flanagan, & Adams, 2006; see Moore, 2008 for review). In the case of CI users, a complete absence of TFS in the signal processing and extremely broad auditory filter widths (or spectral channels) result in diminished *f0* processing. While TFS processing is undoubtedly important to pitch perception, the following discussion will focus on the effects of broadened auditory filter bandwidth on *f0* processing, particularly as it pertains to CI listeners. In particular, CI processing does not allow for resolved harmonics and consequently CI listeners must rely on TEP cues to obtain voice-pitch information. Several studies using both NH and CI listeners lend support for decreased pitch salience using TEP cues.

Consequences of reduced spectral resolution in cochlear-implant processing. Although diminished spectral and/or temporal coding commonly results in poor voice-pitch coding in those with sensorineural hearing loss (SNHL), voice-pitch coding abilities are even poorer among CI users. Those who utilize CIs are subject to considerable spectral distortion, attributed to electrical field overlap (“cross-talk” between electrodes, or channels), and they consequently exhibit severely degraded spectral coding abilities. For example, most CI users listen with eight spectral channels, at best (Fishman et al., 1997;

Friesen et al., 2001). Even if the number of channels in a CI was adequate to resolve harmonics, the device does not transmit TFS cues, and therefore utilization of such cues to aid voice-pitch perception is impossible. Therefore, CI listeners are unable to code voice pitch using spectral mechanisms, and often demonstrate difficulty perceiving talker gender (Fu et al., 2004; Fu et al., 2005; Vongphoe & Zeng, 2005).

Temporal-envelope pitch perception. Recent results from CI listeners exemplify the inefficiency of TEP cues to adequately code *f0* information (Chatterjee & Peng, 2008; Qin & Oxenham, 2005). These results, however, are not the first to quantify the ability of listeners to perceive *f0* information using only temporal cues. For example, early experiments used repeated broad-band noise bursts or amplitude modulated broad-band noise bursts to measure the ability of listeners to gather pitch information in the complete absence of spectral cues (Burns & Viemeister, 1976; Formby, 1985; Hanna, 1992; Grant et al., 1998; Patterson et al., 1978). In theory, this type of experiment should be akin to the resolution of pitch using unresolved harmonics. Recall from previous discussions that, in the case of unresolved harmonics, pitch information is only conveyed in the modulation rate of the temporal envelope and spectral cues are absent. Likewise, sinusoidally-amplitude modulated (SAM) noise or repeated noise bursts also provides a pitch period via the temporal envelope or frequency of bursts while providing no discernable spectral cues.

The earliest experiment of this kind was reported by Miller and Taylor (1948) and investigated the pitch percept of repeated noise bursts. Results showed that listeners exhibited relatively good performance for presentation rates of up to 250-300 bursts/second, at which point performance begins to decline with increasing rates. Subsequent studies using similar methods (Mowbray et al, 1956; Pollack, 1952) also showed a significant decline in performance as the repetition rate was increased.

In a later study, Formby (1985) used 500 ms, 100% SAM broadband noise (500-4000 Hz) bursts. The modulation frequency of the sinusoid was varied systematically for each condition. A two-alternative-forced choice (2-AFC) (method of constant stimuli) task was used to measure the modulation rate DL for four NH listeners. The modulation frequency of the reference signal was varied from 5-550 Hz, while the modulation frequency of the comparison signal was always presented at a higher frequency than that of the reference. Listeners were instructed to indicate which of the two intervals contained the higher modulation frequency. In keeping with previous results (Miller & Taylor 1948), results showed that discrimination performance worsens with increasing modulation frequency. Results among four younger NH (YNH) listeners exhibited averaged Weber fractions of approximately 5, 15, and 50% for standard frequencies of 100, 200, and 400, respectively. Hanna (1992) used similar methodologies and reported comparable findings in three YNH listeners. In both investigations, the frequency discrimination threshold was fairly low (2-3%) for standard modulation frequencies up to 60 Hz. The difference limens for higher modulation frequencies, however, exponentially increased (worsened). For example, at 100 Hz, the DL was approximately 5% or more, and increased to 15-20% at a standard rate of 200 Hz. A similar performance pattern was also reported in Grant et al (1998). Furthermore, Grant et al. (1998) also demonstrated that the ability to *detect* amplitude modulation was not related to the ability to *discriminate* between amplitude modulation frequencies.

Perception of fundamental-frequency by cochlear-implant listeners. CI users are not able to use resolved harmonics in order to hear *f0* information, but recent work suggests that some CI users are able to use modulations in the temporal envelope to obtain voice pitch information (Chatterjee & Peng, 2008; Green et al., 2002; Green et al., 2005). In this case, CI users were utilizing *f0* cues conveyed in the temporal envelope periodicity, which is somewhat analogous to the concept of residue pitch in NH listeners. In both cases, none or few of the harmonics are resolved, and the perception of pitch is driven primarily by the modulation frequency of the envelope.

In an effort to quantify *f0* processing in CI listeners, Qin and Oxenham (2005) measured *f0*DLs using a 2-AFC adaptive procedure in four YNH listeners using NBV harmonic complexes. The standard *f0* was either 130 or 220 Hz, while the comparison *f0* (*f0* + △*f0*) was adapted based on the listener's response. Harmonic complexes were processed using NBV techniques to contain 1, 4, 8, 24 or 40 channels, thereby measuring the contributions of both temporal and spectral cues on *f0*DLs. Unprocessed harmonic complexes were also used.

In keeping with previous findings, results showed that *f0*DLs were best when unprocessed stimuli were used, in which case the average *f0*DL was approximately 1% for both 130 and 220 Hz standard *f0*. In other words, participants only needed about 1.3 or 2.2 Hz change in *f0* to reliably perceive a difference, for 130 or 220 Hz standard *f0*, respectively. However, *f0*DLs tended to worsen with decreasing spectral resolution. For example, the *f0*DL doubled (to approximately 2%) when stimuli contained 40 spectral channels. Listeners needed about 5% difference (6.5 Hz) for 130 Hz *f0* and about a 9%

difference (19.8 Hz) for the 220 Hz *f0* to detect a difference when relying solely on temporal cues (1 spectral channel).

Chatterjee and Peng (2008) measured modulation frequency discrimination (MFD) abilities in CI listeners using direct stimulation methods. In this case, the stimulus is delivered through a custom interface, thereby bypassing the external everyday speech processor and allowing for precise control of the stimulus. Results were highly variable across subjects but showed median MFD thresholds (percent Weber fractions) of CI users were approximately 10-20% for standard rates of 100 and 200 Hz, respectively. These average results somewhat approximate those thresholds obtained in NH individuals tested with SAM noise stimuli (Formby, 1985; Grant et al., 1998; Hanna, 1992) or NBV CI simulations (Qin & Oxenham, 2005), despite significant variability in performance among the CI listener cohort.

Taken together, these results confirm the ability to code pitch using only temporal envelope periodicity cues. While they collectively demonstrate that NH listeners are able to use such cues to gather pitch information, the salience of pitch in this case is less than that promoted by resolved harmonics. Among young listeners, the *f0*DL for resolved harmonics (unprocessed harmonic complex) hovers around 1-2%, while the *f0*DL for unresolved harmonics (or the discrimination of the rate of modulated noise) typically varies from 5-15%, for standard frequencies of 100-200 Hz, respectively. Similarly, results suggest that CI users and those listening to CI simulations also demonstrate a reduced capacity to code *f0* information due to the fact that they must rely on envelope periodicity cues (Chatterjee & Peng, 2008; Laneau, Moonen & Wouters, 2006; Qin & Oxenham, 2005).

Voice-pitch Perception in Older Listeners

While the ability to code complex pitch has been extensively investigated in younger NH and HI listeners, less is known about how aging might impact this capacity. In particular, changes in spectral or temporal coding could theoretically impact older listeners' ability to accurately perceive complex pitch information. In particular, diminished temporal processing skills are commonly cited in older listeners, but interactions between aging and TEP coding have yet to be explored. Several researches have pointed to age-related changes in auditory processing at various levels in the central auditory nervous system to account for senescent declines in temporal processing. While CIs are undoubtedly efficacious in older listeners, potential interactions between aging and TEP processing could have important implications for elderly CI users.

Fundamental-frequency (f0) discrimination using resolved harmonics

Only a few studies to date have investigated the effects of aging on the ability of humans to discriminate between complex tones varying in *f0* value. Moore and Peters (1992) measured *f0*DLs in five younger NH listeners and in ten older listeners with normal-to- near-normal hearing thresholds at frequencies of 2000 Hz and lower. Stimuli were harmonic complexes (containing the first 1-12 harmonics) and reference *f0* values were 50, 100, 200 or 400 Hz. Averaged results show that *f0*DLs in younger listeners varied from approximately 1.5 – 0.7%, from 50-400 Hz, respectively, with performance improving slightly from lower to higher *f0* values. Performance among the older listeners, however, was significantly worse, as average *f0*DLs varied from approximately 4% -

1.6% from 50-400 Hz, respectively; again, performance improved slightly from lower to higher *f0* values.

A later study by Vongpaisal and Pichora-Fuller (2007) measured *f0*DLs using synthesized vowel stimuli (containing all spectral and temporal cues) in 15 younger and 15 older adults, and found that the *f0*DLs for older adults were significantly worse (mean= 1.8 Hz) than that of the younger adults (mean=0.6 Hz). The results from Vongpaisal and Pichora-Fuller (2007) are in agreement others (Moore & Peters, 1992; Qin & Oxenham, 2005; Summers & Leek, 1998) suggesting that YNH individuals require as little as a 1% change in *f0* to perceive a difference in pitch when harmonics are resolved (containing all spectral and temporal cues). It appears, however, that older listeners need as much as two-to-three times the difference in *f0* to achieve comparable discrimination performance on an *f0*DL task, when compared to that of their younger, NH counterparts (Moore & Peters, 1992; Vongpaisal & Pichora-Fuller, 2007). Taken together, these results suggest age-related declines in periodicity coding of resolved harmonics.

Fundamental-frequency (f0) discrimination using unresolved harmonics & envelope periodicity

Behavioral measures. The perception of TEP cues has been somewhat quantified among younger listeners, while less is known about how such abilities might be affected by the age of the listener. The ability of older listeners to simply detect amplitude modulation in broad-band noise seems to be relatively unaffected by the age of the listener (Takahashi & Bacon, 1992), while the influence of aging on the ability of adults to detect changes in the frequency of amplitude modulation, however, remains rather ambiguous. Results from Grant et al. (1998) suggest that the detection and discrimination of amplitude modulation are, in fact, fairly independent abilities (assuming the modulation depth is adequate to maximize the perception of differences in amplitude modulation rate). Therefore, it is possible that while the ability of listeners to simply detect modulation in a signal is fairly robust to aging, the discrimination of salient amplitude modulation frequencies is a separate matter.

In the same study previously mentioned, Moore and Peters (1992) also investigated the ability of younger and older NH and HI listeners to detect changes in *f0*, while manipulating the number of harmonics present in a 12-harmonic complex. In one particular condition, listeners were tested on their ability to detect changes in *f0* when stimuli only contained the highest, harmonics (6-12); in this case, the authors proposed that any discrimination was based on TEP via unresolved harmonics. The performance for older listeners was worse than that of younger listeners, regardless of hearing status, for all reference *f0* values (50-400 Hz). For example, *f0*DLs varied around 1% in YNH listeners and from 1.5-3% for ONH listeners (for 50-400 Hz, respectively). Despite these results, it is debatable if listeners were relying solely on TEP information to perform the task. Several studies have since shown that listeners are able to reconstruct *f0* information from the unresolved harmonics using spectral cues via combination tones (or distortion products), and that, in some cases such spectral cues aid performance unless masked appropriately (Pressnitzer & Patterson, 2001). No such masker was used by Moore and Peters (1992) and it is therefore uncertain to what extent listeners were able to use extraneous cues in either age group. Further, Moore and Peters (1992) used stimuli with the lowest six harmonics removed to demonstrate *f0* discrimination of unresolved harmonics. There is controversy, however, regarding how many harmonics of a complex signal are resolved via cochlear filtering; some suggest that as few as five, but perhaps up to ten harmonics of a complex signal may be considered ‘resolved harmonics’ (Bernstein and Oxenham, 2003; Moore and Ohgushi, 1993; Plomp, 1964; Plomp and Mimpen, 1968). Overall, the results of Moore and Peters (1992) are rather equivocal, and the effect of aging on the pitch of unresolved harmonics (via TEP) remains ambiguous.

Objective measures. While the results of Moore and Peters (1992) are rather inconclusive, subsequent electrophysiologic and neurophysiologic data do support senescent declines in the ability to code TEP. For example Purcell and colleagues (Purcell et al., 2004) measured the envelope following responses (EFRs) in younger and older listeners, and demonstrated reduced amplitude in the older individuals but only for frequencies greater than 100 Hz. The authors hypothesized that such results support decreased temporal acuity in the aging auditory system, and in particular, a deficit in auditory brainstem function. These results are corroborated by subsequent electrophysiologic evidence (Leigh-Paffenroth & Fowler, 2006; Grose et al., 2009) and collectively endorse diminished periodicity coding in elderly listeners, particularly at higher modulation rates (those ≥ 100 Hz).

Psychoacoustic and electrophysiologic results are further confirmed by neurophysiologic data, showing that single-unit responses to SAM noise recorded from the inferior colliculus (IC) also change as a function of age (Walton et al., 2002). Walton and colleagues (Walton et al. 2002) recorded single-unit responses in the IC to SAM noise in young and old CBA mice. Stimuli were 100% amplitude modulated, and modulation frequencies varied from 10-800 Hz. Overall results showed an increase in spike count and phase-locking among older mice, but less frequency specificity compared to those results obtained in younger mice. That is, neurons in older mice were more likely to respond to any stimulus rather than a specific modulation rate. Furthermore, the strongest responses among younger mice occurred at higher modulation frequencies (200-400 Hz), while the strongest responses among older mice occurred at lower modulation frequencies (40-100 Hz). Collectively, behavioral and objective measures suggest that older listeners would experience increased difficulty perceiving voice-pitch information using TEP cues compared to younger listeners.

Possible origins of senescent declines in pitch coding: Afferent and efferent coding

It is hypothesized that age-related declines in temporal periodicity coding could be attributed to either senescent degeneration of afferent neural populations and/or alterations in the efferent inhibitory mechanisms. The following discussion reviews how a degeneration of either pathway could potentially influence temporal envelope pitch coding in older adults.

Loss of afferent information: Neural presbycusis. Schucknecht and colleagues (Schuknecht, 1955; 1989; Schuknecht & Gacek, 1993) were the first to identify four major types of 'presbycusis' which is the general term for hearing deficits associated with the natural processes of aging. One classification in particular, termed *neural presbycusis*, was defined as a loss or degeneration of neurons within the cochlea and cochlear nucleus (Amesen, 1982; Schuknecht, 1989; Schuknecht & Gacek, 1983). Although the term presbycusis is often associated with elevated detection thresholds (i.e., hearing loss), it has been hypothesized that neural presbycusis can exist in the absence of significant hearing impairment as measured by a conventional pure-tone audiogram (Schuknecht, 1989). It is reasonable to conjecture that a reduction in neurons at the level of the cochlea or cochlear nucleus could cause deficits in auditory processing (e.g., temporal processing) attributed to diminished representation of the auditory signal. It is therefore possible, that a decline in TEP coding could be at least partially caused by a loss or de-synchronization of afferent temporal coding.

Review of central mechanisms in pitch perception. The tonotopicity exhibited in the cochlea continues throughout the cochlear nucleus, IC and cortex (Merzenich, Knight, Roth, 1975; Rose, Galambos, & Hughes, 1959), and this finding is commonly used to endorse spectral pitch theories. It is also known, however, that the temporal response properties of auditory neurons are crucial in pitch perception, either playing a role in rate-place, or simply temporal coding (Joris, Schreiner & Rees, 2004). Sub-cortical coding of pitch is seemingly driven by the diverse temporal response properties of cell units. For example, different types of auditory nerve fibers differ in their maximum firing rate; medium spontaneous rate fibers have higher maximum firing rates compared to high spontaneous rate fibers (Joris & Yin, 1992; Wang & Sachs, 1993). At the level of the cochlear nucleus, chopper units, which exhibit consistent and highly regular firing patterns, are thought to promote envelope periodicity information, and more specifically, fusiform cells are thought to aid in envelope coding (Frisina, Walton, & Karcich, 1994; Joris et al., 2004; Kim, Sirianni, & Chang, 1990). Conversely, those units exhibiting primarly-like responses code faster temporal fluctuations and are hypothesized to convey TFS information (Schofner, 1991; 1999). While the coding of temporal fine structure information is undoubtedly important, the remainder of the discussion will focus on the coding of periodicity information.

Neurons throughout the central auditory nervous system (CANS) exhibit diverse band-pass characteristics, a finding which has resulted in the postulation of amplitude modulation filterbank theories (Dau, Kollmeier, & Kholrausch, 1997a; 1997b). Several studies, however, have collectively demonstrated that the coding of amplitude modulation changes significantly as a signal ascends in the CANS. For example, the band-pass characteristics become increasingly complex, and the upper-limit of phase locking decreases while the coding of modulation itself can be enhanced (Frisina, 2001; Frisina, Smith, & Chamberlain, 1985; Javel, 1980; Joris and Yin, 1992; Moller and Rees, 1986; Rees & Moller, 1983; 1987). Because of this decrease in the upper-limit of modulation rate coding, it has been proposed that temporal phase-locking mechanisms present in the lower CANS must be replaced with a mechanism that exploits average firing rate of neurons (Langer & Schreiner, 1988; Moller & Rees, 1986; Rees & Moller 1987).

Furthermore, it is thought that the efferent auditory system and inhibitory mechanism are especially important in the processing of temporal modulations. Specifically, two neurotransmitters are fundamental to inhibitory processes in the CANS: glycine and γ-Aminobutyric acid (GABA) (Caspary, Pazara, Kossl & Faingold, 1987; Caspary, Ling, Turner & Hughes, 2008; Kolston, Osen, Hackney, Ottersen, Storm-Mathisen, 1992). While the comprehensive roles of glycine and GABA are beyond the scope of this paper, several studies have demonstrated their importance to auditory temporal processing. For example, various studies have shown a significant alteration in auditory temporal processing when glycine or GABA receptors are artificially blocked using neurochemicals. Specifically, investigations revealed a loss of synchrony and changes in the band-pass morphology of individual units (Backoff, Shadduck-Palombi, & Caspary, 1997; Backoff, Palombi, & Caspary, 1999). Collectively, they provide further support for the importance of top-down modulation of temporal coding in the CANS, and in fact, leading theories of specto-temporal complex pitch coding (i.e. auto-correlation theories) have been recently adapted to account for the influence of such factors (Balaguer-Ballester, Clark, Coath, Krumbholz, and Denham, 2009).

Age-related changes in auditory temporal processing within the central auditory nervous system. As previously noted, both excitatory and inhibitory mechanisms are responsible for the coding of temporal information in the CANS, and neurochemicals such as glycine and GABA are imperative to this process. In fact, several researchers have hypothesized that changes in the processing of glycine and GABA likely contribute to the noted deterioration of temporal coding in the aging auditory system (Banay-Schwartz, Lajtha, & Palkovits,1989; Caspary, Milbrandt, & Helfert, 1995; Walton et al. 2002; Wang, Turner, Ling, Parrish, Hughes, & Caspary, 2009; Willott, Milbrandt, Bross & Caspary, 1997). For example, several animal studies have reported noticeable declines in GABA synthesis and changes in GABA receptors with advanced age (Caspary et al., 1995; Caspary et al., 2008; Helfert, Sommer, Meeks, Hofstetter & Hughes, 1999; Milbrant, Albin, Turgeon, & Caspary, 1996).

In general, it is believed that inhibitory mechanisms (such as those regulated by glycine and GABA) help to sharpen neural coding, particularly at the level of the IC (Frisina et al., 1994; Caspary et al. 1995; Joris et al., 2004; Langer & Schreiner, 1998;

Walton et al., 2002). This hypothesis is supported by studies in which neurotransmission of glycine or GABA is artificially blocked or restricted in the cochlear nucleus and IC (Backoff et al., 1999; Caspary, Shadduck-Palombi, & Hughes, 2002) thus altering temporal processing at these locations. Consequences of restricting glycine and/or GABA processing include an increase in the total discharge rate, a reduction in the synchrony, and alterations in the band-pass characteristics of neurons (Backoff et al., 1999). These consequences are similar to those observed in aged auditory systems (Schatteman, Hughes, & Caspary, 2008; Walton et al. 2002). Therefore, it is reasonable to conjecture that age-related changes of periodicity coding are likely related to a reduction of neural inhibition in the aged auditory system.

Summary

Evidence from neurophysiologic and electrophysiologic studies support senescent declines in temporal envelope periodicity coding, but such declines have yet to be quantified using behavioral measures. Therefore, the extent to which aging affects this capacity remains unclear. It is hypothesized that changes in afferent and/or efferent processing could potentially affect coding of TEP information. While TEP is of less importance to NH listeners, CI users must rely on this cue to process *f0* information. Therefore, it is possible that older CI users might exhibit increased difficulty coding *f0* information compared to younger CI users.

The aim of the current experiments was to determine the extent to which aging affects complex-pitch discrimination using non-speech stimuli. In Experiment 1, we measured the resolution of TEP in younger and older NH listeners (no spectral cues). Participants were tested in a MFD task, using broadband SAM noise. In Experiment 2, we extended these results by measuring complex-pitch perception as spectral cues were varied using NBV processing techniques. Participants were tested in their ability to detect differences in the *f0* of a complex series, which contained 1, 8, or 24 spectral channels, but unaltered envelope periodicity cues.

Questions and Hypotheses

1. To what extent do younger and older NH listeners differ in their ability to discriminate temporal envelope periodicity when spectral cues are completely absent?
2. Does the effect of age on the ability to process periodicity cues differ as a function of modulation frequency or *f0*?
3. How does adult age affect *f0* discrimination when the availability of spectral cues is systematically varied to approximate cochlear-implant processing?

Based on previous literature (Grose et al., 2009; Leigh-Paffenroth & Fowler, 2006; Purcell et al., 2004) it is hypothesized that older listeners will exhibit a reduced ability to perceive TEP information, however the extent to which age will affect performance is unknown. Previous literature demonstrates that *f0*DLs approximate 1% and 2% in younger and older listeners, respectively (Moore & Peters, 1992; Vongpaisal & Pichora-Fuller, 2007). Therefore, it is possible that older listeners would experience a similar deficit when using TEP cues to discriminate between stimuli. On the other hand, it is possible that there is an interaction between aging and resolution of harmonics in pitch perception: age-related differences in discrimination thresholds might be greater for unresolved compared to resolved harmonics.

Results among younger listeners show that the ability to discrimination TEP cues becomes increasingly poorer as the modulation frequency increases (Formby, 1985; Hanna, 1992). It is hypothesized that the age-related difference in TEP processing is dependent on frequency of the stimulus. Neurophysiologic data show that the band pass characteristics of IC neurons change with age, and in particular, neurons of older mice typically exhibit lower band pass functions compared to those of younger mice (Walton et al. 2002). Electrophysiologic data in humans also support senescent decline of periodicity coding for frequencies greater than 100 Hz (Grose et al., 2009; Purcell et al., 2004). While such evidence is rather tenuous, it does provide support for a possible interaction between age and frequency in TEP processing. It is predicted that age-related differences in TEP processing will be greater for a higher frequency stimulus (i.e. 200 Hz), compared to those observed using a lower frequency stimulus (i.e., 50 or 100 Hz). Lastly, it is important to consider how listeners are able to perceive periodicity cues when such information must be gathered across different frequency bands, as in CI processing. Although it is also of interest to determine the effect of aging on across-frequency periodicity processing, it is difficult to make any predictions regarding this interaction.

Experiment 1: Part 1

Method

Participants

Participants were NH males and females, recruited for placement in two different groups based on their age at the time of testing: younger (ages 21-30; Mean= 23.1, SD=3.21) or older (ages 60-71; Mean=65.6, SD=4.18). The YNH group consisted of ten individuals (six males, four females) and the ONH group consisted of nine individuals (four males, five females). All participants were required to have pure-tone thresholds ≤ 20 dB HL from 250-4000 Hz in the test ear (ANSI, 2004). Table I shows the average audiometric data for each age group, along with the corresponding standard deviations at each frequency tested. All participants in the older age group obtained scores within the normal range on the Mini Mental State Examination (MMSE) (Folstein et al., 1975), a screening test used to identify gross cognitive dysfunction.

Stimuli and procedure

Stimuli were 200 ms, 100% SAM broadband white noise bursts (10 ms rise/fall time). Previous results demonstrate that pitch discrimination using resolved and resolved harmonics reaches asymptotic performance with a stimulus duration of 200 ms (Plack & Carlyon, 1995; White & Plack, 1998; 2003). Signals were generated online (22,100 Hz sampling rate) and the experiment was controlled using the experimental software program *psycon* v. 1.14. Stimuli were output through an external soundcard (Edirol 25-UAEX) and mixer (RaneSM26B), before being delivered through a calibrated circumaural headphone (Sennheiser HDA 200; Frequency response = 20-20,000 Hz). All stimuli were delivered monaurally at 65 dBA in the better-hearing ear. When both ears were deemed equally sensitive, the right ear was used. All testing was conducted in a double-walled soundproof booth and was completed in 3-4, two hour sessions (6-8 hours total).

Participants were tested using a 3-AFC task (method of constant stimuli), in which two of the intervals (random) contained a standard modulation frequency of 50,

Experiment 1-Part 1 Audiometric pure-tone thresholds (dB HL)										
		Frequency (Hz)								
		250	500	1000	2000	3000	4000	6000	8000	PTA
YNH	AVG	9.5	7.5	5	9.5	10	4.5	8.75	6	7.33
	SD	3.68	4.24	3.33	3.68	5.34	4.97	5.17	5.67	2.74
ONH	AVG	13.8	11.6	13.8	11.1	16.6	15	17.7	22.2	11.6
	SD	5.46	4.33	3.63	4.16	3.53	5	3.63	11.2	2.86

Table I. Average audiometric air-conduction pure-tone thresholds (dB HL) at each frequency tested and pure-tone average (PTA=average thresholds at 500, 1000 and 2000 Hz) for each age group. Standard deviations are shown below each respective average value.

100 or 200 Hz, while a third random interval contained the experimental modulation frequency (always greater than the standard frequency). The computer interface displayed button boxes (labeled “1”, “2”, and “3”) simultaneously with each interval in sequence. Using a mouse, listeners were asked to click on the box that contained the “different” sound (experimental modulation frequency).

Each interval contained refreshed noise, so listeners were encouraged to perceive differences in pitch evoked by changes in the modulation frequency itself. For a given run, the reference modulation frequency was held constant, while the experimental modulation frequency was randomly varied from trial to trial. The exact values for the experimental modulation frequency depended on the acuity of each listener and therefore varied across the participants, but listeners were exposed to a wide range of modulation frequencies within a given run. Within each run, one condition was always equal to the reference modulation frequency (50, 100 or 200 Hz), while 6-7 conditions contained an experimental modulation frequency. Each condition within a run was repeated ten times for a total of 70-80 total presentations within a given run. Each listener was provided with one practice run of each standard modulation frequency condition (50, 100 and 200 Hz); during these practice trials, listeners were told if their response was “correct” or “incorrect” via text that appeared on the computer screen so they were able to better understand and learn the task. Feedback was not provided during the test trials. Several listeners (particularly those that were older) found the feedback to be more distracting and it was thought that perhaps this variable might influence performance in an unpredictable manner. In total, 30-80 repetitions were collected for each experimental modulation value.

Results

In order to quantify the performance of each listener, average data for each participant was fitted in Sigmaplot (Version 9.0) using the following equation:

$$f = y0 + \frac{a - y0}{1 + e^{-\frac{x - x0}{b}}}$$

,where $y0 =$ chance performance (0.333), $a - y0$ = asymptotic performance, $x =$ value of independent variable, $x0 =$ threshold (experimental modulation frequency at which performance reaches half of $a - y0$, and $b =$ slope factor (the higher the value of b, the shallower the slope of the psychometric function). From each curve-fit, a threshold and slope were obtained. For each curve-fit reported here, R-squared values ranged between 0.768 and 0.978. The threshold is equal to the point along the abscissa corresponding to the halfway point along the ordinate (between chance performance and 100% correct performance). To better compare threshold data across the three conditions (standard modulation frequency), thresholds (in Hz) were converted into percent Weber fractions (Δ modulation frequency/standard modulation frequency*100).

Psychometric functions obtained using the 50, 100 and 200 Hz standard modulation frequencies are shown in Figures 3a, b, and c (top, middle and bottom panels, respectively). The experimental modulation frequency (Weber fraction) is shown along the abscissa, while the proportion of correct responses is shown along the ordinate. Individual curve fits are shown for each participant, with solid lines/symbols representing data obtained from YNH listeners, and dashed lines/open symbols representing data obtained from ONH listeners. For the 200 Hz condition (Figure 3c), a curve could not be fitted to the data obtained with one ONH listener (shown as a horizontal dotted line on the graph). In order to compare performance across all three conditions, percent Weber

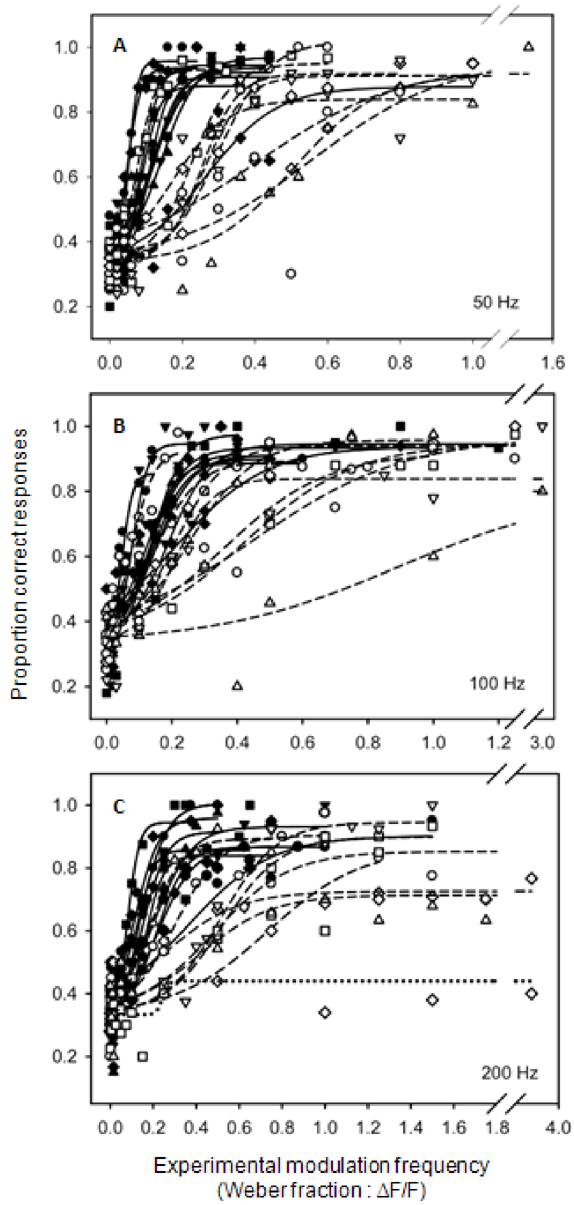

Figure 3. Individual curves fitted using sigmoid equation to derive individual threshold and slope values for each subject, in each condition. The results for reference modulation frequencies of 50, 100, and 200 Hz, are shown in the top (A), middle (B) and bottom (C) panels, respectively. The filled symbols and solid lines represent data obtained among the younger subjects, while the open symbols and dashed lines represent data obtained among the older subjects. The dotted line in panel C (200 Hz) represents data for one ONH subject whose performance could not be reliability fitted with a function.

fractions for each group and condition are shown in Figure 4. The condition (reference modulation frequency) is shown along the abscissa and the thresholds are shown along the ordinate. The filled circles and solid lines represent average data from the YNH group, whereas open squares and dashed lines represent data from the ONH group. Average thresholds and slopes for each group are also provided in Table II.

Effect of age and frequency on discrimination thresholds.

Data were analyzed using SPSS (Version 16.0). Overall, results show that the average performance of the younger listeners exceeded that of the older listeners for each condition (50, 100 or 200 Hz). A split-plot ANOVA was used to test for differences in MFD thresholds (percent Weber fractions) across the two groups (between factor), for each standard modulation frequency (within factor). The assumption of sphericity was violated, and therefore the Greenhouse-Geisser correction was applied to interpret the results. Results showed a significant main effect of standard frequency [$F(1.13, 18.00) = 10.42$, $p<0.01$), and a significant main effect of age [$F(1,16) = 6.48$, $p<0.01$). Contrary to the hypothesis, there was no significant frequency x age interaction. Follow-up testing for the significant main effect of frequency was performed using paired-sample t-tests (with Bonferroni correction) and revealed that performance on the 200 Hz condition was significantly worse than that of the 50 Hz ($t= -2.67$, $p<0.016$) and 100 Hz ($t= -3.82$, $p<0.01$) conditions.

Effect of age and frequency on estimated slope of the performance function

A second split-plot ANOVA was performed to test for differences between the estimated average slope factor of each curve for each standard modulation frequency

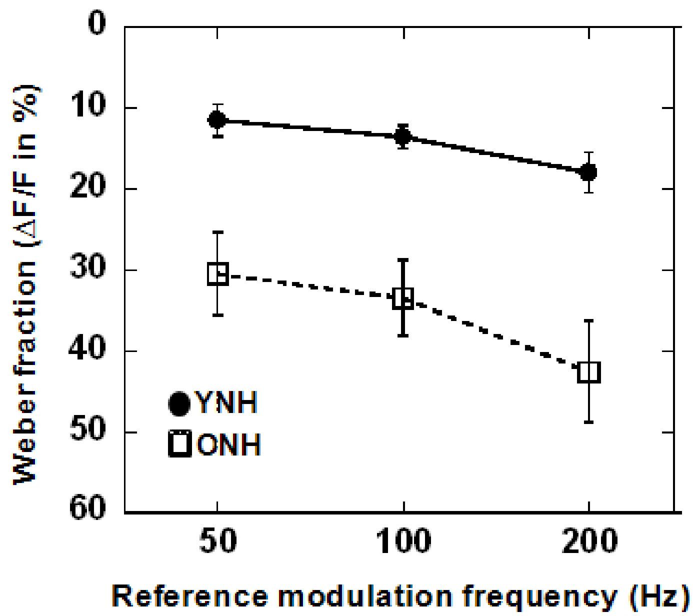

Figure 4. This figure shows the discrimination thresholds as estimated by the fitted curves (ordinate) for each group at each reference modulation frequency (abscissa). The discrimination thresholds are shown in percent Weber fractions (ΔFrequency/Frequency*100) with respect to each reference frequency. The results of the younger listeners are shown using filled circles and solid lines, whereas the results of the older listeners are shown using open squares and dashed lines. Error bars represent ± 1SE from the mean threshold value.

Experiment 1-Part 1 Audiometric pure-tone thresholds (dB HL)										
		Frequency (Hz)								
		250	500	1000	2000	3000	4000	6000	8000	PTA
YNH	AVG	9.5	7.5	5	9.5	10	4.5	8.75	6	7.33
	SD	3.68	4.24	3.33	3.68	5.34	4.97	5.17	5.67	2.74
ONH	AVG	13.8	11.6	13.8	11.1	16.6	15	17.7	22.2	11.6
	SD	5.46	4.33	3.63	4.16	3.53	5	3.63	11.2	2.86

Table II. Average values of the estimated discrimination thresholds (*x0*) shown in percent Weber fractions (ΔFrequency/Frequency*100) and slope factor (*b*), for each age group and reference modulation frequency. A larger slope factor (*b*) suggests a shallower slope. Standard deviations for each average are shown below in parentheses.

(within factor), across the two age groups (between factor). The assumption of sphericity was violated and the Greenhouse-Geisser correction was applied to interpret the results. Results indicate a significant main effect of frequency [$F(1.071,17.14)= 31.531$, $p<0.01$], a significant main effect of age [$F(1, 16)= 6.667$, $p<0.05$], and a significant frequency x age interaction [$F(1.07,17.14) = 4.828$, $p<0.05$). This interaction suggests that estimated slopes became more disparate across groups with increasing reference frequency. For example, greater between-group differences were found between the estimate slope values for the 200 Hz condition compared to the 50 Hz and 100 Hz condition ($t=3.57$, $p=0.016$). In other words, the estimated performance slopes were shallower for the older listeners, particularly for the 200 Hz condition.

Relationship between hearing sensitivity and discrimination thresholds

Stimuli were created with a sampling rate of 22, 100 Hz, and consequently the spectral bandwidth contained information up to approximately 11, 050 Hz. It is therefore possible that between-group differences in performance could have been attributed to disparate high-frequency audiometric thresholds. Thus, Pearson correlation coefficients (two-tailed) were calculated to determine the extent to which Weber fraction thresholds for each reference frequency (50, 100 and 200 Hz) were related to high-frequency pure-tone-averages (HFPTAs; average pure-tone thresholds at 4000, 6000, and 8000 Hz in the test ear) across the two age groups.

Results of the correlation analyses are shown as scatterplots in Figures 5a, b and c , for reference modulation frequencies of 50, 100, and 200 Hz, respectively. The abscissa shows the HFPTA value (in dBHL), while the ordinate shows the discrimination

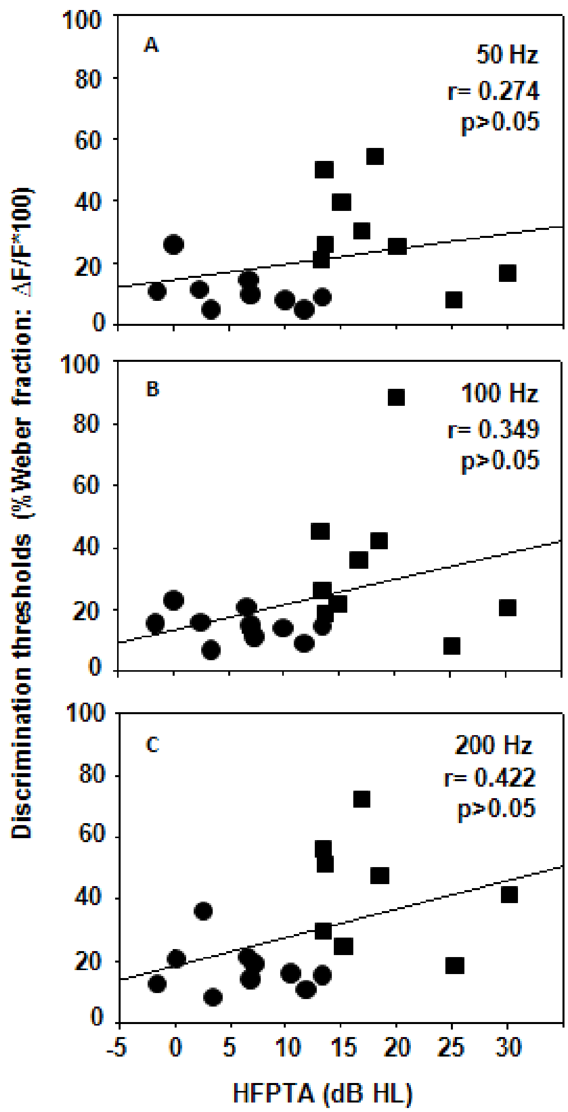

Figure 5. Scatterplots showing the results of the correlation analyses examining the relationship between high-frequency pure-tone averages (HFPTA in dB HL) and modulation frequency discrimination thresholds. Results for reference modulation frequencies of 50, 100 and 200 Hz are shown in the top (A), middle (B) and bottom (C) figures, respectively. Filled circles represent data from the younger listeners, and the filled squares represent data from the older listeners. The correlation coefficient (*r*) and significance values (*p*) for each condition are also provided within each respective plot.

threshold in percent Weber fraction. While there does appear to be some trend between HFPTA and discrimination thresholds, results of the correlation analyses failed to reach statistical significance in any of the conditions. It should also be noted that further correlation analyses were performed within the ONH group alone to determine if there was a significant relationship between discrimination performance and HFPTA; however, no such relationships were present. In fact, results showed a negative slope (better performance with higher HFPTA) which is contradictory to the trend one might expect. Taken together, these results support age-related differences in MFD that are not significantly related to hearing sensitivity.

Experiment 1 – Part 2

Rationale

The results obtained from Experiment 1-Part 1 provide support for declines in processing of TEP with advanced age, and in particular, provide absolute thresholds for the discrimination of TEP in older listeners. Despite the fact that correlation analyses did not demonstrate a significant relationship between MFD thresholds and peripheral hearing sensitivity, it does not mean that between-group differences in hearing sensitivity did not contribute to performance observed on the MFD task (Experiment 1-Part 1), at least to some extent. In the second part of Experiment 1 (Experiment 1-Part 2) we re-tested MFD thresholds on a separate group of participants to determine if between-group differences in high-frequency audiometric thresholds contributed to the performance shown in Experiment 1-Part 1. For this experiment, we used SAM noise low-pass filtered at 4000 Hz so stimuli did not contain high-frequency spectral information. These results will elucidate the effect of aging on within-channel TEP processing while minimizing contributions of peripheral hearing sensitivity.

Method

Participants

Participants were NH males and females, recruited for placement in two different groups based on their age at the time of testing: younger (ages 21-26; Mean= 21.8, SD=2.04) or older (ages 60-77; Mean=66.2, SD=6.25). The YNH group consisted of six individuals (two males, four females) and the ONH group consisted of seven individuals (one male, six females). All participants were required to have pure-tone thresholds≤ 20 dB HL from 250-4000 Hz in the test ear (ANSI, 2004). Table III shows the average audiometric data for each age group, along with the corresponding standard deviations at each frequency tested. All participants in the older age group obtained scores within the normal range on the Mini Mental State Examination (MMSE) (Folstein et al., 1975), a screening test used to identify gross cognitive dysfunction.

Stimuli

Stimuli were 300 ms SAM broad-band white noise bursts (100% modulation depth). As previously mentioned, this duration should produce asymptotic performance (Plack & Carlyon, 1995; White & Plack, 1998; 2003). Each stimulus was created online (44,100 Hz sampling rate) by multiplying low-pass filtered (4,000 Hz cut-off frequency, Chebychev filter, 40 dB/octave) white noise with a SAM envelope of a specified frequency (dependent on the condition). The signal was then low-pass filtered again (Chebychev filter, 80 dB/octave) at 4000 Hz to eliminate all possible high-frequency spectral components introduced by the modulation, and all stimuli were normalized in RMS value to equate loudness. Similar to the signal processing performed in Part 1 of this experiment, stimuli were generated using refreshed noise in each interval. Reference modulation frequencies were 100 or 200 Hz, while comparison modulation frequencies were always higher in frequency. All stimuli were presented at 65 dBA.

Procedure

A two-down, one-up, adaptive 3-AFC procedure was used to measure MFD thresholds (approximating 70.7% correct; Levitt, 1971), in which two of the intervals

Experiment 1-Part 2 Audiometric pure-tone thresholds (dB HL)								
		Frequency (Hz)						
		250	500	1000	2000	3000	4000	PTA
YNH	AVG	8.3	8.3	5.0	6.6	10.0	15	6.6
	SD	6.8	4.1	5.4	6.0	5.0	3.7	3.8
ONH	AVG	16.4	11.4	10.7	10	15	14.2	10.7
	SD	6.2	4.7	3.4	4.0	5.4	5.3	2.5

Table III. Average audiometric air-conduction pure-tone thresholds (dB HL) at each frequency tested and pure-tone average (PTA=average thresholds at 500, 1000 and 2000 Hz) for each age group. Standard deviations are shown below each respective average value.

(random) contained a reference stimulus with modulation frequency value equal to 100 or 200 Hz, while a third random interval contained the experimental value (always greater than the reference frequency). This adaptive procedure differed from the method of constant stimuli employed in Part 1. Although adaptive techniques are more efficient, and therefore commonly preferred over those using a method of stimuli, threshold calculations based on adaptive procedures assume a monotonic performance function. Therefore, it is always prudent to assess the shape of the psychometric function prior to implementing an adaptive procedure. The results from Part 1 allow us to use an adaptive procedure to estimate performance thresholds in this experiment.

The modulation frequency of the experimental stimulus was adapted until a minimum of eight reversals was reached. Adaptive step sizes varied depending on the condition and listener's sensitivity, however step sizes were either 4 and 2 Hz, 2 and 1 Hz, or 1 and 0.5 Hz. In general, smaller step sizes were used for a lower standard value (i.e.,100 Hz) whereas larger step sizes were required for a higher reference value (i.e., 200 Hz). The mean was calculated from the last four reversals of a run, and that value was taken as the MFD threshold. A maximum of 10 reversals or 55 trials was allowed, at which point testing would cease and no mean was calculated for that run.

Participants were tested in a double-walled, sound-attenuating booth using a computer interface. The computer interface displayed boxes (labeled "1", "2", and "3") which appeared sequentially, but simultaneous with the corresponding reference or experimental stimuli. Using a mouse, listeners were asked to click on the box that contained the "different" sound (experimental modulation frequency).The inter-stimulus-interval was 400 ms, and there was no time limit in which the subject was asked to respond. After the subject made a selection, the next sequence of sounds was played 600 ms later.

Subjects first received practice before being tested; that is, one run of every condition was provided as practice and individuals received feedback about their response in the form of text that appeared at the top of the screen (“Correct” or “Incorrect”). Practice runs were presented in a completely random order. After completing the practice run, three complete cycles of conditions were presented in a random order. Feedback was not provided during the test trials. Several listeners (particularly those that were older) found the feedback to be more distracting and it was thought that perhaps this variable might influence performance in an unpredictable manner. These three runs were used to calculate the final mean performance. In cases when the difference between the thresholds of the two means was greater than 15%, another run was performed, and then the average of all four runs was calculated as the final mean.

Results

Results for Experiment 1- Part 2 are shown in Figure 6. The reference modulation frequency is plotted along the abscissa, while the MFD thresholds are plotted along the ordinate in percent Weber fraction (ΔFrequency/Frequency * 100). Average MFD thresholds for the YNH listeners were 9.3% and 11.2%, and the average MFD thresholds for the ONH listeners were 24.1% and 42.7% for reference frequencies of 100 and 200 Hz, respectively. A split-plot ANOVA was performed with one within-group variable [reference modulation frequency (‘frequency’)] and one between-group variable (‘age’). Analysis of the data revealed a significant main effect of frequency [$F(1,11) = 21.02$, p<0.01], but the significant main effect of age was not significant [$F(1,11) = 4.76$, $p=0.52$]. These results are not in agreement with those obtained in Part 1 of this experiment which demonstrated that MFD performance was worse in older compared to younger listeners. The discrepancy between data could be attributed to fact that stimuli in Part1 contained higher frequency spectral information which could have cued better performance among the YNH group. Conversely, it is also possible that the small number of participants in the current experiment did not provide sufficient power to reach statistical significance. Taken together, the results of Experiment 1 do argue for age-related declines in TEP processing. While the results of Experiment 1-Part 1 suggest that hearing sensitivity was not related to performance, the results of Experiment 1- Part 2 argue that between-group differences in high-frequency hearing sensitivity might have contributed to the results. In conclusion, while these results *suggest* that aging effects the ability to discriminate TEP information, definitive conclusions cannot be drawn from the results of Experiment 1.

Summary and Discussion

The results of the current experiments do offer some support for senescent declines in the processing of TEP pitch coding by listeners with normal audiograms. Collective results, however, remain rather inconclusive given the conflicting contributions of peripheral hearing sensitivity observed in Part 1 and Part 2. For example, the results of Experiment 1-Part 1 (broad-band noise) showed that at standard frequencies of 100 and 200 Hz, younger listeners exhibited percent Weber fraction

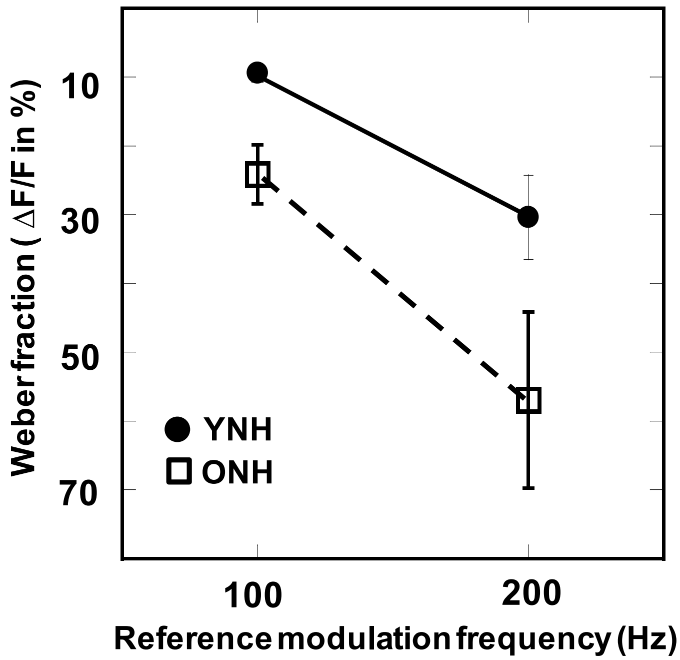

Figure 6. Results of the modulation frequency discrimination (MFD) task in Experiment 1- Part 2. The reference modulation rate is shown along the abscissa, while the discrimination thresholds are shown along the ordinate and expressed in percent Weber fractions (ΔFrequency/Frequency*100). Error bars represent ±1 SE from the mean value.

thresholds of 13% and 17%, respectively. In comparison the ONH listeners exhibited average percent Weber fraction thresholds of 33% and 42%, respectively. Furthermore, the performance slopes estimated among the ONH cohort exhibited a significantly shallower function for all conditions when compared to those estimated in the YNH group and the age x frequency interactions reveals that the age-related differences in slopes are greater at 200 Hz. These results suggest greater internal noise among ONH listeners when discriminating modulation frequencies particularly at higher frequencies ($\geq$ 200 Hz). While there was no significant interaction between age group and discrimination threshold as a function of standard frequency, the analyses of the estimated slopes do support the hypotheses that age-related differences in the perception of TEP pitch depend on the frequency of the modulation. Even though most studies have cited age-related differences in performance using frequencies 100 Hz or greater (Grose et al., 2009), it is possible that using a reference modulation frequency >200 Hz would have evoked appreciable differences in performance across the two age groups.

While correlation analyses in Experiment 1- Part 1 suggest that peripheral hearing sensitivity was *not* related to performance on the MFD task, the results of Experiment 1- Part 2 indicate that age-related differences in high-frequency hearing sensitivity could have indeed contributed to the between-group performance differences obtained in Part 1. Conversely, it is possible that the limited number of participants and large variance in data observed in Part 2 did not produce sufficient statistical power to result in significant differences in MFD performance between the two age groups. Altogether, the individual contributions of age and hearing sensitivity on MFD performance remain uncertain.

While a previous study by Moore and Peters (1992) provided cursory evidence that *f0*DLs using higher harmonics were larger in older, compared to younger, subjects with near NH it is unknown if audible distortion products influenced the results (Pressnitzer & Patterson, 2001). Further, more recent studies have suggested that, in fact, the first ten harmonics can be resolved depending on the *f0* of the complex (Bernstein & Oxenham, 2003). Moore and Peters (1992) used harmonic complex stimuli to measure *f0* discrimination of unresolved harmonics, but only removed the first six harmonics in each complex. Consequently, the results of Moore and Peters (1992) are not definitive, and it is possible listeners could have benefited from spectral cues.

Contrary to the method employed by Moore and Peters (1992), the current investigation used a stimulus that unequivocally did not convey spectral cues. Therefore, it can be safely assumed that discrimination of stimuli was possible by using only those pitch cues available in the temporal envelope which more closely predict cues available to CI users. In fact, the results of the current study would suggest that those obtained by Moore and Peters (1992) underestimated the effects of age on the processing of TEP. In the current study (Experiment 1- Part 1), we found that the MFD thresholds were approximately 30% and 42% for standard rates of 50 and 200 Hz, respectively. Conversely, Moore and Peters (1992) reported *f0*DLs of approximately 10% and 2% for identical reference rates, respectively. It should also be noted that our results were worse for YNH listeners as well, when compared to their results. Such discrepancies are likely attributed to differences in the stimuli between the studies (see discussion in the previous paragraph).

On the whole, these results do lend some support for age-related declines in TEP coding, which corroborate previous findings obtained using electrophysiologic and neurophysiologic measures (Grose et al., 2009; Leigh-Paffenroth & Fowler, 2006; Walton et al. 2002). It is also possible, however, that disparate high-frequency thresholds between the two age groups contributed to the results. It is widely-believed that CI users rely on TEP cues to perceive voice-pitch information, a cue that is crucial to everyday speech perception. While several studies have clearly demonstrated the efficacy of cochlear implantation among older listeners, the results of the current study argue that older CI users *may* exhibit greater difficulty hearing *f0* cues compared to younger CI recipients.

Experiment 2

Rationale

The results from Experiment 1 are thought provoking as they suggest that within-channel (one-channel) TEP coding might be altered with aging. It is possible, however, that between-group differences in hearing sensitivity could have contributed to the results observed in Experiment 1. Therefore, in Experiment 2 we used stimuli that were carefully low-pass filtered to ensure that stimuli did not contain content outside the range of NH for the older group of listeners. Consequently, the results of this experiment should reveal the contributions of aging alone on TEP processing.

Furthermore, those with CIs typically receive up to 8-10 channels of spectral information (Fishman et al., 1997; Friesen et al., 2001). In this case, the modulated envelope is preserved in CIs but current processing strategies result in non-synchronous amplitude modulation phase information across the electrodes. Studies show that, while

CI users are able to gather across-frequency (channel) envelope cues, conflicting phase cues can affect the salience of TEP information (Chatterjee & Oba, 2004).

In fact, recent efforts have been directed at developing compensatory speech processing strategies designed to more accurately code or enhance *f0* information (Geurts & Wouters, 2004; Green, Faulkner, & Rosen, 2004; Vandali, Sucher, Tsang, McKay, Chew, & McDermott, 2005). For example, Vandali and colleagues (Vandali et al., 2005) demonstrated improved *f0* perception using processing strategies designed to enchance TEP cues by increasing the amplitude depth and synchrony of temporal envelope modulation across multiple electrodes. Collectively, these results suggest that processing of pitch via envelope modulation frequency is most salient when information is coincident across frequency channels. Therefore, the results obtained in Experiment 1, a case in which amplitude modulation was conveyed across a broad-frequency band (within-channel condition), may represent 'optimal' coding of TE information in the auditory system. While the results of Experiment 1 demonstrate age-related changes in within-channel TE processing, it is unknown if this magnitude of change would differ for across-channel TE processing.

Experiment 2 measured *f0*DLs in younger and older NH listeners, using unprocessed and NBV harmonic complexes, with a low-pass cut-off frequency of 4000 Hz. The NBV stimuli contained 1, 8, and 24 channels, and reference *f0* values of 100 and 200 Hz were used to approximate the voice-pitches for a male and female talker, respectively. Similar to the algorithms used in CIs, multi-channel NBV processing results in incongruous temporal envelope information across each spectral channel. Therefore, it is anticipated that the results of Experiment 2 will better approximate the ability of younger and older CI listeners to combine TEP cues across multiple channels.

Method

Participants

Participants were NH males and females, recruited for placement in two different groups based on their age at the time of testing: younger (ages 21-26; Mean= 21.9, SD=1.62) or older (ages 60-77; Mean=64.7, SD=5.43). The YNH group consisted of twelve individuals (seven males, five females) and the ONH group consisted of thirteen individuals (three males, ten females). All participants were required to have pure-tone thresholds≤ 20 dB HL from 250 -4000 Hz in the test ear (ANSI, 2004). See Table IV for audiometric data. All participants in the older age group obtained scores within the normal range on the Mini Mental State Examination (MMSE) (Folstein et al., 1975), a screening test used to identify gross cognitive dysfunction.

Stimuli

All signals were created online (44,100 Hz sampling rate) and delivered through a custom graphical user interface developed in Matlab. Stimuli were NBV harmonic complexes, 300 ms in duration. Signals were created by first generating equal-amplitude harmonics between 100-4000 Hz. The *f0* and number of harmonics in each signal varied depending upon the condition, however the highest harmonic was selected so that each stimulus did not contain information above 4000 Hz. For each signal, the *f0* and

Experiment 2 Audiometric pure-tone thresholds (dB HL)									
		Frequency (Hz)							
		250	500	1000	1500	2000	3000	4000	PTA
YNH	AVG	8.3	8.3	7.1	6.0	7.1	9.5	5	7.5
	SD	6.5	4.4	5.8	3.2	4.9	5.2	3.6	4.1
ONH	AVG	13.8	11.5	11.5	10.8	11.9	15	13.0	11.6
	SD	5.8	3.7	3.1	5.1	4.8	4.7	5.2	2.5

Table IV. Average audiometric air-conduction pure-tone thresholds (dB HL) at each frequency tested and pure-tone average (PTA=average thresholds at 500, 1000 and 2000 Hz) for each age group. Standard deviations are shown below each respective average value.

depending upon the condition, however the highest harmonic was selected so that each stimulus did not contain information above 4000 Hz. For each signal, the *f0* and associated harmonics were summed to create a harmonic complex series which was subsequently windowed (Tukey window). Noise-band vocoding was similar to the process described earlier in this paper and in Shannon (1995). Each harmonic series was first band-pass filtered into 1, 8 or 24 channels (Chebychev, 40 dB/octave) depending on the condition, and the unaltered temporal envelope was then extracted from each channel using a Hilbert transform. The division frequencies used for the band-pass filtering were determined by using the logarithmic equation provided in Greenwood (1990) which estimates filter characteristics in the cochlea. Specific values of the division frequencies and bandwidths for each filter are shown in Table V. The TFS within each filter was replaced with a randomly generated noise-band equal in frequency content (Chebychev, 40 dB/octave) and multiplied with the corresponding temporal envelope. All band-pass filters were summed, low-pass filtered at 4000 Hz (Chebychev, 80 dB/oct) and then windowed (Tukey window) to eliminate any unwanted frequency components that could arise from processing. Lastly, all stimuli were equalized in RMS value before presentation.

Procedure.

Procedures were similar to those used in Experiment 1-Part 2 (adaptive MFD task). A two-down, one-up, 3-AFC procedure was used to measure *f0* discrimination (threshold = 70.7%; Levitt, 1971), in which two of the intervals (random) contained a reference stimulus with *f0*

Division frequencies for noise-band vocoding (Hz)						
	Number of channels					
Channel number	24			8		
	Min	Max	BW	Min	Max	BW
1	100	131	31	100	204	104
2	131	165	34	204	352	148
3	165	204	39	352	563	211
4	204	248	44	563	863	300
5	248	297	49	863	1291	428
6	297	352	55	1291	1900	609
7	352	414	62	1900	2766	866
8	414	484	70	2766	4000	1234
9	484	563	79			
10	563	652	89			
11	652	751	99			
12	751	863	112			
13	863	989	126			
14	989	1131	142			
15	1131	1291	160			
16	1291	1470	179			
17	1470	1672	202			
18	1672	1900	228			
19	1900	2155	255			
20	2155	2443	288			
21	2443	2766	323			
22	2766	3130	364			
23	3130	3539	409			
24	3539	4000	461			

Table V. Division frequencies used for noise-band vocoding of harmonic complex stimuli in Experiment 2. The minimum (Min), maximum (Max) and bandwidth (BW) frequency values are given for each channel number in the 24- and 8-channel conditions. The 1-channel condition is not shown.

value equal to 100 or 200 Hz, while a third random interval contained the experimental value (always greater than the reference *f0*). The *f0* value of the experimental stimulus was adapted until a minimum of eight reversals was reached. Adaptive step sizes varied depending on the condition and listener's sensitivity, however step sizes were either 4 and 2 Hz, 2 and 1 Hz, or 1 and 0.5 Hz. In general, smaller step sizes were used for a lower reference value (i.e., 100 Hz) and conditions in which stimuli contained more spectral cues (e.g., unprocessed or 24 channels), whereas larger step sizes were required for a higher reference value (i.e., 200 Hz) and conditions in which stimuli contained fewer spectral cues (e.g., 1 or 4 channels). The mean was calculated from the last four reversals of a run, and that value was taken as the *f0*DL. A maximum of 10 reversals or 55 trials was allowed, at which point testing would cease and no mean was calculated for that run.

Participants were tested in a double-walled, sound-proof booth using a computer interface. The computer interface displayed boxes (labeled "1", "2", and "3") which appeared sequentially, but simultaneous with the corresponding reference or experimental stimuli. Using a mouse, listeners were asked to click on the box that contained the "different" sound (experiment *f0*).The inter-stimulus-interval was 400 ms, and there was no time limit in which the subject was asked to respond. After the subject made a selection, the next sequence of sounds was played 600 ms later.

For the *f0*DL task, subjects first received practice before being tested; that is, one run of every condition was provided as practice and individuals received feedback about their response in the form of text that appeared at the top of the screen ("Correct" or "Incorrect"). Practice runs were presented in a completely random order. After completing the practice run, second and third complete cycles of conditions were presented in a random order. Feedback was not provided during the test trials. Several listeners (particularly those that were older) found the feedback to be more distracting and it was thought that perhaps this variable might influence performance in an unpredictable manner. These two runs were used to calculate the final mean performance. In cases when the difference between the thresholds of the two means was greater than 15%, another run was performed, and then the average of all three runs was calculated as the final mean.

Results

Results on the *f0*DL task are shown in Figure 7, with the 100 and 200 Hz data shown on the top and bottom, respectively. Within each graph, the abscissa represents the degree of spectral degradation/number of channels (U=unprocessed harmonic complex), whereas the

ordinate represents the discrimination threshold in percent Weber fraction (Δ *f0*/reference *f0**100). The filled circles represent data obtained in the YNH listeners, whereas the open squares represent data obtained in the ONH listeners. Exact threshold values are also provided in Table VI.

Data were analyzed using SPSS (version 16.0). A split-plot ANOVA was performed to analyze the data obtained for the *f0*DL task with two within-group factors [number of channels ('channels') and reference *f0* value ('*f0*')], and one between-group factor [Age group ('age')]. The Greenhouse-Geisser correction was used for interpretation of the results when the assumption of sphericity was violated. Analyses

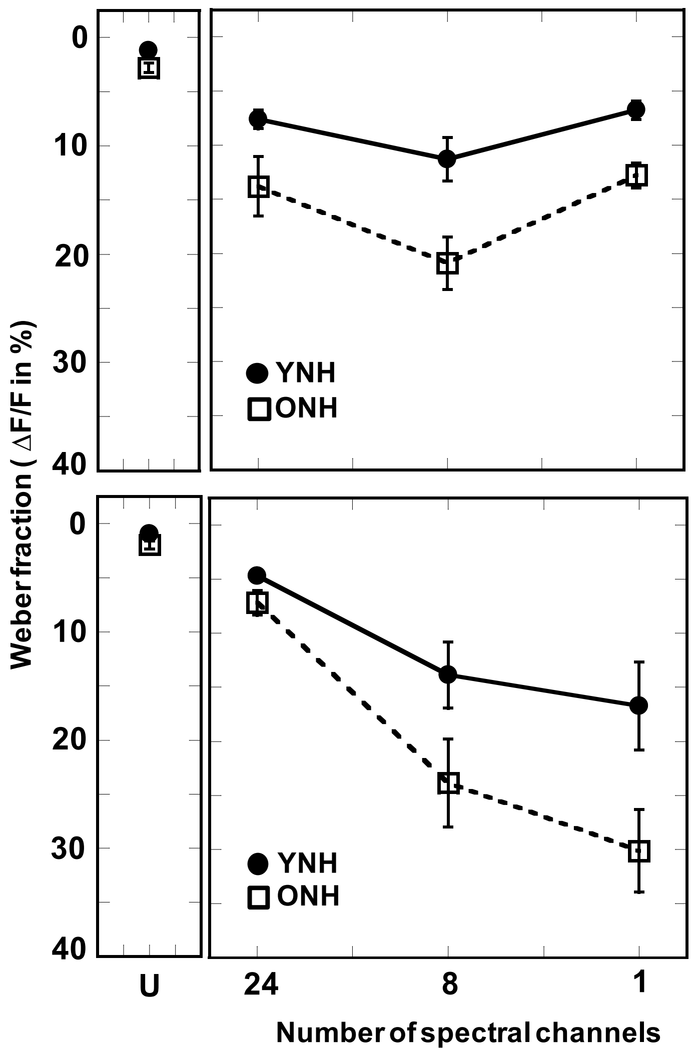

Figure 7. Results of the fundamental frequency (*f0*) discrimination task, with the top and bottom graphs representing data obtained with reference *f0* values of 100 and 200 Hz, respectively. Discrimination thresholds are shown along the ordinate and expressed as percent Weber fractions (ΔFrequency/Frequency*100). The abscissa shows the number of spectral channels made available for each condition. The far left panel in each graph ("U") represents data using unprocessed stimuli. The error bars represent ± 1SE from the average value.

Experiment 2 Fundamental frequency discrimination limens (f0DLs) in percent Weber fractions (Δf0/f0*100)									
		100 Hz				200 Hz			
		U	24	8	1	U	24	8	1
YNH	AVG	1.17	7.61	11.3	6.77	0.93	4.71	13.8	16.7
	SD	0.7	2.9	6.6	2.8	0.53	0.73	10.1	13.4
ONH	AVG	2.79	13.7	20.8	12.7	1.9	7.19	23.8	30.1
	SD	1.6	9.9	8.6	4.2	1.3	4.1	14.7	13.7

Table VI. Absolute values for the average performance on the fundamental frequency (*f0*) discrimination task for each group. Results are shown for each group, at each reference frequency and test condition (number of channels). Those results labeled "Un" represented data for the conditions in which stimuli were unprocessed. Values are expressed as percent Weber fractions (ΔFrequency/Frequency*100). Standard deviations (SD) are also shown.

revealed a significant main effect of channels [F(1.45,33.48)=58.38, p<0.001], a significant main effect of *f0*[F(1,23)=21.06, p<0.001], and a significant main effect of age [F(1,23)=11.26, p<0.01]. Analyses also indicated a significant interaction between channels and age [F(1.45, 33.48)=4.86, p<0.05] and a significant interaction between channels and *f0* [F(1.67, 38.44)=10.94, p<0.001]. In order to parse the overall results, paired or independent t-tests were used whenever appropriate and the Bonferroni correction was applied appropriately to interpret the results.

Follow-up testing for the significant interaction of channel and age showed there was a significant difference between age groups for the 1-channel (t=-2.77, p<0.01), 8-channel (t=-3.51, p<0.01), and unprocessed (t=-3.87, p<0.01) conditions only. This interaction also suggests that, while older listeners exhibit larger *f0*DLs than younger listeners when spectral cues are unaltered ('unprocessed' condition) the age-related disparities in performance are greater when spectral cues are reduced, in most cases. This interaction is apparently demonstrated in Figure 7 and argue that senescent decline in periodicity coding is more severe for unresolved than resolved harmonics.

Follow-up testing for the significant channel and *f0* interaction showed that within the 100 Hz reference condition, performance in all degrees of spectral resolution was significantly different from one another (p<0.006), except when comparing the performance between 1-and 24-channels. Similarly, within the 200 Hz reference condition, performance in all degrees of spectral resolution was significantly different from one another (p<0.006) except when comparing the performance between 1- and 8-channel channels. Lastly, results showed that performance was worse in the 200 Hz compared to the 100 Hz conditions for the 1-channel (t= -5.226, p<0.01) condition, whereas performance was better for the 200 Hz condition for the 24-channel ($t= 3.41$, $p<0.01$) and Unprocessed ($t=2.901$, $p<0.01$) conditions. There was no significant difference between reference *f0* values for the 8-channel condition.

Taken together, results show that the performance of the ONH group was worse than that of the YNH group for nearly all conditions. These findings suggest impaired periodicity coding in older listeners, regardless of the degree to which spectral cues where degraded. While older listeners did exhibit slightly larger *f0*DLs than younger listeners when stimuli were unprocessed (full spectral cues, resolved harmonics), between-group differences were substantially larger for stimuli in which spectral cues were severely degraded (reduced spectral cues, unresolved harmonics). In other words, this interaction suggests that senescent decline in periodicity coding is greater when listeners are forced to rely on TEP information rather than resolved, spectral pitch information.

Summary and Discussion

General Summary

While the results of Experiment 1 suggested that older listeners were less proficient at discriminating within-channel envelope periodicity information, the results of Experiment 2 demonstrated that older listeners experience difficulty processing within- *and* across-channel envelope periodicity information. Furthermore, the between-group differences in TEP processing observed in Experiment 1 could have been influenced by differences in peripheral hearing sensitivity across the two groups; however, age-related differences still persisted in Experiment 2 when stimuli were low-pass filtered to confine spectral information within the limits of NH for both groups. Although it is possible that residual differences between lower-frequency (250-4000 Hz) hearing thresholds in both groups (see Table IV) could have influenced performance, it is unlikely given the supra-threshold presentation level (65 dBA).

Previous studies using non-speech (harmonic complexes) and synthetic speech stimuli (synthetic vowels) revealed that *f0*DLs among older listeners were ~2% whereas those of younger listeners were ~1% when harmonics can be resolved (Moore & Peters, 1992; Vongpaisal & Pichora-Fuller, 2007). The current study corroborates such findings, as average *f0*DLs for unprocessed harmonic complexes were 1.17% (100 Hz) and 0.93% (200 Hz) for YNH listeners. Conversely, average *f0*DLs for unprocessed harmonic complexes were 2.79% (100 Hz) and 1.9% (200 Hz) for ONH listeners. Contrary to the hypothesis, the effect of aging on periodicity coding was independent of modulation frequency as similar age-related deficits were found for 100 and 200 Hz reference stimuli.

The interaction between age and channels suggests that, in general, the performance of the two age groups becomes increasingly divergent as spectral cues are reduced. This pattern is particularly evident for the 200 Hz condition (See Figure 7). Statistical analyses revealed that (collapsed across both reference *f0* values), older listeners exhibited poorer (larger) *f0*DLs for the unprocessed, 1-channel, and 8-channel conditions, but not the 24-channel condition. In the case of the unprocessed condition, it is presumed that all listeners are using resolved harmonics to perceive and discriminate between *f0*s. Conversely, in the case of the 1-channel condition, it is presumed that all listeners are relying solely on TEP to perceive and discriminate between *f0*s. For those remaining conditions (i.e., 8- and 24-channel) one must consider how a given *f0* and its harmonics would be processed given the division frequencies used in the NBV processing (see Table V). By examining the division frequencies provided in Table V, it is logical to conclude that, in most cases, harmonics would remain unresolved for the 8-channel condition. It is then reasonable that age-related differences would be noted for both the 1- and 8- channel conditions. Conversely, it is logical to conclude that the 24-channel condition would result in a combination of resolved and unresolved harmonics. In this case, it is hypothesized that *f0* cues might be confusing to the listener, regardless of listener age. These results suggest an overall decline in periodicity coding with advanced age, for both resolved and unresolved harmonics.

The results of Experiment 2 indicate that older CI users might experience significantly greater difficulty perceiving voice-pitch information compared to younger CI recipients. As previously mentioned, modern CI devices convey approximately 8-10 spectral channels (Friesen et al., 2001; Fishman et al., 1997). Based on these estimates, the results of the present study suggest that older CI users would require approximately 20% difference in *f0* in order to detect a difference in voice-pitch between talkers. Younger CI users, on the other hand, require nearly half that value to obtain comparable performance (See Table VI). These results obtained in YNH listeners closely match those reported in previous studies (Qin & Oxenham, 2005). Older listeners exhibited greater difficulty perceiving *f0* information using primarily temporal cues for both 100 and 200 Hz reference stimuli, and therefore it could be expected that the perception of spectrally degraded male and female voices would be equally affected by aging. Overall, these results confirm previous findings that used neurophysiologic and electrophysiologic methods, and consequently lend further support for senescent decline in temporal envelope pitch processing.

Envelope Rise-time and Temporal Coding

Both Experiments 1 and 2 measured the ability of listeners to discriminate between modulation frequencies based only on differences in the periodicity of the envelope. For example, stimuli in the MFD task (Experiment 1) and *f0*DL 1-channel NBV task (Experiment 2) required that the listener utilize differences in the pitch period of the temporal envelope alone to correctly discriminate between stimuli. In both cases, spectral pitch cues were complete absent. Despite such similarities in the stimuli, observation of the data reveal seemingly better discrimination performance in both age groups for the 1-channel NBV task compared to the MFD task using SAM noise.

In order to demonstrate these differences, results from the MFD (Experiment 1- Part 2) and *f0*DL 1-channel NBV (Experiment 2- Part 1) tasks are shown for comparison in Figure 8. It should be noted that this figure only represents data from the MFD task in Experiment 1- Part 2. Each panel in Figure 8 represents data for each reference frequency (MFD task) or *f0* (*f0*DL task), for each age group. In each panel, the specific conditions are located along the abscissa, and the discrimination thresholds [percent Weber fraction (ΔFrequency/Frequency*100)] are shown along the ordinate. For each box plot, the gray shaded area represents the 25th percentile and 75th percentile, the vertical error bars represent the 10th percentile and the 90th percentile, and the individual black-filled circles represent outlier data points. The solid horizontal line and dashed horizontal line within the gray shaded box represent the median and mean values, respectively.

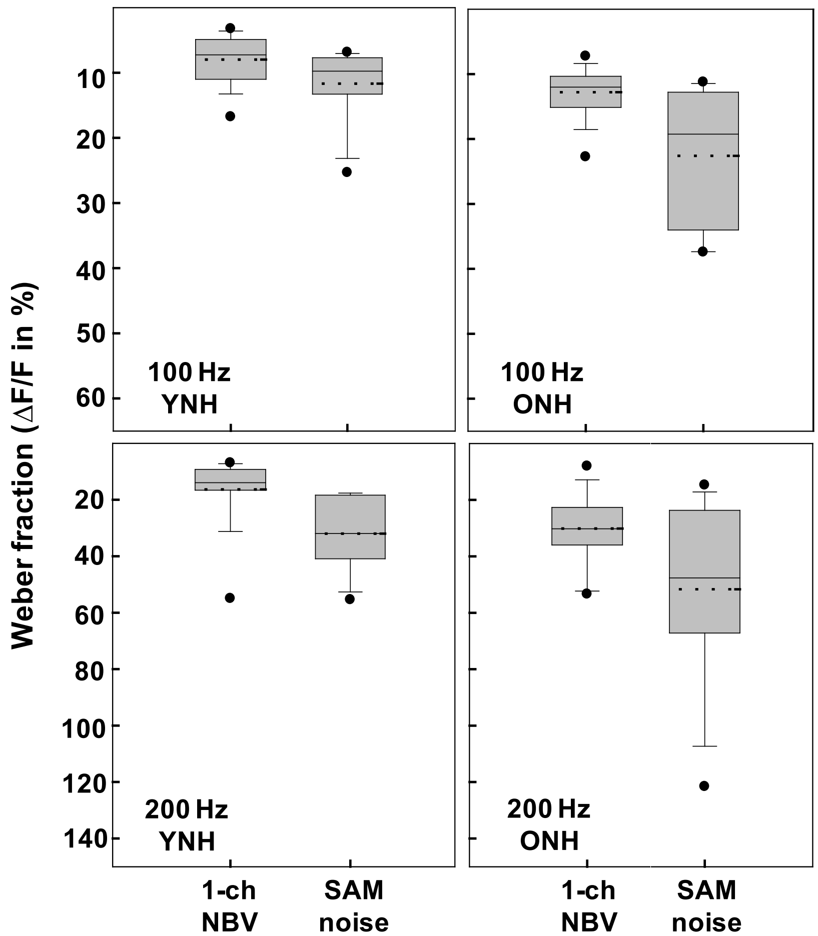

Figure 8. Box-plots representing discrimination thresholds for the 1-channel noise-band vocoded stimuli (1-channel NBV) and sinusoidally-amplitude modulated stimuli (SAM noise). Each plot shows data for reference frequencies of 100 and 200 Hz for each age group. The gray shaded area represents the 25th percentile and 75th percentile, the vertical error bars represent the 10th percentile and the 90th percentile, and the individual black-filled circles represent outlier data points. The solid horizontal line and dashed horizontal line within the gray shaded box represent the median and mean values, respectively. Note that different ordinate scales are used for the 100 and 200 Hz conditions.

As shown in the graph, there was significant within-group variability in performance especially for the ONH listeners on the MFD task. Moreover, this graph demonstrates that the performance of both groups was better with a 1-channel NBV rather than a SAM noise stimulus. To compare the results obtained of Experiment 2 (1-channel NBV *f0*DL) with those of Experiment 1- Part 2 (MFD), a two-way ANOVA was performed with two within-group factors [reference frequency ('frequency') and type of stimulus ('stimulus')]. Results showed a significant main effect of frequency [$F(1,13)=23.24$, $p<0.01$], and a significant main effect of stimulus [$F(1,13)= 15.34$, $p<0.01$]. There were no significant interactions. Overall, these results suggest that the 200 Hz condition was more difficult than the 100 Hz condition for both stimulus types. Further, the MFD task was more difficult than the 1-channel NBV *f0*DL task for each reference modulation frequency.

It is hypothesized that, although both stimulus types provide temporal pitch cues through the pitch period in the envelope, the auditory system processes these signals differently as a result of dissimilarities in the *shape* of the temporal envelope. The amplitude waveforms for the 100 Hz, 1-channel NBV stimuli (Experiment 2) and SAM noise stimuli (Experiment 1- Part 2) are respectively shown in Figures 9a and 9b. Similarly, results for the 200 Hz condition are shown in Figures 10a and 10b. Both represent a signal in which the envelope repeats every 0.01 seconds (100 Hz). It is also apparent, however, that the rise-fall time and peak-amplitude of the envelope differs across the two stimuli, with faster rise-fall times for the 1-channel NBV stimulus.

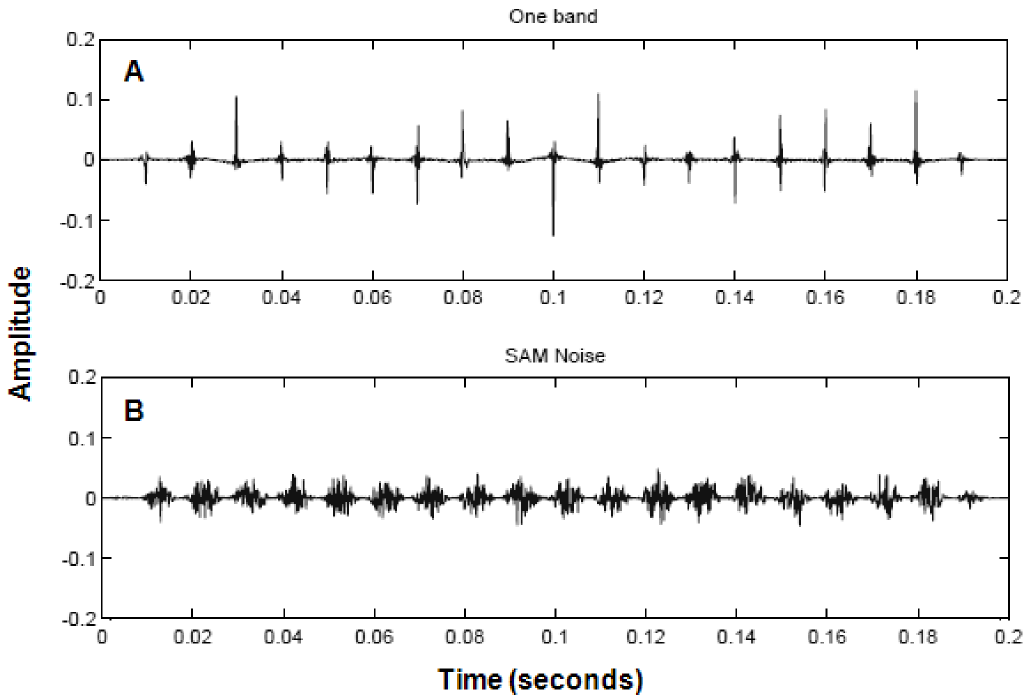

Figure 9. Amplitude waveforms for the 1-channel noise band vocoded stimulus with a fundamental frequency of 100 Hz (A) and a 100 Hz sinusoidally-amplitude modulated noise stimulus (B). The pitch period is equal for both stimuli, but the rise-time of the envelope (envelope shape) and peak-amplitude differ across the two stimuli.

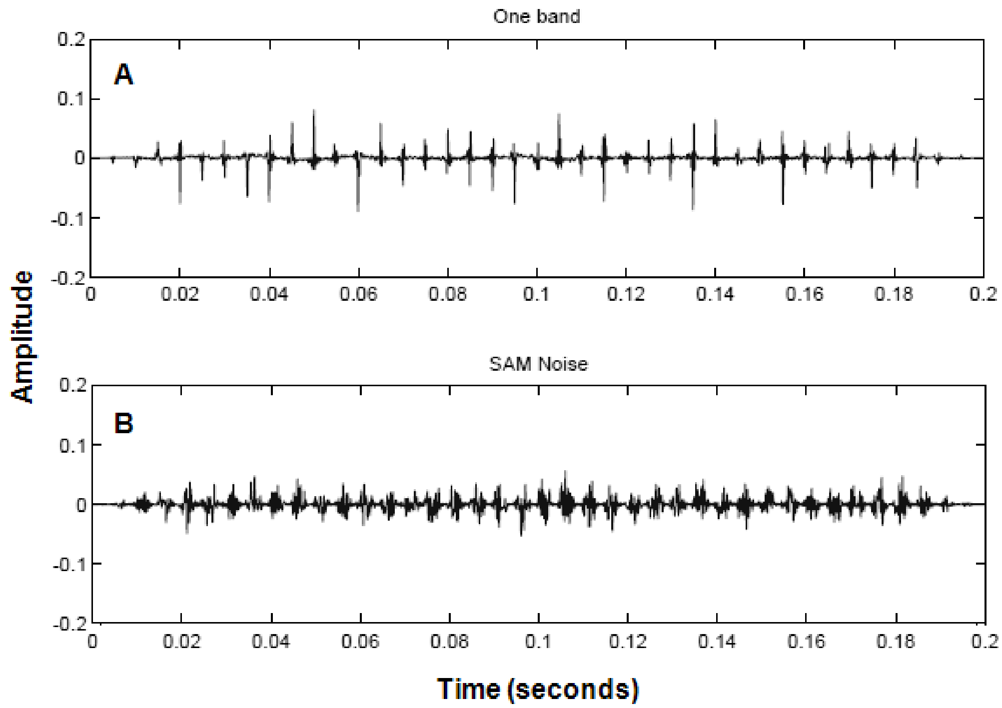

Figure 10. Amplitude waveforms for the 1-channel noise band vocoded stimulus with a fundamental frequency of 200 Hz (A) and a 200 Hz sinusoidally-amplitude modulated noise stimulus (B). The pitch period is equal for both stimuli, but the rise-time of the envelope (envelope shape) and peak-amplitude differ across the two stimuli.

Given this difference, the results of the current study are not altogether unexpected as a wealth of existing literature provides evidence that the precision of temporal coding in the CANS improves with rapid onsets (Burkard, 1991; Gooler & Feng, 1992; Heil, 2001; 2003; Suga, 1971). For example, results of neurophysiologic studies demonstrate higher spike counts and more rapid post-stimulus onset latencies among neurons when the envelope of the stimulus has a faster, rather than slower, rise-time (Gooler & Feng, 1992; Heil, 2001; 2003; Suga, 1971). Likewise, electrophysiologic measures show responses with faster latencies and higher amplitudes when stimuli with more rapid, rather than slower, onsets are presented to the listener (Burkard, 1991; Barth & Burkard, 1993). Taken together, these studies suggest that the auditory system is sensitive to the temporal characteristics of stimulus onset and/or rise-time of the envelope. Specifically, responses typically mirror the temporal properties of the stimulus itself: faster or slower rise-times result in decreased or increased response latencies, respectively. Therefore, the results of the current study suggest that discrimination of envelope frequency is improved when the rise-time of the envelope is rapid. The lack of interaction between age and stimulus suggests that the effect of envelope shape on temporal resolution does not differ as a function of adult age. Overall results, however, were poorer among older listeners.

It is difficult to predict how these results would translate to temporal envelope perception in CI users. For example, transmission of temporal envelope information could differ slightly depending on the processing strategy. Landsberger (2008) measured MFD thresholds in eight CI listeners (age unspecified) using direct stimulation methods while varying the shape of the modulation envelope. The reference stimulus was a 100

Hz pulse train modulated by a sine-wave, square-wave, saw-tooth, or sharpened saw-tooth envelope. Unexpectedly, there was no consistent effect of envelope shape on MFD performance, and in fact, the envelope modulation with the fastest rise time (square-wave) sometimes caused significant decrements in performance. Although the reason for such results is unclear, they do demonstrate, perhaps, that the effects of envelope shape on modulation coding are less important in electrical stimulation. Overall, the results of the current study further confirm the effect of envelope rise-time on temporal coding and provide cursory evidence that such effects are not altered in an aging auditory system, at least for acoustic stimuli.

CHAPTER 3: THE ROLE OF SPECTRAL AND TEMPORAL PERIODICITY CUES IN A VOICE-GENDER IDENTIFICATION TASK

Experiment 3

Rationale

In previous chapters we demonstrated interactions between aging and the perception of spectrally degraded voice-pitch information using non-speech stimuli. It is therefore reasonable to conjecture that older listeners might exhibit greater difficulty when asked to identify, or discriminate between, talker gender. Although voice-pitch, or *f0*, has proven to be an important cue for discerning the gender of a talker, several other cues also contribute to the total perception of voice gender (Klatt & Klatt, 1990; Murry & Singh, 1980). Futhermore, the *f0* of an individual's speech is not static, but rather changes both within and across utterances. Therefore, the purpose of this study was to determine if the age-related declines in degraded *f0* perception observed in Experiments 1 and 2 would persist when naturally uttered speech stimuli were used. Lastly, we also manipulated the amount of temporal envelope information that was available to the listener to determine how older listeners benefit from temporal envelope cues if spectral cues are degraded.

Literature Review

Acoustic cues for voice-pitch information

Several investigations have probed the anatomical, psychological, and production differences between male and female talkers to determine how individuals correctly identify talker-gender. The *f0* of a talker is mainly determined by individual differences in vocal fold length, thickness of vocal folds, as well as the size of the larynx (Titze, 1989). Altogether, these characteristics combine to result in the natural *f0* for a specific talker, with male talkers exhibiting lower *f0* values compared to females of the same age (Childers & Wu, 1991; Peterson & Barney, 1952). Although notable vocal differences are present between talkers of different genders (i.e., breathiness, effort, nasality, and hoarseness) (Klatt & Klatt, 1990; Murry & Singh, 1980), *f0* commonly prevails as the most important cue for assigning a gender to speech stimuli. It is also important to note that these studies also reveal that formant frequency information can also be used to convey information about talker-gender (Klatt & Klatt, 1990; Murry & Singh, 1980).

Gender-identification with spectrally-degraded cues

As previously mentioned, CI listeners are especially susceptible to confusions in gender identification, a deficit which can be attributed to a reduction in spectral resolution. Several studies show that gender recognition is difficult for CI users presumably because they perceive weaker *f0* cues compared to a NH individual (Fu et al., 2004; 2005; Gonzalez & Oliver, 2005; Vongphoe & Zeng, 2005). It is well-known that information about gender-talker is also available using formant structure cues, and in CI users such cues would be available via spectral peaks. Studies have shown, however, that spectral-peak resolution is poor among CI users (Henry & Turner, 2003; Henry, Turner, & Behrens, 2005). In fact, recent evidence suggests that such spectral envelope cues do not strongly promote differentiation of talker-gender in CI users, at least in children using CIs (Kovacic & Balaban, 2009). Rather, it is presumed that CI users gather information about talker-gender using f0 information conveyed in the temporal envelope (Fu et al., 2004; 2005; Kovacic & Balaban, 2009).

Effect of spectral and temporal envelope cues. Fu et al. (2004) measured gender identification in six younger NH subjects (ages 22-30) and 11 adult CI users (ages 39-70). Voice gender discrimination was measured using a closed-set, 2-AFC task. Stimuli consisted of 12 vowels (in a /h/-vowel-/d/ context), varying in talker gender (12 talkers total, six male/six female). Listeners were instructed to identify the talker as a male or female, by using a computer interface and selecting the appropriate button (labeled "male" or "female"). For NH participants, the degree of spectral resolution was varied by providing 4, 8, 16, or 32 spectral channels (Sine-wave vocoding, or SWV). Furthermore, the upper-limit of the temporal envelope was manipulated thereby allowing access to varying degrees of temporal envelope cues; envelope filter cutoff frequencies were 20, 40, 80, 160 and 320 Hz). CI users listened to unprocessed speech via their everyday speech processors.

Overall, results showed a significant effect of spectral degradation on gender identification. Performance was near perfect with 32 channels, however listeners exhibited systematic declines in performance as the number of channels was reduced. Furthermore, performance was strongly influenced by the cut-off frequency of the temporal envelope: Performance improved when the temporal envelope cut-off frequency was higher (ie: 320 Hz), compared that of a lower frequency (ie: 20 Hz). This was particularly true as spectral cues were degraded, suggesting a significant interaction between spectral and temporal cues. Performance of the CI listeners ranged from approximately 70-95%.

Effect of talker-f0 similarity on gender identification. A subsequent study, also conducted by Fu and colleagues (Fu et al. 2005), delved deeper into the spectral and temporal cues involved in gender identification. In particular, two talker sets were used for the study. For Talker-set 1, the *f0*s of the male and female talkers differed by 100 Hz, and in Talker-set 2, *f0* differences between males and females were only 10 Hz. A task identical to that used in the previous investigation (Fu et al. 2004) was implemented for this study, and performance was measured in ten younger NH listeners (ages 22-38) and ten CI users (ages 35-74). Overall, results were similar to those in the previous study (Fu et al, 2004) and suggest that limiting both spectral and temporal cues greatly affected gender identification abilities. Futhermore, the contribution of temporal envelope cues depended on the talker set (1 v. 2). For example, the results using Talker Set 1 (100 Hz difference between the average *f0s* for male and female talkers) were similar to those results obtained in the previous study: when spectral resolution was adequate, the degree of temporal resolution became less important. However, as the number of channels was decreased (ie: 4 channels) listeners required more temporal information to maintain accuracy. This pattern was not maintained for Talker Set 2 (10 Hz difference between male and female talker). In this case, there was a general decline in performance as the number of channels was increased, but unlike those results obtained from Talker Set 1, the addition of temporal envelope cues did not aid the listener when identifying gender. These results suggest that while listeners are able to utilize temporal envelope cues to perceive large disparities in two *f0*s, small differences in temporal envelope cues cannot be adequately conveyed using TEP cues. These results agree with those obtained by Vongphoe and Zeng (2005) who found that CI listeners (and those listening to CI simulations) experience difficulty utilizing temporal envelope cues when required to discriminate between small differences in *f0* of talkers.

A note on sine-wave vocoding. One important detail regarding the above studies (Fu et al., 2004; 2005; Vongphoe & Zeng, 2005) is the fact that a SWV method was employed instead of a NBV method. Both methods are fairly similar, with one exception: after the original signal is divided into a specified number of frequency bands, the TFS is replaced with *a sine-wave carrier* equal to the center frequency of each band. Then, the originally extracted temporal envelope is used to modulate the sine-wave in each band. This method is sometimes used as an alternative to the more popular NBV method, however often results in better performance on several tasks (Dorman, Loizou, and Rainey, 1997; Dorman, Loizou, Fitzke, &Tu, 1998; Faulkner, Rosen & Wilkinson, 2001; Gonzalez & Oliver, 2005), particularly when performance on a task was dependent upon *f0* coding.

For example, Gonzalez and Oliver (2005) measured gender identification in 15 NH listeners (ages 21-30), when listening to either SWV or NBV stimuli. The original stimulus was a Spanish sentence ("How old is your cousin from Barcelona?), recorded by 40 native speakers of Spanish (20 males, 20 females). All stimuli were then manipulated to contain 3- 16 channels, using sine-wave and noise-band vocoding methods. The cut-off frequency of the temporal envelope was maintained at 400 Hz for sine-wave carriers, and 160 Hz for noise-band carriers. Overall, results showed far better performance with the SWV stimuli compared to the noise-band vocoded stimuli. Performance for both types of processors was nearly 100% with 16 spectral channels. However, for three channel stimuli the performance of listeners significantly declined to 65% accuracy for the NBV stimuli, while performance was maintained at 90% for the SWV stimuli. The differences in performance can be readily explained, as a modulation of a sine-wave generates side bands which convey information about the spectral content of the signal. Listeners are able to use the information contained in these side bands as a spectral cue. For this reason, sine-wave carriers are not the best choice when attempting to predict the performance of CI users' ability to perceive *f0* cues, as CI listeners do not have access to such spectral cues. Nonetheless, these studies (Fu et al., 2004; 2005; Gonzalez & Oliver, 2005; Vongphoe & Zeng, 2005) do show that gender and talker identification are difficult tasks among younger listeners when spectral cues are severely limited.

Potential interactions with aging. Despite such findings in younger listeners, the influence of spectral degradation on gender identification among older listeners has yet to be investigated. It is obvious based on the data in the aforementioned studies (Fu et al, 2004; 2005; Gonzalez & Oliver, 2005; Vongphoe & Zeng, 2005) that temporal envelope cues play a substantial role in gender identification, particularly when spectral cues are severely limited. Based on evidence from electrophysiologic and neurophysiologic data (Grose et al., 2009; Leigh-Paffenroth & Fowler, 2006; Purcell et al., 2004; Walton et at. 2002), as well as those results shown in Chapter 2, there is strong evidence that older listeners experience greater difficulty perceiving periodicity cues (using non-speech stimuli) when spectral cues are degraded or absent altogether. Therefore, it might be expected that this deficit exhibited by older listeners would hold true when speech stimuli are used; however, it could also be the case that other cues used to differentiate gender-talker would be sufficient to compensate for the loss of *f0* information. Although there has been some research focused on age-related memory differences in the ability of listeners to recognize a familiar talker (Yonan & Sommers, 2000), the ability of older listeners to discriminate between talker genders has yet to be investigated.

The aim of the third experiment is to quantify the effects of age on gender identification, when spectral and/or temporal cues are degraded. A gender identification task was measured in younger and older NH adults, using stimuli that have been processed using NBV to contain varying amounts of spectral and temporal cues. Stimuli were manipulated to contain 1, 4, 8, 24, or 32 channels. The cut-off frequency of the low-pass filter used to extract the temporal envelope was manipulated so the stimuli varied in the extent of temporal cues (20-400 Hz). Lastly, *f0* values for each talker were systematically manipulated to determine how this variable affects the performance of both younger and older listeners.

Questions and Hypotheses

1. Does the ability to identify talker gender differ as a function of age when spectral cues are reduced or absent?
2. Does the ability to identify talker-gender differ as a function of age when temporal cues are reduced or absent?
3. If the ability to identify talker gender does differ as a function of age, how does the range or variance in voice-pitch between talkers affect performance?

Chapter 2 demonstrated that aging affects the ability to discern spectrally degraded *f0* cues when non-speech stimuli were used. It is hypothesized that such deficits will also be observed using speech stimuli. Differences in *f0* (voice pitch) have been shown to be the most salient cue for determining the gender of a talker (Klatt & Klatt, 1990; Murry & Singh, 1980). Therefore, based on the results of Chapter 2, it is hypothesized that older listeners will experience greater difficulty identifying talker gender when spectral cues are reduced. Previous studies have shown that increasing temporal envelope information improves recognition of talker gender by younger listeners when spectral cues are severely reduced (Fu et al., 2004; 2005). It is hypothesized that age-related deficits in temporal envelope processing will affect the ability of older listeners to use such additional cues, even when they are made available to the listener. Lastly, previous studies in younger listeners show that manipulating the range of voice pitches in a talker set so that they are more similar causes a reduction in the ability to identify talker gender under conditions of spectral degradation (Fu et al., 2005). Although the exact outcome is difficult to predict for older listeners, it is possible that this reduction in performance would be greater for older listeners. That is, older listeners need larger differences in the voice-pitch of talkers to determine talker gender when spectral and/or temporal cues are reduced.

Method

Participants

Ten younger (two males and eight females ages 21-28; Mean=24.3; SD=2.62) and ten older (Six males and four females ages 60-72; Mean= 66.8; SD=4.31) NH listeners participated in the current investigation. For the purposes of this study, NH was defined as air-conduction audiometric thresholds ≤ 20 dB HL from 250 -4000 Hz in the test ear (ANSI, 2004). Average audiometric thresholds and corresponding standard deviations for each group are provided in Table VII. These values reflect data obtained in the test ear only. All participants reported to be in good health and denied a significant history of audiologic or otologic disease. All participants in the older age group obtained scores within the normal range on the Mini Mental State Examination (MMSE) (Folstein et al., 1975), a screening test used to identify gross cognitive dysfunction.

Stimuli

Stimuli were six naturally produced and digitally recorded (sampling rate =22,050 Hz) vowel tokens (/h/vowel/d/ context, Hillenbrand et al., 1995): /æ/, /i/, /oʊ/, /ɒ/, /eɪ/, /u/. The duration of each naturally-uttered vowel differed as was not manipulated for the purposes of this study. Each of the six recorded vowels was spoken by four male and four female talkers, for a total of 48 vowels (6 vowels x 8 talkers). Vowels were selected from a larger corpus consisting of 144 vowels, originally recorded at the House Ear Institute (HEI) and were available within the Computer-Assisted-Research-Training (CAST) software developed by Qian-Jie Fu at HEI.

The *f0* values were analyzed and manipulated using the autocorrelation method (Boersma, 1993) available in Praat (version 5.0.27) (Boersma & Weenink, 2008). Specifically, the average *f0* value of each token was calculated then shifted to match a specified target (within 1 Hz) for a given condition and talker. Stimuli were carefully manipulated so that the target *f0* did not greatly differ from the natural *f0* of the talker. In most cases, stimuli were shifted by less than 20 Hz, and by no more than 40 Hz. The *f0* target values were selected to create two general sets of stimuli for the current investigation: 'Expanded' (EXP) and 'Compressed' (COM). For EXP, *f0* values ranged

Experiment 3 Audiometric pure-tone thresholds (dB HL)									
		Frequency (Hz)							
		250	500	1000	1500	2000	3000	4000	PTA
YNH	AVG	9.5	7	6.6	9.2	5.5	7.5	3.5	6.3
	SD	5.9	4.8	4.1	3.4	5.9	5.9	4.7	3.5
ONH	AVG	15	8.5	12	12	11	13.5	16.5	10.5
	SD	5.2	4.1	4.2	3.4	3.1	3.3	4.1	1.9

Table VII. Average audiometric air-conduction pure-tone thresholds (dB HL) at each frequency tested and pure-tone average (PTA=average thresholds at 500, 1000 and 2000 Hz) for each age group. Standard deviations are shown below each respective average value.

from 100 Hz (lowest *f0* of male voice) to 275 Hz (highest *f0* value of female voice). These values were selected based on pilot data and were meant to test listeners' perception of talker-gender when salience of pitch differences between talkers varied. In the case of the EXP condition, gender identification should be easier since *f0*s are more disparate across talkers. Conversely, in the case of the COM condition, gender identification should be more difficult as *f0*s will be more closely spaced across talkers. Further, for the COM condition, gender identification will likely be more difficult as these target values may be very high or low compared to common *f0* values of each gender. For example, an *f0* of 180 Hz for a male talker is higher than normal.

Each manipulated token was stored on a computer hard drive and then uploaded into the CAST program at the time of testing (see below). Table VIII provides the target *f0* values for each talker in each condition. Based on these values, the average difference in *f0*s between the male and female talkers was 100 Hz and 40 Hz for the EXP and COM conditions, respectively. It is important to point out that while the average pitch was manipulated, the natural pitch contour remained unaltered.

Noise-band vocoding was accomplished online using TigerCIS within the CAST interface. This was done after each of the pitch manipulations described above. The NBV methods used in the current study were comparable to those described in Experiment 2 and in Shannon et al. (1995). Stimuli were first band-pass filtered into 1, 4, 8, 16, or 32 channels depending on the specific condition, using fourth-order Butterworth filters (24 dB/octave); all stimuli were filtered from 100-4000 Hz, and specific division frequencies for each case (*i.e.* number of channels) were determined based on the logarithmic

Experiment 3								
	Fundamental frequency (f0) values of each talker (Hz)							
	Male 1	Male 2	Male 3	Male 4	Female 1	Female 2	Female 3	Female 4
EXP	100	125	150	175	200	225	250	275
COM	150	160	170	180	190	200	210	220

Table VIII. Approximate fundamental frequency (*f0*) values (in Hz) for the male and female talkers used for the gender identification task in Experiment 3. Values are shown for both the Expanded (EXP) and Compressed (COM) conditions.

equation provided by Greenwood (1990). See Table IX for division frequency values. The temporal envelope was then extracted from each frequency band using half-wave rectification and low-pass filtering. The cut-off frequency of the low-pass filter was manipulated and was 20, 50, 100, 200 or 400 Hz, depending on the condition. The original TFS of each filter was then replaced with corresponding band-pass noise (fourth-order Butterworth filters), and the corresponding temporal envelope was used to modulate the noise bands. The modulated noise bands were then summed and then normalized in RMS value to result in the final NBV speech token.

Procedure

Each stimulus was presented in a single interval. Using a computer mouse and screen, the participant was instructed to select “start” in order to hear the first vowel token. Following the presentation, two rectangular boxes appeared on the computer screen: one labeled “Male” and the other “Female”. The subject was instructed to identify the gender of the speaker by selecting appropriate box; there was no time limit for responding. After making a selection the next stimulus was presented 1 second later. This continued until the end of a run, which consisted of 48 vowel tokens (1 repetition x 8 talkers x 6 vowels). Within each run, the 48 vowel tokens were presented in a completely random order. Participants were first presented with a random ordering of all possible conditions [each run was one of 20 conditions (EXP=4 channel conditions x 5 envelope conditions; COM= 5 channel conditions x 4 envelope conditions)] to serve as a practice. Feedback about the response (“Correct” or “Incorrect”) was provided via text on the screen following each response. This presentation cycle was repeated twice again and

Experiment 3 Division frequencies for noise-band vocoding (Hz)												
	Number of channels											
Channel number	32			16			8			4		
	Min	Max	BW	Min	Max	BW	Min	Max	BW	Min	Max	BW
1	100	123	23	100	147	47	100	204	104	100	352	252
2	123	147	24	147	204	57	204	352	148	352	863	511
3	147	174	27	204	272	68	352	563	211	863	1900	1037
4	174	204	30	272	352	80	563	863	300	1900	4000	2100
5	204	236	32	352	448	96	863	1291	428			
6	236	272	36	448	563	115	1291	1900	609			
7	272	310	38	563	700	137	1900	2766	866			
8	310	352	42	700	863	163	2766	4000	1234			
9	352	398	46	863	1058	195						
10	398	448	50	1058	1291	233						
11	448	503	55	1291	1568	277						
12	503	563	60	1568	1900	332						
13	563	629	66	1900	2295	395						
14	629	700	71	2295	2766	471						
15	700	778	78	2766	3329	563						
16	778	863	85	3329	4000	671						
17	863	957	94									
18	957	1058	101									
19	1058	1169	111									
20	1169	1291	122									
21	1291	1424	133									
22	1424	1568	144									
23	1568	1727	159									
24	1727	1900	173									
25	1900	2088	188									
26	2088	2295	207									
27	2295	2520	225									
28	2520	2766	246									
29	2766	3035	269									
30	3035	3329	294									
31	3329	3650	321									
32	3650	4000	350									

Table IX. Division frequencies used for noise-band vocoding of vowel stimuli in Experiment 3. The minimum (Min), maximum (Max) and bandwidth (BW) frequency values are given for each channel number in the 32, 16, 8, and 4 channel conditions. The 1-channel condition is not shown.

scored, and an average percent correct was calculated based on the final two runs of each condition. Feedback was not provided during the test trials. Several listeners (particularly those that were older) found the feedback to be more distracting and it was thought that perhaps this variable might influence performance in an unpredictable manner. In order to reduce the possibility of order effects, half of the listeners were first presented with stimuli from the EXP condition, whereas the other half of the listeners were first presented with stimuli from the COM condition. Prior to being tested using the degraded stimuli for each condition, listeners first listened to one run of unprocessed stimuli; all listeners were required to achieve 90% accuracy on this task before continuing on in the study. No subject was excluded from the study based on this criterion.

Stimuli were output through an external soundcard (Edirol 25-UAEX) and mixer (Rane SM26B), before being delivered monaurally through a calibrated circumaural headphone (Sennheiser HDA 200) at a level of 65 dBA. For most listeners, stimuli were presented to the right ear with the exception of those listeners whose right-ear audiometric thresholds did not meet the criteria for normal hearing (as defined under "Participants" above). In this case, stimuli were presented to the better-hearing (left) ear. Testing was performed in two-hour time blocks, with breaks provided at the listener's discretion. The total duration of the experiment varied from 8-10 hours (4-5 sessions).

Results

Prior to analysis, percent correct scores were converted to rationalized arcsine units (RAUs) (Studebaker, 1985). Results from EXP are shown in Figure 11 while results from COM are shown in Figure 12.

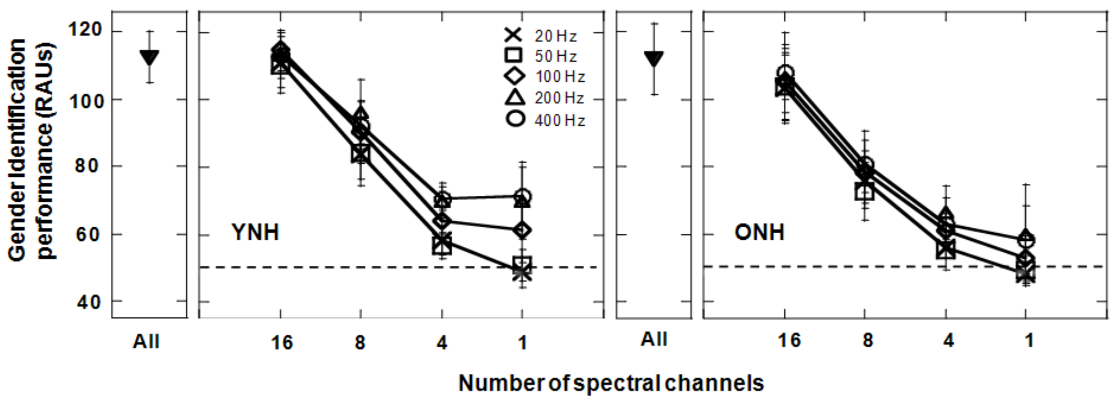

Figure 11. Results from the gender identification task for Expanded (EXP) condition, for each age group. The number of channels is located along the abscissa (U=unprocessed stimuli), while identification performance is located on the ordinate (in RAUs).The dashed horizontal line indicates chance performance (50%). Error bars represent ±1 SD from the mean.

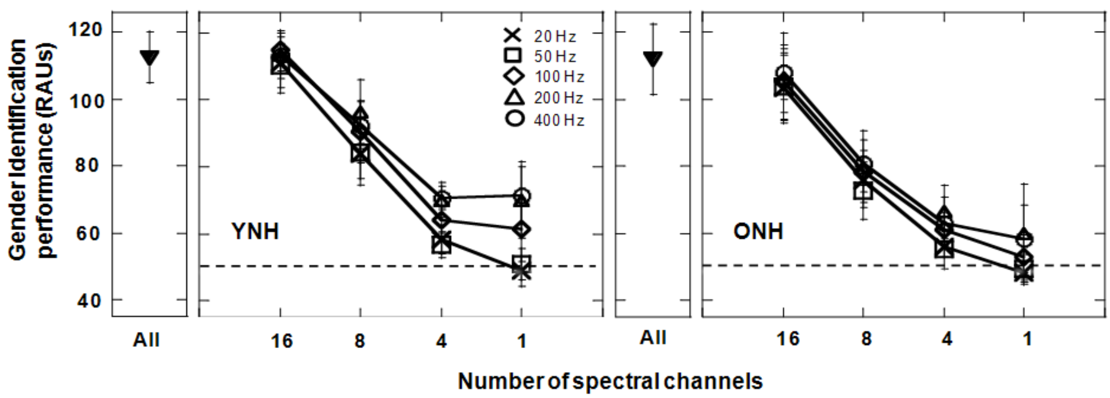

Figure 12. Results from the gender identification task for Compressed (COM) condition, for each age group. The number of channels is located along the abscissa (U=unprocessed stimuli), while identification performance is located on the ordinate (in RAUs).The dashed horizontal line indicates chance performance (50%). Error bars represent ±1 SD from the mean.

The number of channels is shown along the abscissa, while gender identification performance is shown along the ordinate (in RAUs). As expected, performance was better for EXP compared to that of COM. The between-group differences, however, varied within each condition. A split-plot ANOVA was performed within the EXP and COM conditions, to quantify the interactions between age group, as well as availability of spectral and temporal cues. Subject group (age) served as the between-group factor, whereas number of spectral channels (channels) and cut-off frequency of the low-pass temporal envelope filter (envelope) served as the within-group factors. For cases in which the assumption of sphericity was violated, the Greenhouse-Geisser correction was used to interpret the results.

For EXP, there was a significant main effect of channels [$F(3,54) = 441.84$, $p<0.01$] , a significant main effect of envelope [$F(2.10, 37.91) = 33.99$, $p<0.01$], and a significant interaction between channel and envelope [$F(5.13, 92.34)=5.33$, $p<0.01$]. Further, there was a significant main effect of age [$F(1,18) = 16.89$, $p<0.01$], as the average performance of the YNH group exceeded that of the ONH group. Follow-up testing (using multiple paired t-tests with Bonferroni corrections) was performed for the significant channel x envelope interaction.

Overall, the results from this analysis showed that there was trade-off between spectral and temporal cues; that is, as the spectral cues were reduced, temporal cues had a greater influence in performance and conversely as spectral cues are increased, temporal cues have a lesser influence on performance. For example, for the 1-channel condition, performance significantly improved by 16.33 RAUs, on average, from 20 to 400 Hz (t= -

5.94, p< 0.001), whereas there was no effect of temporal envelope in the 16-channel condition.

The ANOVA results for COM show that there was a significant main effect of channels [F (2.63,72) = 579.06, p<0.01), a significant main effect of envelope [F (3,54) = 10.63], and a significant main effect of age [F (1, 18) = 7.55, p<0.05]. There was also a significant age x envelope interaction [F (3,54) = 3.40, p<0.05). Follow-up tests (one-way ANOVA followed by paired t-test with Bonferroni correction) were used to further examine the interaction between age and envelope, and results revealed that was a significant main effect of envelope within the YNH group [F (3,147) = 13.9, p<0.01], but not within the ONH group. More specifically ,the performance of the YNH group improved by 3.88 – 5.84 RAUs, on average, as the temporal envelope cut-off frequency was increased from 50 to 200 Hz, from 50 to 400 Hz, and from 100-200 Hz (p<0.001) . Although these are not large increases in performance, they are significant, which is certainly meaningful given the closeness of *f0* values in the COM condition. Lastly, a two-way ANOVA showed no significant differences across conditions or between age groups when listeners performed the gender identification task using unprocessed stimuli in the EXP and COM conditions.

Discussion

In keeping with previous investigations (Fu et al. 2004; 2005), the listeners who participated in the present study demonstrated a systematic relationship between the utilization of temporal and spectral cues. Spectral cues undoubtedly provided better resolution when coding voice-pitch information, and temporal envelope cues were seemingly extraneous when listeners are provided with 16 or 32 spectral channels.

However, when spectral cues were absent, some listeners were able to utilize temporal envelope cues to perceive voice pitch, albeit to a lesser extent than when all spectral cues were available.

The results presented in the current study expand upon those reported in previous investigations measuring temporal envelope processing in aging adults. Previous neurophysiological, electrophysiological, and psychoacoustic studies have alluded to the presence of senescent decline in temporal envelope processing, while the current study demonstrates that such temporal processing deficits translate to voice pitch coding of natural speech when spectral cues are decreased or absent altogether. Overall, it appears that, while older listeners are able to utilize temporal envelope cues to perceive voice pitch in some cases, the resolution and discrimination of temporal envelope modulation frequency is worse than that of their YNH peers.

In the current study we implemented two main conditions: one in which the range of *f0* values was large (EXP), and the second in which the range of *f0* values was compressed (COM). Based on the results of EXP, older listeners were able to discriminate between voice gender even when the signal was severely spectrally degraded. In doing so, they are presumably coding voice pitch using differences in the modulation frequencies of the residual temporal envelope. Results from the COM condition, however, reveal that older listeners lack the resolution to discriminate between residual temporal envelope modulation frequencies when *f0* values (and consequently amplitude modulation frequencies in the speech envelope) are more similar. These findings corroborate those results reported in Experiments 1 and 2.

To explore the interaction between age, spectral cues and temporal cues in greater depth, the results shown in Figures 11 and 12 are re-plotted in Figure 13. This graph can be thought of as the "YNH advantage". Specifically, a positive bar indicates that the YNH group outperformed the ONH listeners. Figure 13a shows results from EXP, and Figure 13b shows results from COM. It is clear that the younger group outperformed the older group in every condition. The performances of the two groups, however, were more comparable under conditions of greater difficulty, as in COM. Although statistical analyses of a three-way interaction between age, envelope and channels did not reach statistical significance, the patterns shown in Figure 13 suggest a tendency toward such an interaction. Specifically, as spectral cues become compromised, the between-group difference grows as the YNH group utilizes more residual temporal envelope information. Although this inference is made with much caution, it is possible that the pattern is indeed true and the lack of statistical significance can be attributed to a small sample size and high variability. Minimally, Figure 13 serves to demonstrate the intricate and dissimilar response pattern obtained across the two age groups, and an apparent dependence of the results upon the relative weights of spectral and temporal cues.

Collectively, evidence suggests that CI listeners are able to utilize 8-10 spectral channels, at best (Friesen et al., 2001). Looking at the results for 8-channels in the current study, it appears that even younger CI users might experience difficulty coding voice-pitch information, as our YNH participants' performance was 92.53 and 79.42 RAUs for EXP and COM, respectively when all temporal envelope cues were available (*i.e.,*400 Hz cut-off). Older listeners, however, exhibited poorer performance in the 8-channel

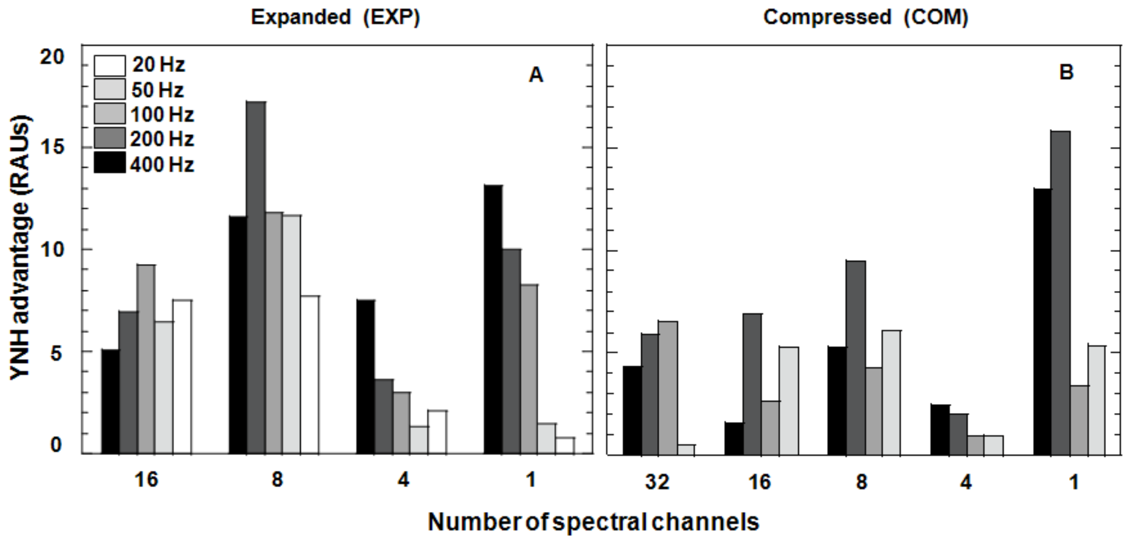

Figure 13. Bar graph representing the gender recognition performance of the younger listeners relative to that of the older listeners (YNH advantage). Results for the Expanded (EXP) and Compressed (COM) conditions are shown in A and B, respectively. The number of channels are shown on the abscissa and the relative performance of the YNH listeners is shown along the ordinate (in RAUs). A positive value (all cases) indicates better performance for the YNH group. The various shades represent different amounts of temporal envelope cues.

condition when compared to their YNH peers, as they achieved average scores of 80.96 and 74.14 RAUs for EXP and COM, respectively (i.e., 400 Hz cut-off). This comparison also exemplifies the reduced ability of the ONH listener to code not only temporal, but spectral cues as well. That is, even in this case when all temporal cues are made available, but spectral cues are still diminished (i.e., 8-channel, 400 Hz), older listeners did not perform as well as the younger listeners. The results of the current experiment substantiate those obtained in Experiments 1 and 2, and collective findings demonstrate changes in the processing of spectro-temporal pitch cues in older listeners. It is therefore reasonable to conjecture that older CI recipients might be at a disadvantage when identifying the gender of a target stimulus.

CHAPTER 4: EFFECTS OF AGING ON SENTENCE RECOGNITION WITH SPECTRALLY DEGRADED SPEECH

Experiment 4

Rationale

In previous chapters we demonstrated that aging affects one's ability to process *f0* information, particularly when stimuli are spectrally degraded. Nearly all conversations, however, take place in complex environments in which the target speech is not presented in isolation. It has been repeatedly shown that *f0* cues are important when segregating competing talkers (Brokx & Nooteboom, 1982; Brungart, 2001). Likewise, NH listeners also demonstrate 'release from masking' when target speech is masked by a modulated signal, as opposed to a steady-state signal (Brungart, 2001; Festen & Plomp, 1990). It appears, however, that even a slight diminution in spectral resolution can hinder one's ability to use either of these cues to aid in performance, and conversely, evidence has shown that a modulated masker may actually increase overall masking of the signal (Baer & Moore, 1993; 1994; ter Keurs, Festen & Plomp, 1992; 1993). While the effect of aging on speech recognition in noise has been previously studied to a great extent, the potential effect of aging on the ability to understand a spectrally degraded masked signal has yet to be investigated. The purpose of the current experiment was to determine the extent to which older listeners are able to understand speech in various types of interferers and to quantify the degree to which spectral degradation may interfere with their ability to do so.

Literature Review

Energetic and informational masking

There are very few occasions in our daily lives when we attempt to listen to a target speaker in complete quiet. We are constantly required to ignore other distracting talkers in order to correctly understand our target message, a task commonly referred to as 'simultaneous segregation'. Simultaneous segregation abilities are influenced by both *energetic* and *informational masking* (EM and IM, respectively). Energetic masking occurs if the acoustic energy in signal A overlaps with some or all of the acoustic energy in signal B, thereby reducing the perception or recognition of signal B. It is thought that EM largely reflects the purported "peripheral" masking at the level of the auditory filters in the cochlea: if the peripheral excitation patterns of signal A and B are similar or identical then EM will occur. The exact definition of IM is a matter of great debate (See Durlach, Mason, Kidd, Arbogast, Colburn, & Shinn-Cuningham, 2003 and Watson, 2005 for reviews), but generally includes masking that cannot be attributed to EM alone. Examples include stimulus uncertainty (Durlach et al., 2003; Watson & Kelly, 1981) or acoustic or linguistic similarity between the target and masker (linguistic uncertainty or masking) (Brungart, 2001; Brungart et al., 2001).

Both EM and IM can significantly affect one's ability to focus on a target sentence in the presence of a competing masker, but the degree to which EM and IM interfere with comprehension depend on the type of masker present (e.g., single talker, multi-talker, or noise). For example, speech masked by a steady-state noise will be subject to significant EM, but IM will not play a role. When speech is masked by a single-talker, EM is small but IM can be significant if semantics from the target and masker are confused (linguistic uncertainty or linguistic masking). As the number of talkers increase in a masker, EM also increases but IM decreases since the masker speech becomes incomprehensible. When the masker is a modulated noise, IM will not occur but listeners will be subject to some effects of EM. Although these general patterns are true in those with NH, it is also known that those with CIs are differentially affected by IM and EM components within a masker.

Effect of masker type on speech perception in normal-hearing listeners

A difference in voice pitch (*f0*) is one of the strongest cues that aids a listener in parsing two or more simultaneous stimuli and selectively attending to a single talker. Several studies have shown that in the healthy auditory system, even small discrepancies between the *f0*s of a target and competing signal (i.e., 2 or 3 semitones) can aid a listener in separating the two simultaneous signals (Brokx & Nooteboom, 1982; Darwin, Brungart & Simpson, 2003). For example, Darwin et al. (2003) systematically manipulated the *f0* of both a single target and single masker, and measured sentence identification using a closed-set task (Coordinated Response Measure, or CRM). Nine younger NH individuals listened to the target and masker sentences at varying target-to-masker ratios (TMRs). Overall, results showed that performance significantly improved when the target and masker *f0*s differed by as little as 2 semitones. Performance (collapsed across all TMRs) indicated that target word identification improved by almost 30% as the target and competing *f0s* were transitioned from identical values (0 semitones apart) to those one octave apart (12 semitones) from one another.

Several studies have demonstrated that NH listeners are actually able to hear better in the presence of another talker, versus when a non-speech, energetic steady-state masker is present (Brungart, 2001; Festen & Plomp, 1990). Although it might be intuitive to suspect that with a single-talker masker, IM would result in significant linguistic masking of the signal, these studies show that a competing talker produces less masking compared to a steady-state masker depending on the TMR (Brungart, 2001; Festen & Plomp, 1990). A single-talker masker provides less EM compared to a steady-state noise masker, and the temporal and modulations of the masker allow the listener to "glimpse" the signal between the peaks of noise in order to better hear the target stimulus. This is commonly referred to as a "release from masking" effect. Typically, the ability of younger NH listeners to correctly perceive speech is most affected when a steady-state noise masker is present; however, they benefit from release from masking when an AM noise is used. Performance further improves with the use of a different talker masker (different *f0* than target speech), as listeners are able to take advantage of both *f0* differences and temporal modulations (Brungart, 2001; Festen & Plomp, 1990). This is not to say that IM is completely absent when speech is masked by competing speech; for example, linguistic masking can occur if words are confused across the target and masker (Lew & Jerger, 1991; Tun, Kane & Wingfield, 2002).

Results from a classic study by Festen and Plomp (1990) exemplify the effect of masker type (noise versus speech) on sentence recognition abilities of younger NH listeners. Young NH listeners achieved an average TMR of -4.7 dB to obtain 50% accuracy when asked to identify a target sentence (male or female talker) in the presence of steady-state noise. Average performance among younger NH listeners improved with the use of amplitude modulated noise as the masker (-8.4 TMR), and improved further with the use of an opposite sex single-talker as the masker (reversed speech) (-11.4 TMR). These results reflect a clear pattern of performance related to the type of masker used as the interfering signal. Subsequent studies support these findings, suggesting that young NH listeners are able to obtain release from EM when a masker is modulated, particularly when the modulated masker is another talker different in *f0* than the target speech (Brungart, 2001; Darwin et al., 2003).

Although listeners exhibit significant release from masking when a single interfering talker is present, results are quite different when multiple talkers are present and performance varies depending on the number of multiple talkers used for the background masker. Miller (1947) was one of the first to report on how the number of background talkers affected speech recognition abilities among NH listeners. Listeners were asked to understand a male talker (conversation) in the presence of 1, 2, 4, 6, or 8 other male interferers. Results showed that recognition of the target stimulus was easiest in the presence of a single masker, decreased with two talkers, and was the most difficult with 4-8 talkers. Overall, the presence of 4-8 talkers increased (worsened) the speech-recognition-threshold (SRT) by approximately 12 dB compared to the SRT of a single-talker condition (see Figure 9, Miller, 1947). Simpson and Cooke (2005) also found that performance tends to be best with one talker, and worst when 4-9 talkers are used (maximum EM and/or IM). In this case, performance is worse than when a steady-state noise masker is used (Danhauer & Leppler, 1979). These results suggest that the effects of IM and EM are maximal when at least 8 talkers are present in the background.

Taken together, these results suggest that the speech recognition in noise performance of younger NH individuals is best when another competing talker serves as the background interferer, particularly when *f0*s differ across the target and masker stimuli. Conversely, performance is poorest when a steady-state noise masker or multiple talkers (at least 4-8) are present. Performance is commonly intermediate when a modulated noise is used.

Effect of masker type on speech understanding in hearing-impaired and cochlear implant listeners: A focus on spectral resolution

Although various maskers elicit different amounts speech masking in YNH listeners, the same pattern is not observed among those with hearing loss. Results typically show that listeners with HI need an overall higher TMR in order to obtain scores equivalent to NH listeners when attempting to listen to a target signal in noise (Dirks, Morgan, & Dubno, 1982; Festen & Plomp, 1990; Plomp & Mimpen, 1979). For example, the Festen and Plomp (1990) study previously mentioned (Festen & Plomp, 1990) also measured speech recognition in noise performance among 20 listeners with moderate sensorieneural hearing loss. Results showed that overall performance was worse for HI listeners, compared to those with NH. Unlike NH listeners, the HI cohort performed equally for all three masker conditions (steady-state noise, modulated noise, and opposite-sex single talker) and did not exhibit a release from masking effect. Subsequent studies lend support to their findings (Jin & Nelson, 2006; Summers & Molis, 2004), and further investigations point to declines in both temporal and spectral resolution as a cause for the decline in performance (Baer & Moore, 1993; 1994;

Houtgast & Festen, 2008; Jin & Nelson, 2006; Moore, 2008; ter Keurs et al., 1992; 1993).

A reduction in temporal resolution might affect one's ability to listen in the dips of the masker, but others have argued that adequate spectral resolution is essential when performing such a task (Baer & Moore, 1993; 1994; ter Keurs et al., 1992; 1993). This argument has been further substantiated by CI research. Multi-channel CIs afford most listeners with excellent speech recognition abilities in quiet, as most listeners are able to achieve nearly perfect identification of sentences when presented without a masker (Dorman et al., 1998; Friesen et al., 2001; Shannon et al., 1995; Shannon, Fu & Galvin, 2004). However, even at high (easy) TMRs, most CI listeners exhibit significant difficulty understanding speech in noise (Fetterman & Domico, 2002; Firszt, et al., 2004). Furthermore, unlike NH individuals who benefit from a modulated masker (noise or speech), the performance of CI listeners does not improve in modulated noise, and in fact, sometimes becomes worse compared to performance obtained when a steady-state masker is present (Cullington & Zeng, 2008; Fu & Nogaki, 2005; Nelson, Jin, Carney, & Nelson, 2003; Nelson & Jin, 2004; Qin & Oxenham, 2003; Stickney et al., 2004). Lastly, the ability to parse two competing talkers based on *f0* differences is significantly worse among CI users (Cullington & Zeng, 2008; Qin & Oxenham, 2003; Stickney et al., 2004).

Qin and Oxenham (2003) used NBV techniques to estimate the effects of noise on speech understanding among the CI population. Participants consisted of 32 young NH listeners. Individuals were presented with target sentences [Hearing In Noise Test (HINT) sentences] spoken by a male talker (*f0*=110.8 Hz), which were processed to contain 4, 8 or 24 spectral channels; unprocessed stimuli also were presented. Speech recognition abilities were measured in several masker conditions: 1) male talker, *f0*=114.4, 2) female talker, *f0*=129.4, 3) steady-state noise, or 4) modulated noise. The SRT was calculated using a two-parameter sigmoid model, estimating the point at which identification was 50% accurate.

As expected, performance with the unprocessed speech was best (-11 dB TMR) when the masker was of the opposite sex (female), and slightly worse (-10 dB TMR) when the masker was of the same sex (male). Performance worsened to -9 dB TMR with the modulated masker, and to -6 dB TMR with the steady-state noise. When stimuli were spectrally degraded, higher TMRs were required for all masker types, but results were dependent upon the degree of spectral resolution (lower TMRs for 24 channel < 8 channels < 4 channels). Listeners did not benefit from release from masking for any of the NBV conditions, unlike those results obtained with the unprocessed stimuli. For example, in the 4 channel condition, listeners required 15 dB TMR and 16 dB TMR for the steady-state and modulated noises, respectively, in order to obtain 50% accuracy. However, TMRs were nearly 18 dB for both of the speech maskers (male and female). Additional studies (in both CI users and those listening to CI simulations) show a general trend for increased masking with a modulated signal when speech is spectrally degraded (Cullington & Zeng, 2008; Stickney et al., 2004).

Subsequent investigations (Stickney et al., 2004) suggest that individuals listening to CI simulations are able to benefit from disparate *f0s* across talkers when they are sufficiently different from one another. Stickney and colleagues (Stickney et al., 2004) measured sentence identification in 25 young NH listeners. Sentences were spoken by a male talker (*f0*=108 Hz) and were spectrally degraded (NBV) to contain 1, 2, 4, 8 spectral bands; unprocessed speech was also presented. Target sentences were presented in quiet, in steady-state speech-spectrum-shaped noise, or in the presence of the same male talker, another male talker (*f0*= 136 Hz), or a female talker (*f0*=219 Hz). The same sentence was used for the competing speech masker for all presentations. The TMR was systematically varied with each condition. Overall, results indicate that performance was best with the steady-state masker regardless of the degree of spectral distortion or *f0* of the competing talker. However, listeners' performance did improve when the speech masker was female, compared to the performance obtained when the speech masker was the same male talker. For the 8-channel condition, listeners' performance improved by 5 to 40% (depending on the SNR) when the female masker was used instead of the same male talker. However, this effect was not present in the 4-channel condition. These results suggest that CI listeners may be able to utilize differences in voice pitch to separate competing talkers, if spectral resolution is adequate.

Although these results support the long-standing view that the release from masking benefit is affected by spectral degradation, recent evidence suggests that such claims may be rather misleading. Specifically, several studies (including all of those mentioned above) demonstrate that HI and CI listeners require a higher TMR than YNH listeners to obtain equal performance when speech is masked by a steady-state noise. Further findings suggest that the release from masking benefit in NH listeners is greatest for more adverse (negative TMR) listening situations (Festen & Plomp, 1990; Oxenham & Simonson, 2009). Recall that release from masking benefit is calculated by subtracting performance when a modulated masker is used from that when a steady-state masker is used. Therefore, the lack of release from masking benefit in HI and CI listeners could be attributed to the elevated TMR required when a steady-state masker is present.

In fact, a recent investigation by Bernstein and Grant (2009) demonstrated that most of the reduction in release from masking benefit noted among HI listeners can be attributed to differences in baseline performance using a steady-state masker. Specifically, HI listeners require elevated TMR to achieve equal performance to NH listeners when a steady-state masker is used, and therefore exhibit reduced fluctuating masker benefit. Results from Bernstein and Grant (2009) did reveal, however, that approximately 1-4 dB of the reduced release from masking benefit observed in those with HI could *not* be accounted for by baseline differences in speech masked by a steady-state interferer. This residual loss, therefore, is likely caused by underlying changes in spectral and/or temporal resolution in HI listeners. Taken together, these results suggest that, while some loss of release of masking benefit in HI and CI listeners may be attributed to changes in spectral and/or temporal processing, the magnitude of such deficits may be far less than previously estimated.

Effects of aging, hearing impairment, and cochlear implantation on speech perception in noise

Previous sections reviewed how speech perception is affected when stimuli are masked by various types of interferers, and how such perception differs in younger NH, HI, and CI users. Several findings reveal that aging may cause further deficits when listening to speech in noise, regardless of peripheral hearing status, and overall patterns of performance with various types of maskers may differ between younger and older listeners. Investigators have researched the combined effects of aging and SNHL on speech perception in noise, but little is known about the ability of older CI users to understand speech in adverse conditions. The following section reviews current knowledge regarding cochlear implantation in elderly listeners as well as the effects of aging alone on masked speech perception. Overall, these findings argue a need for better understanding of how older adults perceive spectrally degraded speech stimuli, particularly in adverse listening conditions.

It is apparent from the CI and CI-simulation literature gathered among younger individuals, that the decreased spectral resolution provided by a CI causes difficulty when listening in noise. Based on results from Chapters 2 and 3, older listeners experience difficulty perceiving TEP when spectral cues are reduced. However, it is unknown how such results would carry over to everyday speech communication. Overall, the interactions between aging, cochlear implantation, and the ability to listen in noise are not well understood.

Effects of adult age on speech recognition in cochlear-implant listeners. Studies have shown that cochlear implantation among older adults is quite beneficial as they report improved quality of life and exhibit adequate speech recognition abilities when stimuli are presented in quiet (Chan, Tong, Yue, Wong, Leung, Yuen, & van Hasselt, 2007; Chatelin et al., 2004; Leung et al., 2005; Noble, Tyler, Dunn & Bhullar, 2009; Pasanisi et al., 2003). Despite these general benefits, much of what we know about CI processing in older adults is rather superficial. Investigations do suggest that older listeners have difficultly perceiving severely spectrally-degraded phonemes (Schvartz et al., 2008), and a recent study by Noble et al (2009) also lends support for reduced binaural CI advantage among elderly listeners. One particular issue when measuring the effects of aging on cochlear implantation outcomes is the fact that age often covaries with duration of deafness. As a result, some studies reveal that duration of deafness is a stronger predictor than age at implantation (Fetterman & Domico, 2002; Leung, et al., 2005). Regardless, it is important to understand how performance is affected by aging alone; using NBV simulations affords this analysis. To date, it remains unknown how older listeners would perform in more challenging situations, such as when listening in noise, a task in which even those older listeners with NH experience difficulty (Dubno, Dirks & Morgan, 1994; Dubno, Ahlstrom, & Horwitz, 2002; 2003; Gifford, Bacon, & Williams, 2007; Gordon-Salant & Fitzgibbons, 1995; Hargus & Gordon-Salant, 1995; Helfer & Wilbur, 1990; Tun & Wingfield, 1999).

Review of speech understanding in noise among older normal-hearing listeners. There is a wealth of information regarding the effects of aging on speech recognition in noise. Overall, results suggest that, in most cases, the speech recognition performance of older listeners is only slightly worse than that of younger listeners when a steady-state noise is present (Dubno et al., 2002; 2003; Stuart & Phillips, 1996; Takahashi & Bacon, 1992). Greater age-related disparities in performance are noted when speech is masked by a modulated or interrupted signal (speech or noise) (Dubno et al., 2002; 2003; Helfer & Freyman, 2008; Rossi-Katz & Arehart, 2009; Stuart & Philips, 1996; Takahashi & Bacon, 1992; Tun & Wingfield, 1999). It has been hypothesized that such findings can be attributed to declines in temporal processing, increased susceptibility to IM, or both among older listeners.

a) *Temporally modulated maskers.* In theory, a modulated masker should provide a "release from masking" so that listeners are able to listen between the amplitude peaks of the masker (Festen & Plomp, 1990). Individuals must also possess adequate temporal resolution in order to effectively process the amplitude fluctuations. It has been hypothesized that the inability to process such temporal fluctuations underlies older listeners' difficulty in modulated noise backgrounds (Dubno et al. 2002; 2003; Gifford et al., 2007; Gustafsson & Arlinger, 1994; Hygge, Ronnberg, Larsby & Arlinger, 1992; Pichora-Fuller & Souza, 2003). Although it is not completely understood why older listeners have difficulty processing rapid temporal fluctuations, some studies in both animals and humans suggest that recovery from forward masking may be prolonged with advanced age, consequently influencing older listeners' ability to listen in the presence of a fluctuating masker (Dubno et al., 2003; Gifford et al., 2007; Gleich, Hamann, Kittel, Klump, &Strutz, 2007; Walton, Orlando, & Burkard, 1999). For example, Dubno et al. (2003) showed significant correlations between speech recognition in interrupted noise and forward-masked thresholds in younger and older NH listeners.
b) *Linguistic and semantic masking.* Speech maskers have been shown to be more detrimental to the speech recognition abilities of older adults, compared to steady-state or modulated noise (Helfer & Freyman, 2008; Tun &

Wingfield, 1999). Although modulated noise may provide temporal masking of the target signal, meaningful speech stimuli promote an additional degree of confusion. This type of masking [linguistic uncertainty or masking (Simpson & Cooke, 2005)] occurs because it is more difficult to discriminate between two like signals (i.e., two meaningful sentences) compared to two different signals (i.e., one meaningful sentence and one non-linguistic stimulus such as noise). Some research suggests that linguistic masking greatly contributes to the inability of older listeners' to suppress a non-target speech signal (Rossi-Katz & Arehart, 2009; Tun et al., 2002), while other data suggest that the susceptibility of listeners to the effects of linguistic masking is comparable across adult age (Agus, Akeroyd, Gatehouse, & Warden, 2009; Li, Daneman, Qi, & Schneider, 2004).

For example, Tun and colleagues (Tun et al., 2002) measured sentence perception in younger and older adults with NH sensitivity from 250-3000 Hz. Target sentences varied in semantic and syntactic accuracy and were spoken by a female talker, and distracter sentences were either meaningful sentences or random word strings spoken by a different female talker. Results showed that the performance of both younger and older listeners was worse in the presence of a speech masker compared to performance in quiet. However, the performance of the younger listeners did not depend on the type of speech distracter (meaningful sentences v. random word strings) whereas the performance of the older listeners was worse with the former compared to the latter. In other words, older listeners were more susceptible to speech distracters that were semantically and syntactically appropriate compared to those that were not so, whereas younger listeners were equally susceptible to both types of distracters. This was particularly true when the target stimuli were less semantically and syntactically appropriate. The results of Agus et al. (2009), however, demonstrated that the effect of linguistic masking was approximately 3-4 dB for both younger and older listeners. Both groups showed greater linguistic masking at lower (more difficult) TMRs, and both groups exhibited little or no linguistic masking at TMRs greater than -4 dB. Taken together, these results suggest that the interaction between IM and aging remains uncertain.

c) *Segregating simultaneous speech using differences in voice-pitch*. To date, few studies have systematically investigated the effect of aging on the ability to use *f0* differences between the talker and masker when identifying target speech stimuli. In a recent investigation, Helfer and Freyman (2008) measured the ability of younger and older listeners to identify a target sentence in the presence of a simultaneous competing talker. The degree of hearing sensitivity varied across subjects: all younger listeners had NH, whereas the hearing sensitivity of the older listeners ranged from NH to a moderate high-frequency hearing loss. Stimuli consisted of topic-based, contextual and semantically simple sentences. Stimuli for the target sentences were naturally-uttered sentences spoken by a female talker (*f0* unspecified), while the background sentences were spoken by two naturally recorded male speakers or two female speakers, different than that of the female speaker of the target sentence (*f0*s unspecified). A third condition consisted of a speech-shaped modulated masker. Listeners were evaluated on their ability to identify two or three target words from the target sentence as the TMR was varied (-8, -4, 0 and 4 dB).

Results indicated that the overall performance of the younger listeners exceeded that of the older listeners. The performance of the younger group was dependent upon the masker, with the worst performance occurring for the same-sex (female) masker. For example, for the most difficult condition (-8 dB TMR) the average performance of the younger listeners was approximately 30% correct for the female maskers, but improved to approximately 65% correct with the speech-shaped masker, and again to approximately 85% correct for the male maskers. For the same condition (-8 dB TMR), the performance of older listeners was approximately 10, 50, 55% correct for the female, speech-shaped, and male maskers, respectively. Recent results from Rossi-Katz and Arehart (2009) also demonstrated that younger listeners received greater benefit from differences in talker gender compared to older listeners. Contrary to these results, Vongpaisal and Pichora-Fuller (2007) reported concurrent vowel identification in younger and older listeners and showed equal benefit of *f0* separation between competing vowels in both age groups.

d) *Speech recognition in multi-talker backgrounds*. The ability of ONH listeners to understand speech in the presence of other speech is affected by the number of talkers present in the background. Overall, several studies have also demonstrated that older listeners perform worse than younger listeners when listening to speech in a multi-talker background (Dubno et al., 1994; Gordon-Salant & Fitzgibbons, 1995; Hargus & Gordon-Salant, 1995; Helfer & Wilbur, 1990). As previously mentioned, the speech recognition abilities of YNH listeners declines monotonically as the number of background talkers is increased from 1 to 8 (Miller, 1947; Simpson & Cooke 2005). Results obtained by Tun and Wingfield (1999) would suggest that the number of talkers differentially affects younger and older listeners. Younger (ages 18-22) and older (ages 61-79) listeners were asked to recall short sentences spoken by a female talker, when the background noise was either a single talker (male), two talkers (two different male talkers), 20-talker babble, or white noise. While both groups had pure-tone thresholds of 25 dB or less from 250-3000 Hz, slight differences were present between the audiometric thresholds of the listeners in each age group. Stimuli were presented at two levels, 0 and -6 dB TMR.

As expected, results showed that both younger and older listeners exhibited greater difficulty at the -6 dB compared to the 0 dB TMR. Overall, the performance of the older group was worse than the younger group for every condition at both TMRs. However, the pattern of performance did depend on the age of the listener. For example, both groups obtained nearly 100% accuracy when sentences were presented in quiet. The performance of the younger group was maintained at a high level (nearly 100%) in the presence of one talker, whereas the performance of the older group decreased by an average of 20 and 30%, for the 0 and -6 dB TMRs, respectively. Both age groups performed equally when a steady-state noise was used as the masker. The performance of the younger listeners for the two-and 20-talker conditions was significantly better (approximately 60-65%) compared to that of the older group (approximately 40%). These results suggest differential effects of energetic and informational masking among younger and older listeners. Overall, both groups performed equally well in quiet and when speech was masked with a steady-state noise. Older listeners performed worse than younger listeners when speech was masked with a single- or multi-talker signals. It is also possible, however, that age-related differences in performance were caused by slightly poorer audiometric hearing thresholds among the older cohort.

In summary, results show that the performance of older listeners is only slightly worse than that of younger listeners when stimuli are masked by a steady-state interferer. Older listeners also show a lessened ability to listen in the presence of a fluctuating noise (Dubno et al. 2002; 2003; Festen & Plomp, 1990; Gifford et al., 2007; Gustafsson & Arlinger, 1994), and do not benefit from masking release to the same extent as their NH peers. Results also suggest that older listeners can benefit from the differential *f0* cue when the pitch of a simultaneous background sentence is different than that of the target sentence, but not as much as their younger counterparts (Helfer & Freyman, 2008; Humes, Lee, & Coughlin, 2006; Rossi-Katz & Arehart, 2009). Further, informational (linguistic) masking can sometimes interfere with the ability of older listeners to hear target speech in the presence of another talker (Tun et al. 2002) and deficits are also present in older listeners when speech is masked with multiple talkers (Dubno et al., 1994; Gordon-Salant & Fitzgibbons, 1995; Hargus & Gordon-Salant, 1995; Helfer & Wilbur, 1990; Tun & Wingfield, 1999). Overall, older listeners commonly experience difficulty understanding speech in noise, but the extent to which various maskers differentially affect perception is only somewhat understood.

Speech recognition in relation to cognitive factors and peripheral hearing sensitivity

Several studies have attempted to identify additional factors, other than chronological age, to help account for the variance noted among adults when performing speech recognition in noise tasks. Specifically, researchers have been most interested in determining the extent to which hearing loss and cognition play a role in speech perception, particularly when listening in difficult situations. Altogether, findings confirm that both of these factors do contribute to speech perception abilities in older listeners, but the extent to which each of these plays a role is uncertain.

Results presented thus far exemplify the deleterious effects of aging on speech recognition, and it is only logical that the addition of hearing loss would further promote speech understanding difficulties in this population. Several studies have demonstrated that hearing impairment will result in greater difficulty understanding speech in noise compared to older NH listeners (Akeroyd, 2008; Amos & Humes, 2007; Dubno et al., 1994; Gordon-Salant & Fitzgibbons, 1995; Hargus & Gordon-Salant; Humes & Roberts, 1990; Humes & Christopherson, 1991; Humes, 2007; van Rooij & Plomp, 1990; 1992). Humes (2007) reported that approximately 65-90% of age-related differences in speech understanding in noise arise from loss of audibility, and a similar estimate was also reported in van Rooij and Plomp (1990). Other studies have also used predictions based on the Articulation Index (AI) or the Speech Intelligibility Index (SII) to determine the extent to which individuals' hearing thresholds may affect speech perception in noise (Dubno et al., 2002; Hargus & Gordon-Salant, 1995; Humes, 2002). Altogether, these studies demonstrate that even small differences in peripheral hearing sensitivity can contribute to listeners' speech perception in noise abilities.

Although peripheral hearing sensitivity contributes to older listeners' reduced speech in noise understanding, there remains a proportion of the variance in performance that cannot be accounted for by auditory thresholds (Dubno et al., 2002; Humes & Christopherson, 1991; Humes, 2007; van Rooij & Plomp, 1990). This residual variance in performance is most commonly thought to arise from differences among the cognitive capacities in listeners (Akeroyd, 2008; Craik, 2007; Humes, 2007; Pichora-Fuller, Schneider, & Daneman, 1995). A review by Craik (2007) explains the intricate levels of attention processing that occur beyond the auditory periphery. In brief, interpretation of information includes both "bottom-up" (stimulus driven) and "top-down" (cognitive driven) processing, and declines in perception will occur if either of these pathways is adversely affected. Perceptually speaking, individuals are able to actively attend to one "channel" of information. When competing stimuli are present in the auditory scene, we must selectively attend to the stimulus of choice, while ignoring irrelevant stimuli (Hasher & Zacks, 1988). It is believed that working memory function is central to the fluidity of this process. In brief, working memory is defined as the ability to store and manipulate visual and/or auditory information (Baddeley, 1986). It has been hypothesized that aging alters working memory function, and cognitive inhibitory mechanisms also change as a function of age (Craik, 2007; Hasher & Zacks 1988). Based on this hypothesis, older listeners would exhibit a reduced capacity to attend to a target stimulus in the presence of background noise, and cognitive changes would contribute to reduced speech in noise understanding in this population (Craik, 2007; Schneider, Li & Daneman, 2007).

Various studies have used different measures to assess cognitive and working memory function among the participants. While there is no one standard measure to assess working memory or inhibitory abilities, some tests are more commonly used than others. For example, the Wechsler Adult Intelligence Scale [(WAIS), Wechsler, 1997] consists of three subtests which are commonly used by researchers and clinicians to quickly but reliably assess one's working memory abilities: forward digit span, backward digit span, and letter-number sequencing (for the specifics of each test, refer to page 111). All three tests require that listeners attend to the auditory stimulus (i.e., series of numbers or letter), remember the sequence and then repeat back the sequence in a specified manner. Humes et al. (2007), for example, used the digit span tasks (forward and backward) as an estimate of working memory function and found that performance on the digit span tasks was strongly correlated with younger NH and older HI listeners' ability to perform a selective attention task (i.e., attending to one auditory stimulus while ignoring another); conversely, age and hearing status were not correlated with performance.

Along a similar vein, we also demonstrated in a previous study that spectrally degraded vowel recognition in quiet was more related to working memory abilities than chronological age or hearing sensitivity in middle-aged and older NH individuals

(Schvartz et al. 2008). While other measures do exist, simple digit span tasks seem to capture some facet of working memory function related to complex speech understanding. Taken together, these studies demonstrate the importance of considering differential hearing sensitivity and cognitive abilities, such as working memory, when investigating the abilities of younger and older individuals to understand speech in adverse conditions.

Summary

Collectively, we know that aging affects the processing of speech in noise, and that results are dependent on the type of masker used. Furthermore, spectral degradation interferes with the ability of younger listeners to use *f0* or modulation cues in order to aid in sentence recognition performance. Although there is little effect of age on the ability of CI users to understand sentences in quiet, nothing is known regarding older CI listeners' ability to understand speech in noise. It has been shown that measures of cognitive function (i.e., working memory) are related to one's ability to listen to degraded or masked speech. In some cases, such cognitive measures are better predictors of performance than chronological age or hearing status. There is also some evidence that peripheral hearing sensitivity plays a role when measuring speech recognition in noise. The purpose of the current experiment was to better understand the interactions between aging and the ability to understand spectrally degraded speech masked by noise. A secondary goal was to determine if performance on a spectrally-degraded speech recognition in noise task was at all predicted by one's working memory function. Speech recognition abilities were measured in younger and older NH individuals, when stimuli were unprocessed (full spectrum) and spectrally degraded (i.e., 24 channels). It should be noted, however, that the 24-channel stimulus is mildly degraded, does not represent the degree of spectral degradation inherent to CI processing. However, the pilot testing revealed that reducing the number of channels to approximate CI listening (i.e., 8-10 channels) resulted in extremely poor performance on the speech in noise tasks.

The target speech was a male or female talker. Three masker conditions were used: single male talker (different than the target speech), 10-talker babble (male), and a speech envelope modulated (SEM) noise. An adaptive procedure was employed to determine the SNR at which the listener is able to achieve approximately 70.7% accuracy on the sentence identification task (Levitt, 1971).

This choice of target and masker was designed to investigate contributions of various factors to speech recognition in noise abilities. First, we aimed to measure the interaction between aging and spectral degradation when listening in noise. Second, we sought to investigate the ability of younger and older listeners to use three cues known to differentially affect speech recognition in noise performance: *f0* differences between the target and masker, listening in the dips of a fluctuating masker, and linguistic IM. Specifically, we hypothesize the following: when the signal is masked by another competing talker, listeners are subject to linguistic masking but can take advantage of amplitude dips in the masker envelope and differences in the *f0*s of the talker and masker. When the masker is 10-talker babble (male), listeners cannot listen in the dips of the masker envelope, but are no longer subject to the effects of linguistic masking; listeners may, however, be able to use *f0* cues. When the masker is SEM noise, listeners can listen in the dips of the masker envelope and are not subject to linguistic masking, but cannot take advantage of differences in the *f0*.

Questions and Hypotheses

1. Do spectrally-degraded speech recognition in noise abilities differ in younger and older listeners?
2. If aging does affect one's ability to understand speech masked by noise when stimuli are spectrally degraded, does age-related performance also differ depending on the masker (single-talker, multi-talker or modulated noise)?
3. Is speech recognition in noise performance among adults correlated with measures of peripheral hearing sensitivity or cognitive function?

Based on previous studies, it is expected that older listeners will perform worse than their younger counterparts when unprocessed speech stimuli are masked by noise (Dubno et al., 1994; 2002; 2003; Gordon-Salant & Fitzgibbons, 1995; Helfer & Wilbur, 1990; Stuart & Phillips, 1996; Takahashi & Bacon, 1992; Tun & Winfield, 1999). Studies have demonstrated that older listeners are less able to utilize temporal fluctuations in the masker (Dubno et al., 2002; 2003; Gifford et al., 2007; Gustafsson & Arlinger, 1994) and differences in the voice-pitch of competing talkers (Helfer & Freyman, 2008; Tun & Wingfield, 1999). Some studies also demonstrate that older listeners perform poorer than younger listeners when speech is masked by multiple talkers (Dubno et al., 1994; Gordon-Salant & Fitzgibbons, 1995; Hargus & Gordon-Salant, 1995; Helfer & Wilbur, 1990).

Furthermore, previous studies have demonstrated that the ability of older listeners to understand speech in noise is somewhat dependent on the masker. For example, Tun and Wingfield (1999) showed equal performance across adult age groups when speech was presented in quiet, and only slight differences in performance when speech was masked by a steady-state noise. The performance of the older listeners, however, was significantly poorer in the presence of a single-talker, and then slightly worse again when a multi-talker masker was used. Overall, it is expected that the results of the current study will also show interactions between aging and masker type for speech in noise performance.

It has been shown that aging can affect the perception of spectrally- degraded phonemes presented in quiet (Schvartz et al., 2008), and therefore it is also expected that such age-related deficits would also occur when spectrally-degraded speech is masked by noise. While the exact interaction is difficult to predict, it is possible that there will be interactions between aging and masker type when stimuli are degraded. In particular, it has been demonstrated that, while a modulated masker typically provides release from masking in unprocessed speech, the same masker will often result in increased masking when stimuli are spectrally degraded (Qin & Oxenham, 2003). It is possible that aging will interact with this effect.

Lastly, it has been demonstrated that both peripheral hearing sensitivity and cognition are somewhat related to the age-related differences in speech in noise performance that are commonly observed in adult listeners (Akeroyd, 2008; Humes, 2007; van Rooij & Plomp, 1990; 1992). It is therefore important to determine the extent to which of these factors contribute to speech in noise performance when stimuli are both unprocessed and spectrally degraded. Previous findings are mixed with regards to the extent to which each of these variables contribute to performance in older adults, however, it is hypothesized that either or both of these factors (peripheral hearing sensitivity and cognition) will be related to the performance of adults when measured on a speech recognition in noise task.

Method

Participants

Participants were NH males and females, recruited for placement in two different groups based on their age at the time of testing: younger (ages 21-26; Mean= 23.3, SD=1.63) or older (ages 60-76; Mean=66.6, SD=6.05). The YNH group consisted of ten individuals (three males, seven females) and the ONH group consisted of nine individuals (two males, eight females). All individuals were required to have pure-tone thresholds≤ 25 dB HL from 250-4000 Hz in both ears (ANSI, 2004). See Table X for audiometric data. All participants in the older age group obtained scores within the normal range on the Mini Mental State Examination (MMSE) (Folstein et al., 1975), a screening test used to identify gross cognitive dysfunction.

Stimuli

Target sentences. Stimuli were naturally-uttered, digitally recorded, IEEE-Harvard sentences (IEEE, 1969) spoken by a male ($f0 \approx 100$ Hz) or female ($f0 \approx 240$ Hz) talker. Sentences were originally recorded at HEI (sampling rate = 22,050 Hz), and are available with the iSTAR research interface developed by Qian-Jie Fu, Ph.D. at HEI. The IEEE sentence corpus consists of 720, low-to-moderate context sentences.

			Experiment 4 Audiometric pure-tone thresholds (dB HL)						
			Frequency (Hz)						
			250	500	1000	2000	3000	4000	PTA
YNH	Right	AVG	6.5	5.5	3	3.5	5	3.5	4
		SD	4.1	4.3	5.8	4.7	5	4.1	4.38
	Left	AVG	8	3.5	3	4.5	9	4.5	3.6
		SD	4.2	3.3	5.8	5.9	4.1	4.9	4.36
ONH	Right	AVG	10	8	10.5	11.2	11	15	9.8
		SD	5.7	5.8	4.9	6.2	4.5	5	2.7
	Left	AVG	10.5	9.5	9.5	12.5	10	17.5	9.5
		SD	5.9	5.5	4.3	2.7	4.0	5.2	3.0

Table X. Audiometric pure-tone air conduction thresholds for the two groups of listeners that participated in Experiment 4. Average values and standard deviations are shown individually for each ear and at each frequency tested. Pure tone averages (PTA= average threshold at 500, 1000 and 2000 Hz) are also given.

Maskers. Three different types of maskers were used in the present study: Single-talker (male), Ten-talker babble (male), and SEM noise. Descriptions and construction of each masker are provided below.

a) *Single-talker masker*. The single-talker male masker (S_M) consisted of continuous discourse passages taken from a pre-recorded reading of “Schulz and Peanuts: A Biography” (An autobiography of Charles Schulz) written by David Michaelis. Passages were selected at random, and were spliced into six minute sections using Adobe Audition (version 1.5); a total of 20, six-minute tracks were created and stored on computer hard-disk. Further processing was performed on each of the 20 tracks in order to analyze and remove exceedingly long (>100 ms) silent intervals contained in the original recording. The final tracks consisted of concatenated, meaningful sentence passages (with no long pauses). Each track was then adjusted in RMS level to equate that of all other stimuli (target sentences and other maskers).

b) *Ten-talker babble masker*. The ten-talker male babble masker (B_M) was constructed by randomly selecting ten of the single, six-minute tracks described above (S_M tracks). All ten, six-minute tracks were mixed down into a single file using Adobe Audition, then adjusted in to a specified RMS value.

c) *Speech envelope modulated masker*. Speech envelope modulated noise (SEM noise) was constructed by selecting one of the single-talker, six-minute tracks at random, and then processing the file using a single-channel NBV method. This process resulted in a band-pass white noise (100-4000 Hz) modulated by the extracted, original speech envelope. Noise-band vocoding was accomplished off-line using TigerCIS (available through TigerSpeech Technologies). The original signal was first band-pass filtered (100-4000 Hz, 4^{th} order Butterworth), then the temporal envelope was extracted using low pass filtering and half-wave rectification (400 Hz, low-pass filter, 4^{th} order Butterworth). The original TFS was replaced with random white noise [band-passed to from 100-4000 Hz (4^{th} order Butterworth)], then multiplied with the extracted temporal envelope. The final signal was then manipulated using Adobe Audition to match a target RMS value. The same SEM noise track was used for all testing.

Measures of verbal working memory. Three verbal working memory tasks were performed: Forward digit span (FDS), backward digit span (BDS) and letter-number sequencing (LNS). These tests are standardized cognitive measurements taken from the Wechsler Adult Intelligence Scale (WAIS, Version IV), and are aimed at estimating memory span and/or working memory abilities among adults. Stimuli for these tasks consist of a string of digits and/or letters, spoken in sequences that vary in length. All stimuli were recorded with a Marantz Professional recorder (PMD660) using a 44,100 Hz sampling rate, then transferred and stored on computer disk. The recorded materials were edited using Adobe Audition (version 1.5). Each test was edited to contain 1 second inter-stimulus-interval (ISI) to standardize the timing of presentation across all participants. Edited stimuli were stored on computer disk.

Noise-band vocoding. Noise-band vocoding was conducted using software made available by TigerSpeech Technologies. Target stimuli (IEEE sentences) were processed online using TigerCIS, available within the iSTAR research interface. Masker stimuli (S_M or B_M) were processed off-line using the same software, then stored on computer disk. However, identical parameters were used for all NBV processing throughout the study. First, all stimuli were band-pass filtered from 100-4000 Hz using a 4th order Butterworth filter. The stimulus was then divided up into 24 spectral channels, and noise band vocoded using identical methods as described in Experiment 3. For conditions in which stimuli were unprocessed, TigerCIS was implemented to band-pass filter the stimulus from 100-4000 Hz.

General procedures: speech recognition testing

All stimuli for the speech recognition tests were presented using the iSTAR research interface. This interface allows the user to present stimuli at a standard, calibrated level in quiet, or allows the user to implement an adaptive SRT procedure in which the background distracter is presented at a constant specified level, and the target sentence stimulus is adapted according the listener's response. All testing was completed in a double-walled sound proof booth.

For all tasks, the listener heard a short (500 ms) 1000 Hz tone prior to the start of each sentence. This tone was meant to cue the listener's attention to the upcoming sentence stimulus. Following the end of the tone, a sentence played 500 ms later and was delivered through a single loudspeaker (Tannoy Reveal) located at 0 degrees azimuth at the level of the listener's head. Each masker was presented continuously throughout the run, and was delivered through the same speaker as the target sentence (co-located at 0° azimuth). Each participant was asked to repeat back the sentence as clearly as possible; there as no time limit for responding.

The tester sat outside of the soundbooth during the entire test session. Stimulus presentation and scoring were performed using the iSTAR computer interface. A microphone (Shure SM58), located inside the soundbooth, transmitted the participants response, which was heard by the administrator through headphones using a monitor system (Samson S-monitor) connected to the response microphone on the inside of the soundbooth. The iSTAR program displayed each sentence (text) on the tester's screen outside the soundbooth as it is being presented to the listener. After the listener gave his/her verbal response, the tester then selected the words that were spoken correctly by clicking on the typed words that appeared on the test screen. Once all of the appropriate words were selected, the tester selected "next" in order to present the next sentence. This was repeated until a single condition was completed.

Scoring. For the purpose of this experiment, the IEEE sentences were scored as correct or incorrect based on key words correct, and an adaptive procedure was used to determine SRT. Listeners needed to correctly identify all key words in the sentence for the presentation level to decrease in the following sentence (two-down, one-up adaptive procedure). Key words were defined as nouns, adjective, verbs and adverbs; pronouns, conjunctions, and propositions were not scored, therefore substitutions or deletion of such words were permitted. Additionally, key words were scored with some leniency; words were not scored as incorrect due to errors in tense, plurality or word order. It was decided that this scoring method was reasonable given the difficult nature of the sentences and test conditions.

Practice. All listeners received practice prior to being scored on any task. First, listeners were presented with IEEE sentences in quiet and a total of 30 sentences were presented for the following four conditions: Male unprocessed, Female unprocessed, Male 24-channel, and Female 24-channel. Listeners did not receive feedback about the accuracy of their response, and the presentation of the sentences was block randomized (without replacement) within each of the four conditions. Based on previous findings (Friesen et al., 2001), it was expected that all listeners would perform this task with good accuracy for both the unprocessed and 24-channel conditions. Following the practice in quiet, listeners were then presented with four practice runs of the spectrally degraded, SRT task. These conditions were presented in a pseudo- random order: listeners heard at least one presentation of each type of masker, and two of the runs contained unprocessed stimuli while the remaining two runs contained a 24-channel condition. The order of these conditions, however, was randomly varied across all listeners.

Test protocol. Following the practice conditions, each condition was presented twice and stimuli were presented in a block randomized order, according to masker-type. That is, within each type of masker, four different conditions (Male unprocessed, female unprocessed, Male 24-channel and Female 24-channel) were presented in a random order. Masker order was also randomized. Four different masker conditions were presented in total: Quiet, S_M, B_M, and SEM noise. One complete run was performed for all possible conditions [16 total: Four target conditions (male or female, unprocessed or 24-channel) x Four masker conditions], and then the second run of all possible conditions was performed in the same block-randomized fashion (but in a different random order compared to the first run). The average performance of the two runs was calculated to determine the final score in each condition. Total test time (including practice) varied from 6-8 hours depending on the listener. Breaks were provided whenever necessary.

a) *Sentence recognition in quiet*. A total of 30 sentences were presented for each talker condition (four total). Sentences were presented at a calibrated level of 65 dBA. Sentence lists were selected at random, without replacement.

b) *Sentence recognition in noise*. An adaptive SRT task was performed for each of the three masker conditions. The masker level remained constant at a 60 dBA, while the level of the target sentence adapted according to the listener's response. A 2-down, 1-up procedure was used to track responses, and the resulting SRT represented the 70.7% correct identification point on a psychometric function (Levitt, 1971). The listener had to repeat back all key words correctly in order for the TMR to decrease. Otherwise, the TMR was increased for the following presentation. A minimum of eight reversals, but not more than ten reversals or a total of 55 sentences were used to calculate the final threshold. An initial step size of 4 dB was used for the first two reversals, but was reduced to 2 dB for the remainder of reversals; the final threshold was calculated from the last four reversals. Target sentences were randomly selected by the computer program for each presentation (with replacement). Therefore, it was anticipated that performance might improve because of increased familiarity of the target sentences. A maximum TMR level of 25 dB SL (re: masker level of 60 dBA) was chosen to limit loudness discomfort for all listeners; this level was chosen after pilot testing which indicated that this generous TMR level was loud enough to optimize SRT performance, and levels above this point began to become uncomfortable for some listeners. If an SRT could not be calculated at this high level, then a default value of 25 dB was entered for the purpose of data analysis.

c) *Verbal working memory testing*. All verbal working memory tasks were presented following completion of the speech recognition testing. The order of presentation was identical for all listeners: FDS, BDS, then LNS. For the FDS test, listeners are presented with a sequence of digits and are asked to repeat the entire sequence back in the order that they heard it. For the BDS task, listeners are presented with a sequence of digits, and are asked to repeat the entire sequence back in the reverse order of presentation. For the LNS task, listeners are presented with a sequence of letters and numbers (in a random order) and are asked to first repeat back the numbers back in numerical order, and then the letters back in alphabetical order. For each task, listeners were instructed to take guesses when they are not sure and response time limit was not imposed. A total of 16, 14, and 21 points were possible for the FDS, BDS, and LNS, respectively.

Results

Speech recognition in quiet

Overall, results suggest that IEEE sentence recognition in quiet was similar across both age groups and did not differ as a function of talker-gender (male or female). Results are shown in Figure 14. A split-plot ANOVA was used to analyze speech recognition in quiet across the two age groups. The analysis contained two within group factors [number of channels ('channels') and talker gender ('gender')] and one between group factor [age group ('age')]. The analysis revealed a main effect of channels [$F(1,18) = 17.72$, $p<0.01$], suggesting that average performance (collapsed across both groups) was better for the unprocessed stimuli (99.33%) compared to the 24- channel stimuli (94.92%). The main effects of talker-gender or listener age were not significant.

Speech recognition in noise

Results of the speech recognition in noise task are shown in Figure 15. Performance on the SRT task proved difficult for some listeners, particularly when stimuli were spectrally degraded. While overall results were better in younger listeners and for unprocessed stimuli (both groups), the effect of masker type was dependent upon the degree of spectral degradation. Differential masking effects were shown for unprocessed stimuli, but not for spectrally degraded stimuli. The results did not show any systematic interaction between aging and masker type, regardless of spectral resolution.

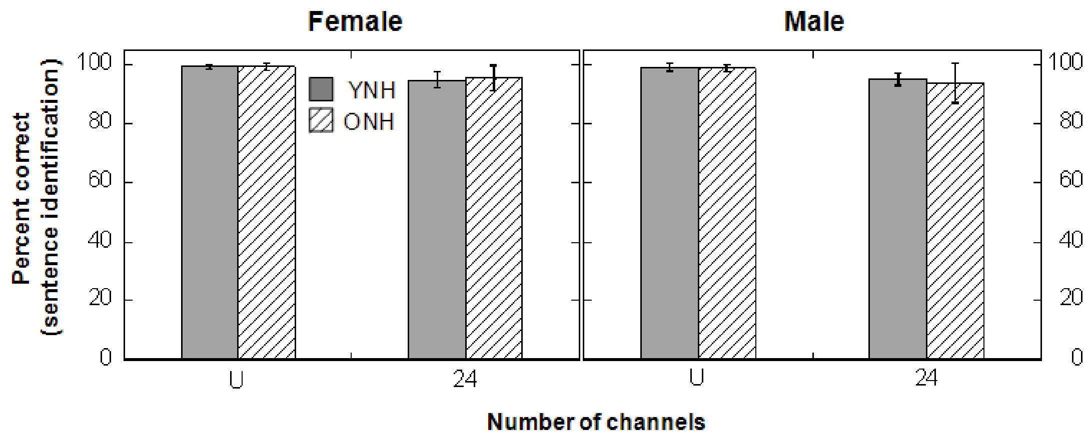

Figure 14. Bar graph representing sentence recognition in quiet performance of the YNH and ONH listeners. The number of channels is shown along the abscissa, and the percent correct scores (sentence identification) are shown along the ordinate. Results from the female talker are shown on the left, and results from the male talker are shown on the right. The error bars represent ±1 SD from the average value.

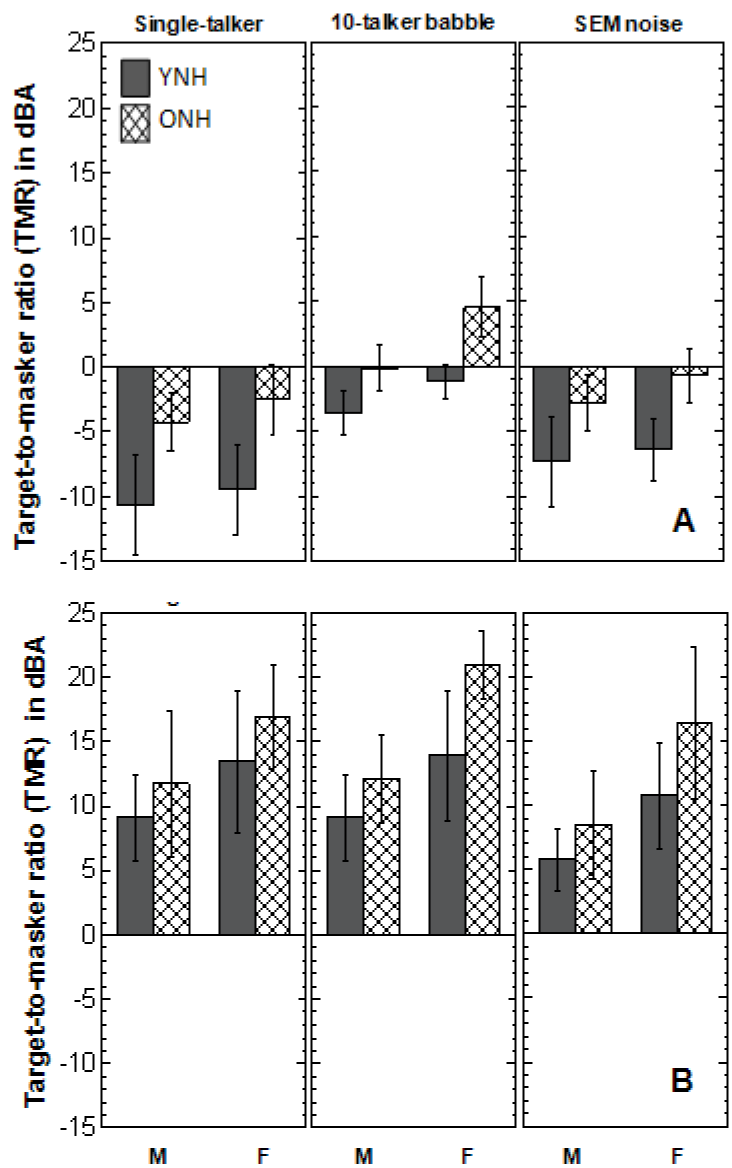

Figure 15. Bar graph representing average group data on the speech recognition in noise task. Results from the unprocessed speech condition are shown on the top (A), while results from the 24-channel speech condition are shown on the bottom (B). The masker type (Single-talker, 10-talker babble, or SEM noise) is indicated at the top of the graph. The talker-gender of the target stimulus (M=Male; F=Female) is shown along the abscissa. The TMR required for the speech recognition threshold is shown along the ordinate in dBA. Error bars represent ±1 SD from the average value.

A split-plot ANCOVA was used to analyze speech recognition in noise performance across the two age groups. To examine the possibility that differences in hearing sensitivity across the two age groups could have contributed to SRT measures, the PTA of both ears served as a covariate in the analysis. The aim of the analysis was to determine if the SRT measures varied as function of age and/or masker aside from any influences of peripheral hearing sensitivity. The ANCOVA was performed with three within-group factors [number of channels ('channels'), masker type (masker), and talker gender ('gender')], and one between-group factor [age group ('age')]. The Greenhouse-Geisser correction was used to interpret results in cases in which the assumption of sphericity was violated.

Overall, results suggest that while spectral degradation, talker gender, and masker influenced the results, the age of the listener had relatively little influence on performance, once PTA was entered as a covariate factor. The ANCOVA analysis indicated a significant main effect of channels [$F(1,17)=20.63$, $p<0.01$), a significant main effect of gender [$F(1,17)=35.23$, $p<0.01$, a significant interaction between gender and age [$F(1,17)=5.35$, $p<0.05$, and a significant interaction between channel and masker [$F(2,24= 4.96$, $p<0.05$). The main effect of age was not significant [$F(1,17)=3.85$, $p=0.06$] .

The significant main effect of channels indicated that overall performance was better for unprocessed speech (average TMR across groups and maskers = -3.68), when compared to performance obtained for the 24-channel conditions (Average TMR across groups and maskers= 12.41). The significant main effect of gender revealed that performance was better for the male speaker (average SNR = 2.31) compared to that of the female speaker (average TMR=6.41). These results are somewhat surprising, as it is expected that the opposite would be true; that is, a wealth of literature provides evidence that when a talker and masker are consistent in gender, then the result is greater masking when compared to instances in which the talker and masker genders are different (Brungart, 2001; Brungart, et al., 2004; Helfer & Freyman, 2008).

Follow-up testing was performed to further analyze the significant interactions of gender x age and channel x masker. For the following analyses, multiple t-tests with Bonferroni correction were performed as appropriate. Consistent with the analysis reported above (main effect of gender), follow-up testing of the gender x age interaction (using paired two-tailed t-tests) revealed that performance was better for the male talker compared to the female talker within both the YNH ($t= -7.92$, $p<0.01$) and ONH ($t= -9.49$, $p<0.01$) groups. For the YNH group, the average TMR for the male and female talkers were 0.11 (SD=8.35) and 3.12 (SD=9.96), respectively. For the ONH group, the average TMR for the male and female talkers were 3.93 (SD=8.07) and 8.42 (SD=10.22), respectively. Taken together, these results reveal that while both groups performed better with the male masker, the younger group performed more comparably with both maskers compared to the older group. More specifically, the female masker resulted in 3.01 and 4.49 dB more masking than the male masker, for the YNH and ONH groups, respectively.

Follow-up testing for the significant channel x masker interaction showed that, on average, performance with each type of masker depending on the spectral resolution available to the listener. Analysis using paired two-tailed t-tests revealed that each masker produced TMRs significantly different from one another ($p<0.016$), but only for the Unprocessed stimuli. Average TMRs for the three masker-types in this case were -6.70 dB (S_M), -3.52 dB (SEM noise) and 1.71 dB (B_M). In the case of 24-channel stimuli, all of the maskers provided an equal amount of masking; average TMRs for the three masker-types in this case were 9.17 dB (SEM noise), 10.61 dB (S_M), and 12.11 dB (B_M). Further analyses confirmed that performance in each of the masker conditions was better for the Unprocessed stimuli compared to that of the 24-channel stimuli ($p<0.016$). Overall, these results further support previous findings suggesting that, while different maskers produce various degrees of masking when stimuli are unprocessed, even a mild spectral degradation of stimuli results in higher (worse) SRT thresholds and decreased performance compared to unprocessed stimuli, regardless of the masker.

Relationship with verbal working memory

The results of the ANCOVA suggest that performance on the SRT tasks did not differ across the two age groups, once differences in PTA were entered into the analysis as a covariate. Previous investigations, however, have revealed that working memory abilities can also contribute to performance on a speech recognition in noise test (see Akeroyd, 2008 for review). Therefore, we chose to also analyze subjects' performance on the SRT tasks in relation to their scores on the three working memory measure: FDS, BDS, and LNS.

Working memory scores for each age group. Performance on each of the working memory tests are shown in Table XI for each age group. Raw scores and z-scores are shown for each measure. As shown in the table, YNH and ONH listeners performed

		FDS	BDS	LNS
YNH	Raw score	10.1 (2.37)	8.3 (1.76)	14.6 (2.63)
	Z-score	-0.02	0.08	0.49
ONH	Raw score	10.2 (2.20)	8 (2.05)	11.7 (2.54)
	Z-score	0.02	-0.08	-0.49

Table XI. Raw scores and Z-scores for the three tests of working memory [forward digit span (FDS), backward digit span (BDS) and letter-number sequencing (LNS)] measured in both age groups (YNH and ONH). Standard deviations for each raw score are shown underneath each value in parentheses.

similarly on each of the tests. Previous studies demonstrate mixed findings regarding effects of aging on these measures (Dobbs & Rule, 1989; Wingfield, Stine, Lahar & Aberdeen, 1988). It is possible that the lack of between- group differences can be attributed to a small sample size for each group. There was little difference between the three measures used to assess WM, as indicated by the inter-correlations between the measures of WM shown in table XII. For this reason, the z-scores for each of the WM tests were averaged to form a combined WM scale, AVG-WM.

Working memory as a predictor of performance on the speech-in-noise tasks. Correlations and linear multiple regression analysis were performed to determine the extent to which peripheral hearing sensitivity and measures of WM help to account for the variance on speech-in-noise performance. Although several conditions of speech-in-noise testing were performed, we chose to average the results within each of the spectrally degraded conditions; that is, two independent variables were explored during these analyses: average TMR for speech-in-noise tasks using unprocessed stimuli (AVG-UN) and average TMR for speech-in –noise tasks using 24-channel NBV stimuli (AVG-24). Prior to regression analyses, exploratory correlation analyses (Two-tailed, Pearson product- moment coefficients) were performed to examine the inter-correlations between the following variables: Average binaural PTA (PTA), AVG-WM, AVG-UN, and AVG-24. Results are show in Table XII. In keeping with the results of the ANCOVA,

	PTA	AVG-WM	AVG-UN	AVG-24
PTA	---	r= -.323 p= .164	r= .669 p= .001	r= .490 p=.028
AVG-WM		---	r= -.223 p= .344	r= -.444 p= .049
AVG-UN			---	r= .656 p= .002

Table XII. Inter-correlation coefficients (*r* values) and level of significance (*p* values) tested prior to performing the regression analysis. Variables were PTA, AVG-WM (averaged z-scores for three tests for working memory), AVG-UN (average SRTs for all unprocessed stimuli), and AVG-24 (average SRTs for all NBV stimuli).

results of the correlation analyses show that PTA is significantly correlated with both speech recognition measures (AVG-UN and AVG-24). Further, results also suggest that AVG-WM is correlated to AVG-24, but not AVG-UN.

Because the results of the correlation analysis suggested that both peripheral hearing sensitivity and WM abilities are related to performance on the 24-channel speech-in-noise task, a stepwise multiple regression analysis was performed to determine the extent to which PTA and AVG-WM could collectively explain the variance on the performance of listeners on this task. The results of the correlation analyses showed that PTA and AVG-WM were somewhat equivalently related to AVG-24; therefore, PTA and AVG-WM were entered into one block for the stepwise multiple regression analysis ($p<0.05$ to enter; $p>0.10$ to remove). In keeping with the results of the correlation analysis, the results of the stepwise multiple regression analysis show that PTA helps to explain a small portion of the performance variance ($R^2 = .240$, $p=.028$) and that AVG-WM did not contribute significantly to the equation ($t= -1.52$, $p=.147$). Taken together, the results of the correlation and regression analysis suggest that PTA was the most significant predictor of performance on the speech-in-noise tasks, irrespective of spectral resolution (unprocessed or 24-channel NBV stimuli).

Discussion

While previous chapters demonstrated the effects of aging on sequential segregation of voice-pitch, the results of the current experiment do not indicate that the same effect is present in the case of simultaneous segregation. These findings are consistent with those reported by Qin and Oxenham (2005) who found that the performance of YNH listeners when performing a sequential *f0*DL task improved from 1 to 24 channels. Simultaneous vowel segregation, however, remained rather poor even when stimuli were manipulated to contain as many as 24 channels. These results confirm that, while temporal envelope periodicity cues may adequately convey *f0* information when stimuli are presented in isolation, across- and within-channel envelope masking interferes with envelope periodicity coding when stimuli are presented simultaneously. Furthermore, the impact of aging on such a task remains unclear. While significant between group differences were apparent, slight differences in peripheral hearing sensitivity confound the data. No relationship was found between working memory abilities and speech in noise recognition.

Speech-in-noise perception of unaltered stimuli

For unprocessed stimuli, the three maskers produced differential degrees of masking, with the S_M condition providing the best SRTs in both groups, and the B_M condition producing the worst SRTs in both groups. These results are also consistent with those of previous studies (Brungart, 2001; Festen & Plomp, 1990; Qin & Oxenham, 2003) all of which exhibit a similar trend: a steady-state masker (e.g., steady-state noise or multi-talker babble) provides the greatest degree of masking, a competing talker typically provides a lesser degree of masking, and a modulated-noise results in an intermediate degree of masking.

Previous studies show that age does not affect speech perception when stimuli are presented in quiet, but the addition of noise causes deleterious performance in older listeners. In the current study, we did not find differences between the two age groups when unprocessed IEEE sentences were presented in quiet. Multiple studies have demonstrated that, while performance is equal across adult age groups when listening to speech in a steady-state masker, older listeners do not receive as much release from masking as younger listeners Dubno et al., 2003; Gifford et al., 2007). Based on the results of previous studies, potential interactions between aging and simultaneous *f0* coding are not straightforward (Helfer & Freyman, 2008; Vongpaisal & Pichora-Fuller, 2007). Although the performance of older listeners was generally poorer than that of younger listeners for each type of masker, findings from the present study do not support an interaction between aging and speech in noise recognition. Statistical analyses (ANCOVA) revealed that any ostensible interaction between masker-type and age group could not be shown above and beyond differential peripheral hearing sensitivity. The covariation between aging and peripheral hearing sensitivity, however, makes such an analysis difficult to interpret. These results therefore, do not confirm nor contradict a relationship between aging and speech in noise understanding.

One objective of the experiment was to investigate the potential interactions between aging and masker type to further define differential contributions of energetic and informational masking. While it is possible that such an interaction does exist to some extent, the current study did not unveil any clear pattern. Lastly, it is commonly acknowledged that speech is better understood when masked with a talker of the opposite, as opposed to the identical, gender (Brungart, 2001; Brungart et al, 2001; Darwin et al., 2003) and differences in *f0* are salient cues used to identify talker-gender (Childers & Wu, 1991). This effect, however, was not demonstrated in the current study. The reason for this finding is unclear as the *f0* values used in the current study were sufficiently different from one another to promote segregation. It could be, however, the

the *f0* values of the male target (*f0*≈ 100 Hz) and male masker (*f0*≈ 140 Hz) were too dissimilar from one another and were therefore easy to separate. Previous studies have demonstrated that competing stimuli spoken by different talkers of the same gender can often easily be segregated (Brungart, 2001; Brungart et al., 2001).

Spectral degradation and speech-in- noise perception

Similar to previous studies, the results of the present experiment clearly demonstrate the deleterious effect of spectral degradation on speech-in-noise perception (Cullington & Zeng, 2008; Friesen et al., 2001; Qin & Oxenham, 2003; Stickney et al., 2004). This effect was seemingly independent of age, as the performance of both younger and older listeners was equally affected by spectrally degrading the target and masker stimuli, after accounting for differences in peripheral hearing sensitivity. Unlike previous studies, all three types of maskers provided essentially equivalent amounts of masking in both age groups. For example, Qin and Oxenham (2003) found that, with 24-channels of spectral information, listeners performed best with modulated noise and worst with a single-talker masker. Results with the steady-state masker were intermediate. The results of the current study do approach this trend, as we found that the average performance (collapsed across groups) for each of the maskers was 9.17 dB (SEM noise), 10.61 dB (S_M), and 12.11 dB (B_M), however these differences were not statistically significant. Furthermore, Qin and Oxenham (2003) reported SNRs ranging from, approximately, -3 to +1 dB (24-channel condition), while the performance of listeners in the current study was comparatively much worse (range of approximately +6 to +21 dB). While this discrepancy is surprising, target stimuli for the Qin and Oxenham (2003) study were selected from the HINT corpus (Nilsson, Soli, & Sullivan, 1994) which is comprised of sentences that are simpler and high-context compared to IEEE sentences. It should be noted that for the unprocessed conditions, our results were more comparable to those of Qin and Oxenham (2003), suggesting that stimulus-dependent contextual effects might be augmented when stimuli are degraded. Furthermore, their SRT approximated 50% correct identification, whereas the SRT in the current study approximated 70.7% correct identification. It is possible that these variables contributed to the differences in performance noted across the two studies.

Correlating speech perception with measures of working memory

Although previous studies noted significant correlations between one's ability to perceive speech in the presence of a competing signal and estimates of cognitive function (i.e., working memory) the present investigation did not find a significant relationship between these two variables, either for unprocessed or NBV stimuli. A review by Akeroyd (2008) provides a comprehensive examination of the literature regarding this interaction, and similar to other studies' (van Rooij & Plomp, 1990; 1992) comprehensive findings, support a link between cognitive processes and understanding speech in adverse situations. This meta-analysis, however, also demonstrates that results are highly variable and some studies did not uncover any such association. Several measures exist that aim to assess working memory function; it is possible that the digit and letter span tasks chosen for the current study were not the most appropriate measures for the given task.

Taken together, these results suggest a complex interplay between peripheral and central processes when assessing speech in noise performance. In the case of the current investigation, we found that differential performance between younger and older listeners was primarily driven by slight between-group differences in peripheral hearing sensitivity, even though all listeners had *clinically* normal hearing. The ANCOVA analysis (with PTA entered as a covariate) that was used for the current study poses a problem since one would expect very strong correlations between PTA and age. Therefore, it is not certain that age-related differences in performance were captured by this type of analysis. Overall, the effect of aging and spectral degradation on speech in noise recognition remains unresolved.

CHAPTER 5: COMPREHENSIVE DISCUSSION AND SUMMARY

Summary of Results

The current series of experiments sought to better understand possible interactions between aging and the processing of spectrally degraded voice-pitch information. In particular, the results of the current study are intended to predict performance among younger and older adults who utilize CIs in everyday life. Altogether, these results demonstrate senescent decline in degraded voice-pitch coding, particularly when listeners must rely on temporal envelope periodicity cues alone and thereby suggest that older CI recipients might experience a similar deficit at least when initially implanted with the device.

Experiment One: Modulation Frequency Discrimination

Part one

The first part of Experiment 1 measured the ability of younger and older NH listeners to discriminate between sequentially presented SAM noises based on differences in modulation frequency (MFD task). It was hypothesized that older listeners would experience greater difficulty when performing this task, due to age-related decreases in TEP coding. A 3AFC method of constant stimuli design was implemented, and individual results were fitted with psychometric functions in order to obtain values for discrimination thresholds, as well as a slope. The results were highly variable across subjects, but demonstrated poorer average performance among the older cohort. Older listeners exhibited shallower slopes, on average, than did the younger listeners, especially for the 200 Hz condition. Although correlation analyses suggested that high-frequency hearing sensitivity was not related to performance across the two groups, it is still possible that this factor did influence results to some extent.

Part two

We repeated the MFD task (Part one) using low-pass filtered (4,000 Hz cut-off) noise stimuli in order to determine if HFPTA did indeed contribute to the results. A 3-AFC adaptive procedure was used to measure MFD performance. Results showed that while there was a general trend for older listeners to exhibit poorer MFD thresholds compared to younger listeners, there was not a significant between-group difference in performance. This is likely attributed to the reduced number of participants and high variability. These results also suggest, however, that the results of the initial experiment (Part One with unfiltered noise) could have been influenced by between-group differences in high-frequency hearing sensitivity. Nevertheless, the results of Experiment 1 do provide tenuous evidence that discrimination of stimuli based on differences in the frequency of the temporal envelope alone is affected by the age of the adult listener.

Experiment Two: Fundamental-Frequency Discrimination

Experiment 2 was designed to extend from those results obtained in Experiment 1 by measuring across-channel TEP processing in adults. This experiment implemented NBV processing strategies to manipulate the number of spectral channels available to the listener, while the temporal envelope information was unaltered, thereby closely mimicking CI processing. Furthermore, all stimuli were low-pass filtered at 4,000 Hz to limit the influence of differences in hearing sensitivity between the two age groups. An adaptive 3AFC paradigm was used to measure *f0*DLs using NBV harmonic complexes. It was expected that performance would be poorest with one spectral channel and also similar to those results obtained in Experiment 1 as both tasks required the listener to rely solely on temporal envelope cues to discriminate between sounds. We hypothesized, however, that *f0*DLs would be poorer among older listeners, particularly when stimuli were spectrally degraded.

In general, the results of this experiment revealed smaller *f0*DLs for younger listeners, regardless of the degree of spectral resolution and reference *f0* condition. Consistent with previous literature, performance is generally best when stimuli are unprocessed (full spectral cues), but systematically declines as spectral resolution is reduced. While this pattern was true for both age groups, there was also a significant interaction between age and number of available channels. Specifically, between-group differences in *f0*DLs were smaller with greater spectral resolution and larger when stimuli were degraded. In other words, older listeners exhibited poorer performance than younger listeners when forced to rely more on TEP to glean *f0* information. Although the results from Experiment 1 suggested that aging might affect TEP processing, between-group differences in high-frequency hearing thresholds could have contributed to the results. The results from Experiment 2, however, showed that discrimination of TEP was affected by aging when stimuli were controlled for differences in high-frequency hearing sensitivity between groups. Taken together, these results suggest that processing of periodicity information is adversely affected as a consequence of aging, particularly when listeners must rely on periodicity information in the temporal envelope (i.e., unresolved harmonics).

Experiment Three: Gender Identification

Although the results of Experiments 1 and 2 lend support for senescent decline in spectrally degraded voice-pitch coding, it was difficult to determine how such results would translate to performance when speech stimuli were used. Further, while the results of the previous experiments demonstrated the interaction between aging and reduction of spectral cues, the current experiment also investigated how older listeners use varying degrees of temporal envelope information, when spectral cues are kept constant. A single interval gender recognition task was performed in which recorded vowel stimuli (spoken by a male or female talker) were spectrally and/or temporally degraded using NBV processing. In one case, the *f0*s of the talker set were similar (COM), and in the other case, the *f0*s of the talker set were more disparate (EXP).

Younger listeners outperformed older listeners, but only for conditions in which stimuli were spectrally degraded. For the EXP condition, older listeners performed worse at each level of spectral degradation, but were able benefit from additional temporal envelope periodicity information to some extent. For the COM condition, older listeners continued to perform worse than the younger listeners at each level of spectral degradation; however, the performance of younger listeners, but not older listeners, improved as temporal envelope information was increased. These results are consistent with those of Experiments 1 and 2 and provide further evidence for declines in periodicity pitch processing among older listeners.

Experiment Four: Speech-in-Noise Recognition

The results of Experiments 1, 2 and 3 demonstrated that age-related declines in the processing of periodicity information hinder spectrally degraded voice-pitch discrimination and identification when stimuli are presented in isolation. The current experiment, however, measured how younger and older listeners are able to use *f0* cues to parse competing, *simultaneous* speech when stimuli were degraded. Furthermore, listening in noise has been shown to be a difficult task for both older listeners, as well as those with CIs (regardless of age). It was unknown if aging would further complicate perception with a CI when listening to speech in noise. Target sentences (spoken by a male or female talker) were presented in the presence of another competing talker (S_M), 10-talker babble (B_M), or SEM noise stimuli. In keeping with previous literature, we also expected that each type of masker would cause different degrees of masking across the two age groups. It was unknown how spectral degradation would affect performance with various maskers within each age group. Results showed that, while older listeners generally needed a higher TMR in order to perform equally well to the younger subjects, between-group differences could largely be accounted for by peripheral hearing sensitivity for both unprocessed and NBV stimuli. For the current study, it was not found that simultaneous parsing of *f0* information was at all influenced by the age of the listener, regardless of the degree of spectral resolution or type of masker.

Overall Summary of the Present Results

Taken together, the results of the current experiments suggest that, in some cases, aging alters the coding of periodicity information. This seems to be particularly true when spectral cues are reduced or absent, and listeners must therefore rely on periodicity information in the temporal envelope to glean *f0* information. In general, the findings support our hypothesis of reduced TEP perception in the aging auditory system. The results of Experiments 1 and 2 demonstrate that aging affects both within-and across-channel TEP coding of non-speech stimuli, while the results of Experiment 3 reveal that such age-related changes carry over to the perception of speech stimuli, a case in which *f0* cues are dynamic and are also accompanied by additional cues to aid in discerning voice-pitch. The results of Experiment 3 also provide further support for declines in temporal envelope processing with advanced age, as older listeners showed a reduced capacity to use additional TEP information when spectral cues were reduced. Although Experiments 1-3 generally lend support for senescent decline in TEP processing, the results of Experiment 4 do not reinforce this claim. In Experiment 4, we did not find that aging alters the recognition of masked speech, regardless of masker type and degree of spectral resolution. Conversely, we found that the ability to segregate simultaneous signals was related to peripheral hearing sensitivity. Although we hypothesized that masked speech recognition might be related to cognitive function, we found no such correlation in the present study.

It should also be noted that, although older participants exhibited *clinically* normal audiometric thresholds, slight discrepancies in hearing sensitivity did exist among this cohort when compared to those of the younger participants. In most cases, stimuli were manipulated or presented in a manner in which we would not anticipate that age-related differences in performance would significantly contribute to performance. For example, in most cases stimuli were low-pass filtered to contain only spectral content in which older listeners did have clinically normal audiometric thresholds (≤ 4000 Hz). Further, all stimuli were presented at an appropriate sensation level with respect to the hearing thresholds of all listeners. We did find that hearing sensitivity contributed to performance in Experiment 4 (speech-in-noise task), but do not expect that such contributions would hold true for Experiments 1-3 (stimuli presented in isolation). Regardless, it is always possible that age-related differences in hearing sensitivity, and not aging alone *per se*, produced the generally poorer performance of the older listeners.

In summary, these results do suggest that aging affects the perception of periodicity coding when non-speech and simple speech stimuli are presented sequentially or in isolation. Although it is possible that the same is true when contextual stimuli are presented simultaneously, the findings in the current study do not support such a hypothesis.

Applicability of Noise-band Vocoding to Estimate Performance in Cochlear Implant Listeners

Effect of cochlear implant processing strategy on pitch perception

Several studies have demonstrated that, in most cases, results obtained using NBV techniques closely match those obtained in actual CI users (Cullington & Zeng, 2008; Friesen et al., 2001; Fu & Nogaki, 2005; Laneau et al., 2006; Stickney et al., 2004; Vongphoe & Zeng, 2005). There are, however, limitations to using CI simulations to predict performance of CI users, particularly when estimating pitch perception. For example, Laneau et al. (2006) measured pitch perception tasks such as electrode discrimination and *f0*DLs in four CI users and five NH listeners who listened to NBV stimuli. Four different processing strategies were tested, and the principle difference between each type of strategy was the filter-width characteristics. Overall, results showed that, on the whole, results using CI simulations were fairly comparable to those obtained in CI users. The results, however, were dependent upon the processing strategy. For the *f0*DL task, the CI users performed better using some strategies, while the NH listeners performed better using other strategies. When stimuli were manipulated to independently test the use of place and temporal cues, there were some strategies in which use of place and temporal cues was better in NH compared to CI listeners, and in other strategies the opposite was true. Stone and colleagues (Stone, Fullgrabe, & Moore, 2008) also noted that bandwidth impacts the saliency of TEP cues.

The overall significance of the data reported in Laneau et al (2006) is that the ability to perceive *f0* information was significantly affected by several specific differences across the processing strategies. The authors also reported notable intersubject variability in the performance of both CI and NH listeners. Lastly, the authors demonstrated that compression characteristics can significantly alter the ability to perceive *f0* information in the temporal envelope. Specifically, CI processing requires considerable amplitude compression due to the exceedingly small dynamic range of CI users. Results suggest that lack of compression in acoustic simulations results in an underestimate of modulation sensitivity compared to that observed in actual CI users. In other words, it is possible that because compression was not manipulated in the current study, that our results underestimate the ability of CI users to hear voice-pitch via temporal envelope cues. It is unknown, however, how aging would influence this interaction.

Comments on spectral shifting

Cochlear implants are designed to take advantage of the tonotopic organization in the cochlea which promotes place coding of pitch information. Surgical limitations, however, prevent the electrode array from being fully inserted into the cochlea and in most cases, the most apical electrode is only inserted about 20-25 mm from the base. According to Greenwood's frequency-place map (Greenwood, 1990), the most apical electrode in such instances would be stimulating a site in the cochlea primed to receive information ranging from approximately 500 Hz (25 mm insertion) to 1000 Hz (20 mm insertion). Regardless of the insertion depth, CI device processes and transmits those frequencies most important for speech understanding (e.g., 100-8000 Hz).This thereby creates a spectral mismatch between the analysis band (frequency content of incoming stimuli) and the carrier band (electrode-to-place alignment in the cochlea), and is commonly referred to as *spectral shifting*.

Numerous studies have illustrated the adverse consequences of spectral shifting on pitch perception and speech understanding in CI users and those listening to CI simulations (Baskent & Shannon, 2003; 2004; 2005; 2007; Fu & Shannon, 1999; Laneau et al., 2006; Schvartz et al., 2008). In particular, Schvartz et al (2008) used showed that phoneme recognition in middle-aged and older listeners was more susceptible to spectral shifting compared to younger listeners, when NBV stimuli were used to estimate performance. Other studies have demonstrated, however, that the negative impact of spectral shifting is most is detrimental to performance immediately after implantation and listeners are able to adapt over time (Fu, Shannon & Galvin, 2002; Fu, Nogaki & Galvin, 2005; Li & Fu, 2007).

The results of the Experiments 1, 2 and 3 suggest that even without spectral shifting older listeners exhibit difficulty perceiving TEP cues. It is possible that the addition of spectral shifting would cause even greater age-related disparities in performance on such a task. The results of the present study thereby might be more predictive of the performance among CI users after sufficient adaptation has occurred, and they may actually underestimate voice-pitch coding problems exhibited by older listeners immediately after implantation. Likewise, it is possible that larger between-group differences would have been noted in Experiment 4 if stimuli were spectrally shifted.

Impact of prolonged deafness and neural degeneration

Lastly, one particular difficulty of using NBV stimuli to the predict performance of actual CI users is the co-variation between age at implantation and duration of deafness. Research has shown that duration of deafness can result in deleterious performance among CI users (Blamey, Arndt, Bergeron, Bredberg, Brimacombe, Facer, Larkey et al., 1996; Leung et al., 2005; Shea et al., 1990), which is assumed to originate from cellular and neural degeneration at many levels in the auditory system (See Leake & Rebscher, 2004, for review). Obviously, the influence of neural degeneration cannot be conveyed when using CI simulations. Therefore, it is possible that aging coupled with consequences of prolonged hearing loss would create poorer performance among elderly CI users. For example, it has been demonstrated in animals that prolonged deafness can decrease temporal phase locking at the level of the IC (Snyder, Leake, Rebscher, & Beitel, 1995). It is therefore logical that aging coupled with neural degeneration subsequent to long-term auditory deprivation will increase temporal processing deficits in the CANS. Lastly, the current study only used older listeners with good peripheral hearing sensitivity and therefore this posed a restriction on the upper limit on the age of participants. It is logical that performance in older listeners would continue to decline with increasing age.

Summary and Future Directions

While the current study helps to further explain possible interactions between aging and cochlear implantation, there are several questions which remain to be addressed. First, the current experiments only estimated performance using CI simulations, but did not substantiate results using actual CI users. While CI simulations afford strict control of experimental variables, performance in actual CI users is certain to be much more variable. Therefore, the effect of age at implantation on voice-pitch perception should be further confirmed using CI users. Second, it is important to understand how perception would change with practice and training. From a clinical perspective, it is crucial to determine how age might impact the absolute potential of adult CI users if given adequate opportunities to improve performance. Lastly, it was hoped that results from Experiment 4 would reveal how aging affects speech perception in challenging environments, even when good performance is observed under quiet conditions. Although the results of Experiment 4 were rather inconclusive, understanding how age might impact performance in unfavorable listening conditions remains an important issue since such situations are most common in everyday communication environments.

Overall, the results of the current study do lend support for age-related declines in voice-pitch coding and suggest altered spectro-temporal coding of pitch perception in older adults. It is hypothesized that these results reflect changes in modulation processing in the CANS that are associated with aging. These results also have significant implications for the rehabilitation of older CI users and perhaps argue for targeted rehabilitation among this cohort. Despite such findings, work is needed to further elucidate how aging impacts the perception of spectrally-degraded stimuli and the degree to which this interaction affects adult CI recipients.

REFERENCES

Abel, S. M., Krever, E. M., & Alberti, P. W. (1990). Auditory detection, discrimination and speech processing in ageing, noise-sensitive and hearing-impaired listeners. *Scandinavian Audiology, 19*, 43-51.

Agus, T. R., Akeroyd, M. A., Gatehouse, S., & Warden, D. (2009). Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise. *Journal of the Acoustical Society of America, 126*, 1926-1940.

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*, S53-S71.

Ameson, A. (1982). Presbycusis-a loss of neurons in the human cochlear nuclei. *Journal of Laryngology and Otology, 96*, 503-511.

Amos, N. E. & Humes, L. E. (2007). Contribution of high frequencies to speech recognition in quiet and noise in listeners with varying degrees of high-frequency sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research, 50,* 819-834.

ANSI (2004). American National Standard Specification for Audiometers. ANSI S3.6-2004. (American National Standards Institute, New York).

Arehart, K. H. (1994) Effects of harmonic content on complex tone fundamental frequency discrimination in hearing-impaired subjects. *Journal of the Acoustical Society of America*, *95*, 3574-3585.

Backoff, P. M., Palombi, P. S., & Caspary, D. M. (1999). GABA- and glycinergic inputs shape coding of AM in chilchilla cochlear nucleus. *Hearing Research, 134*, 77-88.

Backoff, P. M., Shadduck-Palombi, P., Caspary, D. M. (1997). Glycinergic and GABAergic inputs affect short-term suppression in the cochlear nucleus. *Hearing Research, 110*, 155-163.

Baddeley, A. D. (1986). *Working memory*. New York, NY: Oxford University Press.

Baer, T. & Moore, B. C. J. (1993). Effects of spectral smearing on the intelligibility of sentences in noise. *Journal of the Acoustical Society of America, 94*, 1229-1241.

Baer, T. & Moore, B. C. J. (1994). Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *Journal of the Acoustical Society of America, 95, 2277-2280.*

Balabuer-Ballester, E., Clark, N. R., Coath, M., Krumbholz, K. & Denham, S. L. (2009). Understanding pitch perception as a hierarchical process with top-down modulation. *PLoX Computational Biology, 5*, e1000301.

Banay-Schwartz, M., Lajtha, A., & Palkovits, M. (1989). Changes with aging in the levels of amino acids in rat CNS structural elements. II. Taurine and small neutral amino acids. *Neurochemical Research*, *14*, 563–570.

Barth, C. D. & Burkard, R. (1993). Effects of noise burst rise time and level on the human brainstem auditory evoked response. *International Journal of Audiology, 32,* 225-233.

Baskent, D., & Shannon, R. V. (2003). Speech recognition under conditions of frequency-place compression and expansion. *Journal of the Acoustical Society of America, 113,* 2064-2076.

Baskent, D., & Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners. *Journal of the Acoustical Society of America, 116*, 3130-3140.

Baskent, D., & Shannon, R. V. (2005). Interactions between cochlear implant electrode insertion depth and frequency-place mapping. *Journal of the Acoustical Society of America, 117*, 1405-1416.

Baskent, D., & Shannon, R. V. (2007). Combined effects of frequency compression-expansion and shift on speech recognition*. Ear and Hearing, 28*, 227-289.

Bernstein, J. G. & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 125*, 3358-3372.

Bernstein, J. G. & Oxehnam, A. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *Journal of the Acoustical Society of America, 113*, 3323-3334.

Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., Larky, J., Lindstrom, B., Nedzelski, J., Peterson, A., Shipp, D., Staller, S., and Whitford, L. (1996). Factors affecting auditory performance of post-linguistically deaf adults using cochlear implants. *Audiology and Neurotology, 1,* 293–306.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, 17*, 97-110. University of Amsterdam.

Boersma, P. & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.0.34). [Computer Program]. Retrieved: October 1, 2008, from http://www.praat.org.

Brokx, J. P. L., & Nootebohm, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10,* 23–36.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109*, 1101-1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America, 110*, 2527-2538.

Burkard, R. (1991). Effects of noise burst rise time and level on the gerbil brainstem auditory evoked response. *Audiology, 30*, 47-58.

Burns, E. M., & Viemeister, N. F. (1976). Nonspectral pitch. *Journal of the Acoustical Society of America, 60*, 863-869.

Caspary, D. M., Ling, L., Turner, J. G., & Hughes, L. F. (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *Journal of Experimental Biology, 211*, 1781-1791.

Caspary, D. M., Milbrandt, J. C. & Helfert, R. H. (1995). Central auditory aging: GABA changes in the inferior colliculus. *Experimental Gerontology, 30*, 349-360.

Caspary, D. M., Pazara, K. E., Kossl, M., Faingold, C. L. (1987). Strychnine alters the fusiform cell output from the dorsal cochlear nucleus. *Brain Research, 417*, 273–282.

Caspary, D. M., Shadduck-Palombi, P. S., & Hughes, L. F. (2002). GABAergic inputs shape responses to amplitude modulated stimuli in the inferior colliculus. *Hearing Research,* 168-173.

Chan, V., Tong, M., Yue, V., Wong, T., Leung, E., Yuen, K., & van Hasselt, A. (2007). Performance of older adult cochlear implant users in Hong Kong. *Ear and Hearing, 28*, 52S-55S.

Chatelin, V., Kim, E. J., Driscoll, C., Larkey, J., Polite, C. Price, L., & Lalwani, A. K. (2004). Cochlear implant outcomes in the elderly. *Otology and Neurology, 25*, 298-301.

Chatterjee, M. & Oba, S. I. (2004). Across- and within- channel envelope interactions in cochlear implant listeners. *Journal of the Association for Research in Otolaryngology, 5*, 360-375.

Chatterjee, M. & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research, 235*, 143-156.

Chatterjee, M. & Shannon, R. V. (1998). Forward masked excitation patterns in multielectrode electrical stimulation. *Journal of the Acoustical Society of America, 103*, 2565-2572.

Childers, D. G. & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *Journal of the Acoustical Society of America, 90*, 1841-1856.

Cooper, W. E. (1983). The perception of fluent speech. *Annals of the New York Academy of Sciences, 405*, 48-63.

Craik, F. I. M. (2007). The role of cognition in age-related hearing loss. *Journal of the American Academy of Audiology, 18*, 539-547.

Cullington, H. E. & Zeng, F. G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant and implant simulation subjects. *Journal of the Acoustical Society of America, 123*, 450-461.

Danhauer, J. L., & Leppler, J. G. (1979). Effects of four noise competitors on the California Consonant Test. *Journal of Speech and Hearing Disorders, 44*, 354-362.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America, 114,* 2913-2922.

Dau, T., Kollmeier, B., & Kohlrausch. A. (1997a). Modeling audiotyr processing of amplitude modulation I. Detection of masking with narrow-band carriers. Journal of the Acoustical Society of America, 102, 2892-2905.

Dau, T., Kollmeier, B., & Kohlrausch. A. (1997b). Modeling audiotyr processing of amplitude modulation II. Spectral and temporal integration.. Journal of the Acoustical Society of America, 102, 2906-2919.

de Cheveigné, A. (2005). Pitch Perception Models. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper. (Eds)., *Pitch: Neural Coding and Pitch Perception* (pp. 169-233). New York, New York: Springer.

Dirks, D. D., Morgan, D. E. & Dubno, J. R. (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorders, 47*, 114-123.

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging, 4*, 500-503.

Dorman, M. F., Loizou, P. C., Fitzke, J. & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *Journal of the Acoustical Society of America, 104,* 3583-3585.

Dorman, M. F., Loizou, P. C. & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America, 102,* 2403-2411.

Dubno, J. R., Dirks, D. D. & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America, 76*, 87-96.

Dubno, J. R., Horwitz, A. R. & Ahlstrom, J. B (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *Journal of the Acoustical Society of America, 111*, 2897-2907.

Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2003). Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. *Journal of the Acoustical Society of America, 113*, 2084-2094.

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking. *Journal of the Acoustical Society of America, 113*, 2984-2987.

Faulkner, A., Rosen, S., & Wilkinson, L. (2001). Effect of the number of channels and speech-in-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. *Ear and Hearing, 22*, 431-438.

Festen, J. M. & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America, 88*, 1725-1736.

Fetterman, B. L., & Domico, E. H. (2002). Speech recognition in background noise for cochlear implant patients. *Otolaryngology-Head and Neck Surgery, 126*, 257-263.

Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., Runge-Samuelson, C., & Wackym, P. A. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear and Hearing, 25*, 375-387.

Fishman, K. E., Shannon, R. V., & Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *Journal of Speech, Language and Hearing Research, 40,* 1201-1215.

Fitzgibbons, P. J., & Gordon-Salant, S. (1994). Age effects on measures of auditory duration discrimination. *Journal of Speech and Hearing Research, 37*, 662-670.

Fitzgibbons, P. J., & Gordon-Salant, S. (1998). Auditory temporal order perception in younger and older adults. *Journal of Speech and Hearing Research, 41*, 1052–1060.

Folstein MF, Folstein SE, McHugh PR (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research 12*, 189–198.

Formby, C. (1985). Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners. *Journal of the Acoustical Society of America, 78*, 70-77.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America, 110*, 1150-1163.

Frisina, R. D. (2001). Subcortical neural coding mechanisms for auditory temporal processing. *Hearing Research, 158*, 1-27.

Frisina, R. D., Smith, R. L. & Chamberlain, S. C. (1985). Differential encoding of rapid changes in sound amplitude by second-order auditory neurons. *Experimental Brain Research, 60*, 417-422.

Frisina, R.D., Walton, J. P., & Karcich, K. J. (1994). Dorsal cochlear nucleus single neurons can enhance temporal processing capabilities in background noise. *Experimental Brain Research, 102*, 160–164.

Fu, Q-J. (2002). Temporal processing and speech recognition in cochlear implant users. *Neuroreport, 13,* 1635-1639.

Fu, Q-J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology, 5*, 253-260.

Fu, Q-J., Chinchilla, S., Nogaki, G., & Galvin, J. J. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *Journal of the Acoustical Society of America, 118*, 1711-1718.

Fu, Q. J. & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology, 6*, 19-27.

Fu, Q-J., Nogaki, G., & Galvin, J. J. (2005). Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology, 6*, 180-189.

Fu, Q-J., & Shannon, R. V. (1999). Effects of electrode configuration and frequency allocation on vowel recognition with Nucleus-22 cochlear implant. *Ear and Hearing, 20*, 332-344.

Fu, Q-J. & Shannon, R. V. (2002). Frequency mapping in cochlear implants. *Ear and Hearing, 23*, 339-348.

Fu, Q-J., Shannon, R. V. & Galvin, J. J. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *Journal of the Acoustical Society of America,122*, 1664-1674.

Gifford, R. H., Bacon, S. P. & Williams, E. J. (2007). An examination of speech recognition in a modulated background and of forward masking in younger and older listeners. *Journal of Speech, Language, and Hearing Research, 50*, 857-864.

Gleich, O., Hamann, I., Kittel, M. C., Klum, G. M. & Strutz, J. (2007). Forward masking in gerbils: the effect of age. *Hearing Research, 223*, 122-128.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America, 54*, 1496-1516.

Gonzalez, J., & Oliver, J. C. (2005). Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *Journal of the Acoustical Society of America, 118*, 461-470.

Gooler, D. M. & Feng, A. S. (1992). Temporal coding in the frog auditory midbrain: the influence of duration and rise-time on the processing of complex amplitude-modulated stimuli. *Journal of Neurophysiology, 67*, 1-22.

Gordon-Salant, S. & Fitzgibbons, P. J. (1995). Recognition of multiply degraded speech by young and elderly listeners. *Journal of Speech and Hearing Research, 38,* 1150-1156.

Grant, K. W., Summers, V., & Leek, M. R. (1998). Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners. Journal *of the Acoustical Society of America, 104,* 1051-1060.

Green, T., Faulkner, A., & Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling in cochlear implants. *Journal of the Acoustical Society of America,* 112, 2155-2164.

Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice-pitch in continuous interleaved sampling cochlear implants. *Journal of the Acoustical Society of America, 116*, 2298-2310.

Green, T., Faulkner, A., Rosen, S., & Macherey, O. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *Journal of the Acoustical Society of America, 118,* 375-385.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *Journal of the Acoustical Society of America, 87*, 2592-2605.

Grose, J. H., Hall, J. W., & Buss, E. (2006). Temporal processing deficits in the pre-senescent auditory system. *Journal of the Acoustical Society of America, 119*, 2305-2315.

Grose, J. H., Mamo, S. K., & Hall, J. W. (2009). Age effects in temporal envelope processing: Speech unmasking and auditory steady state responses. *Ear and Hearing, 30*, 568-575.

Guerts, L. & Wouters, J. (2004). Better place-coding of the fundamental frequency in cochlear implants. *Journal of the Acoustical Society of America, 115*, 844-852.

Gustafsson, H. A. & Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *Journal of the Acoustical Society of America, 95*, 518-529.

Hanna, T. E. (1992). Discrimination and identification of modulation rate using a noise carrier. *Journal of the Acoustical Society of America, 91*, 2122-2128.

Hargus, S. E. & Gordon-Salant, S. (1995). Accuracy of speech intelligibility index predictions for noise-masked younger listeners with normal hearing and for elderly listeners with hearing impairment. *Journal of Speech and Hearing Research, 38*, 234-243.

Hasher, L., Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. In G. H. Bower, (ed). *The Psychology of Learning and Motivation* (pp. 193-225). New York: Academic Press.

Heil, P. (2001). Representation of sound onsets in the auditory system. *Audiology & Neuro-otology, 6*, 167-172.

Heil, P. (2003). Coding of temporal onset envelope in the auditory system. *Speech Communication, 41,* 123-134.

Helfer, K. S. & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing, 29*, 87-98.

Helfer, K. S. & Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech and Hearing Sciences, 33*, 149-155.

Helfert, R. H., Sommer, T. J., Meeks, J., Hofstetter, P., & Hughes, L. F. (1999). Age-related synaptic changes in the central nucleus of the inferior colliculus of Fischer-344 rats. *The Journal of Comparative Neurology, 406*, 285-298.

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *Journal of the Acoustical Society of America, 113*, 2861-2873.

Henry, B. A., Turner, C. W., & Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. *Journal of the Acoustical Society of America, 118*, 1111-1121.

Houtgast, T. & Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International Journal of Audiology, 47*, 287-295.

Houtsma, A. J. M., and Smurzyuski, J. (1990). Pitch identification and Discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America, 87*, 304-310.

Humes, L. E. (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *Journal of the Acoustical Society of America, 112*, 1112-1132.

Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older listeners. *Journal of the American Academy of Audiology, 18*, 590-603.

Humes, L. E. & Christopherson, L. (1991). Speech identification difficulties of hearing-impaired elderly persons. *Journal of Speech and Hearing Research, 34*, 686-693.

Humes, L. E., Lee, J. H. & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *Journal of the Acoustical Society of America, 120*, 2926-2937.

Humes, L. E. & Roberts, L. (1990). Speech-recognition difficulties for the hearing impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research, 33*, 726-735.

Hygge, S., Ronnberg, J., Larsby, B., & Arlinger, S. (1992). Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds. *Journal of Speech and Hearing Research, 35*, 208-215.

Javel, E. (1980). Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones. *Journal of the Acoustical Society of America, 68*, 133-146.

Jin, S. H. & Nelson, P. B. (2006). Speech perception in gated noise: the effects of temporal resolution. *Journal of the Acoustical Society of America, 119*, 3097-3108.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews, 84*, 541-577.

Joris, P. X. & Yin, T. C. T. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *Journal of the Acoustical Society of America, 91*, 215-232.

Kim, D. O., Sirianni, J. G., Chang, S. O. (1990). Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: analysis with autocorrelation/power spectrum. *Hearing Research, 45*, 95–113.

Kinsella, K., & Velkoff, V. A. (2001). U.S. Census Bureau, Series P95/01-1, *An Aging World: 2001*, U.S. Government Printing Office, Washington, DC, 2001.

Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*, 820-857.

Kolston, J., Osen, K. K., Hackney, C. M., Ottersen, O. P., Storm-Mathisen, J. (1992). An atlas of glycine- and GABA-like immunoreactivity and colocalization in the cochlear nuclear complex of the guinea pig. *Anat Embryol (Berl), 186*, 443–465.

Kovacic, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *Journal of the Acoustical Society of America, 126,* 762-775.

Landsberger, D. M. (2008). Effects of modulation wave shape on modulation frequency discrimination with electrical hearing. *Journal of the Acoustical Society of America (Express Letters), 124*, EL21-EL27.

Laneau, J., Moonen, M. & Wouters, J. (2006). Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlea r implants. *Journal of the Acoustical Society of America, 119*, 491-506.

Langer, G., & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat I. Neuronal mechanisms. *Journal of Neurophysiology, 60*, 1799-1822.

Leake, P. A. & Rebscher, S. J. (2004). Anatomical considerations and long-term effects of electrical stimulation. In F-G Zeng, A. N. Popper, R. R. Fay (Eds). *Cochlear implants: Auditory prostheses and electric hearing.* (pp. 101-147). New York, NY: Springer-Verlag.

Leigh-Paffenroth, E. E., & Fowler, C. G. (2006). Amplitude-modulated auditory steady-state responses in younger and older listeners. *Journal of the American Academy of Audiology, 17*, 582-597.

Leung, J., Wang, N-Y., Yeagle, J. D., Chinnici, J., Bowditch, S., Francis, H. W., & Niparko, J. K. (2005). Predictive models for cochlear implantation in elderly candidates. *Archives of Otolaryngology, Head and Neck Surgery, 131*, 1049-1054.

Levitt, H . (1971). Transformed up-down methods in psychophysics. *Journal of the Acoustical Society of America, 49*, 467-477.

Lew, H., & Jerger, J. (1991). Effect of linguistic interference on sentence identification. *Ear and Hearing, 12,* 365-367.

Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *Journal of Experimental Psychology: Human Perception and Performance, 30*, 1077-1091.

Li, T., & Fu, Q-J. (2007). Perceptual adaptation to spectrally shifted vowels: Training with nonlexical labels. *Journal of the Association for Research in Otolaryngology, 8,* 32-41.

Licklider, J. C. R. (1951). A duplex theory of pitch perception (reproduced in Schubert, 1979, 155-160), *Experientia, 7*, 128-134.

Linke, C. E. (1973). A study of pitch characteristics of female voices and their relationship to vocal effectiveness. *Folia Phoniatrica (Basel), 25,* 173-185.

Lister, J., Besing, J., and Koehnke, J. (2002). Effects of age and frequency disparity on gap discrimination. *Journal of the Acoustical Society of America, 111*, 2793–2800.

Loizou, P. C. (1998). Mimicking the human ear: An overview of signal-processing strategies for converting sound into electrical signals in cochlear implants. *IEEE Signal Processing Magazine, September,* 101-130.

Meddis, R. & Hewitt, M. J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America, 89,* 2883-2894.

Meddis, R. & Hewitt, M. J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity. *Journal of the Acoustical Society of America, 91*, 233-245.

Merzenich, M. M., Knight, P. L., & Roth, G. L. (1975). Representation of the cochlea within primary auditory cortex in the cat. *Journal of Neurophysiology, 38,* 231-249.

Milbrandt, J. C., Albin, R. L., Turgeon, S. M., & Caspary, D. M. (1996). GABA receptor binding in the aging rat inferior colliculus. *Neuroscience, 73,* 449-458.

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin, 44*, 105-129.

Miller, G. A., and Taylor, W. G. (1948). .The perception of repeated bursts of noise. *Journal of the Acoustical Society of America, 20*, 171-182.

Moller, A. R. & Rees, A. (1986). Dynamic properties of the responses of single neurons in the inferior colliculus of the rat. *Hearing Research, 24*, 203-215.

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

Moore, B.C.J., Glasberg, B.R., Flanagan, H.J., Adams, J. (2006). Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure. *Journal of the Acoustical Society of America*, 119, 480–490.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America, 77*, 1853-1860.

Moore, B. C. J. & Ohgushi, K. (1993). Audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America, 93*, 452-461.

Moore, B.C. J., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *Journal of the Acoustical Society of America, 91*, 2881-2893.

Mowbray, G. H., Gebhard, J. W., & Byham, C. L. (1956). Sensitivity to changes in the interruption rate of white noise. *Journal of the Acoustical Society of America, 1*, 106-110.

Murry, T. & Singh, S. (1980). Multidimentional analysis of male and female voices. *Journal of the Acoustical Society of America, 68*, 1294-1300.

Nelson, P. B., & Jin, S. H. (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America, 115,* 2286-2294.

Nelson, P. B., Jin, S. H., Carney, A. E. & Nelson, D. A. (2003). Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America, 113*, 961-968.

Noble, W., Tyler, R. S., Dunn, C. C. & Bhullar, N. (2009). Younger- and older-age adults with unilateral and bilateral cochlear implants: Speech and spatial hearing self-ratings and performance. *Otology & Neurotology, 30*, 921-929.

Ostroff, J. M., McDonald, K. L., Schneider, B. A., & Alain, C. (2003). Aging and the processing of sound duration in human auditory cortex. *Hearing Research, 181*, 1-7.

Oxenham, A. J., & Simonson, A. M. (2009). Masking for low- and high-pass filtered speech in the presence of noise and single-talker interference. *Journal of the Acoustical Society of America, 125*, 457-468.

Pasanisi, E., Bacciu, A., Vincenti, V., Guida, M., Barbot, A., Berhenti, M. R., & Bacciu, S. (2003). Speech recognition in elderly cochlear implant recipients. *Clinical otolaryngology and allied sciences, 28*, 154-157.

Patterson, R. D., Johnson-Davis, D. & Milroy, R. (1978). Amplitude-modulation noise: The detection of modulation versus the detection of modulation rate. *Journal of the Acoustical Society of America, 63*, 1904-1911.

Pearson, J. D., Morrell, C. H., Gordon-Salant, S., Brant, L. J., Metter, E. J., Klein, L. L., & Fozard, J. L. (1995). Gender differences in a longitudinal study of age-associated hearing loss. *Journal of the Acoustical Society of America, 97*, 1196-1205.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in study of the vowels. *Journal of the Acoustical Society of America, 24,* 175-184.

Phillips, S. L., Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. H. (1994). Auditory duration discrimination in young and elderly listeners with normal hearing. *Journal of the American Academy of Audiology*, *5*, 210–215.

Pichora-Fuller, M. K., Schneider, D. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97,* 593-608.

Pichora-Fuller, M. K. & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42*, 2S11-2S16.

Plack, C. J. & Carlyon, R. P. (1995). Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. *Journal of the Acoustical Society of America, 98*, 1355-1364.

Plack, C. J., & Oxenham, A. J. (2005). Overview: The psychophysics of pitch. In C. J. Plack, A. J. Oxenham, R. R. Fay, A. N. Popper (Eds). *Pitch: Neural coding and perception*. (pp. 7-55). New York, NY: Springer.

Plomp, R. (1964). The ear is a frequency analyzer. *Journal of the Acoustical Society of America, 108*, 696-705.

Plomp, R. (1967). The pitch of complex tones. *Journal of the Acoustical Society of America, 41*, 1526-1533.

Plomp, R. & Mimpen, A. M. (1968). The ear is a frequency analyzer II. *Journal of the Acoustical Society of America, 43*, 764-767.

Plomp, R. & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *Journal of the Acoustical Society of America, 66*, 1333-1342.

Pollack,I . (1952). Auditory flutter. *American Psychologist, 6* , 544-554.

Pressnitzer, D., & Patterson, R. D. (2001). Distortion products and the pitch of harmonic complex tones. In: D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, R. Schoonhoven, (Eds). *Physiological and Psychophysical Bases of Auditory Function. Maastricht*: (pp. 97-104). Ashaker.

Purcell, D. W., John, S. M., Schneider, B. A., & Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope following responses. *Journal of the Acoustical Society of America, 116,* 3581-3593.

Qin, M. K. & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America, 114,* 446-454.

Qin, M. K. & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and Hearing, 26*, 451-460.

Rees. A., & Moller, A. R. (1983). Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hearing Research, 10*, 301-330.

Rees, A. & Moller, A. R. (1987). Stimulus properties influencing the responses of the inferior colliculus neurons to amplitude-modulated sounds. *Hearing Research, 27*, 129-143.

Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *Journal of the Acoustical Society of America, 42*, 191-198.

Rose, J. E., Galambos, R., & Hughes, J. R. (1959). Microelectrode studies of the cochlear nuclei of the cat. *Bulletin of the Johns Hopkins Hospital, 104*, 211-251.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society (London), 336*, 367-373.

Rossi-Katz, J. & Arehart, K. H. (2009). Message and talker identification in older adults: effects of task, distinctiveness of talkers' voices and meaningfulness of the competing message. *Journal of Speech, Language, and Hearing Research, 52*, 435-453.

Schatteman, T. A., Hughes, L. F., & Caspary, D. M. (2008). Aged-related loss of temporal processing: Altered responses to amplitude modulated tones in the rate dorsal cochlear nucleus. *Neuroscience, 154*, 329-337.

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18*, 559-572.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D. & Lamb, M. (1994). Gap detection and the precedence effect in young and old adults. *Journal of the Acoustical Society of America, 95*, 980-991.

Schouten, J.F. (1940). The residue and the mechanism of hearing. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A., 43*, 991-999.

Schuknecht, H. (1955). Presbycusis. *Laryngoscope, 65*, 402-419.

Schuknecht, H. (1989). Pathology of presbycusis. In: J. Goldstein, H. Kashima, C. Koopman (Eds.), *Geriatric Otorhinolaryngology*, (40-44). Philadelphia: B. C. Decker.

Schuknecht, H. & Gacek, M. (1993). Cochlear pathology in presbycusis. *Annals of Oto-Rhino-Laryngology, 102*, 1-16.

Schvartz, K. C., Chatterjee, M., & Gordon-Salant, S. (2008). Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners. *Journal of the Acoustical Society of America.*

Shaddock-Palombi, P., Backoff, P. M., & Caspary, D. M. (2001). Responses of young and aged rat inferior colliculus neurons to sinusoidally amplitude modulated stimuli. *Hearing Research, 153,* 174-180.

Shannon, R. V. (1989). Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. *Journal of the Acoustical Society of America, 85*, 2587-2592.

Shannon, R. V. (1990). Forward masking in patients with cochlear implants. *Journal of the Acoustical Society of America, 88,* 741-744.

Shannon, R. V. (1993). Quantitative comparison of electrically and acoustically evoked auditory perception: implications for the location of perceptual mechanisms. *Progress in Brain Research, 97*, 261-269.

Shannon, R. V., Fu, Q-J., & Galvin J. J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-laryngologica Supplementum, 552*, 50-54.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekilid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303-304.

Shea, J. J., Domico, E. H., & Orchik, D. J. (1990). Speech recognition ability as a function of duration of deafness in multi-channel cochlear implant patients. *The Laryngoscope, 100,* 223-226.

Shepherd, R. K., Roberts, L. A., & Paolini, A. G. (2004). Long-term sensorineural hearing loss induces functional changes in rat auditory nerve. *European Journal of Neuroscience, 20*, 3131-3140.

Shofner, W. P. (1991). Temporal representation of ripple noise in the anteroventral cochlear nucleus of the chinchilla. *Journal of the Acoustical Society of America, 90*, 2450-2466.

Shofner, W. P. (1999). Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: quantitative analysis of neural autocorrelograms of PL and chopper units. *Journal of Neurophysiology, 81*, 2662-2674.

Simpson, S. A. & Cooke, M. (2005). Consonant identification in N-talker babble is a non-monotonic function of N (L). *Journal of the Acoustical Society of America, 118*, 2775-2778.

Snell, K. B. (1997). Age-related changes in temporal gap detection. *Journal of the Acoustical Society of America, 101*, 2214–2220.

Snell, K. B. & Frisina, D. R. (2000). Relationships among age-related differences in gap detection and word recognition. *Journal of the Acoustical Society of America*, *107*, 1615-1626.

Snyder, R. L., Leake, P. A., Rebscher, S. J. & Beitel, R. E. (1995). Temporal resolution of neurons in cat inferior colliculus: effects of neonatal deafening and chronic intra-cochlear electrical stimulation. *Journal of Neurophysiology, 73*, 449-467.

Snyder, R. L., Vollmer, M., Moore, C. M., Rebscher, S. J., Leake, P. A., & Beitel, R. E. (2000). Responses of inferior colliculus neurons to amplitude modulated intracochlear electrical pulses in deaf cats. *Journal of Neurophysiology, 84*, 166-183.

Stickney, G. S., Assmann, P. G., Chang, J., & Zeng, F. G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *Journal of the Acoustical Society of America, 122*, 1069-1078.

Stickney, G. S., Zeng, F. G., Litovsky, R. & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *Journal of the Acoustical Society of America, 116*, 1081-1091.

Stone, M. A., Fullgrabe, C., & Moore, B. C. J. (2008). Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region. *Journal of the Acoustical Society of America, 124,* 2272-2282.

Stuart, A. & Phillips. D. P. (1996). Word recognition in continuous and interrupted broadband noise by younger normal-hearing, older normal-hearing, and presbyacusic listeners. *Ear and Hearing, 17*, 478-489.

Suga, N. (1971). Responses of inferior collicular neurons of bats to tone bursts with different rise times. *The Journal of Physiology, 217*, 159-177.

Summers, V. & Leek, M. R. (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech, Language, and Hearing Research, 41*, 1294-1306.

Summers, V. & Molis, M. R. (2004). Speech recognition in fluctuating and continuous maskers: effects of hearing loss and presentation level. *Journal of Speech, Language and Hearing Research, 47,* 245-256.

Takahashi, G. A. & Bacon, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech and Hearing Research, 35*, 1410-1421.

ter Keurs, M., Festen, J. M., & Plomp, R. (1992). Effect of spectral envelope smearing on speech reception. I. *Journal of the Acoustical Society of America, 91*, 2872-2880.

ter Keurs, M., Festen, J. M., & Plomp, R. (1993). Effect of spectral envelope smearing on speech reception II. *Journal of the Acoustical Society of America, 93*, 1547-1552.

Terhardt, E. (1974). Pitch, consonance and harmony. *Journal of the Acoustical Society of America, 55*, 1061-1069.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America, 85,* 1699-1707.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology of Aging, 17*, 453-467.

Tun, P. A. & Wingfield, A. (1999). One voice too many: adult age differences in language processing with different types of distracting sounds. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 54*, 317-327.

van Rooij, J. C. G. M. & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *Journal of the Acoustical Society of America, 88*, 2611-2624.

van Rooij, J. C. C. M. & Plomp, R. (1992). Auditive and cognitive factors in speech perception by elderly listeners:III. Additional data and final discussion. *Journal of the Acoustical Society of America, 91,* 1028-1033.

Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W., & McDermott, H. J. (2005). Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies. *Journal of the Acoustical Society of America, 117,* 3126-3138.

von Helmholtz, H. (1877). *On the sensations of tone (English translation A.J. Ellis, 1885, reprinted 1954)*. New York, Dover.

Vongpaisal, T., & Pichora-Fuller, M. K. (2007). Effect of age on F0 difference limen and concurrent vowel identification. *Journal of Speech, Language, and Hearing Research, 50*, 1139-1156.

Vongphoe, M. & Zeng. F. G. (2005). Speaker recognition with temporal cues in acoustic and electric hearing. *Journal of the Acoustical Society of America, 118*, 1055-1061.

Walton, J. P., Orlando, M., & Burkard, R. (1999). Auditory brainstem response forward-masking recovery functions in older humans with normal hearing. *Hearing Research, 127*, 86-94.

Walton, J. P., Simon, H. & Frisina, R. D. (2002). Age-related alterations in the neural coding of envelope periodicities. *Journal of Neurophysiology, 88,* 565-578.

Wang, H., Turner, J. G., Ling, L., Parrish, J. L., Hughes, L. F., Caspary, D. M. (2009). Age-related changes in glycine receptor subunit composition and binding in dorsal cochlear nucleus. *Neuroscience, 160*, 227-239.

Wang, X., & Sachs, M. B. (1993). Neural encoding of single-formant stimuli in the cat I. Responses of auditory nerve fibers. *Journal of Neurophysiology, 70*, 1054-1075.

Watson, C. S. (2005). Some comments on informational masking. *Acta Acoustica, 91,* 502-512.

Watson, C.S. & Kelly, W.J. (1981). The role of stimulus uncertainty in the discrimination of auditory patterns. In *Auditory and Visual Pattern Recogntion*, D.J. Getty and J.H. Howard (Eds). Hillsdale: LEA.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale—3rd Edition (WAIS-3®) San Antonio, TX: Harcourt Assessment.

White, L. J. & Plack, C. J. (1998). Temporal processing of the pitch of complex tones. *Journal of the Acoustical Society of America, 103*, 2051-2063.

White, L. J. & Plack, C. J. (2003). Factors affecting the duration effect in pitch perception for unresolved complex tones. *Journal of the Acoustical Society of America, 114,* 3309-3316.

Wightman, F. L. (1973). The pattern transformation model of pitch. *Journal of the Acoustical Society of America, 54*, 407-416.

Willott, J. F., Milbrandt, J. C., Bross, L. S., & Caspary, D. M. (1997). Glycine immunoreactivity and receptor binding in the cochlear nucleus of C57BL/6J and CBA/CaJ mice: effects of cochlear impairment and aging. *The Journal of Comparative Neurology, 385*, 405-414.

Wingfield, A., Stine, E. A. L., Lahar, C. J, & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Experimental Aging Research, 14*, 103-107.

Yonan, C. A. & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology of Aging, 15*, 88-99.