A Comprehensive Comparative Hearing Aid Study: Evaluating the Neuro-Compensator Relative to Wide Dynamic Range Compression

A COMPREHENSIVE COMPARATIVE HEARING AID STUDY: EVALUATING THE NEURO-COMPENSATOR RELATIVE TO WIDE DYNAMIC RANGE COMPRESSION

BY

JEFF S. BRUCE, BCS

A THESIS

SUBMITTED TO THE DEPARTMENT OF PSYCHOLOGY, NEUROSCIENCE &

BEHAVIOUR

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2013)	McMaster University
(Psychology)	Hamilton, Ontario, Canada

TITLE:	A Comprehensive Comparative Hearing Aid Study: Evaluating the Neuro-Compensator Relative to Wide Dynamic Range Compression
AUTHOR:	Jeff S. Bruce Bachelor of Computer Science, Joint Honours Computer Science and Psychology University of Waterloo, Waterloo, ON, Canada
SUPERVISORS:	Dr. Sue Becker, Dr. Laurel Trainor
NUMBER OF PAGES:	xvi, 162

This thesis is dedicated to making hearing aids better.

Abstract

This Master's thesis presents results from two clinical hearing aid studies. Wide dynamic range compression (WDRC), a hearing aid amplification algorithm widely used in the hearing aid industry, is compared against a novel hearing aid called the Neuro-Compensator (NC), which employs a neural-based amplification algorithm based on a computational model of the auditory periphery. The NC strategy involves preprocessing an incoming auditory signal, such that when the signal is presented to a damaged cochlea, auditory nerve output is reconstructed to look similar to the auditory nerve output of a healthy cochlea for the original auditory signal.

The NC and WDRC hearing aid technologies are compared across a multitude of auditory domains. Objective measures of speech intelligibility in quiet and in noise, music perception, sound localization, and subjective measures of sound quality are obtained.

It was hypothesized that the NC would restore more normal auditory abilities across auditory domains, due to its proposed strategy of restoring more normal auditory nerve output. Results from the clinical hearing aid studies quantified domains in which the NC was superior to WDRC, and vice versa.

Acknowledgements

There are many people who helped me at some point during the course of my studies here at McMaster University. First and foremost, I would like to thank both of my supervisors, Dr. Sue Becker and Dr. Laurel Trainor, for being exemplary mentors. I have grown so much in the past few years as a result of their guidance, and for this I am very grateful. I must also extend a thank you to Dr. Larry Roberts, for his generosity in lending his lab space to me to perform hearing evaluations, thus allowing me to recruit participants for my experiments. I surely would not have had the same success without my labmates, who often provided me with helpful suggestions when I presented at lab meetings, and made the lab a fun place to work. Thank you to Dave Thompson, and Dan Bosnyak for your technical support, and Elaine Whiskin, for always lending a hand and taking care of all the lab essentials. I am thankful for the considerable assistance I received from two undergraduate assistants, Ranya Amirthamanoharan, and Katelyn Muir, who helped run participants through many experiments.

I must also thank the industry partner who I collaborated with during this project, VitaSound Audio Inc., for the time they spared to meet with me to provide guidance on how to properly conduct my experiments. From VitaSound in particular, I would like to thank Gora Ganguli, Philippe Pango, Karl Wiklund, Mario Augustyniak, Sam

Mok, Karen Lalonde, and Clare Daley for their assistance throughout the course of the project, and their contribution of resources. I am deeply grateful for the assistance of Peter Chmielewski, the hearing instrument specialist who I worked with throughout the study, who put in many hours programming the hearing aids for the participants in all of my experiments, as well as consulted with the participants, and provided me with a basic, practical education of hearing aids.

Last, I thank my Mom and Dad for their consistent support, encouragement, wisdom, and love, and my brothers, David, Neil, and Daniel, for paving the path.

This work was funded by NSERC 5-38048 and MITACS 5-98517.

Notation and Abbreviations

3AFC = three-alternative forced-choice

AN = auditory nerve

ANOVA = Analysis of Variance

APHAB = Abbreviated Profile of Hearing Aid Benefit

BM = basilar membrane

BTE = behind-the-ear

CF = characteristic frequency

CVC = consonant vowel consonant

dB = decibel

dB HL = decibels (Hearing Level)

dB(A) = A-weighted decibel

DNR = digital noise reduction

DPOAE = distortion-product otoacoustic emission

DSL = Desired Sensation Level

DSP = digital signal processor

FC = feedback cancellation

HA = hearing aid

HINT = hearing in noise test

HIS = hearing instrument specialist

HRTF = head-related transfer function

hVd = consonant vowel consonant syllable in the form h-V-d

IHC = inner hair cell

ICRA = International Collegium of Rehabilitative Audiology

ISI = inter-stimulus interval

ITC = in-the-canal

NAL = National Acoustic Laboratories

NC = Neuro-Compensator

OHC = outer hair cell

REIG = real-ear insertion gain

REM = real-ear measurement

RIC = receiver-in-canal

SD = standard deviation

SEM = standard error of the mean

SNHL = sensorineural hearing loss

SNR = signal-to-noise ratio

SPL = sound pressure level

SRT = speech reception threshold

VCV = vowel consonant vowel

WDRC = Wide Dynamic Range Compression

Contents

List of Figures

Chapter 1

Sound and Hearing

1.1 The Nature of Sound

A sound source, such as clapping two hands together in the air, causes the displacement of air particles. The displacement of air particles at such a location forces the displaced particles closer to other air particles (condensation), resulting in an increase in the density of air. There is then a reactive force, analogous to releasing a compressed spring, which causes the air particles to bounce away from each other (rarefaction), resulting in a decrease in the density of air at that location. Sound waves propagate through a medium such as air, or water, through repeated condensation and rarefaction, and can be understood as a series of rapid changes in the density of a medium.

Sound is often analyzed in terms of amplitude and frequency. Amplitude is the magnitude of vibration of a sound source (how much air is displaced), measured in decibels (dB), and frequency is the rate of vibration of a sound source (how often the air is displaced), measured in Hertz (Hz). The perceptual correlates of amplitude and frequency are loudness, and pitch, respectively. The frequency range of human hearing for young adults is 20-20000 Hz, and the sensitivity for high frequencies generally reduces as a natural progression of aging, known as age-related hearing loss, or presbycusis, which is the most common cause of hearing loss.

Sound is an important subject of study because it helps us extract useful information from our environments. Sound can warn us of potential dangers in our environments (e.g. fire alarms, traffic noise around a street corner, the growl of a predator in the jungle), as well as signal potential rewards (e.g. a prey's movements in a forest, the voice of a potential mate). Most obviously, sound is a useful tool for communication through the mode of speech. It is not surprising that, with hearing loss, it becomes more difficult to communicate with others, but it may surprise some to learn that hearing loss may often lead to social isolation (Weinstein & Ventry, 1982), an increased probability of depression and anxiety (Tambs, 2004), and an increased risk of dementia (Lin, 2011).

Furthermore, speech is not the only mode of communication that relies on sound; there is also music. Participating in musical activity has all sorts of demonstrated benefits; it has been shown to facilitate group trust and cooperation (Anshel & Kipper, 1988), to be associated with greater cognitive abilities (Schellenberg, 2004), to prevent stress-induced anxiety (Knight & Rickard, 2001), and to regulate one's mood (Saarikallio & Erkkilä, 2007).

1.2 The Peripheral Auditory System

The human auditory pathway can be schematized as a hierarchy of processing blocks, organized from lower to higher levels, with connections between the blocks.

The processing blocks which respond to sound input sooner are commonly referred to as lower level, and are where basic features of the sound, such as amplitude, and frequency, are encoded. More complex features, such as timbre, the formation of auditory objects, and detecting an auditory change, are governed by higher level processes and are largely mediated by later processing blocks. For the most part, this thesis concerns itself only with the lower level auditory structures, often called the auditory periphery. The auditory periphery is defined as all of the auditory structures and connections peripheral to the auditory brainstem in the auditory pathway, which includes the ear and auditory nerve. In this thesis, little discussion will be given to the higher auditory structures.

1.2.1 Outer and Middle Ear

The peripheral auditory system is divided into three regions: the outer ear, the middle ear, and the inner ear. An illustration of all regions can be seen in Figure 1.1. The auditory canal and pinna make up the outer ear system (Moore, 2007). When sound is directed at the ear, it is first shaped by interactions with the head and outer ear, which is well described by a head-related transfer function (HRTF) (Wiener & Ross, 1946). The HRTF describes how frequencies are amplified and attenuated for different orientations of incoming sound, due to the natural ear canal and pinna resonances. Several measurements of transfer functions have been taken over the years, and some of these data can be seen in Figure 1.2 (Mehrgardt & Mellert, 1977). For the most part, all of the transfer function measurements show the same overall shape, in that frequencies from 2-6 kHz are amplified by the natural resonances of the pinna and ear canal, with an additional peak occurring at around 10 kHz. It should be noted that this frequency response is not only dependent on the direction of the incoming sound, but also the frequency. Medium and high frequencies are especially modified by the head, torso, and pinna (Moore, 2007), and in fact, the variations in the spectrum of the sound due to this shaping can be used to localize sound.

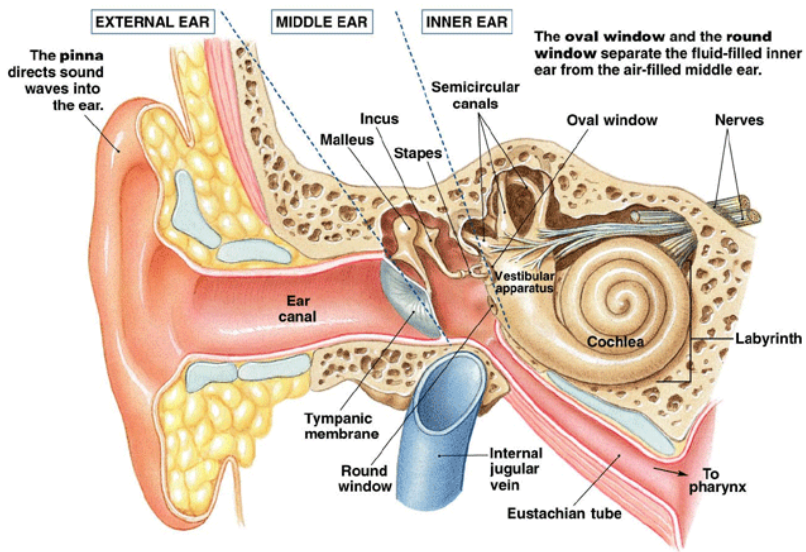

Figure 1.1: Anatomy of the ear. The outer, middle, and inner ear and their components. (Retrieved June 1, 2013, from: http://www.directhearingaids.co.uk/index.php/33/how-hearing-balance-work-together/)

Once sound has traveled down the auditory canal, it reaches the tympanic membrane, causing it to vibrate. The tympanic membrane constitutes the beginning of the middle ear system. Upon the tympanic membrane's vibration, it sets into motion a chain of three tiny bones (ossicles), named the incus, malleus, and stapes. The stapes presses against the oval window of the cochlea, which has a surface area roughly 27 times smaller than that of the tympanic membrane (Moore, 2007), concentrating the acoustic energy into a smaller region through an impedance-matching

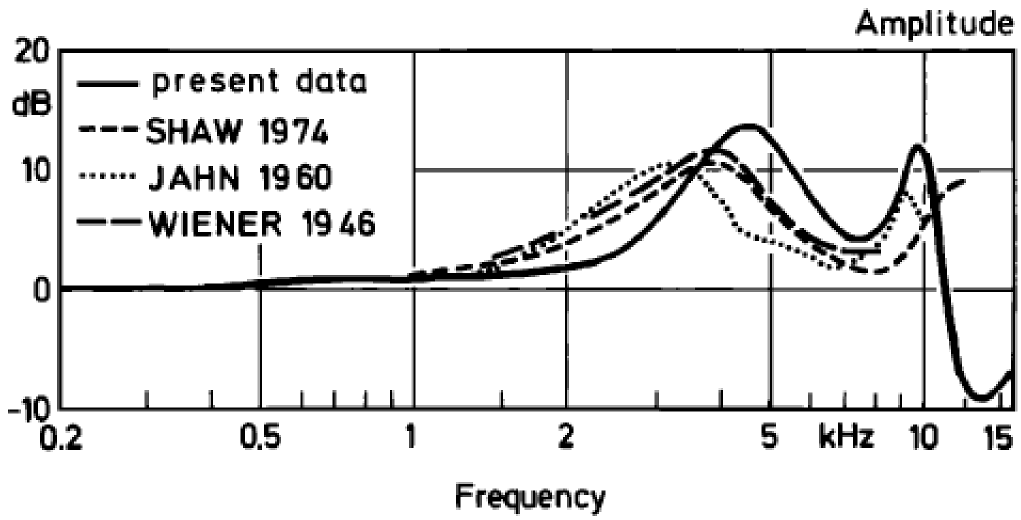

Figure 1.2: Outer ear transfer function data obtained from different investigations. A transfer function describes how different frequencies are amplified/attenuated as a result of traveling through the system. (Reprinted from Mehrgardt and Mellert (1977)).

process. Impedance-matching in this context essentially means that the bones of the middle ear reduce the reflections from the oval window that would otherwise occur if the sound transmission was through air, thereby maximizing the power transferred through the system.

Similar to the outer ear, the middle ear alters the frequency response of the incoming sound in a systematic fashion. In particular, sounds between 500-5000 Hz are the most efficiently transmitted (see Figure 1.3), which is an important range for speech perception. Also shown in Figure 1.3 is how the phase of the sound is altered by traveling through the middle ear system. Notice how higher frequencies tend to be delayed longer than lower frequencies.

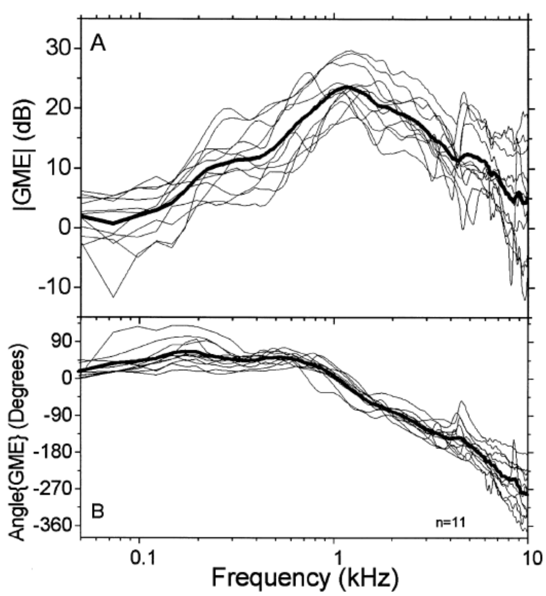

Figure 1.3: Middle ear transfer function plotted for 11 ears (average in bold). *Top:* Magnitude response; the middle ear most efficiently transmits middle frequencies. *Bottom:* Phase response; higher frequencies are delayed more than lower frequencies. (Reprinted from Aibara et al. (2001)).

Thus far, both the outer and middle ear systems have been briefly described. For the most part, both of these systems are linear; a linear system transforms an input abiding by two steadfast principles. If an input to a linear system is scaled by a factor k, the output must also be scaled by a factor k (the scaling principle); and if two independent inputs, x and y, are presented to a linear system to produce separate outputs, $o(x)$ and $o(y)$, then the response to $x + y$ summed together must equal $o(x) + o(y)$ (the superposition principle). In the next section, the inner ear will be described, otherwise known as the cochlea, and in many cases we will observe that the cochlea behaves nonlinearly.

1.2.2 The Cochlea

The cochlea is a spiral shaped organ containing a series of fluid filled chambers, called the scala media, scala tympani, and scala vestibuli. It begins with the oval window, which is the outermost shared boundary with the middle ear. When the stapes of the middle ear presses upon the oval window, it causes the fluid in the scala vestibuli to flow to the scala tympani and eventually forces the outward movement of another opening in the cochlea, called the round window. The scala media is separated from the scala vestibuli by Reissner's membrane, and the scala media is separated from the scala tympani by the basilar membrane. The movements of the oval window and the round window result in energy being absorbed by the basilar membrane, which causes it to vibrate, deflecting perpendicularly to its surface in the form of a traveling wave. The traveling wave gradually grows in amplitude from the base of the cochlea to its maximal point of excitation (different for each frequency), at which point its amplitude quickly decreases. The basilar membrane responds preferentially to different frequencies along its length, due to its elastic properties. It is narrower and stiffer at the base, which high frequencies more easily resonate with, and it is wider at the apex, which lower frequencies preferentially resonate with (these physical properties are collectively referred to as the passive mechanism of the basilar membrane). Thus, the basilar membrane is tonotopically arranged, encoding high frequencies at its base and low frequencies at its apex (see Figure 1.4), and each spot on the basilar membrane has a frequency it is most sensitive to, called a characteristic frequency (CF). Another structure, called the tectorial membrane, forms a boundary with the basilar membrane, and between these two boundaries there are hair cells and nerve fibers which make up the organ of Corti (see Figure 1.5).

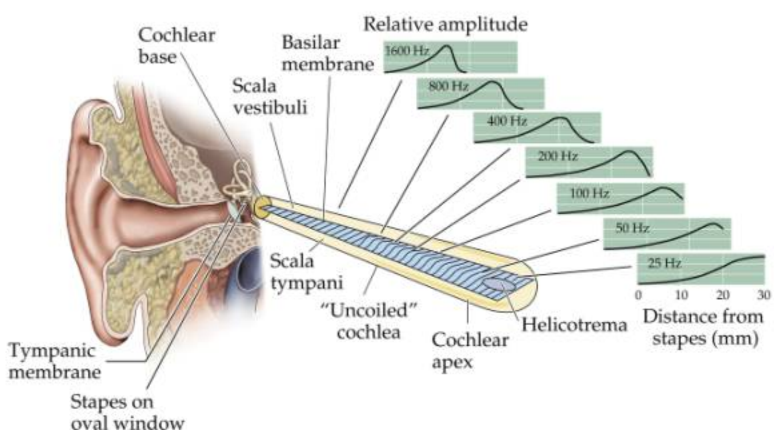

Figure 1.4: The "uncoiled" cochlea. The basilar membrane is tonotopically arranged; basal regions respond preferentially to high frequencies, and more apical regions respond preferentially to low frequencies. (Retrieved June 1, 2013, from: http://www.rci.rutgers.edu/ uzwiak/AnatPhys/Audition.htm).

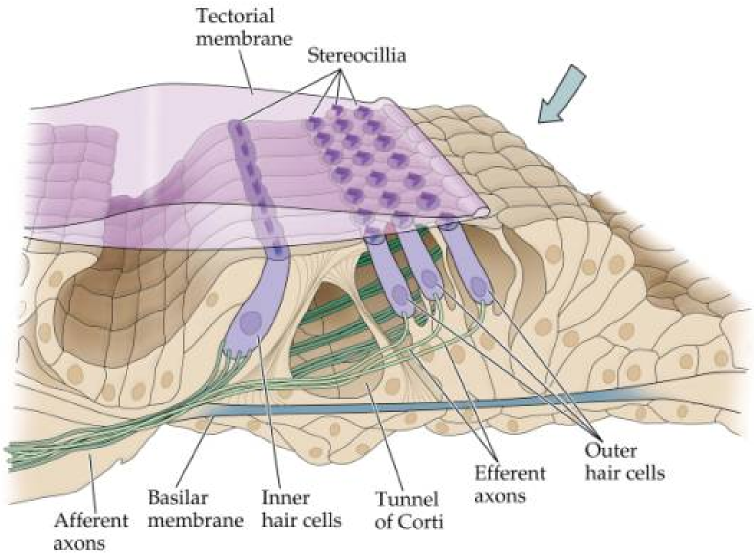

Figure 1.5: The organ of Corti holds the inner and outer hair cells and the auditory nerve fibers. (Retrieved June 1, 2013, from: http://www.rci.rutgers.edu/ uzwiak/AnatPhys/Audition.htm).

Organ of Corti

The organ of Corti rests upon the basilar membrane, inside the scala media. It holds the outer hair cells (OHCs) and inner hair cells (IHCs), which are structurally different from one another and serve different functional purposes. There are also nerve fibers which connect to these cells in the organ of Corti. The nerve fibers are either afferent or efferent; the afferent nerve fibers are primarily innervated by the IHCs, and carry electrical impulses to higher auditory centres, while the efferent fibers' axons connect primarily to the OHCs. Many of the efferent fibers originate from the superior olivary complex in the auditory brainstem. The OHCs have a motor component and can change their length and stiffness, which actively shapes the basilar membrane response (this property is referred to as the active mechanism of the basilar membrane). The tuning of the basilar membrane is thus a combination of a passive and active mechanism, and each mechanism alone cannot account for the basilar membrane's large vibration amplitude and narrow region of resonance (Pickles, 2008).

The OHCs are connected to the tectorial membrane, the other boundary of the organ of Corti, while the IHCs are not. When the basilar membrane vibrates, it causes fluid motion in the scala media, while also causing the tectorial membrane to move back and forth in a shearing motion. On the top of the OHCs and IHCs are tiny hairs called stereocilia, which are moved back and forth by the fluid motion in the scala media, or in the case of the OHCs, the shearing motion of the tectorial membrane. When the basilar membrane deflects perpendicularly, the IHC stereocilia are moved back and forth, which results in a flow of electrical current through the cell, and the subsequent delivery of electrical impulses to the brain via the afferent nerve fibers described earlier. Like the basilar membrane, each afferent nerve fiber has a CF, or a frequency for which it is most sensitive.

Nonlinear Cochlear Mechanics

In this section, the many nonlinearities of the cochlea will be described to provide a segue into how hearing aids have attempted to ameliorate the symptoms of hearing loss.

To begin, in a healthy ear, the basilar membrane behaves in a nonlinear manner, such that its amplitude of vibration grows nonlinearly with the intensity of the stimulus. This is often referred to as the compressive nonlinearity of the basilar membrane, and can be illustrated with an input-output function (see Figure 1.6). At low sound pressure levels (SPLs) and very high SPLs, for a spot along the basilar membrane with a given CF, the input-output relation is near linear (a slope near 1), whereas for moderate SPLs, the input-output relation is compressive (a slope less than 1). Interestingly, when the stimulus frequency deviates from the CF, the input-output response becomes progressively more linear (see Figure 1.7).

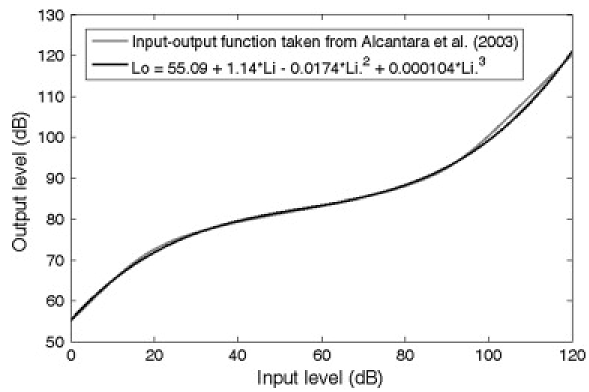

Figure 1.6: Compressive nonlinearity of the basilar membrane. For soft and loud sounds at a fixed place on the basilar membrane with a given CF, the input-output response is linear. For moderate level sounds, the input-output response is compressive. (Reprinted from Deroche and Culling (2011)).

Another psychoacoustic phenomenon, called two-tone suppression, introduces a nonlinearity in the way the basilar membrane responds to complex sounds. In essence, when two pure tones with different frequencies are presented simultaneously, one tone (the suppressor) may suppress the basilar membrane response to the other tone (the probe), particularly when the suppressor is greater in intensity (see Figure 1.8), and close to the probe tone frequency (see Figure 1.9).

Similar to two-tone suppression, when two tones are presented simultaneously and their frequencies are not too far apart from each other, distortion products are generated at linear combinations of the two frequencies (2f1 - f2, f2 - f1, etc.). These distortion products even behave like separate sinusoidal tones, causing vibration on

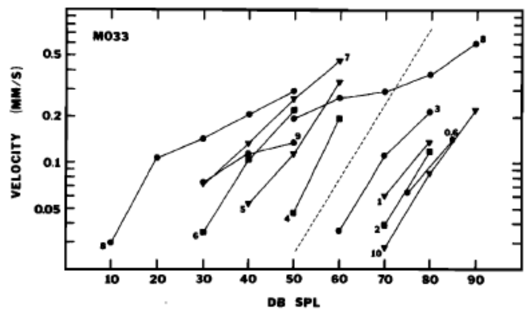

Figure 1.7: Separate input-output curves plotted for several different frequencies. All measurements were taken at the same place on the basilar membrane, which had a CF of 8 kHz. The input-output response is progressively more linear for frequencies which deviate from the CF. (Reprinted from Robles et al. (1986)).

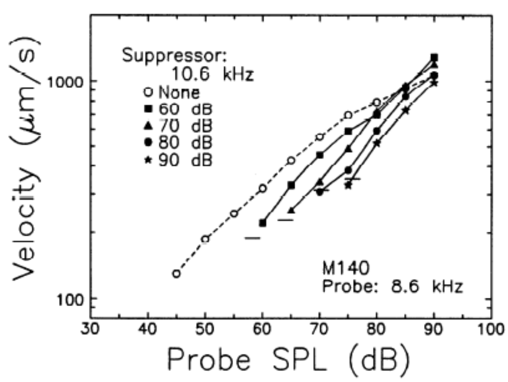

Figure 1.8: Basilar membrane amplitude velocity as a function of the probe tone level (in dB SPL). When the level of the suppressor tone is increased, the level of the probe tone required to reach criterion velocity of the basilar membrane also increases. (Reprinted from Ruggero et al. (1992)).

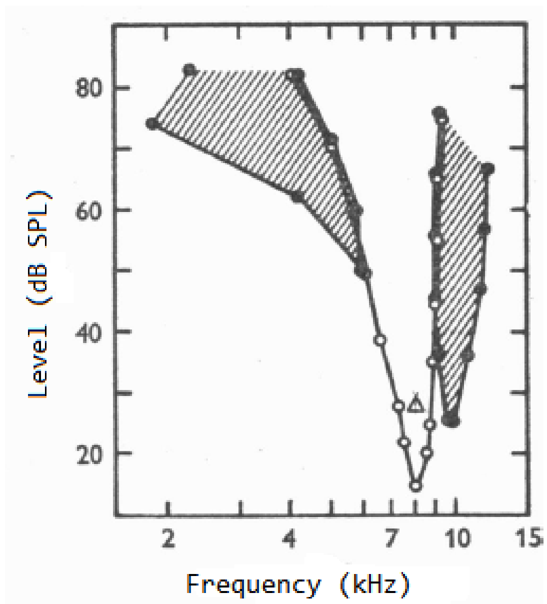

Figure 1.9: The effect of two-tone suppression on an individual tuning curve (white circles), of a cat auditory neuron with a CF of 8 kHz. The neuron was stimulated with the probe tone at a given intensity (indicated by the triangle), and a suppressor tone was varied in frequency and intensity. The area above the white circles corresponds to an increase in the discharge rate, in excess of 20% above the spontaneous rate. Tones within the shaded region reduced the discharge rate from the probe tone by 20%, while not producing an excitatory response when simply presented alone. (Reprinted from Arthur et al. (1971)).

the basilar membrane at locations with those CFs (Moore, 2007).

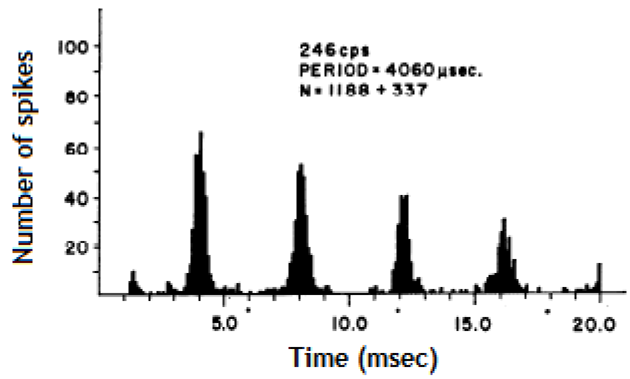

Figure 1.10: An illustration of phase locking. Auditory nerve fibers fire at a particular phase of the stimulus waveform. For a tone with a frequency of approximately 250 Hz, and a corresponding period of 4 msec, the peak firing rate occurs every 4 msec. (Adapted and reprinted from Rose et al. (1967)).

One last nonlinear component to cochlear mechanics is the concept of phase locking. When an AN fiber responds to a tone presented at its CF, it responds in a predictable way, firing at a particular phase in the waveform (see Figure 1.10). Phase locking only occurs for sufficiently low frequency stimuli (less than 5 kHz, according to Pickles (2008)); it is stimulus dependent, not AN dependent, meaning that an AN fiber with a high CF is able to phase lock to sufficiently low frequency stimuli.

1.3 Hearing Loss

Hearing loss is widespread, affecting 11% of the US population to some degree (Kochkin, 2009). This proportion is likely to grow in the coming decade due to an aging population, increasing noise pollution (Goines & Hagler, 2007), and the pervasive use of personal audio devices, which afford listeners the opportunity to expose themselves to potentially damaging sound levels on a more frequent basis.

Hearing loss is generally classified as either conductive, sensorineural (SNHL), or mixed, depending on the nature and location of the damage. A conductive loss is defined as any problem in the conduction of sound waves from the outer to the inner ear, and can be caused by factors such as wax buildup, an accumulation of fluids, the degradation of the ossicles, or a perforated ear drum. It is characterized by significantly higher air-conduction thresholds than bone-conduction thresholds. Sensorineural impairments are generally typified by the degradation of OHCs or IHCs or the malfunction of AN fibers in the inner ear, but sometimes they may involve dysfunction in higher auditory centres. This thesis only concerns itself with SNHL, the most prevalent form of hearing loss.

To quantify one's level of hearing loss, one generally visits a hearing clinic to have a hearing test administered by an audiologist or a hearing instrument specialist. During the hearing test, the participant is seated in a sound booth wearing headphones, and the audiologist plays pure tones of different frequencies through the headphones, one by one, using an instrument called an audiometer. Depending on the available equipment, the participant signals that they can hear a pure tone by pressing a button or by raising their hand. The intensity of the pure tone is adaptively adjusted to find the threshold at which the listener can consistently hear the pure tone. At the end of the test, using the data collected, the audiologist creates an audiogram, which quantifies the level of hearing loss in units of dB HL for the listener at specific frequencies, generally ranging from 250 Hz to 8000 Hz (see Figure 1.11). dB HL stands for decibels Hearing Level, which is a standard used to define how many more decibels one requires over a normal, young listener, in order to hear a sound at a particular frequency. The most common shape of hearing loss is a sloping hearing loss, such that the higher frequencies are more damaged than the lower frequencies.

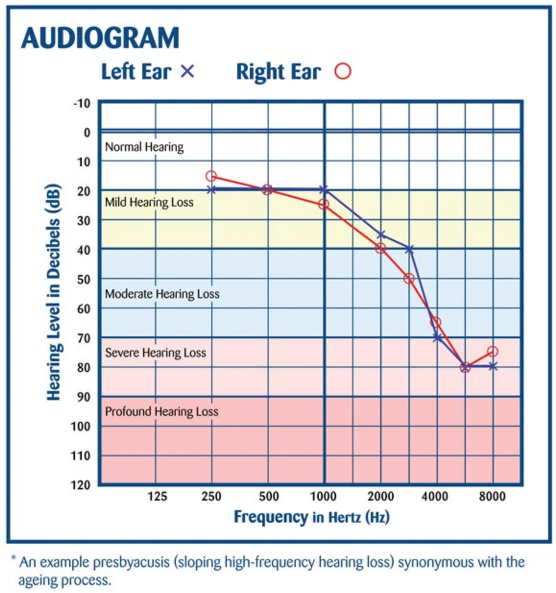

Figure 1.11: An audiogram with presbycusis, sloping from low to high frequency. Hearing loss is graded from mild to severe on the dB HL scale, measured at specific frequencies ranging from 250 to 8000 Hz. Both the left and right ears are plotted. (Retrieved June 1, 2013, from: http://auditoryneuroscience.com/acoustics/clinical_audiograms).

With OHC loss, there is a loss of frequency sensitivity and selectivity, which can be observed as a greater threshold of activation for an individual AN fiber, and a broader tuning curve, respectively (see Figure 1.12). The corresponding effect on perception is a loss of audibility and clarity, and sounds become especially susceptible to masking in noise. Another way of illustrating the effect of the loss of OHCs is to observe that the input-output response of the basilar membrane shifts to the right and behaves more linearly (see Figure 1.13).

IHC loss, contrary to OHC loss, does not affect the compressive nonlinearity of

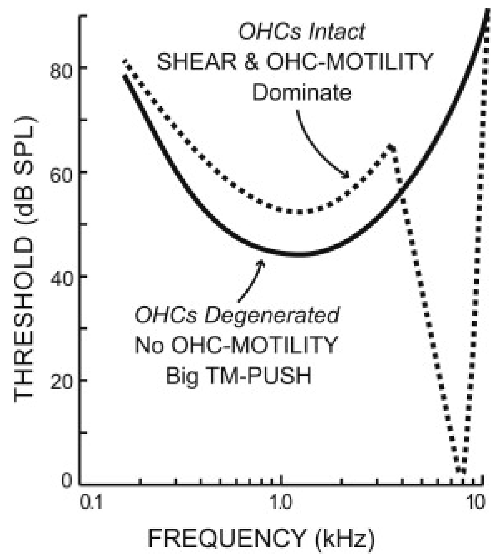

Figure 1.12: Stylized tuning curve with and without OHC loss. When OHCs are intact, the tuning curve is sharply tuned. Losing the OHCs broadens the tuning curve, and results in a loss of sensitivity due to the failing active mechanism which amplifies the basilar membrane response at low intensities. (Reprinted from Guinan Jr. (2012)).

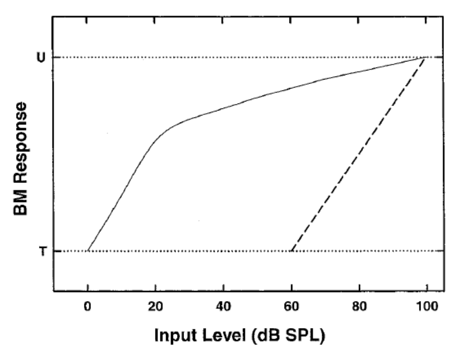

Figure 1.13: The basilar membrane response amplitude following OHC loss. A healthy basilar membrane response exhibits nonlinearity (solid line), while loss of OHCs results in a more linear response (dashed line). On the ordinate, T represents threshold and U represents uncomfortable loudness. (Reprinted from Oxenham and Bacon (2003)).

the basilar membrane, as is indicated by a normal cochlear microphonic and intact distortion product otoacoustic emissions (DPOAEs) (Trautwein, Hofstetter, Wang, Salvi, & Nostrant, 1996). IHC loss also does not significantly affect the threshold and tuning of remaining AN fibers (Wang et al., 1997). However, loss of IHCs does reduce the output of the cochlea, as measured by a reduction in the amplitude of the compound action potential (Trautwein et al., 1996; Qiu, Salvi, Ding, & Burkard, 2000). Interestingly, in the Qiu et al. paper, despite a reduced amplitude for the compound action potential and inferior colliculus potential, some animals showed an increase in the auditory cortex potential amplitude, reflecting an increased gain in the central auditory system. In fact, the increased gain in the central auditory system has been proposed as a compensatory mechanism for reduced input from the cochlea, and could be the mechanism driving the perception of tinnitus (Noreña, 2011), which is often described as a phantom ringing or hissing sensation. Recent evidence from chinchillas has shown that significant loss of IHCs results in very minor audiometric threshold shifts, and therefore audiograms may not provide much information about IHC loss (Lobarinas, Salvi, & Ding, 2013).

Temporal coding in the cochlea is altered with SNHL. Two ways of measuring temporal properties of an AN fiber are to look at its phase response and its group delay. The phase response describes how the signal is delayed as a function of frequency, and group delay is the total length of time it takes for the signal to be processed by the auditory filter. With a broadened tuning curve, the phase response becomes shallower (see Figure 1.14), and there is a shorter group delay (Shi, Carney, & Doherty, 2006). Also, there is psychophysical evidence suggesting that SNHL affects one's ability to use temporal fine-structure cues for speech perception (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Hopkins, Moore, & Stone, 2008), which is not entirely surprising, given that the temporal coding properties of individual AN fibers are altered with SNHL.

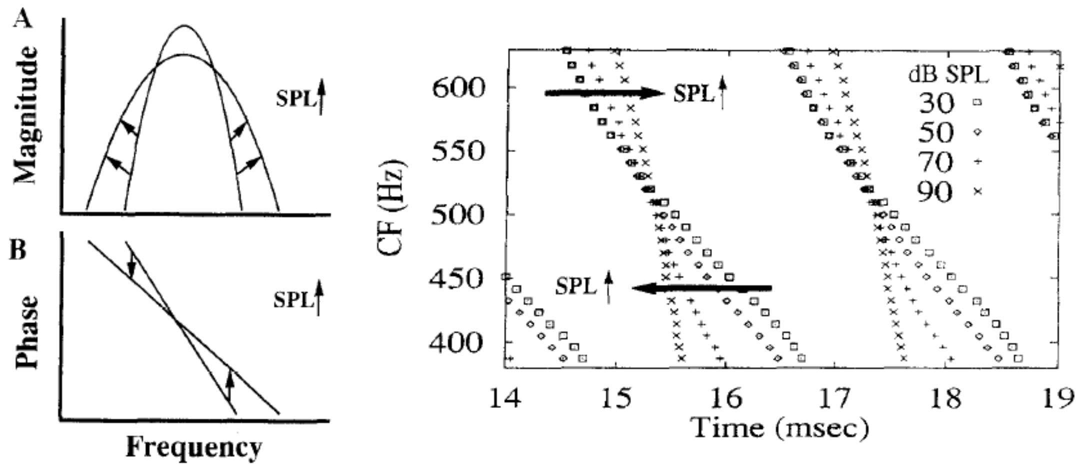

Figure 1.14: *A:* with increasing stimulus level, an AN fiber's tuning curve broadens, similar to what happens with OHC loss. *B:* with increasing stimulus level, the slope of the phase response of an AN fiber becomes shallower. *Right:* with increasing stimulus level, the group delay decreases. (Reprinted from Carney (1994)).

As mentioned in the previous section, phase locking is an important nonlinearity of a healthy ear. In ears with SNHL, however, the tuning curves of AN fibers are shifted toward lower CFs and become broader. The result is that these shifted AN fibers begin to phase lock and synchronize to lower frequency stimuli, which is called an upward spread of synchrony (Miller, Schilling, Franck, & Young, 1997). This presents a problem for speech perception, since the formants are very important for speech intelligibility (Peterson & Barney, 1952; Kiefte, Enright, & Marshall, 2010); higher formants may not be properly encoded in ears with SNHL, which could result in confusability between the formants, and therefore reduced speech intelligibility.

In summary, SNHL is a pervasive disorder and can be caused by dysfunction of any of the following: IHCs, OHCs, or AN fibers. With SNHL, the response properties of individual AN fibers are altered in numerous ways, disrupting the neural code at the level of the AN. Disruptions in the neural code may degrade speech intelligibility, which is a common complaint of those with SNHL, especially in the presence of background noise. Other auditory abilities, such as music perception and sound localization, may be similarly affected by a disrupted neural code in the auditory nerve. The purpose of this thesis is to compare two hearing aid algorithms on their efficacy of restoring these auditory abilities; one algorithm proposes to restore a neural code closer to normal, and the other algorithm is a standard prescription in the hearing aid industry. The next chapter introduces these hearing aid algorithms, among other hearing aid components and features, in more detail.

Chapter 2

Hearing Aids

2.1 Introduction

Hearing aids, quite simply, are prescribed to help listeners hear desired sounds in their everyday listening environments. For most people who are fit with hearing aids, these desired sounds are primarily speech, so the main goal for wearing hearing aids is to restore one's ability to hear and understand speech. However, one can imagine that other signals are critical to restore in many cases. Take, for example, the case of a musician who needs to strain to hear particular instruments or frequencies. In the musician's case, hearing loss may affect his ability to perform on his instrument, and it is quite conceivable that for this person, restoring music perception ought to take priority over restoring speech perception. Additionally, two people with the same audiogram may have very different tolerances for loud sounds and speech in noise outcomes. Prescribing the same amount of amplification for these two people may not be the best strategy. The underlying theme here, is that there is not a one-fits-all solution when prescribing hearing aids, as every individual has their own needs.

For this very reason, it is common for hearing aid wearers to require one or several adjustments to their hearing aid fitting; the general best practice is to start off with an average best fit, and then perform one or several adjustments over the course of weeks or months to optimize an individual fitting, based on the patient's feedback.

Hearing aids are tasked with the difficult problem of amplifying sound just enough to restore speech intelligibility, while suppressing unwanted sounds (i.e. noise), for which the definition varies from person to person. Hearing aids do not perfectly restore normal hearing, and there are still many returns for refund on hearing aids as well as hearing aids sitting unused in dresser drawers (Kochkin et al., 2010). Thus, there is much room for improvements to hearing aid technology, although hearing aid satisfaction has been slowly increasing (Kochkin, 2010).

In the 1980's, most hearing aids were linear, meaning that an equal amount of amplification was given to an input signal regardless of the level of the incoming input signal. Once hearing science began to unravel and understand the issue of loudness recruitment (rapid growth of loudness in hearing-impaired subjects), compression was added to hearing instruments in order to amplify soft sounds more than loud sounds. As hearing science progresses, additional signal processing techniques will be added to hearing aids, with the hope of steadily increasing hearing aid satisfaction.

2.1.1 Physical Components

A hearing aid is made up of one or several small microphones, a battery, a speaker (often called a receiver), internal components which alter the original auditory signal, and a shell which houses all of these components in a small space. An incoming acoustic signal is picked up by the microphone and transduced into electricity, which is then altered in some way by the internal components (the signal is generally split into different frequency channels, and given amplification proportional to the amount of hearing loss in each channel), and finally the speaker transduces the signal back into a sound pressure wave in the ear canal.

There have been considerable advances in hearing aid technology over the past couple of decades, most dramatically in the way that sound is processed, amplified, and compressed. In the early 1980's, all hearing aids were analog, meaning that physical transistors and other electrical components acted as the amplifiers, filters, and compressors. Modern hearing aids, however, are essentially all digital, meaning that they have digital signal processors (DSPs). Unlike analog hearing aids, which alter the signal electrically with electrical components, digital hearing aids turn the acoustic signal into a string of numbers with an analog-to-digital converter, alter the signal mathematically with the DSP, and turn the signal back into sound with a digital-to-analog converter. DSPs offer much more flexibility in the way of programming hearing aids and implementing more advanced signal processing algorithms, which has resulted in an explosion of features that are starting to become standard in modern hearing aids.

2.1.2 Features

Many technologies found on modern day hearing aids have existed for decades in one form or another, such as directional microphones, but it is only with the advent of DSPs that these technologies have been able to be exploited in order to improve hearing aid outcomes.

As mentioned previously, digital hearing aids are programmable. Several programs can be installed on a single hearing aid (one for speech in quiet, one for music, etc.), and the user can change programs by pressing a button on the hearing aid, or by using a remote control. Many modern hearing aids automatically adapt to their environments, switching between programs depending on the acoustic input.

Directional microphones are either implemented as one or several microphones, and are the only hearing aid technology that have been shown to improve speech intelligibility in noise, when used in optimal settings (McCreery, Venediktov, Coleman, & Leech, 2012). Some newer directional microphone technologies track the location of the noise source and change their polar plot accordingly (adaptive directionality).

Noise reduction is another signal processing technique that is commonly seen on modern hearing aids. Plomp (1983) discovered that noise-like stimuli have different temporal/envelope properties than speech-like stimuli, in that the rate of amplitude modulation of noise-like stimuli tends to be higher than speech, and the depth of amplitude modulation tends to be lower than speech. Once this discovery was made, hearing aid algorithm designers exploited these differing characteristics between speech and noise, creating noise reduction schemes which reduced amplification in frequency channels where speech was statistically unlikely to be present. In modern hearing aids, there are many different implementations of noise reduction (Bentler & Chiou, 2006). One problem with noise reduction is that everyone's definition of noise varies, so a noise reduction system that works well for one person will not necessarily work well for another person. In a review of many different clinical studies investigating noise reduction, it was found that noise reduction did not improve nor degrade speech understanding (McCreery et al., 2012). Despite the lack of evidence for noise reduction improving speech intelligibility, there is emerging evidence that it can reduce cognitive effort (Sarampalis, Kalluri, Edwards, & Hafter, 2009), thereby improving hearing aid fitting outcomes.

Similar to some other audio applications, one pervasive engineering problem with hearing aids is the problem of feedback. With the microphone and speaker placed so close in proximity, and the near impossibility of forming a perfect seal with the ear canal, it is very difficult to prevent sound emanating from the speaker from leaking out of the ear canal and back into the microphone, creating a feedback loop. One workaround that the industry has devised is to equip hearing aids with feedback cancellation algorithms. The most popular forms of feedback cancellation work by detecting tonal stimuli, and subsequently reducing amplification in the frequency region(s) containing the tonal stimuli by using a notch filter, or by adding a signal which is opposite in phase to the feedback signal (Parsa, 2006). Although feedback cancellers work well to suppress feedback, sometimes, for tonal sounds, the feedback canceller can produce what are called entrainment artifacts. Entrainment artifacts are perceived as an additional tone or echo added in by the feedback canceller in order to attempt to cancel out a feedback-like signal. Entrainment can even occur with non-feedback, tonal-like stimuli, such as the sound of a musical instrument, or alarms and beeps.

Another feature which has existed for some time but is just now becoming commonplace, is known as frequency lowering. The main concept behind this technology is to shift higher frequency speech information to lower frequencies, where hearing loss is generally less severe. As mentioned in a previous section, most hearing losses are greater in extent in the high frequencies; sometimes it is challenging to provide enough amplification in these regions to restore speech cues, and frequency lowering addresses this problem by shifting speech cues to lower frequencies that otherwise cannot be restored at higher frequencies. In some cases, amplification in regions with severe hair cell loss can even decrease speech intelligibility in quiet (Vickers, Moore, & Baer, 2001). Other studies have found high frequency amplification not to be detrimental to speech intelligibility (Cox, Johnson, & Alexander, 2012). Either way, there is evidence that frequency lowering may help restore speech intelligibility (Simpson, Hersbach, & McDermott, 2005; Glista et al., 2009).

As DSPs become faster and increase in memory capacity, the number of hearing aid features will continue to grow. Audiologists and hearing instrument specialists are in some sense forced to keep pace with the advent of new features, or they risk falling behind the curve, potentially losing business as a result. There is a need for hearing aid algorithms to build in these features implicitly, which would make it easier for audiologists and hearing instrument specialists to keep up with new technologies, therefore simplifying the fitting process both for hearing practitioners as well as patients.

2.2 Hearing Aid Prescriptions

2.2.1 Overview

As discussed in the previous chapter, the first step in programming a hearing aid is to measure the audiogram of a patient. With the audiogram, a prescription can be generated for a given hearing aid using the hearing aid manufacturer's fitting software. Prescription gain formulas differ across hearing aid manufacturers, but most are generally derived from either the National Acoustic Laboratories (NAL) formulas (Byrne & Dillon, 1986; Byrne, Parkinson, & Newall, 1990; Dillon, 1999; Keidser, Dillon, Flax, Ching, & Brewer, 2011), or the Desired Sensation Level (DSL) formulas (Cornelisse, Seewald, & Jamieson, 1995; Scollie et al., 2005), which are two families of empirically derived prescription formulas that have been developed over the past couple decades. There have been many revisions to the NAL and DSL fitting formulas over the years, which first started out as linear prescriptions, and have since been revised to be nonlinear prescriptions in order to accommodate compression, a common signal processing technique found on essentially all modern hearing aids. Most fitting software packages allow the clinician to select NAL or DSL directly, in addition to the manufacturer's prescription.

A hearing aid prescription specifies the target amount of gain that the hearing aid ought to prescribe for a given audiometric loss, as a function of frequency and intensity. Decades of research have gone into optimizing the amount of gain as a function of frequency and intensity for the full range of hearing losses. Usually, 3 target output curves are plotted on the fitting graph; one for soft speech, one for average speech, and one for loud speech. Figure 2.1 illustrates an example of a particular fitting software's target curves. In Figure 2.1, the red lines plot the target gain curves as a function of frequency, and the blue lines plot the estimated gain delivered by the hearing aid. In this particular example, the blue lines are lower than the red lines on the y-axis, meaning that the estimated gain delivered by the hearing aid is less than that prescribed. The highest curve for each color corresponds to an input of 40 dB SPL, the middle curve is for an input of 65 dB SPL, and the bottom curve is for an input of 90 dB SPL. The fact that the soft, average, and loud inputs have different curves indicates that compression is being prescribed (low level sounds

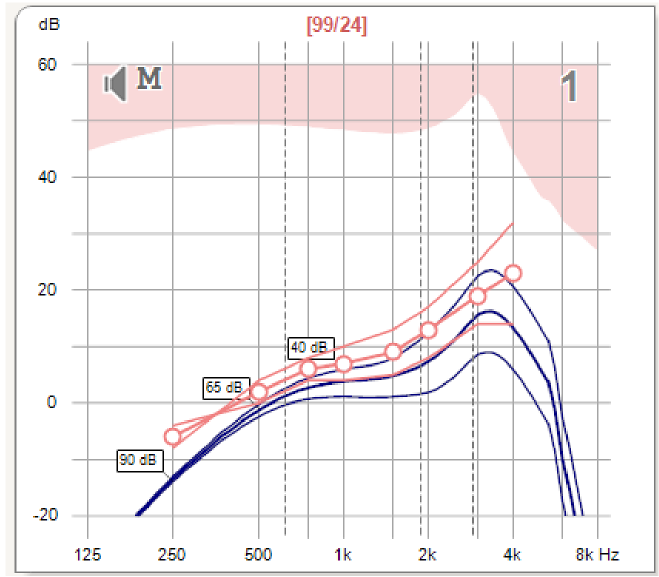

Figure 2.1: Target curves for a single hearing aid. The red lines plot the amount of gain (y-axis) prescribed by the fitting software, for soft speech (40 dB SPL), average speech (65 dB SPL), and loud speech (90 dB SPL). The blue lines plot the estimated gain of the hearing aid as a function of frequency (x-axis). For this particular fitting, the estimated gain is less than the target curves prescribe, meaning that the hearing aid is under-fit. Compression can be seen, in that the soft speech gain curves are giving more gain than the average and loud speech gain curves.

require more gain than high level sounds, in order to compensate for the loss of OHCs, which effectively amplify low level sounds). Using the software, clinicians can adjust the amount of gain as a function of frequency.

The focus of generally all hearing aid prescription formulas is to normalize loudness in individual frequency channels, or across frequency channels, while maximizing speech intelligibility. Wide dynamic range compression (WDRC), which can be implemented as one out of many prescription formulas, is a nonlinear amplification strategy used to map the larger dynamic range of normal hearing onto the smaller dynamic range of a hearing-impaired person, as illustrated in Figure 2.2. The essence

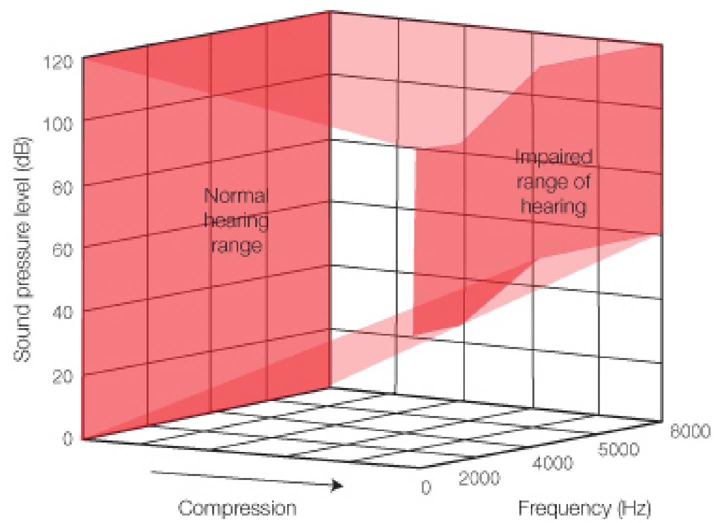

Figure 2.2: WDRC and dynamic range. People who are hearing-impaired have a reduced dynamic range (soft sounds are inaudible, but loud sounds are just as loud as with normal hearing). WDRC applies amplification in different frequency bands independently, with compression, and maps the dynamic range of normal hearing onto the dynamic range of impaired hearing. (Retrieved June 1, 2013, from: http://www.gnresound.ca/professionals/technology-and-innovation/surround-sound-by-resound/wide-dynamic-range-compression)

of WDRC is that it gives more amplification to quiet sounds than to loud sounds, reducing the amount of loudness that would result from amplifying loud sounds, while increasing the audibility of soft sounds (mimicking the compressive nonlinearity of the basilar membrane for normal hearing). Some studies have demonstrated speech intelligibility benefits when using a WDRC scheme in place of a linear gain strategy (Jenstad, Seewald, Cornelisse, & Shantz, 1999; Humes et al., 1999). The hearing aid comparison study reported on in this thesis used WDRC as the standard for comparison, given its widespread acceptance, accessibility, and influence on the hearing aid industry.

2.2.2 Problems With Prescription Formulas

There are two main problems with current prescription formulas, despite the success that these formulas have had with restoring normal loudness perception.

The first problem is that WDRC treats each frequency channel independently, largely ignoring frequency by frequency relationships, which exist in most complex sounds. A complex sound is any sound that is made up of more than one frequency, and the vast majority of sounds encountered in everyday environments are complex. The consequence of ignoring frequency by frequency interactions in an amplification algorithm is that distortion is added to the original signal by altering frequency-specific interactions. Frequency-specific interactions carry important information, and when altered the result can be a reduction in speech intelligibility and sound quality. One example comes from a study by Kiefte and Kluender (2005), who showed that altering the intensity relationships of vowel formants reduces vowel intelligibility. Music offers another example of how ignoring frequency by frequency interactions can degrade intelligibility or quality. When a musical instrument sounds a note, a fundamental frequency is generated, as well as several harmonics produced at integer multiples of the fundamental frequency. Timbre, the musical quality that allows one to differentiate one musical sound from another with the same pitch, loudness, and duration, depends on attack times, the spectral energy distribution of the instrument, and the temporal envelope synchronicity across partials (Grey, 1977). Applying amplification unequally across all frequency bands distorts the spectral energy distribution, and obscures the timbre of an instrument, especially when compression is activated. Such a distortion in timbre alters the information contained in the musical passage, and can affect the perceived quality of this music. Given how a musical signal's timbre can be altered by applying unequal amplification across frequencies, it follows that other timbres are also likely modified (e.g. vocal timbres).

The second problem with standard prescription formulas is that they only attempt to optimize one kind of signal; speech. Optimizing loudness for speech does not necessarily optimize loudness for other complex sounds. For example, while the correlation between the physical and perceptual intensity of speech is high, for some musical instruments it is low (bass, cello), especially for instruments that have multiple harmonics which fall in the same critical band, resulting in less loudness summation (Chasin & Russo, 2004). Additionally, sound localization is an important ability that is enabled by having two ears, pinna filtering, and complex signal processing in the auditory system, and is performed by calculating interaural time and level differences (ITD, ILD) between the ears (Middlebrooks & Green, 1991). Standard prescription formulas do not consider sound localization, and ITD and ILD cues could be distorted by these formulas. In fact, several studies have demonstrated that hearing aids actually impair sound localization abilities relative to wearing no hearing aids at all (Noble & Byrne, 1990; Van den Bogaert, Klasen, Moonen, Van Deun, & Wouters, 2006; Keidser et al., 2006), so there is likely room for improvement to current hearing aid prescriptions in restoring these cues for sound localization.

Clearly, there exists a need for a single hearing aid algorithm which is capable of restoring as many of the abilities that normal hearing listeners take for granted, such as accurate speech intelligibility in quiet and noise, high fidelity music perception, and proper sound localization. The purpose of this thesis is to investigate a novel hearing aid called the Neuro-Compensator (NC), which exploits a neural-based amplification algorithm (Bondy, Becker, Bruce, Trainor, & Haykin, 2004) that could potentially provide many benefits that other hearing aids fail to provide. At present, there is anecdotal evidence of improvements in speech intelligibility, music perception, and sound localization abilities in users fit with the NC.

2.3 Neuro-Compensator

The NC (Becker & Bruce, 2002; Bondy et al., 2004; Chen, Becker, Bondy, Bruce, & Haykin, 2005) uses neurophysiological models of the normal and impaired auditory periphery (Bruce, Sachs, & Young, 2003), which are used to estimate auditory nerve output for a given sound input. The output of the models for a given stimulus are plotted as neurograms, which are essentially plots of neural spiking rates as a function of AN fiber CFs. The block diagram in the left panel of Figure 2.3 illustrates the basic idea of the NC. An input to a normally hearing ear (X) is processed by the normal auditory periphery (H) and responds with normal neural output (Y). In a damaged auditory periphery (one with SNHL), when the same input (X) is presented to the ear, the damaged auditory periphery ($\hat{H}$) does not recreate normal neural output but rather results in a distorted neural output ($\hat{Y}$). By pre-processing the input (X) to the damaged auditory periphery ($\hat{H}$) with the NC algorithm, perceptual distortions are minimized by recreating a more normal neural output (Y).

The details of the auditory periphery model can be seen in the right panel of Figure 2.3. First, the input is processed by the Wiener and Ross (1946) head related transfer function and the middle ear filter. The control path (OHCs) then modulates the signal path (IHCs) with a wideband, nonlinear, time-varying band-pass filter and OHC nonlinearity and a low-pass filter, controlling the behaviour of the narrowband signal path BM filter. The signal path then mimics traveling wave and filter properties

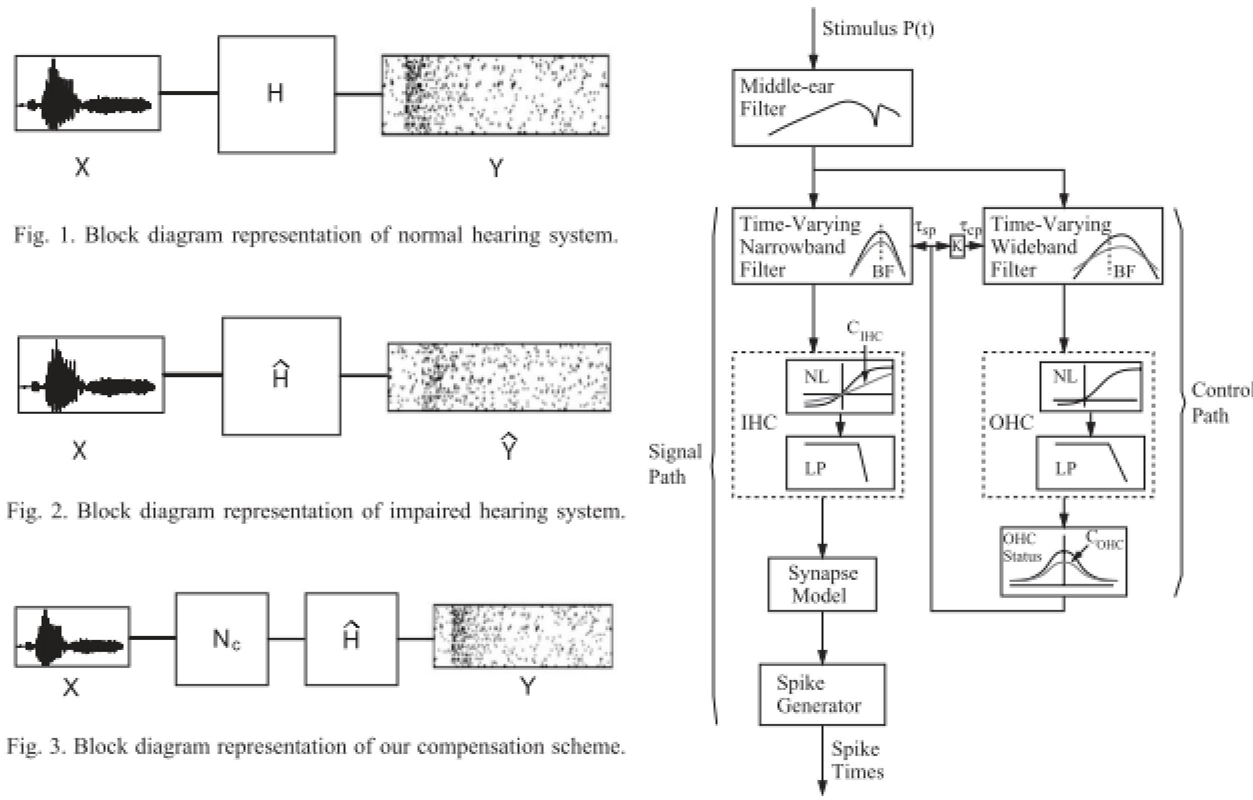

Figure 2.3: The Neuro-Compensator concept. *Left:* The Neuro-Compensator block preprocesses an input acoustic waveform (X), such that when the processed input is fed into a damaged auditory nerve ($\hat{\text{H}}$), a near normal auditory nerve output (Y) is produced. *Right:* Schematics of the Bruce et al. (2003) auditory nerve model. (Both figures were reprinted from Bondy et al. (2004)).

of the BM, and generates spikes in the auditory nerve with spontaneous and driven activity using spike generation and auditory nerve refractoriness. The Bruce et al. (2003) model effectively models many nonlinear characteristics of the cochlea, such as the compressive input-output response of the BM, masking effects like two-tone suppression, synchrony capture in the normal and damaged ear, and shifted tuning curves with OHC damage.

The NC is fit to a particular hearing loss by supplying an audiogram, and training adjustable weights which are present in the NC gain term, for a corpus of training samples (speech). The general functional form of the gain term for the NC is divisive normalization as proposed by Schwartz and Simoncelli (2001), although the coefficients are learned. The gain prescribed for a particular frequency channel depends on the output power in the channel (P_i), normalized by a weighted sum of power in the other frequency channels (P_j) multiplied by the trainable weights (v_{ij}), plus a constant (σ) to prevent gain from exceeding a certain limit (see Equation 2.1, below).

$$G_i = \frac{P_i}{\sum_j (v_{ij} P_j) + \sigma} \tag{2.1}$$

On each iteration of the training algorithm, the output of the normal and aided impaired auditory nerve models are compared, and the weights in the NC gain terms (v_{ij}'s in the formula) are adjusted using a modified version of a stochastic optimization algorithm called Alopex (the original version of Alopex was proposed by Unnikrishnan and Venugopal (1994)). Currently, the error metric used to compute the difference between the normal and aided-impaired neurograms is a simple absolute difference between the two neurograms. Over several training iterations, the overall error between neurograms is minimized and an optimal set of weights (v_{ij}) in the gain term are obtained for a given individual. The logic is that the optimal set of weights will reduce perceptual distortions and perhaps restore some nonlinear behaviour of the cochlea because the compensated neural output is restored to be as close to normal as possible.

Chapter 3

Methods for Comparing Hearing Aids

The current chapter describes the development and application of the multitude of behavioural and subjective tasks that were utilized to compare the WDRC and NC hearing aids. These tasks were selected to cover the general auditory domains of speech intelligibility, music perception, sound localization, and sound quality.

3.1 Speech Intelligibility

At the very outset of this research program, the goal was to compare the NC to WDRC on as many different important auditory domains as possible. With hearing loss, there is a loss of speech intelligibility, in that speech becomes less audible and less clear. Restoring one's ability to communicate with others is the most common reason for obtaining hearing aids, so a significant part of this research program sought to quantify how well the WDRC and NC hearing aids restored speech cues.

3.1.1 Consonant Vowel Consonant (CVC)

Vowels obviously make up a very significant part of speech, and so it is important to probe the question of whether vowel recognition is superior with one hearing aid technology versus another. A convenient task to study vowel perception, which has been widely used in the past, is called a CVC task. In a CVC task, CVC words are presented aloud, and participants must identify which CVC word was presented out of a large set of CVC stimuli (the CVC words differ only on the vowel portion of the words). CVC tasks are used to study vowel perception, as opposed to using whole sentences, because they allow the experimenter to manipulate only the vowel, ignoring any sentence-specific context effects.

CVC words of the form hVd have been used extensively in the context of CVC experiments. The reasons for the commonplace use of hVd stimuli is that the /h/ sound does not significantly affect the upcoming vowel formants, and most of the hVd words are real English words, making the experimental setup easier to get accustomed to (Potter & Steinberg, 1950).

Before programming the hVd experiment, research was done to estimate the difficulty of an hVd task for aided hearing-impaired participants. Recognition of vowels embedded in hVd stimuli by normally hearing subjects in quiet is very high (95% at 77 dB(A) with 12 response alternatives; Hillenbrand, Getty, Clark, and Wheeler (1995)), and aided hearing-impaired subjects are also able to achieve high scores. A study by Ferguson and Kewley-Port (2002) had normal hearing and hearing-impaired subjects listen to and discriminate monosyllabic CVC stimuli in multi-talker babble to determine whether vowels were more easily discriminated in clear speech than conversational speech. In this study, aided hearing-impaired subjects were able to achieve greater than 90% on vowel discrimination in quiet when the speech materials were presented at 70 dB SPL with 10 response alternatives. Given the relative ease with which vowels are understood in quiet, even for aided hearing-impaired participants, the hVd task in the current study included a noise condition to make the task more difficult in case ceiling effects were present in the quiet condition.

3.1.2 Vowel Consonant Vowel (VCV)

Consonants differ from vowels, in that they generally have less acoustic energy (Hamill & Price, 2008), yet they contain proportionally greater high frequency energy, and contribute more to speech intelligibility (Hamill & Price, 2008). Since they are lower in level and contain more high frequency information than vowels, consonants are more vulnerable to reduced intelligibility with hearing loss, and the consequences of reduced intelligibility of consonants are more dire for overall speech recognition. Given the importance of consonants for speech understanding, and their vulnerability with hearing loss, it is not surprising that there have been many hearing aid studies examining how different hearing aid processing strategies affect consonant recognition.

Experiments that investigate consonant recognition often choose to implement a VCV task, which is essentially a mirror-image of the CVC task described earlier, for consonants. Only the consonant in the VCV stimulus is experimentally manipulated (sometimes different vowels are used), and any context effect of sentence is eliminated since the word is presented in isolation.

Like the CVC task, research was done before programming the VCV task to estimate the difficulty of VCV tasks for aided hearing-impaired subjects. One study conducted by Vickers et al. (2001) amplified 65 dB SPL VCV stimuli (21 consonants, 3 vowels) according to the Cambridge prescription formula (Moore & Glasberg, 1998), and included participants with moderately-severe hearing losses. Overall recognition performance was typically around 50-80%. An experiment by Walden, Grant, and Cord (2001) looked at the effect of amplification and speechreading on consonant recognition for the hearing-impaired, for generally mild-moderate hearing losses. Fourteen consonants were used with 1 vowel (/æ/), and the VCV stimuli were presented in quiet in a sound booth. When VCV stimuli were presented at 50 dB SPL at the participant's head while they wore hearing aids, overall recognition scores were close to 80%. Since the degree of hearing losses to be included in the study was not known in advance, a quiet and a noise condition were programmed for the CVC task used in this thesis in case floor or ceiling effects were present for one of the conditions.

3.1.3 Hearing In Noise Test (HINT)

Although CVC and VCV stimuli are useful for determining what spectrotemporal qualities of speech are restored with one hearing aid technology versus another, these tasks are not without disadvantages. The main disadvantage of using CVC and VCV stimuli to assess speech recognition is that these tasks have low ecological validity. The stimuli used in CVC and VCV tasks are rarely encountered in everyday environments, and they ignore any effect of context and continuity in understanding a sentence, which can play a large role in speech understanding. Ultimately, most real-world speech is spoken in phrases and sentences, often in situations with multiple, sometimes non-stationary, noise sources. Therefore, for a comprehensive comparison of two hearing aid technologies, it is good to have a combination of realistic speech tests and more controlled speech tests.

Many tasks have been created to assess real-world speech intelligibility, and one such task is the HINT (Nilsson, Soli, & Sullivan, 1994). The HINT is an industry standard speech recognition test that has participants listen to audio clips of sentences embedded in speech-shaped noise (SSN), and repeat back the sentences that they heard. By adaptively adjusting the presentation level of the sentence while keeping the noise level constant, the procedure effectively zeros in on a participant's speech reception threshold (SRT), defined as the speech-to-noise ratio (SNR) at which the participant correctly repeats back 50% of the sentences. The sentences are organized into lists of 10, and the different lists are balanced across phonemic content, naturalness, difficulty, and reliability. The HINT sentences were adapted from the Bamford-Kowal-Bench (BKB) sentence set (Bench & Bamford, 1979), excluding sentences with British idioms that could be unfamiliar to American English speakers, and using only sentences up to 8 syllables long in order to reduce any influence of memory on SRT measurements.

3.2 Music Perception

Chapter 1 already discussed many benefits of music, and Chapter 2 described how hearing aid amplification algorithms have focused primarily on restoring normal loudness and intelligibility of speech. Not very many hearing aid studies consider how music perception may be affected by wearing hearing aids, in addition to the primary goal of restoring speech intelligibility.

3.2.1 Mistuned Harmonic

Detection of a mistuned harmonic is a widely-used method in the field of music perception. With this method, one can obtain a threshold for the minimum amount of mistuning of a harmonic that an individual can reliably detect, which is a measure of frequency resolution. It is perhaps obvious why the mistuned harmonic is important in the study of music perception, as accurate pitch perception depends on fine frequency resolution and timbre depends in part on the relation between harmonics (Grey, 1977). A less obvious application of the method is in the domain of auditory scene analysis (Bregman, 1994), as partials which are related by integer multiples of the fundamental frequency tend to be grouped together into a single auditory object (Hartmann, 1996).

Hearing aids may distort the harmonic relations by applying amplification and compression disproportionately to different frequency channels. Degradation of the signal may impair auditory abilities, such as frequency resolution or the detection of inharmonicity. There are no experiments to the author's knowledge which apply the mistuned harmonic method to the study of hearing aid subjects, and very little, if any, with hearing-impaired subjects. Perhaps there is sparse research on the topic because of the difficulty of testing hearing aids using tonal stimuli (due to feedback and entrainment). Based on the mistuned harmonic's importance in music perception research, it certainly should be applied to hearing aid research if possible, and this thesis aimed to apply this method to the study of hearing aids.

3.2.2 Timbre Perception

Timbre is the multidimensional musical quality that allows one to differentiate two dissimilar instruments playing the same note, with the same loudness, and subjective duration. It is multidimensional in the sense that there are many contributing factors to the perception of timbre (Grey, 1977; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995), although the three main factors which characterize timbre include the spectral shape of the sound, the spectral variation, and the rise time. Timbre perception research has mostly focused on altering different attributes of complex tones (Plomp & Steeneken, 1969; Samson & Zatorre, 1994), but there have been more recent studies using real (Emiroglu & Kollmeier, 2008) and synthesized (Rahne, Böhme, & Götze, 2011) instruments. Attention has been given to different special populations in timbre perception research, such as infants (Clarkson, Clifton, & Perris, 1988; Trehub, Endman, & Thorpe, 1990), patients with unilateral temporal lobe excisions (Samson & Zatorre, 1994; Samson, Zatorre, & Ramsay, 2002), and hearing-impaired subjects (Emiroglu & Kollmeier, 2008); much attention has been given to patients with cochlear implants (Gfeller & Lansing, 1991; Gfeller, Witt, Mehr, Woodworth, & Knutson, 2002; McDermott, 2004); yet surprisingly little research has examined patients with hearing aids (Looi, McDermott, McKay, & Hickson, 2008). Considering the absence of research in this area, and the importance of timbre for music perception and appreciation, we decided to compare the NC to WDRC on a novel timbre discrimination task, described in the next chapter.

3.2.3 Gap Detection

Thus far, the discussion has focused on the ability of hearing aids to restore normal speech intelligibility, frequency resolution and timbre perception, but one important element of sound remains: time. Sound is spectrotemporal by nature, so the temporal processing of sound should figure prominently in the discussion, and is a field in and of itself. Several procedures have been developed to investigate temporal aspects of auditory perception and the methods have been applied in the study of different special populations, such as the hearing-impaired (Glasberg, Moore, & Bacon, 1987), populations of different ages (Schneider & Hamstra, 1999; Smith, Trainor, & Shore, 2006), and other species (Church, Getty, & Lerner, 1976; Giraudi-Perry, Salvi, & Henderson, 1982). Temporal resolution has been studied using duration discrimination (Abel, 1972), masking approaches (Glasberg et al., 1987), and gap detection paradigms (Shailer & Moore, 1983), with varying stimulus characteristics. In the experiments contained within this thesis, a gap detection paradigm was selected as it has been used in the past with hearing-impaired subjects (Fitzgibbons & Wightman, 1982; Glasberg et al., 1987) and subjects with hearing aids (Moore, Glasberg, Alcántara, Launer, & Kuehnel, 2001), which allowed for a comparison with the data collected in this thesis. Participants with hearing impairments generally perform just as well as those with normal hearing on gap detection tasks which use pure tones (Moore & Glasberg, 1988), but for narrow-band noises the hearing-impaired perform much worse (Fitzgibbons & Wightman, 1982). Loudness recruitment is one explanation for why hearing-impaired subjects have larger gap thresholds than normals with narrow-band noises but not pure tones, in that "dips" in the noise are confused for the gap to be detected (Glasberg & Moore, 1992). Likewise, hearing aids may not restore normal temporal resolution due to an inadequate amount of amplification (Nelson & Thomas, 1997), and outcomes may depend on the type of compression used (Moore et al., 2001), as well as other factors. Thus, with the well-established methodology of gap detection, we aim to compare the NC with WDRC on their capacity to restore normal temporal processing abilities.

3.3 Sound Localization

Sound localization is an important auditory ability which contributes to communication (Bronkhorst & Plomp, 1988) and survival, and is dependent on binaural processing of intensity, timing, and spectral information. The processes involved for horizontal and vertical localization are not the same (Middlebrooks & Green, 1991), and the focus of this experiment was on frontal horizontal localization, due to the ease with which the experiment could be set up, and the abundance of prior research on horizontal localization.

3.3.1 Horizontal Localization

Interaural time and level difference (ITD, ILD) cues are the main contributors to localization in the horizontal plane. In particular, ITD cues are used in the localization of low frequency sounds, whereas for high frequency localization, ILD cues are used due to the presence of a head shadow and the unreliability of ITD cues because of the short wavelength of the signal.

There is some indication in the literature that hearing aids impair sound localization (Noble & Byrne, 1990; Van den Bogaert et al., 2006; Keidser et al., 2006).

Hearing aid delay, compression, and directional microphones may all contribute to impaired sound localization with hearing aids, though Keidser et al. (2006) claim that directional microphones impair horizontal localization the most out of the features listed. Thus, there is room for improvement in sound localization for hearing aids, and it is the aim of this experiment to assess how well the NC technology restores horizontal localization relative to WDRC, controlling for features such as directional microphones that may significantly affect localization performance.

3.4 Subjective Measures

While objective measures, such as those described above, can be used to assess objective hearing aid performance in controlled settings, ultimately a hearing aid user's preference matters a great deal, as well. Subjective measures were included in this thesis in order to fill this void.

3.4.1 Abbreviated Profile of Hearing Aid Benefit (APHAB)

The APHAB (Cox & Alexander, 1995) is a condensed version of the Profile of Hearing Aid Performance (PHAP; Cox and Gilmore (1990)), and is widely used in hearing aid research. It is composed of 24 items, and for each item, the participant must circle a letter from A to G, indicating how often a particular statement is true of them (A = 99% of the time, B = 87% of the time, C = 75% of the time, D = 50% of the time, and so on). The 24 items are further subdivided into 4 different 6-item subscales (Ease of Communication, Reverberation, Background Noise, and Aversiveness of Sounds). There is both an unaided and an aided section for the

APHAB. It can be self-administered, but may also be employed in clinical research settings to compare different fittings and technologies, or the same fitting, over time. Thus, it lends itself well to the structure of the present study.

3.4.2 Speech and Music Quality

There exist other hearing aid studies comparing subjective quality measures for speech and music across different hearing aid technologies (Boike & Souza, 2000; Larson et al., 2000; Davies-Venn, Souza, & Fabry, 2007), but often in these studies the different technologies are programmed onto the same hearing aid. Since a version of the NC and WDRC could not be programmed onto the same hearing aid for the purpose of the experiments contained in this thesis, it was impossible to make a side by side comparison of sound quality. Instead, a general sound quality questionnaire was drafted, which asked the participant to report perceived music quality and overall sound quality under different listening conditions. The sound quality questionnaire is attached in the Appendix, and is described in more detail in the next chapter.

Chapter 4

NC Experiment 1

4.1 Design

The purpose of the first experiment was to compare the NC to WDRC using a realistic sample of WDRC hearing aids. Thus, we decided to recruit participants who currently wore WDRC hearing aids, and lent out a set of NC hearing aids for each participant to wear.

Figure 4.1 shows a week by week schematic of the study design. At the very beginning of the study, participants were issued a freeform diary, and were encouraged to describe their experiences of various sounds in their everyday environments. Specifically, they were asked to take note of any unusual sounds, any instances of feedback, and whether any sounds were too loud or too quiet. Half of the participants were randomly assigned to begin the study wearing their own WDRC hearing aids, while the other half were fit with NC hearing aids. After a two week period of wearing the hearing aids (deemed the first adaptation period), if a participant started out with

NCs, they were given an adjustment to the NCs by a trained hearing instrument specialist (HIS). After the NCs had been adjusted, participants wore the hearing aids for another two weeks (the second adaptation period) before coming in for two, two-hour test sessions, spaced one week apart. Otherwise, if a participant was assigned to wear WDRC hearing aids first, they waited two weeks before coming in for the two, two hour test sessions.

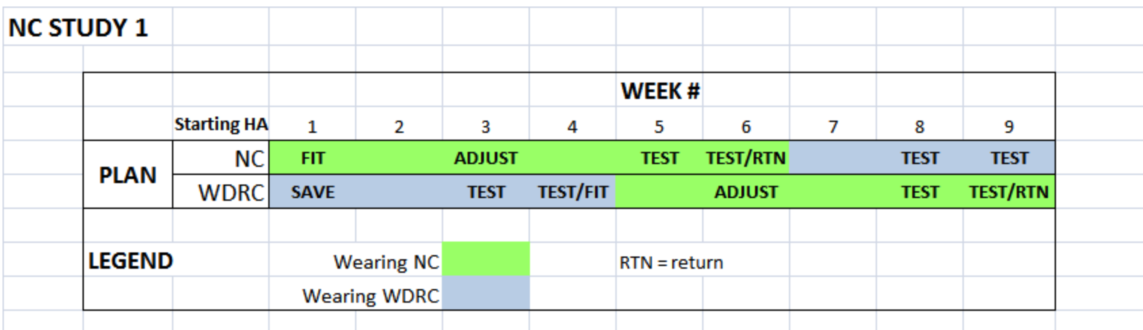

Figure 4.1: NCStudy1 design. Half of the participants started the study wearing WDRC hearing aids, while the other half were fit with NCs at the outset. Adjustments were only given to the NCs, as the participants did not require adjustments to their own set of WDRC hearing aids, which they had already adapted to over months and years of wearing them. Halfway through the study, participants began wearing the second set of hearing aids. Two test sessions were conducted for each pair of hearing aids.

Adjustments were guided by observations that participants had recorded in their diaries, as well as from verbal reports given by the participant at the time of the adjustment. Following the second test session, participants who first wore NCs returned the set of NCs, and participants who first wore WDRC were given a set of NCs which had been preprogrammed to fit their loss.

4.2 Participants

4.2.1 Demographics and Prior Experience

A total of 8 participants qualified for the study (3 male, 5 female), with an average age of 76 (SD = 7.7). All participants had prior experience wearing hearing aids for several years, and spoke English as their primary language. The hearing aids that participants wore coming into the study were from assorted manufacturers (Oticon, Phonak, Siemens), and were in different styles (mostly behind-the-ear (BTE) or receiver-in-the-canal (RIC); 2 participants wore in-the-canal styles (ITC)). The average age of the WDRC hearing aids being worn in the study was 3 years (SD = 2).

4.2.2 Audiograms

Prior to their inclusion in the study, participants visited McMaster University for a hearing test; air conduction thresholds for both ears and bone conduction thresholds for the worse ear were measured. A trained student performed the measurements, following the British Society of Audiology recommended procedure for pure-tone audiometry (British Society of Audiology, 2004) as closely as possible.

To qualify for the study, participants were required to have symmetrical sensorineural hearing loss and currently wear 2 hearing aids. More specifically, participants were included on the basis of the following criteria: 1) no difference greater than 10 dB HL between ears on the calculated pure tone average hearing loss (PTA; the average of 0.5, 1, and 2 kHz), 2) no difference greater than 30 dB HL at any audiometric frequency between the two ears, and 3) no difference greater than 15 dB HL between air conduction and bone conduction thresholds at any audiometric frequency. Due to the limited supply of subjects, one participant was included who had an asymmetrical hearing loss, according to our definition (subject 05, Figure 4.2).

The first session required participants to visit McMaster University, in order for the HIS to hook up their existing hearing aids so that their old audiogram data could be read and extracted. If they were assigned to wear NC hearing aids first, they were given a pair of NCs at the end of the session. The purpose of collecting the old audiogram data was to allow for a fair comparison between the NC and WDRC. Hearing loss changes over time, and several of the participants had not had their hearing retested or their hearing aids adjusted for months, and in some cases, years. Thus, programming the NCs using a newer, fresher audiogram, would have introduced a bias in favor of the NC. For two participants, we were unable to read the old audiogram data programmed on their existing hearing aids, and thus had no choice but to use the new audiogram data to program the NCs. Figure 4.2 plots both the new and the old audiogram data for each participant.

4.2.3 Subjects Analyzed

Despite initially recruiting 8 subjects, only 5 subjects were suitable for data analysis (subjects 2, 3, 4, 7, and 8). Subject 1 was removed from data analysis because he/she was unable to complete some of the tasks. Subject 5 was removed from analysis because of the asymmetrical nature of his/her hearing loss. Subject 6 dropped out of the study after a few weeks of participation, so his/her dataset was incomplete.

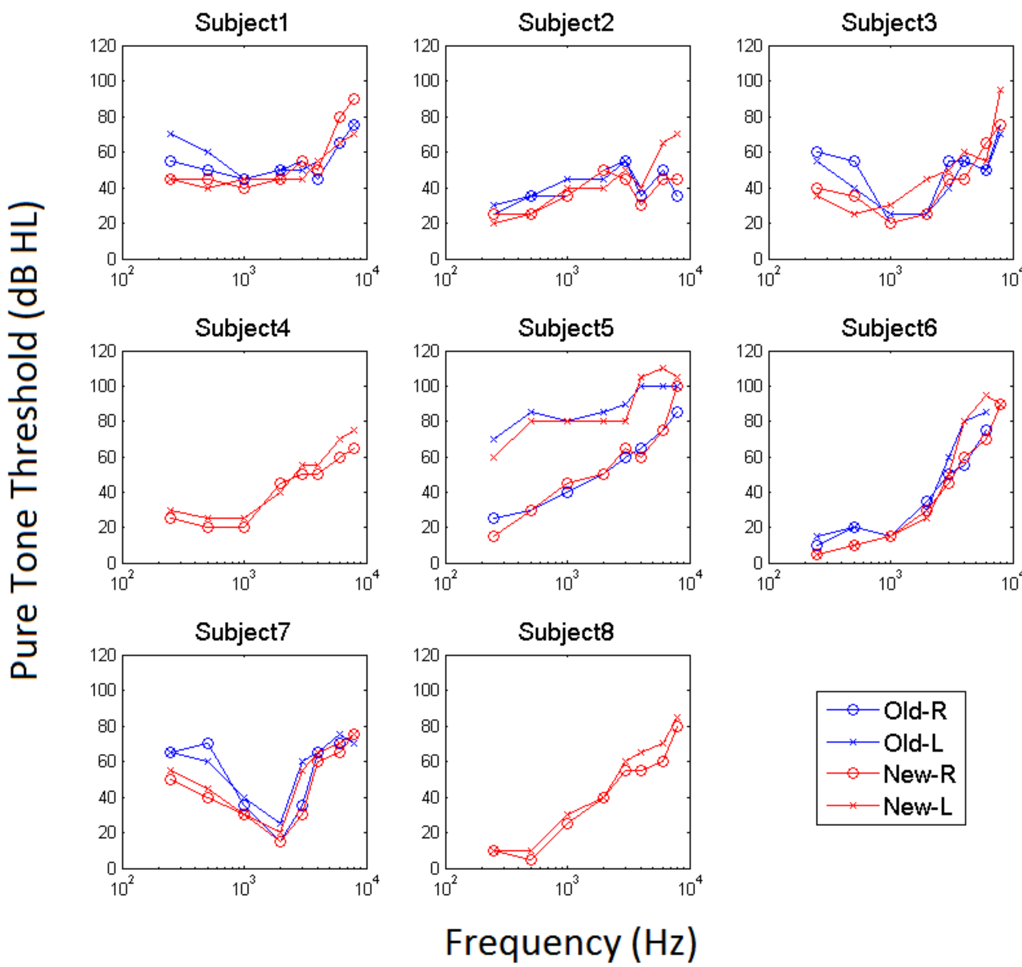

Figure 4.2: Old and new audiograms for all subjects. New audiograms are plotted in red, old audiograms in blue, x's denote measurements obtained for the left ear, and o's denote measurements obtained for the right ear. For the most part, the new and old audiograms matched quite closely. Subject 5 had an asymmetrical hearing loss. The HIS was unable to read audiogram data from subjects 4 and 8 due to software incompatibilities with their devices.

4.3 Testing Room and General Procedures

The two-hour test sessions were conducted in a sound attenuating chamber. For all of the experiments, stimuli were either prerecorded or generated using Matlab (exception: the HINT was programmed with Visual Basic 6). Stimuli were presented through two Tucker-Davis-Technologies Programmable Attenuators (TDT PA5s; one for the left channel and one for the right channel), routed through a TDT Signal Mixer (TDT SM5) to create a mono signal, amplified with a Hafler P1000 Trans.Ana 110 watt amp, and subsequently routed to the appropriate loudspeaker using a TDT RV8 and PM2. There were 7 speakers (Audio Video Methods P73), arranged in a 90° arc for the purpose of testing sound localization. All of the other experiments used only one speaker, located directly in front of the participant (0° azimuth).

All SPL measurements were performed with a Bruel & Kjaer 2239A sound level meter located at the centre of where the participant's head would be located in space. To ensure consistency in the participant's location in the room, tape markings were made on the floor, and a table rested within these tape markings. For the hVd, VCV, mistuned harmonic, timbre perception, and gap detection experiments, participants responded using a touchscreen (Planar PT1510MX), which was propped up on the table in a consistent location. For the other experiments (sound localization, HINT), the experimenter rested the touchscreen in his/her lap, and input the participants' responses. Participants sat 1.5 meters away from the speakers, with their ears level with the centre of the speaker cones (cone diameter of 17 cm). Figure 4.3 shows the room setup for the sound localization experiment.

In the following sections, stimuli, procedures, and results will be presented one

Figure 4.3: Sound attenuating chamber - experimental setup. With the exception of the sound localization experiment, participants faced speaker A for all experiments.

by one for each task, organized by auditory domain (speech intelligibility, music perception, sound localization, subjective measures). Analyses for all experiments were performed in R (R Core Team, 2013). Where appropriate, the Huynh-Feldt correction for deviations from sphericity ($\tilde{\epsilon}$) was used to compute p values from F tests on within-subjects factors (Maxwell & Delaney, 2004). In data tables, degrees of freedom and sums of squares for both the numerator (effect of interest) and denominator (error) are reported, along with F and p values, and a measure of effect size, generalized eta squared (abbreviated *ges*).

4.4 Speech Intelligibility

4.4.1 Hearing In Noise Test (HINT)

Procedure

Participants sat in a sound attenuating chamber facing the speaker at a 0 ° azimuth. The speech-shaped noise (SSN) was measured to be 65 dB(A), and this presentation level remained constant during the course of the experiment. For each session (one for each set of hearing aids), 2 lists of sentences were used, taking care not to use the same lists in the second session as the first session. The presentation level of the first sentence was initially attenuated such that its presentation level was 10 dB less than the noise, and was increased by 4 dB until it was recognized correctly by the participant, at which point it was decreased by 4 dB. For the next 3 sentences, if the participant correctly (incorrectly) recognized the sentence, the sentence presentation level was decreased (increased) by 4 dB. Similarly, the presentation level of each of the remaining 16 sentences was adaptively adjusted based on whether the previous response was correct or incorrect, but instead a 2 dB step was used. To obtain a SRT measurement for a participant (SNR at 50% correct), the presentation levels of the fifth sentence to the twentieth sentence, including the level of a would-be twenty-first sentence, were averaged. The SNR was computed by subtracting the noise level from the average sentence presentation level.

Results

A within-subjects one-way ANOVA with a 2-level factor (hearing aid) was performed, with SRT as the dependent measure. There was no significant effect of hearing aid ($F(1,4) = 0.42$, $p = 0.551$). Group means and standard errors are plotted in Figure 4.4, and individual SRT means and standard deviations are plotted in Figure 4.5. Individual differences in speech intelligibility in noise were far greater than any difference between the two hearing aids.

4.4.2 Consonant Vowel Consonant (CVC)

Stimuli

CVC stimuli were extracted from the Hillenbrand et al. (1995) corpus, which contains 12 vowel sounds (/i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u, ʌ, ɝ, e, o/) in hVd syllables spoken by 45 men, 48 women, and 46 children. All stimuli were normalized to the same root-mean square level. We selected 10 talkers in total (5 men, 5 women) and 11 hVd pronunciations for a total of 110 CVC tokens (10 talkers * 11 syllables). /hod/ was removed from the test set, since it sounded too similar to /hawed/ and could have resulted in ambiguity in the analysis of responses.

Multi-talker babble was used as a noise condition that featured 100 talkers in a

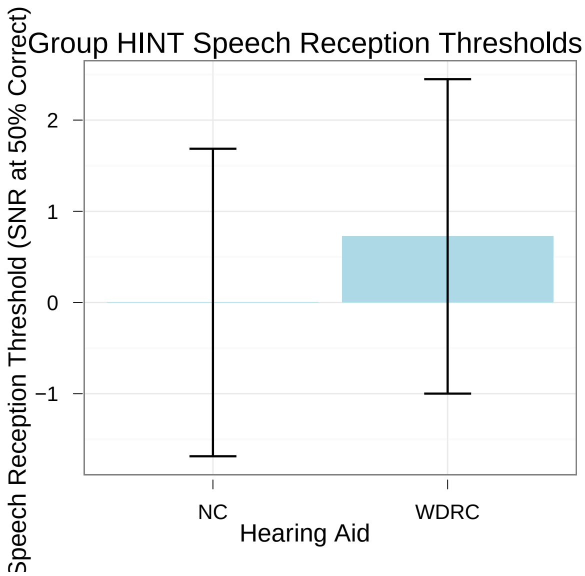

Figure 4.4: NCStudy1 group results on the HINT. Error bars represent ± 1 SEM. Speech reception thresholds were calculated by subtracting the noise level from the average presentation level of the last 16 sentences presented. There was no significant difference in performance between NC and WDRC (F(1,4) = 0.42, p = 0.551).

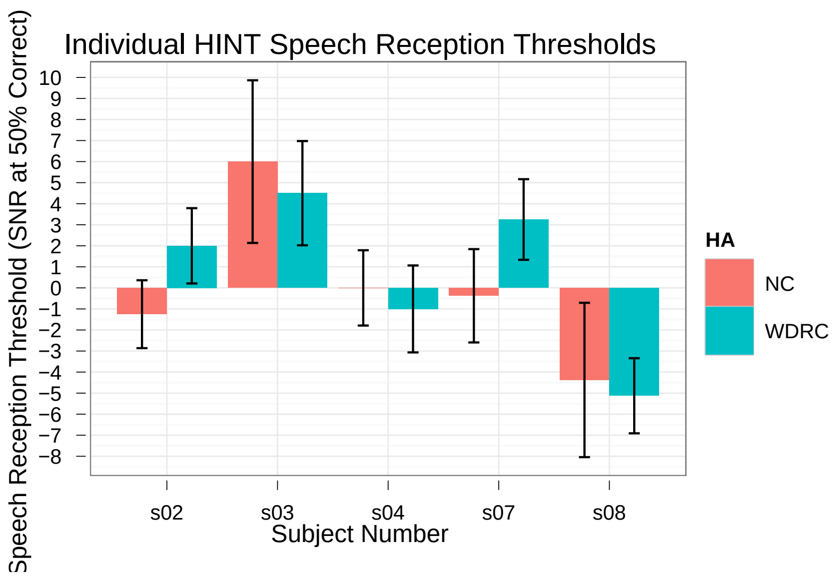

Figure 4.5: NCStudy1 individual results on the HINT. Error bars represent ± 1 SD. Speech reception thresholds were calculated by subtracting the noise level from the average presentation level of the last 16 sentences presented. Most participants performed approximately equally between the NC and WDRC, however, it appears that subjects 02 and 07 performed considerably better with the NC.

cafeteria. The babble was downsampled from its original sampling rate to 16000 Hz to match the sampling rate of the Hillenbrand et al. stimuli.

Procedure

This task took approximately 25 minutes to complete. The participant sat in the sound attenuating chamber and listened to CVC stimuli either in quiet or embedded in multi-talker babble, presented through the speaker at a 0° azimuth.

The CVC tokens had an average presentation level of 67.5 dB(A), and for the noise condition, the noise level was set to 62.5 dB(A). In the noise condition, a 2000 ms clip of noise was randomly selected from the multi-talker babble source on each trial, and the CVC stimulus began anywhere in the range of 600 ms to 1000 ms after noise onset, with equal probability (rectangular distribution). The noise sample was cosine windowed with a rise and fall of 100 ms. Following stimulus delivery on a given trial, the participant pressed on the word they heard (had, hawed, hayed, head, heard, heed, hid, hoed, hood, who'd, hud) using the touchscreen. After responding, there was a 1000 ms period of silence before the next trial began. Feedback was not given on correct or incorrect responses.

There were a total of 5 blocks, each containing 2 presentations of each word in both quiet and noise, resulting in 44 presentations per block (2 talkers * 2 conditions * 11 words), and a grand total of 220 trials for the whole experiment. The presentation order was completely random within a block, and all CVC tokens were used twice (once for quiet, once for noise).

There was no training session, although prior to beginning the experiment, the experimenter asked the participant to read the words on the touchscreen out loud. If one of the tokens was pronounced incorrectly, the experimenter verbalized the correct pronunciation and asked the participant to repeat it. Breaks between blocks were self-paced.

Results

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and condition (quiet or noise). The dependent measure was the percentage of hVd tokens correctly identified. The only significant effect was condition ($F(1,4) = 19.23$, $p = 0.011$), such that the noise condition was more difficult than the quiet condition. Detailed statistical results may be seen in Table 4.1. Group means and standard errors are plotted in Figure 4.6, and individual means are plotted in Figure 4.7.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	4	1.5	18.8	0.32	0.603		0.001
cond	1	4	464.0	96.4	19.23	0.011	*	0.241
HA:cond	1	4	4.1	26.9	0.62	0.477		0.003

Table 4.1: hVd Within-Subjects ANOVA

Although there were no significant differences between hearing aids on the overall percentage of hVd tokens correctly identified, there were fairly consistent differences between the hearing aids on individual phonemes. Figure 4.8 plots a confusion matrix both for NC and WDRC averaged over both noise conditions, to illustrate the difficulty of identifying individual phonemes and dominant confusions made in the experiment. Each value in the confusion matrix is the percentage of times a given response was made, given the stimulus that was presented. Adding all values across a single row will total 100%, to within rounding error. Figure 4.9 plots the difference

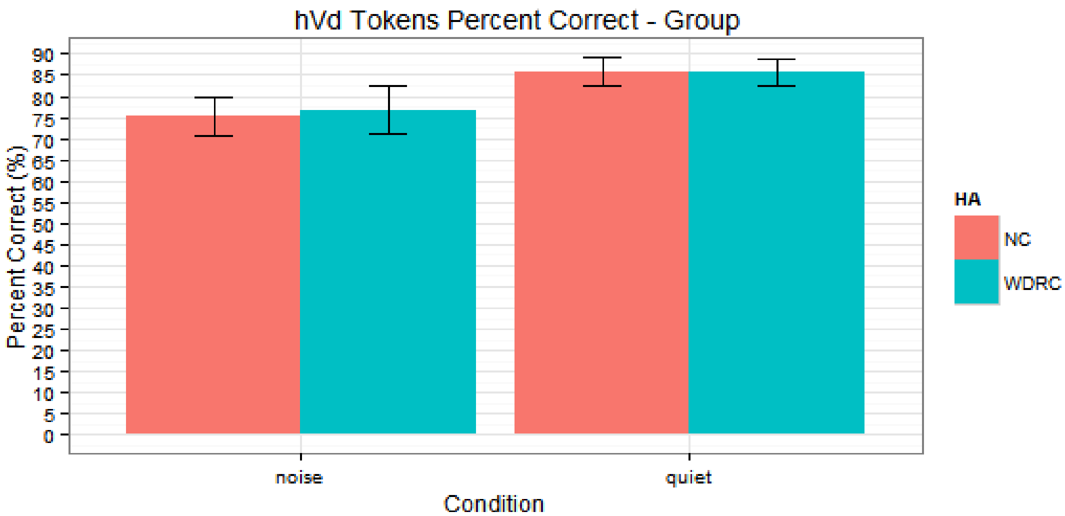

Figure 4.6: NCStudy1 hVd token percent correct - group. Error bars represent ± 1 SEM. There was no significant difference in performance between NC and WDRC on the percentage of hVd tokens correctly identified (F(1,4) = 0.32, p = 0.603).

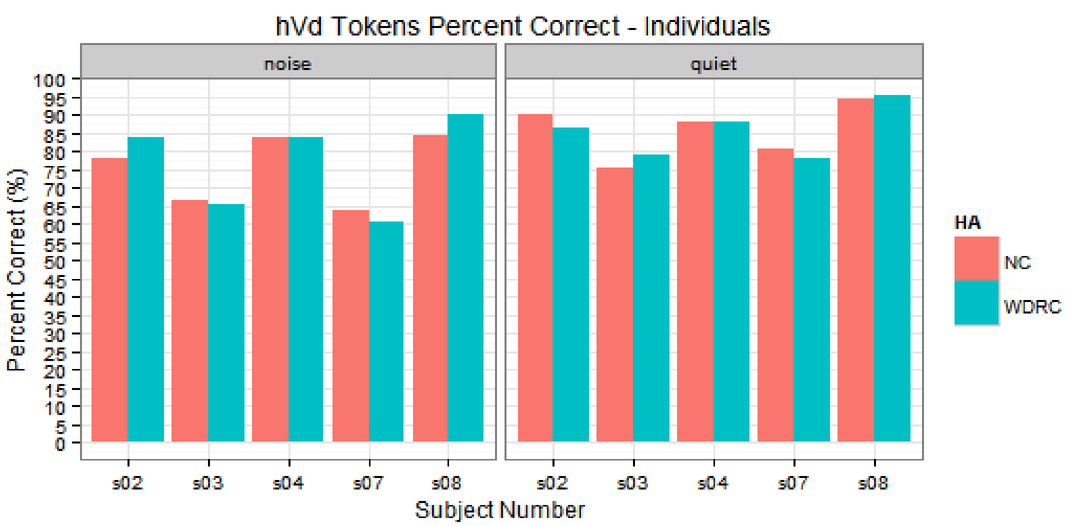

Figure 4.7: NCStudy1 hVd token percent correct - individuals. Some participants correctly identified more hVd tokens with WDRC, while others performed better with NC.

between the NC and WDRC confusion matrices separately for both the quiet and noise conditions.

4.4.3 Vowel Consonant Vowel (VCV)

Stimuli

The stimuli used for the VCV experiment were extracted from the training set found in The Interspeech 2008 Consonant Challenge (Cooke & Scharenborg, 2008). The VCV stimuli contained in this corpus were professionally recorded by trained British talkers with no strong regional accents, and normalized to the same root-mean square level. For our experiment, we selected 6 talkers with complete data sets (3 male, 3 female), 21 consonant sounds (/b/, /d/, /g/, /p/, /t/, /k/, /s/, /sh/, /f/, /v/, /th/, /ch/, /z/, /zh/, /h/, /m/, /n/, /w/, /r/, /y/, /l/), 1 vowel (/æ/), 1 stress (first syllable), and 2 conditions (noise and quiet), for a grand total of 252 VCV tokens (6 talkers * 21 consonant sounds * 1 vowel * 1 stress * 2 conditions).

Procedure

Similar to the CVC task, the whole procedure took approximately 25 minutes. The participant sat in the sound attenuating chamber while VCV stimuli were presented through a loudspeaker, with the same positioning as that used in the HINT and CVC experiments. On the table in front of the participant sat a touchscreen displaying a grid of consonants and example pronunciations, representing all possible responses the participant could choose from. After the presentation of a VCV stimulus, the participant selected the consonant from the grid that corresponded best with the consonant that they heard. The presentation level of the VCV nonsense syllables was

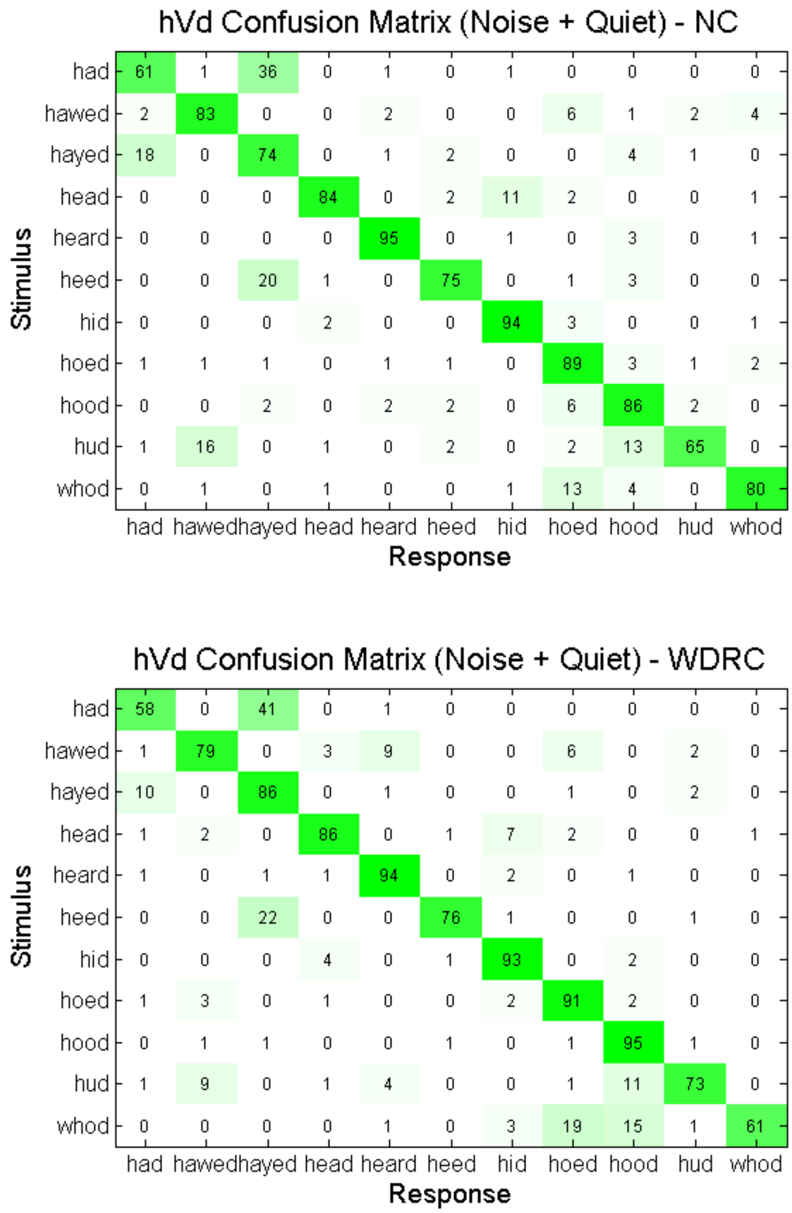

Figure 4.8: NCStudy1 hVd overall confusion matrices. Data shown are for the noise and quiet conditions amalgamated together. *Top*: hVd confusion matrix for NC. *Bottom*: hVd confusion matrix for WDRC. For many of the syllables, performance is at near ceiling. Few differences exist between NC and WDRC; the words where WDRC and NC differ the most are /hayed/, /hood/, /hud/, and /who'd/.

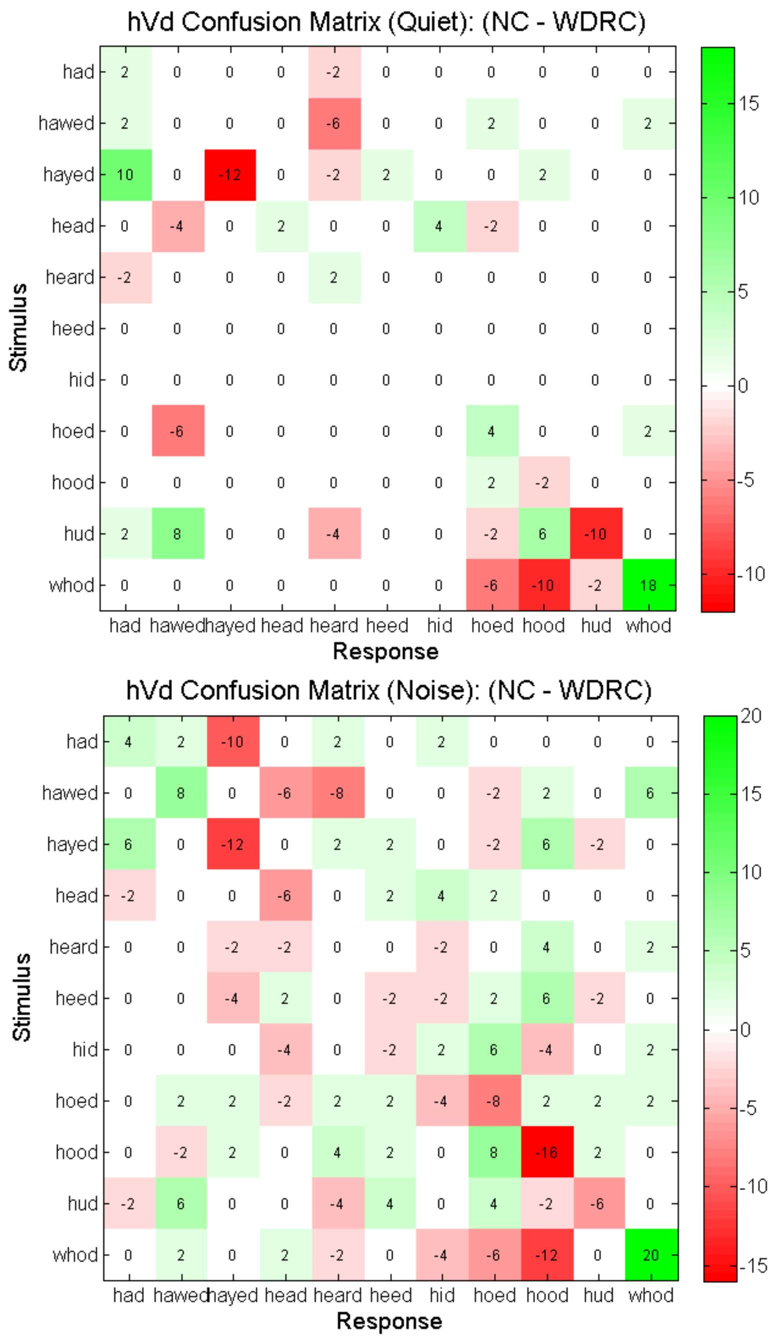

Figure 4.9: NCStudy1 difference between hVd confusion matrices (NC - WDRC). *Top:* Difference between hVd confusion matrices for the quiet condition. *Bottom:* Difference between hVd confusion matrices for the noise condition. Each tile represents NC % response - WDRC % response. In quiet, NC seems to restore /who'd/ better than WDRC, while WDRC favors /hayed/ and /hud/. In noise, NC restores /who'd/ better than WDRC, while WDRC restores /hayed/ and /hood/ better.

fixed at 67.5 dB(A), and the noise was fixed at 62.5 dB(A).

In the noise condition, a 2000 ms clip of noise was randomly selected from the multi-talker babble source on each trial, and the CVC stimulus began anywhere in the range of 600 ms to 1000 ms after noise onset, with equal probability (rectangular distribution). The noise sample was cosine windowed with a rise and fall of 100 ms. Following stimulus delivery on a given trial, the participant pressed on the VCV token that they heard, using the touchscreen. After responding, there was a 1000 ms period of silence before the next trial began. Feedback was not given on correct or incorrect responses.

There were a total of 6 blocks, each containing 1 presentation of each VCV token in both quiet and noise, resulting in 42 presentations per block (1 talker * 2 conditions * 21 words). The presentation order was completely random within a block, and all VCV tokens were used twice (one for quiet, one for noise).

There was no training session, although prior to beginning the experiment, the experimenter asked the participant to read the words on the touchscreen out loud. If one of the tokens was pronounced incorrectly, the experimenter verbalized the correct pronunciation and asked the participant to repeat it.

Results

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and condition (quiet or noise). The dependent measure was the percentage of VCV tokens correctly identified. The only significant effect was a main effect of condition ($F(1,4) = 168.62$, $p < 0.001$), such that the noise condition was much more difficult than the quiet condition. Detailed statistical results may be seen in Table 4.2. Group means and standard errors are plotted in Figure 4.10, and individual means are plotted in Figure 4.11.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	4	99.1	200.8	1.97	0.233		0.037
cond	1	4	2687.0	63.7	168.62	0.000	***	0.509
HA:cond	1	4	8.2	24.3	1.34	0.311		0.003

Table 4.2: VCV Within-Subjects ANOVA

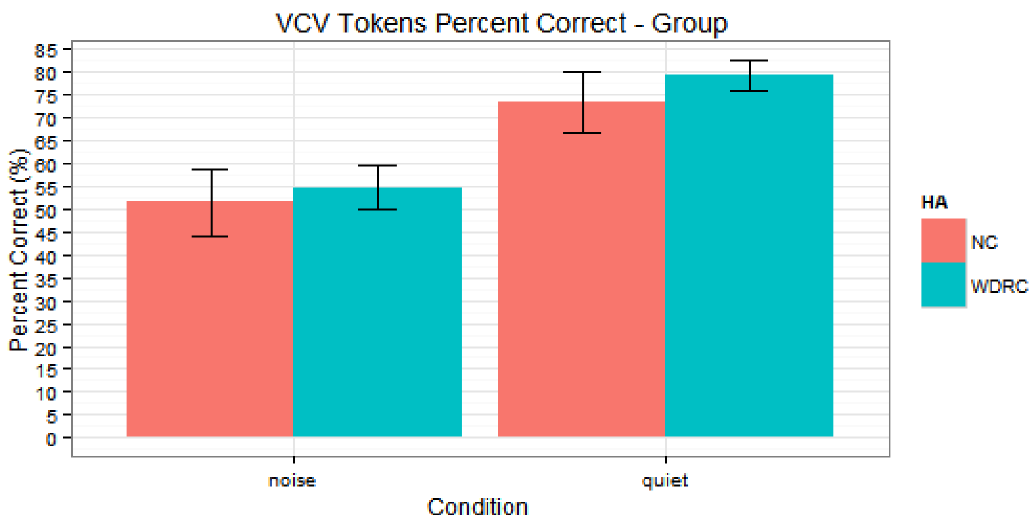

Figure 4.10: NCStudy1 VCV token percent correct - group. Error bars represent ± 1 SEM. There was no significant difference in performance between NC and WDRC on the percentage of VCV tokens correctly identified (F(1,4) = 1.97, p = 0.233).

Although there were no significant differences between hearing aids on the overall percentage of VCV tokens correctly identified, there were consistent differences between the hearing aids on individual phonemes. Figure 4.12 plots a confusion matrix both for NC and WDRC averaged over both the quiet and noise conditions, to illustrate the difficulty of individual phonemes in the experiment. Figure 4.13 plots

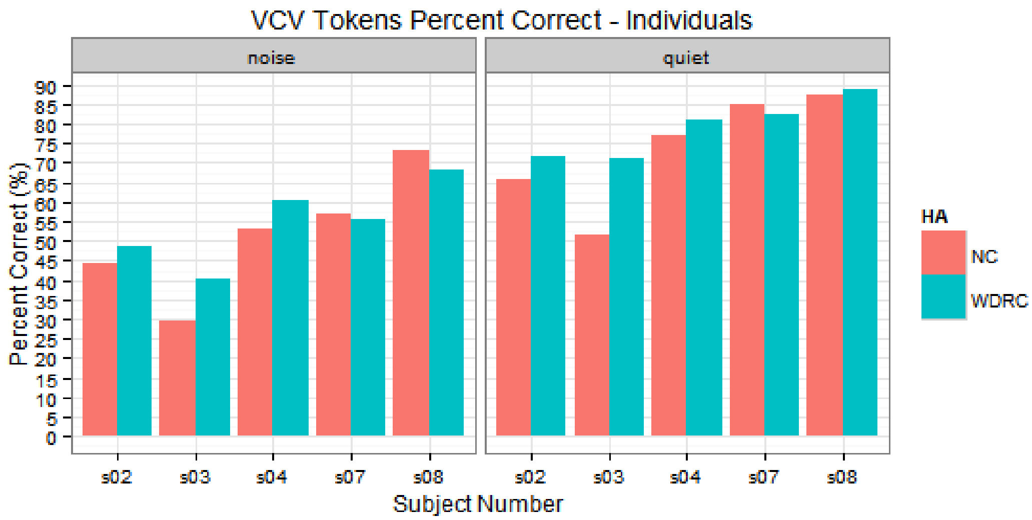

Figure 4.11: NCStudy1 VCV token percent correct - individuals. There were some individuals who performed better with WDRC (in particular, subject 03), while others did better with NC.

the difference between the NC and WDRC confusion matrices for both the quiet and noise SNR conditions.

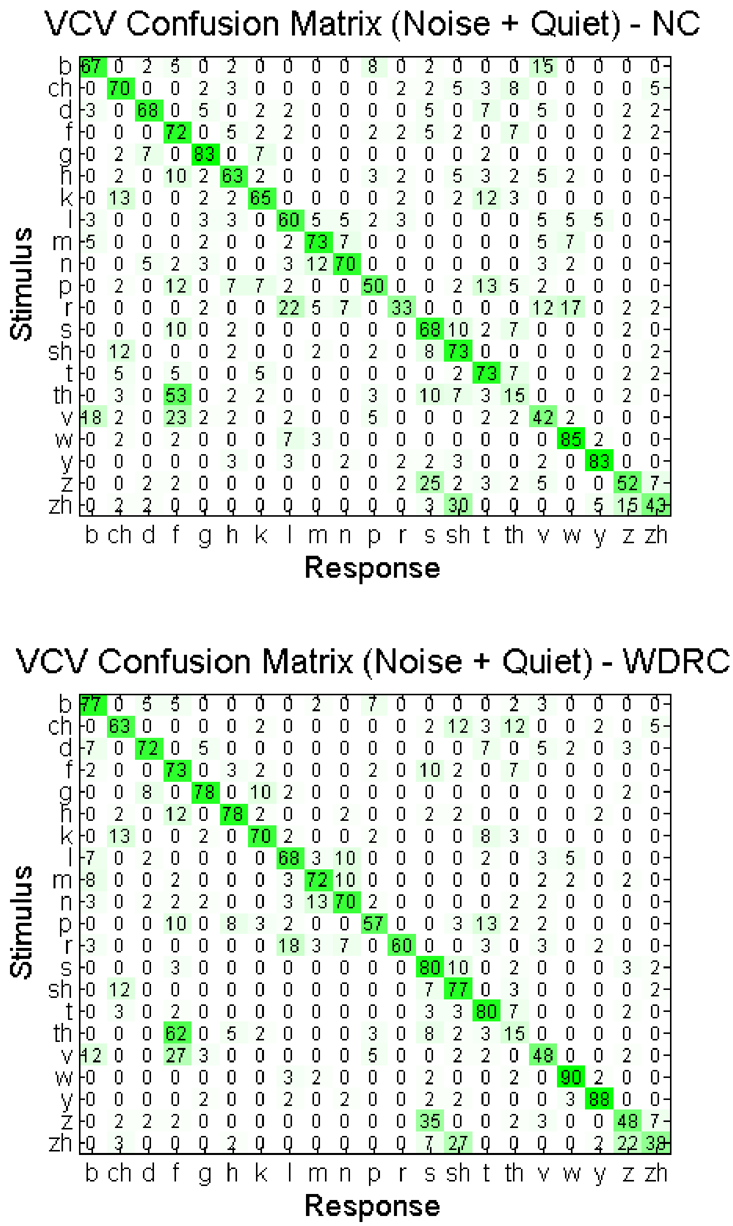

Figure 4.12: NCStudy1 VCV overall confusion matrices. Data shown are for the noise and quiet conditions amalgamated together. *Top*: VCV confusion matrix for NC. *Bottom*: VCV confusion matrix for WDRC.

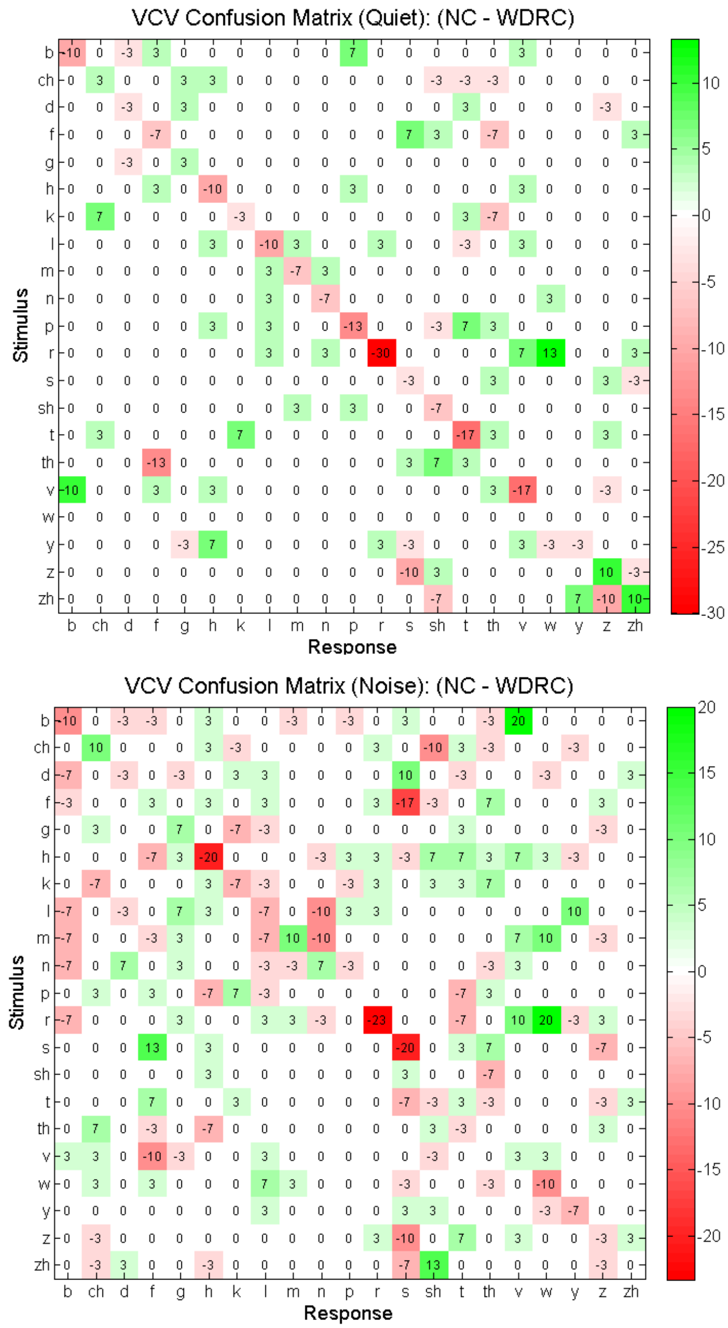

Figure 4.13: NCStudy1 difference between VCV confusion matrices (NC - WDRC). *Top:* Difference between VCV confusion matrices for the quiet condition. *Bottom:* Difference between VCV confusion matrices for the noise condition. The NC had consistent trouble with a few consonants across both the noise and quiet conditions (/b/, /h/, and /r/).

4.5 Music Perception

4.5.1 Mistuned Harmonic

Stimuli

Complex tones with 10 harmonics (fundamental frequency of 200 Hz or 600 Hz) of random phase and equal intensity were created in Matlab. The stimulus was presented at a level of 65 dB SPL measured in absence of the participant at the location of where the participant's head would be in space. The complex tones were presented in a background of white Gaussian noise set at 55 dB(A) (resulting in a SNR of +10 dB). The level of the noise was decided upon by presenting the complex tones and noise to the NC to see if any feedback or entrainment could be detected by the experimenter. Lower SNR values corresponded with less entrainment, so we selected the approximate maximum SNR which produced no audible entrainment.

Procedure

A three-alternative forced-choice (3AFC) adaptive staircase design (Levitt, 1971) with a 2-down, 1-up rule was used to estimate the mistuned harmonic threshold. A trial consisted of three, 500 ms presentations of a complex tone with an inter-stimulus interval (ISI) of 500 ms. One of the 3 complex tones in a trial had the third harmonic mistuned in the upward direction, starting with a (maximum) mistuning of 30%, and it was the task of the participant to identify "which tone sounded different than the other two" by selecting the corresponding button on a touchscreen.

There were 2 blocks in the experiment, and the blocks differed with respect to the fundamental frequency of the complex tone. The first block was always the 200

Hz tone, and the second block was always the 600 Hz tone. Each block consisted of 12 reversals (defined as a transition from at least 2 correct responses to an incorrect response, or vice versa). The percentage mistuning was altered by a factor of 1.4 according to the 2-down, 1-up rule. The mistuned harmonic threshold for a given block was computed by averaging the thresholds measured at the last 8 reversals in the block.

Continuous white Gaussian noise played throughout the presentation of the complex tones, including a 500 ms segment before the first complex tone and a 500 ms segment after the last complex tone. Both the complex tones and the white Gaussian noise were cosine ramped with a linear rise/fall time of 10 ms.

Feedback on task performance (correct or incorrect) was given to participants on every trial via the touchscreen, and was displayed for 2000 ms. Following the feedback, there was a silent period for 1000 ms preceding the onset of the next trial. A practice session featuring 6 trials was included before the first experimental block to ensure that the participant understood the instructions. The practice session was repeated until participants admitted that they understood the task, and until they exceeded 50% correct on the practice trials. For the vast majority of subjects, only 1 practice session was required.

Results

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and tone (200 Hz or 600 Hz). The dependent measure was the mistuned harmonic threshold, computed by averaging the last 8 reversals on each block. There were no significant main effects or an interaction. Detailed statistical results may be seen in Table 4.3. Group means and standard errors are plotted in Figure 4.14, and individual means and standard deviations are plotted in Figure 4.15.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	4	47.5	67.6	2.81	0.169		0.076
harm	1	4	2.3	136.2	0.07	0.810		0.004
HA:harm	1	4	8.1	115.9	0.28	0.623		0.014

Table 4.3: Mistuned Harmonic Within-Subjects Factorial ANOVA

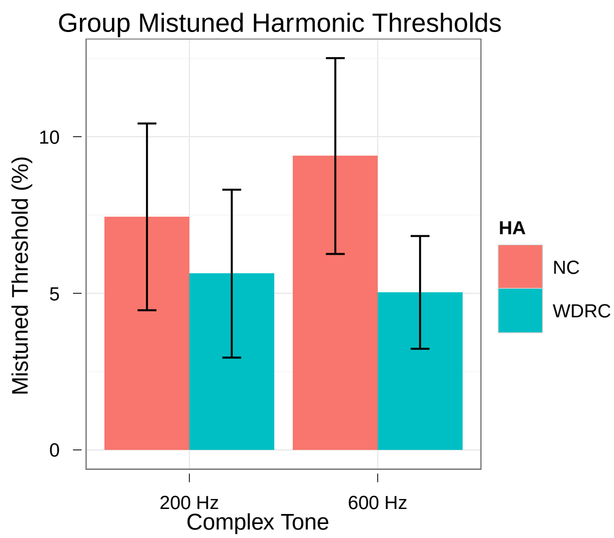

Figure 4.14: NCStudy1 mistuned harmonic - group thresholds. Error bars represent ± 1 SEM. There was no significant difference between NC and WDRC (F(1,4) = 2.81, p = 0.169).

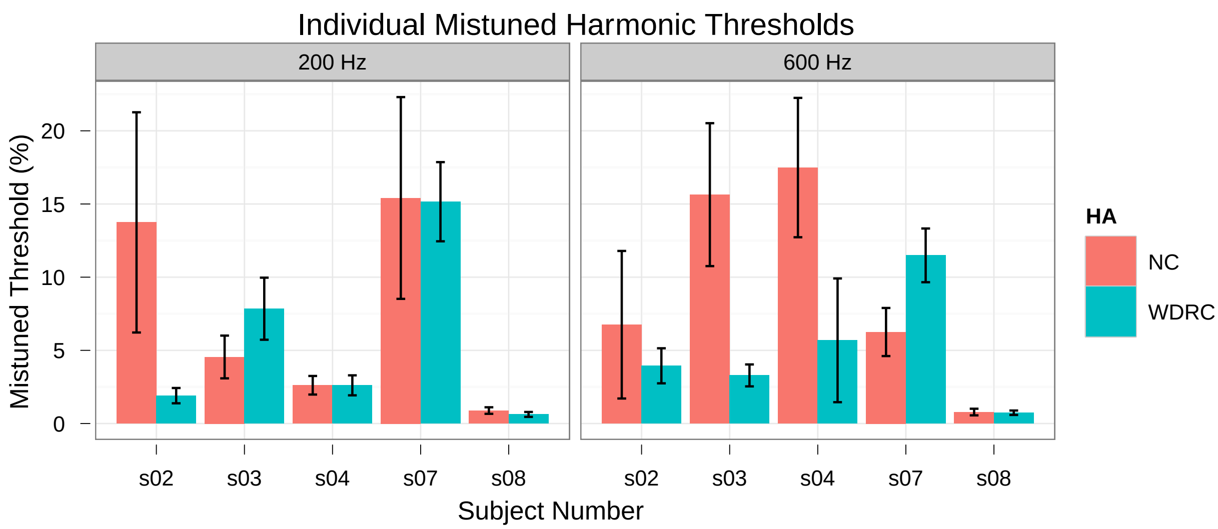

Figure 4.15: NCStudy1 mistuned harmonic - individual thresholds. Error bars represent ± 1 SD. There are huge individual differences on this task.

4.5.2 Timbre Perception

Stimuli

A real recording of a tenor saxophone extracted from the Real World Computing Music Database (Goto, Hashiguchi, Nishimura, & Oka, 2008) was used as the stimulus for the experiment. The saxophone recording is of one note (F5), sounding for a duration of 953 ms. The sound was calibrated to a presentation level of 65 dB SPL. In order to prevent entrainment, the stimulus was played in the presence of white Gaussian noise at a SNR of +20 (the noise was measured to be 45 dB(A)).

Procedure

A 3AFC adaptive staircase design (Levitt, 1971) with a 2-down, 1-up rule was used to estimate the threshold for detecting a change in timbre. A trial consisted of 3 presentations of the saxophone stimulus (zero-padded to be 1000 ms in length) with an ISI of 1000 ms. One of the saxophone notes in a trial had one of its harmonics altered in intensity, beginning by it being completely removed, and it was the task of the participant to identify “which saxophone note sounded different than the other two” by selecting the corresponding button on a touchscreen. Altering the intensity of a single harmonic changes the overall level of the instrument, and to compensate for this alteration the power of all other harmonics were multiplied by a constant to equate acoustic power across all three note presentations.

There were two blocks in the experiment, and the blocks differed with respect to the harmonic that was altered. The first block altered the fundamental, and the second block altered the second harmonic. These harmonics were chosen as they had the most acoustic power and therefore allowed for the most attenuation. Each block consisted of 12 reversals (defined as a transition from at least 2 correct responses to an incorrect response, or vice versa). Intensity was altered by an additive 20% according to the 2-down, 1-up rule, and once 4 reversals had occurred, the percentage intensity step size was changed to 10%. The harmonic intensity threshold for a given block was computed by averaging the thresholds measured at the last 8 reversals in the block.

Continuous white Gaussian noise played throughout the presentation of the saxophone notes, including a 1000 ms segment before the first note and a 1000 ms segment after the last note. The white Gaussian noise was ramped with a linear rise/fall time of 10 ms.

Feedback on performance (correct or incorrect) was given to participants on every trial via the touchscreen, and was displayed for 2000 ms. Following the feedback, there was a silent period for 1000 ms preceding the onset of the next trial. A practice session featuring 6 trials was included before the first experimental block to ensure that the participant understood the instructions. The practice session was repeated until participants admitted that they understood the task, and until they exceeded 50% correct on the practice trials. For the vast majority of subjects, only 1 practice session was required.

Results

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and harmonic (fundamental or second harmonic). The dependent measure was the harmonic intensity threshold. The only significant effect was a main effect of harmonic ($F(1,4) = 39.96$, $p = 0.003$), such that participants were more sensitive to changes in intensity of the fundamental than the second harmonic. However, there was a trending main effect of hearing aid ($F(1,4) = 6.16$, $p = 0.068$), such that participants tended to perform better with NC than WDRC. Detailed statistical results may be seen in Table 4.4. Group means and standard errors are plotted in Figure 4.16, and individual means and standard deviations are plotted in Figure 4.17.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	4	446	289	6.16	0.068	.	0.146
harm	1	4	1684	169	39.96	0.003	**	0.392
HA:harm	1	4	5	1019	0.02	0.901		0.002

Table 4.4: Timbre Perception Within-Subjects Factorial ANOVA

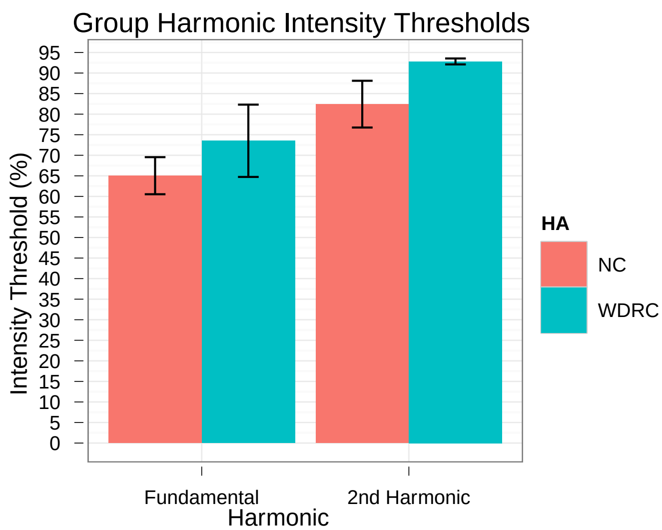

Figure 4.16: NCStudy1 timbre perception - group thresholds. Error bars represent ± 1 SEM. Participants tended to more accurately discriminate changes in timbre with the NC ($F(1,4) = 6.16$, $p = 0.068$), and intensity changes to the fundamental were easier to discriminate than changes to the second harmonic ($F(1,4) = 39.96$, $p = 0.003$).

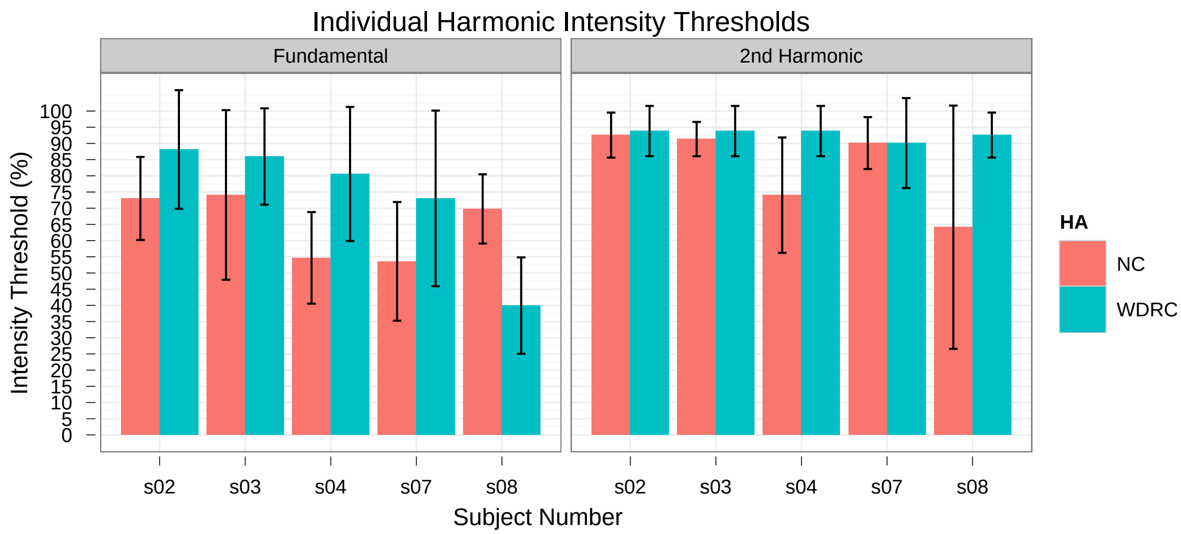

Figure 4.17: NCStudy1 timbre perception - individual thresholds. Error bars represent $\pm$ 1 SD. For all cases except the fundamental for subject 08, participants were either equally as good, or better, at discriminating intensity changes with the NC.

4.5.3 Gap Detection

Stimuli

Noise bursts were generated in Matlab with a bandwidth of 50 Hz and a centre frequency of 1000 Hz. The noise bursts were 500 ms, cosine ramped with a rise/fall of 10 ms, and one of the noise bursts in a trial had a silent gap in the middle. A cosine ramp was used to gate the gap on with a time frame of 1 ms, and the same time frame for gating the gap off. The filtering and signal processing introduced noticeable spectral splatter, which can affect detection of the gap via off-frequency listening. To prevent off-frequency listening, the gap stimuli were presented in white Gaussian noise. The level of the noise was set at 65 dB(A), and the level of the noise burst was 70 dB(A).

Procedure

A 3AFC adaptive staircase design (Levitt, 1971) was used in conjunction with a 2-down, 1-up rule and an ISI of 500 ms. Overall, a trial was 3500 ms in length, starting with white Gaussian noise only and ending with white Gaussian noise only, with the onset of the first noise burst beginning 500 ms into the stimulus. It was the task of the participant to determine "which one of the three sounds sounded different from the other two", by pressing the appropriate button on the touchscreen.

The gap started with a (maximum) duration of 100 ms, and its duration was altered by a factor of 1.4 according to the 2-down, 1-up rule, continuing on for 12 reversals. Reversals were defined as a transition from at least 2 correct responses to an incorrect response, or vice versa. After the fourth reversal, the factor by which the gap duration was altered changed to 1.2.

Three separate blocks were run to obtain a more accurate estimate of the gap threshold, with a self-paced break in the middle of each block. Feedback on performance (correct or incorrect) was given to participants on every trial via the touchscreen, and was displayed for 2000 ms. Following the feedback, there was a silent period for 1000 ms preceding the onset of the next trial. The gap detection threshold for a given block was computed by averaging together the last 8 reversals in the block.

A practice session featuring 6 trials was included before the first experimental block to ensure that the participant understood the instructions. The practice session was repeated until participants admitted that they understood the task, and until they exceeded 50% correct on the practice trials. For the vast majority of subjects, only 1 practice session was required.

Results

A within-subjects ANOVA was performed with hearing aid type as the only factor. The three thresholds measured during a session (one per block) were averaged together to obtain an average threshold for that session. In the end, there was no significant effect of hearing aid type ($F(1,4) = 0.065$, $p = 0.812$), indicating that neither hearing aid conferred a temporal processing advantage. Group means and standard errors are plotted in Figure 4.18, and individual means and standard deviations are plotted in Figure 4.19.

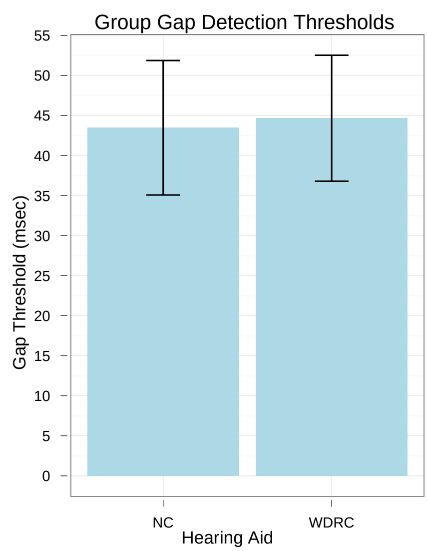

Figure 4.18: NCStudy1 gap detection - group thresholds. Error bars represent $\pm$ 1 SEM. There was no significant difference in performance between the hearing aid types ($F(1,4) = 0.065$, $p = 0.812$).

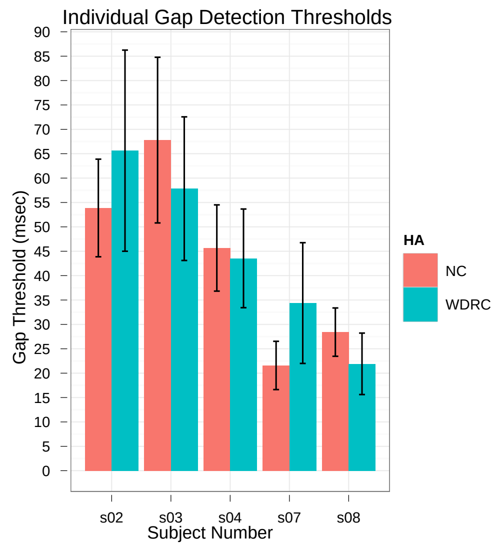

Figure 4.19: NCStudy1 gap detection - individual thresholds. Error bars represent ± 1 SD. There were some individuals who performed better with WDRC, while others did better with NC.

4.6 Sound Localization

4.6.1 Horizontal Localization

Stimuli

The stimuli used in this experiment were similar to those used in Van den Bogaert et al. (2006). One-third octave noise bands centred at 500 Hz (low-frequency) and 3000 Hz (high frequency) with a duration of 200 ms were created in Matlab. The first 50 ms and the last 50 ms of the stimuli were cosine windowed to provide a gradual rise and fall in amplitude. Both a low frequency and a high frequency condition were used to test the preservation of ITD and ILD cues by the hearing aids, respectively. Additionally, a broadband, telephone sample was included in a separate condition to test more realistic sound localization. The telephone stimulus was 1000 ms in duration, and contained peaks in its spectrum both in the low frequency range (less than 1500 Hz) and the high frequency range (greater than 1500 Hz), so it contained both ITD and ILD cues.

Procedure

This task took 25 minutes to complete. Both the experimenter and the participant sat in the sound attenuating chamber, with the participant seated 1.5 meters away from an array of 7 speakers suspended in a 90° arc (15° separation between speakers). Participants were tested at two different orientations (one facing the leftmost speaker, and the other facing the rightmost speaker, in order to test both left and right hemifields).

The stimuli were all presented at 65 dB SPL, measured at the head of the participant. Presentation order was completely random, and four trials of each stimulus were presented at every speaker in each condition, for a total of 168 trials per session (7 speakers * 4 trials * 3 stimulus types * 2 orientations). There were 6 blocks of trials (3 stimulus types * 2 orientations) and the presentation order of these blocks were fixed to provide a more controlled comparison of the NC and WDRC technologies. Adopting "left" and "right" as terminology for facing the leftmost and rightmost speakers respectively, the order of conditions in the experiment was phone-right, low-right, high-right, phone-left, low-left, high-left.

Following each stimulus presentation, the participant indicated with a laser pointer which loudspeaker they thought the sound came from by pointing at a speaker labelled from A-G in the speaker arc. The experimenter coded the subject's response, and wore sound-isolating headphones (ANSI S3.19 Workhorse) to prevent him/her from biasing the subject's response. Subjects were instructed to remain facing straight for the duration of the experiment, rested their chin on a chin rest while sounds were playing, and were encouraged at the outset of the experiment to disengage from the chin rest if they liked, only when making their response (and after cessation of the stimulus).

A practice session consisting of 3 trials (the low frequency stimulus) was delivered before starting the phone-right condition, presenting first "to the speaker directly in front of" the subject, then "to the speaker to the far left", and finally "to a speaker somewhere in the middle."

Results

A within-subjects ANOVA was performed, with three within-subjects factors, hearing aid (WDRC or NC), stimulus (low freq, high freq, or phone), and angle (0 to 90 ° in 15 ° increments). There were too few subjects to test the sphericity assumption of the ANOVA, so results should be interpreted with caution. The dependent measure was the error (|Stimulus° - Response°|). There was a main effect of stimulus ($F(2,8) = 13.71$, $p = 0.003$), such that the high frequency stimulus was much more difficult than the phone or low frequency stimuli. There was also a main effect of angle ($F(6,24) = 6.66$, $p < 0.001$), and an interaction of stimulus and angle ($F(12,48) = 2.11$, $p = 0.034$). Detailed statistical results may be seen in Table 4.5. Group means and standard errors are plotted, as a function of stimulus in Figure 4.20, as a function of angle in Figure 4.21, and combined in Figure 4.22. Figure 4.23 also shows the distribution of responses on each condition for both NC and WDRC.

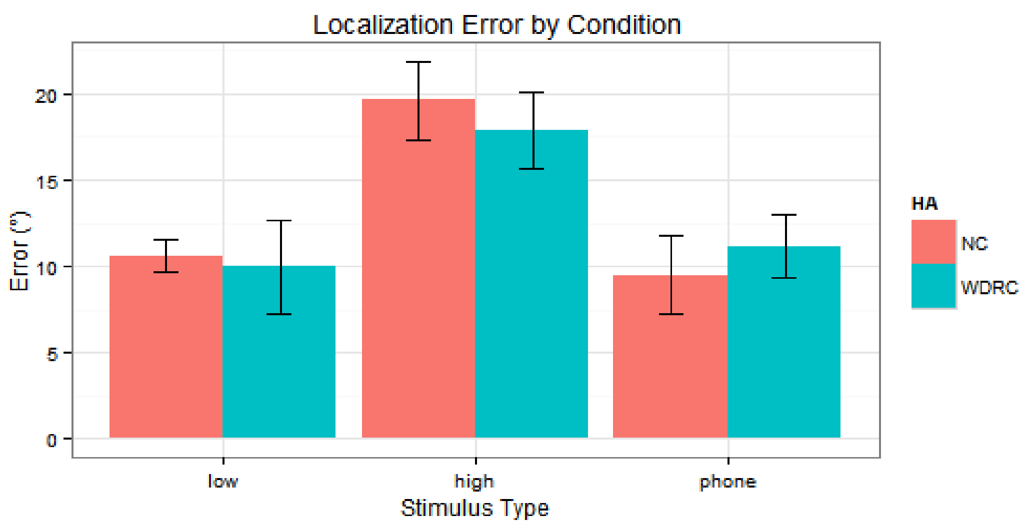

Figure 4.20: NCStudy1 sound localization results - stimulus type. Error bars represent ± 1 SEM. There was no significant difference between NC and WDRC (F(1,4) = 0.02, p = 0.903), but there was a main effect of stimulus (F(2,8) = 13.71, p = 0.003), such that the higher frequency sound was more difficult to localize.

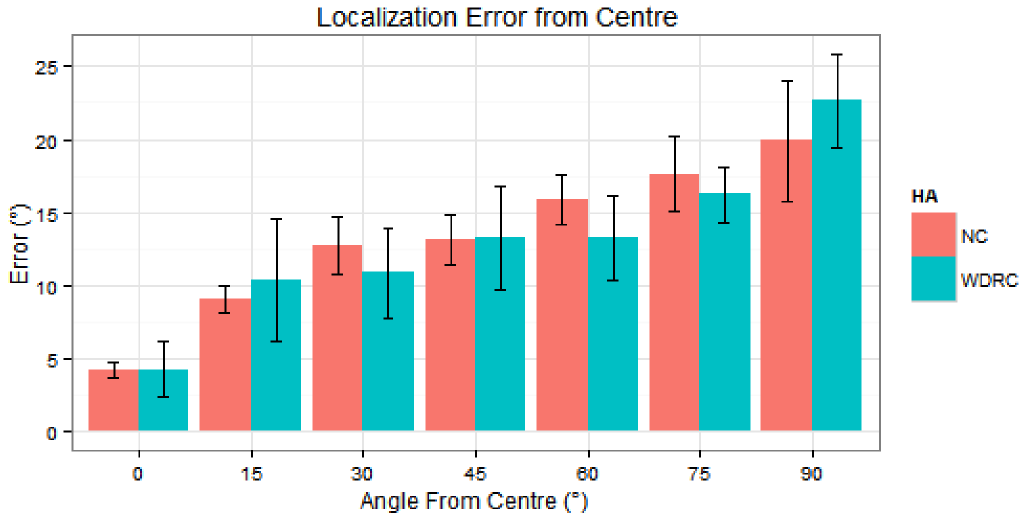

Figure 4.21: NCStudy1 sound localization results - angle. Error bars represent ± 1 SEM. There was a main effect of angle (F(6,24) = 6.66, $p < 0.001$). Generally speaking, the further the stimulus is presented from centre, the greater the error.

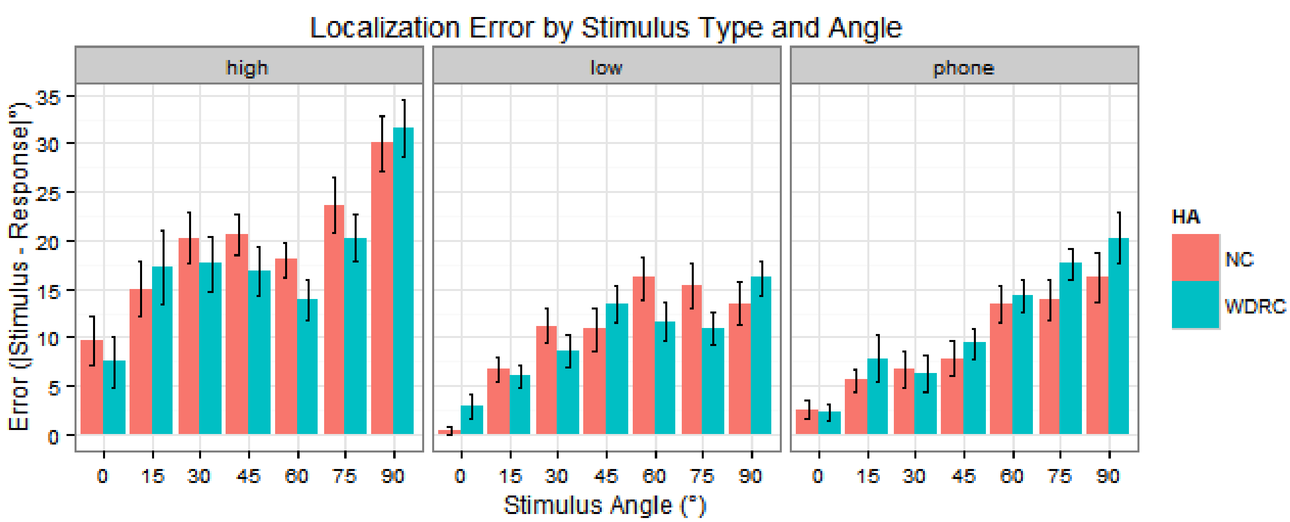

Figure 4.22: NCStudy1 sound localization results - stimulus type and angle. Error bars represent ± 1 SEM.

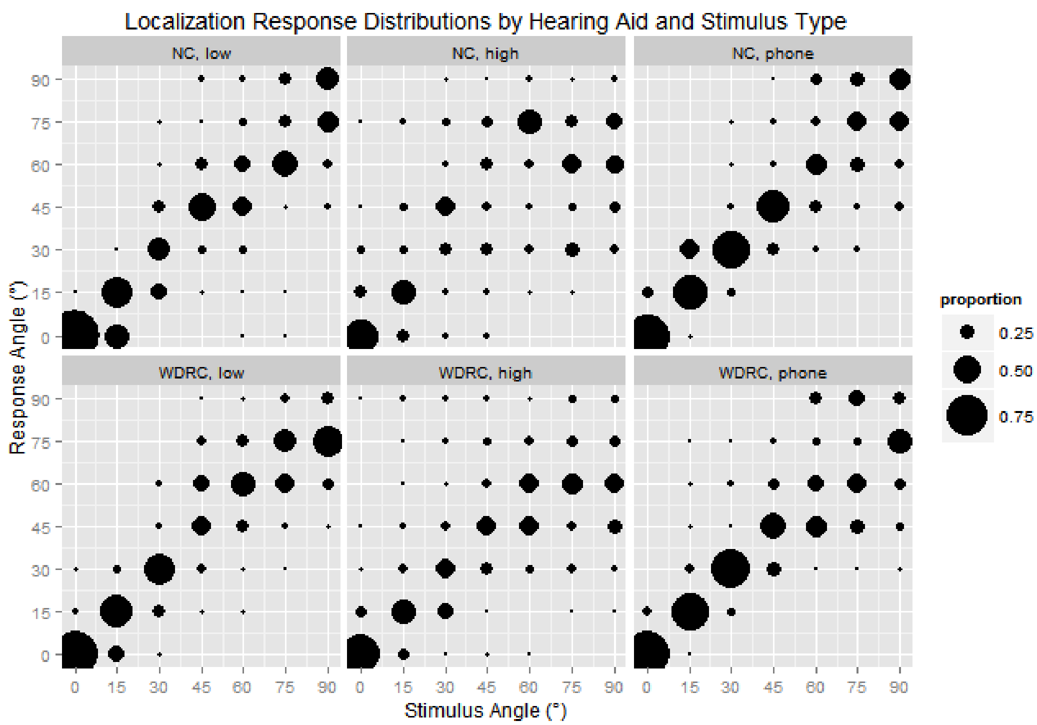

Figure 4.23: NCStudy1 sound localization response distributions. Circle area is proportional to number of responses at a particular stimulus, response configuration.

ANOVA								
Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	4	3	778	0.02	0.903		0.000
stim	2	8	3312	966	13.71	0.003	**	0.277
angle	6	24	5235	3144	6.66	0.000	***	0.377
HA:stim	2	8	107	588	0.73	0.513		0.012
HA:angle	6	24	158	645	0.98	0.463		0.018
stim:angle	12	48	783	1484	2.11	0.034	*	0.083
HA:stim:angle	12	48	156	1034	0.60	0.828		0.018

Table 4.5: Sound Localization Within-Subjects ANOVA

4.7 Subjective Measures

4.7.1 Abbreviated Profile of Hearing Aid Benefit (APHAB)

Procedure

To compare the NC and WDRC using this inventory, we administered the APHAB as the very last task during the second test session for both the NC and WDRC trial periods. Completion of this inventory took approximately 10 minutes.

Results

There was no easy way of looking at group results for the APHAB, because the APHAB software outputs an unaided and aided graph for each subject, as opposed to raw data. APHAB results are reported for NCStudy2, but they were too difficult to tabulate for NCStudy1.

4.7.2 Sound Quality

Procedure

The sound quality questionnaire was administered to assess dimensions of music quality on a 5 point scale (fullness, naturalness, pleasantness, clarity), and a final composite score was computed by averaging the scores on these scales together. The overall sound quality component of the questionnaire asked participants to rate how pleasant/unpleasant sound was for a variety of situations (a talker in quiet, car radio, etc.). The questionnaire is attached in the Appendix.

The questionnaire was administered during the second test session for each hearing aid trial period.

Results

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and questionnaire subtype (music or overall). The dependent measure was a composite score of sound quality, computed by averaging sound quality scores across each questionnaire subtype. The only significant effect was a main effect of questionnaire subtype ($F(1,4) = 15.42$, $p = 0.017$). Detailed statistical results may be seen in Table 4.6. Group means and standard errors are plotted in Figure 4.24.

4.7.3 Feedback Form

Results

Participants were given a questionnaire at the very end of the study, asking them to report which set of hearing aids they preferred for a variety of listening situations,

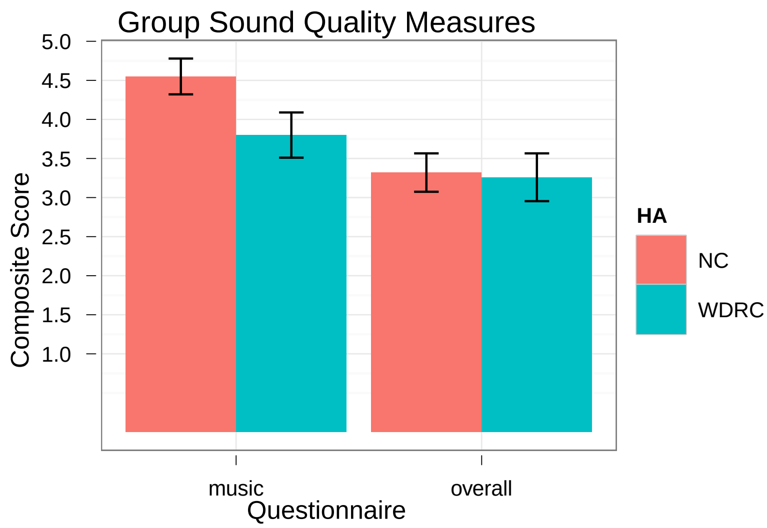

Figure 4.24: NCStudy1 sound quality group results. Error bars represent ± 1 SEM. There was no significant main effect of hearing aid ($F(1,4) = 1.16$, $p = 0.342$).

and to give an overall preference of hearing aid type. Table 4.6 presents the number of participants who preferred NC and WDRC for each listening situation.

Situation	NC Count	WDRC Count
Music	4	1
Speech in quiet	4	1
Speech in noise	3.5	1.5
Television	3	2
Telephone	4.5	0.5
Overall	3.5	1.5

Table 4.6: Listening Situation Preferences

4.8 Discussion

In this section, the results of the present experiment are synthesized as best as possible. However, there were several limitations in the study design, which limited our conclusions. First, the study limitations will be discussed, followed by a synopsis of results in each auditory domain, so that the reader may interpret the conclusions drawn from each auditory domain with the study limitations in mind.

4.8.1 Limitations

First, it is uncertain as to how much amplification was being delivered by the NC and WDRC hearing aids as a function of frequency, since real ear measurements were not performed. Although the fitting software offered by hearing aid manufacturers allows clinicians to see the amount of amplification given by hearing aids, these visualizations are not always completely accurate (Hawkins & Cook, 2003). Real ear measurements offer a method that clinicians can use to verify the amount of gain given by a hearing aid. In the real ear measurement procedure, a tiny probe microphone is placed in a participant's ear close to the tympanic membrane, and the sound level of a given stimulus (usually some sort of speech shaped noise) is measured with the hearing aid in the participant's ear, and without the hearing aid in the participant's ear. By calculating the difference between these two values, a measure of real-ear insertion gain is computed. Not knowing the gain provided by each hearing aid type is a problem, because if one hearing aid type was consistently delivering more gain than the other, it could have made the test stimuli more audible for one hearing aid type, potentially biasing performance on many of the psychoacoustic tasks.

Second, the NC and WDRC hearing aids differed in many respects, such as the age of the hearing instruments, the quantity and quality of adjustments that the hearing aids received, and the features programmed onto the hearing aids. The WDRC hearing aids were much older than the NC hearing aids, so it is possible that the NC hearing aids had newer processors and electrical components which could have influenced task performance. On the other hand, participants had had several adjustments performed on their WDRC hearing aids, so hearing aid fittings could have been more fine-tuned for the WDRC hearing aids. One other complication is that the specialist fine-tuning the NC aids was not the same specialist as those that participants worked with to fine-tune their own WDRC hearing aids. There was no way to control for specialist skill in the current study. One last major difference between the NC and WDRC hearing aids were the features that were programmed onto the hearing aids. The NC hearing aids were programmed with 2 programs: program 1 had feedback cancellation only, and program 2 had feedback cancellation and directional microphones to help with speech in noise in everyday environments. Most of the WDRC hearing aids had automatic programs, which automatically decide when to activate features such as noise reduction, and directionality. All of the hearing aids in the study had feedback cancellation systems, and most of the WDRC hearing aids had noise reduction enabled as well.

Third, there could be selection bias in the recruiting process for the study. It is conceivable that out of the whole population of eligible participants for the study (all hearing aid users with 2 functioning hearing aids, with symmetric sensorineural hearing loss), there may have been a tendency for those who were dissatisfied with their current WDRC hearing aids to sign up to participate in the study. Alternatively, the reverse could be true. Perhaps those who are satisfied with their current hearing aids would be more open to try new hearing aids, due to positive prior experiences with hearing aids.

Fourth, there could have been adaptation effects contaminating the data. There are data which suggest that it can take an extended period of time to adapt to a hearing aid fitting (Brooks, 1996). The adaptation period for the NC may not have been long enough, limiting the NC's effectiveness. Despite asking participants at the outset of the study to wear the hearing aids as much as possible during the course of the study, there was also no way of knowing how often participants were wearing the hearing aids on a daily basis. Certainly, participants would not have adapted as well to the NC hearing aids had they not worn the hearing aids consistently throughout the trial period. An added complication related to this, is that the WDRC hearing aids were always accessible to participants (we could not confiscate the participants' own hearing aids), whereas the NC hearing aids were only accessible during the portion of the study for which they were wearing the NC hearing aids.

Fifth, the hearing aids were programmed on old audiograms for 3 of the participants, and for 2 of the participants only the new audiograms could be used to program the hearing aids. Programming the hearing aids on the old audiograms could have potentially compromised the NC training algorithm, which expects correct audiogram data. However, using newer audiograms to program the NC for 2 participants could have biased the performance of these individuals in favor of the NC.

Sixth, the study was not double-blind, and therefore it is possible that subject responses were biased systematically for one hearing aid or the other. For example, it is conceivable that participants may have responded in a manner so as to please the experimenter, which is a well-known problem in psychological and clinical research, called demand characteristics.

4.8.2 Conclusions from NCStudy1

NCStudy1 was confounded with many variables that were outside of the experimenter's control, making any interpretation of the results quite difficult. Instead, a summary of results on each auditory domain is presented here. Results from NCStudy1 and NCStudy2 will be synthesized and compared in the following chapters.

Speech Intelligibility

There were fairly consistent differences between the NC and WDRC hearing aids in terms of individual phonemes correctly recognized. The NC was not as good at restoring particular consonants, such as /b/, /h/, and /r/. Vowel recognition was more equal between hearing aids; there was a tendency for /who'd/ to be restored better with the NC, while recognition of /hayed/ was superior with WDRC. These differences are unlikely to be due to any effect of noise reduction enabled in WDRC but not NC, since the differences were present in both the quiet and noise conditions.

It would be premature to read too much into these results, because of the many confounds outlined in the Limitations section, above. The real test is whether these results are replicated in NCStudy2. An important point is that despite consistent differences between the NC and WDRC hearing aids on the recognition of individual phonemes, one hearing aid did not outperform the other on a more realistic speech in noise test.

Music Perception

There was no difference in temporal perception between WDRC and the NC as determined by the gap detection task, and no difference in frequency resolution as determined by the mistuned harmonic task. There was, however, a trend for better intensity resolution of particular harmonics for the NC hearing aid.

Again, these results should be interpreted with caution. It would be overly simple to claim that the NC restores timbre perception due to it restoring a more normal neural signal at the level of the auditory periphery. To test this hypothesis, an analysis would be incomplete without computing neurograms for the saxophone stimulus, both for an impaired ear preprocessed by the NC and WDRC. If the result holds, there are other potential reasons for why the NC was superior for timbre perception. One potential reason is that the NC, due to its gain function, enhances frequency channels with high acoustic power, and attenuates frequency channels with low acoustic power. Given that the harmonics which were altered in the timbre perception task were the harmonics in the saxophone stimulus with the greatest acoustic power, to allow for the most attenuation, it is possible that each step in intensity change was greater for the NC, making the task easier from trial to trial. Another possible difference between the hearing aids which could account for performance is compression. If WDRC was applying a greater amount of compression than NC in a given frequency channel, then the step in intensity change would have been greater for the NC. One last possibility is that although participants did not complain of entrainment or feedback being present in any of the music perception tasks (reporting an additional tone, or a whistling), it is possible that feedback/entrainment could have contaminated results for particular subjects, affecting one set of hearing aids more than the other.

Sound Localization

There were no statistically significant differences between the NC and WDRC hearing aids for sound localization. However, the localization error by stimulus type and angle, as well as the response distributions by hearing aid and stimulus type, seems to show some trend of worse localization performance by the NC for high frequency sounds. Perhaps not enough gain was given to high frequency sounds for the NC. It is difficult to draw any conclusions here, given the many limitations of the present study, noted above.

Subjective Measures

There were no differences between the NC and WDRC on the sound quality measure. However, the feedback form at the end of the study seemed to indicate that participants preferred the NC overall. This feedback form was vulnerable to demand characteristics; the study was not double-blind, so participants could have reported a preference for the NC simply to please the experimenter.

Chapter 5

NC Experiment 2

5.1 Design

The purpose of the second experiment was to compare the NC to a generic WDRC hearing aid, while controlling for as many other factors as possible that could contribute to a person's preference of hearing aid. The hearing aids were equated in physical appearance, hardware (both used the GA3280 microprocessor platform), and any supplemental features, such as noise reduction and feedback cancellation. The implementation of the noise reduction system on the NC and WDRC hearing aids was identical, but there were slight differences in the implementation of the feedback canceller, though they used the same resources on the microprocessor.

Figure 5.1 shows a week by week schematic of the study design. The study was double-blind, such that both the experimenter and the participant were unaware of which set of hearing aids a participant wore at any given instant. At the very beginning of the study, participants were issued a structured diary. The structured diary was of a similar format to that used in Moore, Marriage, Alcántara, and Glasberg

(2005), and there were four forms in the diary to be filled out on particular dates. These forms asked participants to rate the loudness and clarity of target speech and the loudness of background noise for six different situations (e.g. car, having a meal with 2-3 people, etc.), as well as to report how many hours per day they wore the hearing aids. In addition to the structured part of the diary, there was also a freeform section, where participants were asked to take note of any unusual sounds, any instances of feedback, and whether any sounds were too loud or too quiet.

NC STUDY 2

		WEEK #											
	Starting HA	1	2	3	4	5	6	7	8	9	10	11	12
PLAN	NC	FIT		ADJUST		TEST	TEST/RTN	FIT		ADJUST		TEST	TEST/RTN
	WDRC	FIT		ADJUST		TEST	TEST/RTN	FIT		ADJUST		TEST	TEST/RTN
LEGEND		Wearing NC				RTN = return							
		Wearing WDRC											

Figure 5.1: NCStudy2 design. Half of the participants were fit with WDRC hearing aids at the outset of the study, while the other half were fit with NCs. Adjustments were given 2 weeks after the initial fit. Halfway through the study, participants returned their current set of hearing aids in exchange for the other set of hearing aids, which were fit to their loss. Two test sessions were conducted for each pair of hearing aids.

Roughly half of the participants (6) were randomly assigned to begin the study wearing the WDRC hearing aids, while the other half (5) were fit with NC hearing aids. After a two week period of wearing the hearing aids (deemed the first adaptation period), the participants visited McMaster, and were given an adjustment to their hearing aids by a trained hearing instrument specialist (HIS). After the adjustment, participants wore the hearing aids for another two weeks (the second adaptation period) before coming in for two, two-hour test sessions, spaced one week apart.

Adjustments were guided by observations that participants had recorded in their diaries, as well as from verbal reports given by the participant at the time of the adjustment. Following the second test session, participants who first wore NCs returned the set of NCs, and were lent out a set of WDRC hearing aids, while participants who first wore a set of WDRC hearing aids were given a set of NCs.

5.2 Participants

5.2.1 Demographics and Prior Experience

A total of 11 participants completed the study (7 male, 4 female), with an average age of 70.5 (SD = 9.6). None had prior experience wearing hearing aids, and all spoke English as their primary language. The hearing aids used for the study were receiver-in-the-canal (RIC) style.

5.2.2 Audiograms

Prior to their inclusion in the study, participants visited McMaster University for a hearing test; air conduction thresholds for both ears and bone conduction thresholds for the worse ear were measured. A trained student performed the measurements, following the British Society of Audiology recommended procedure for pure-tone audiometry (British Society of Audiology, 2004) as closely as possible. Tympanometry was also performed, to screen for normal middle ear function, as indicated by a Type A tympanogram according to the Jerger classification.

To qualify for the study, participants were required to have symmetrical sensorineural hearing loss, which was defined as: 1) no difference greater than 10 dB HL between ears on the calculated pure tone average (PTA; the average of 0.5, 1, and 2 kHz), 2) no difference greater than 30 dB HL at any audiometric frequency between the two ears, and 3) no difference greater than 15 dB HL between air conduction and bone conduction thresholds at any audiometric frequency. Figure 5.2 plots audiogram data for each participant.

5.2.3 Real Ear Measurements (REMs)

REMs provide an illustration of how much gain is being delivered at the ear drum by the hearing instrument, in response to a particular sound sample. In order to get an idea of the amount of gain being delivered by the NC and WDRC hearing aids, to help interpret the results of the experiment more accurately, REMs were collected at the end of the study but only for some participants, due to time and resource constraints. Typically, REMs are obtained on the person who is fit with hearing aids, in order to verify the amount of gain being delivered by the hearing aids. We did not have the resources at McMaster to perform these measurements while participants were enlisted in the study, so the measurements were instead performed on the author of this thesis while he wore hearing aids set to the same loss configuration and adjustments that study participants received (using the save files of participants on the fitting software). A qualified HIS administered the REM procedure at HearAtLast in Burlington, Ontario, using a Medrx Avant system. Figures 5.3 and 5.4 show real-ear insertion gain (REIG) for two subjects in the study, for WDRC (green curve) and for the NC (blue curve), for the left ear only. For these subjects, gain was fairly equal between the ears for both the NC and WDRC (right ear not shown). Another subject's REMs for the left ear are shown in Figure 5.5, with WDRC (blue curve) and the NC (green curve). For this subject, asymmetric gain was given between the left

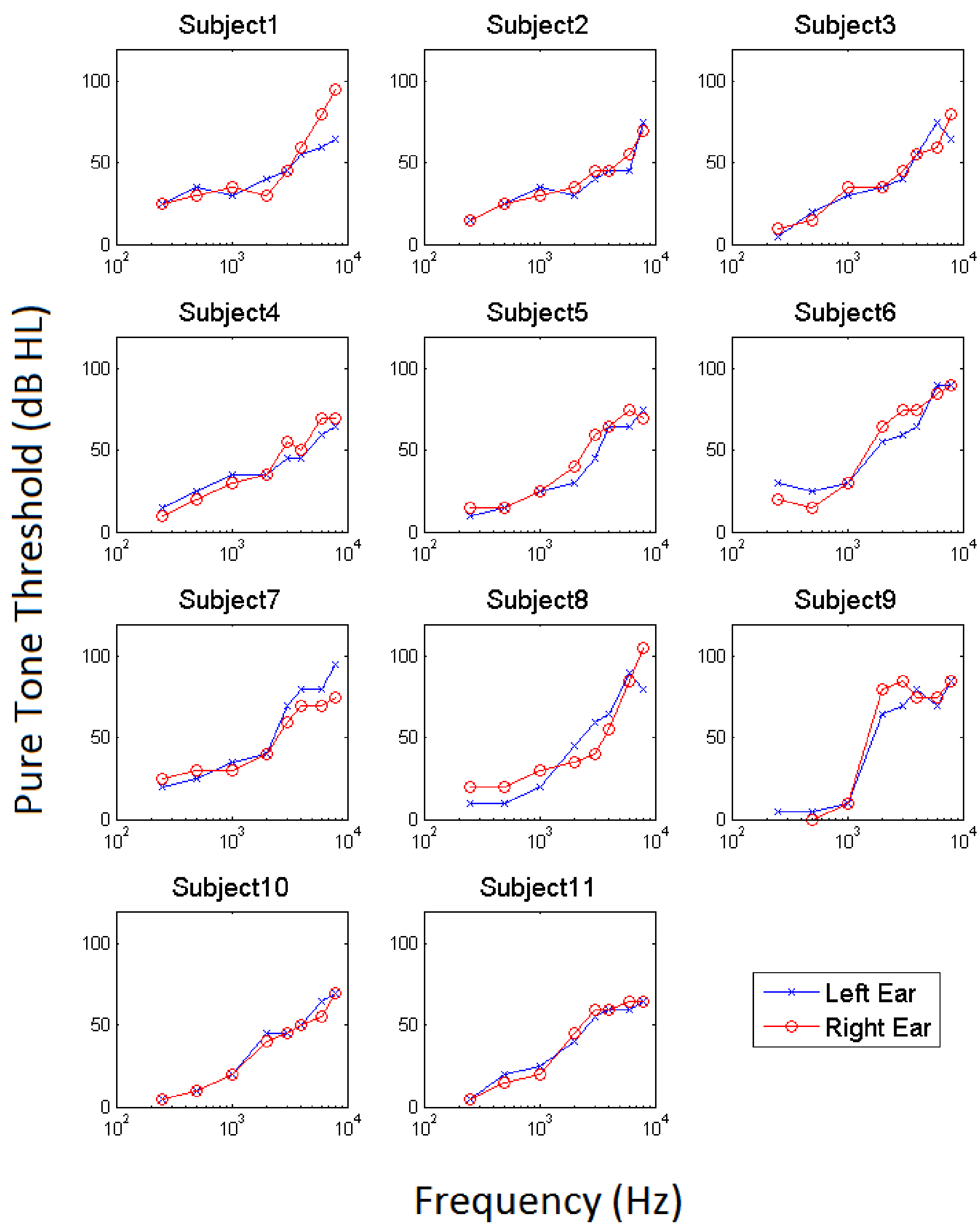

Figure 5.2: Audiograms for all subjects. Right ear is plotted in red o's, left ear in blue x's. All subjects had mild or moderate symmetrical sensorineural hearing losses.

and right ears for the NC, but not WDRC. REMs were obtained using ICRA noise as the sound sample, presented at 65 dB SPL. The spectrum of the ICRA noise is plotted in Figure 5.6.

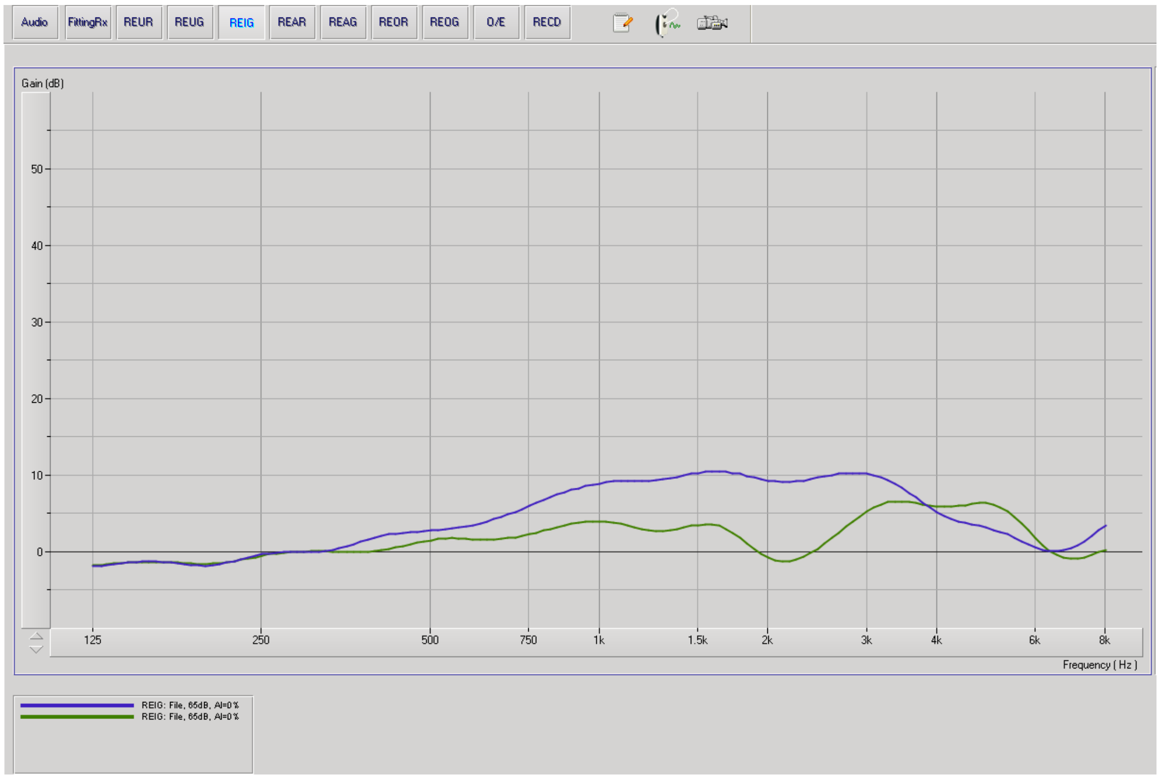

Figure 5.3: Insertion gain for subject 04. Only the left ear is plotted. Frequency is plotted on the x-axis, and insertion gain on the y-axis. NC insertion gain is plotted in blue, and WDRC insertion gain is plotted in green. More gain was being delivered overall with the NC.

As can be seen, overall, the NC was providing significantly more gain than the WDRC hearing aids, at least for the ICRA stimulus. In particular, the NC provided much more low-mid frequency gain than the WDRC hearing aids. Due to the nature of the NC, which adjusts gain dynamically in each frequency channel based on the amount of power, not only in its own frequency channel (WDRC does this), but other

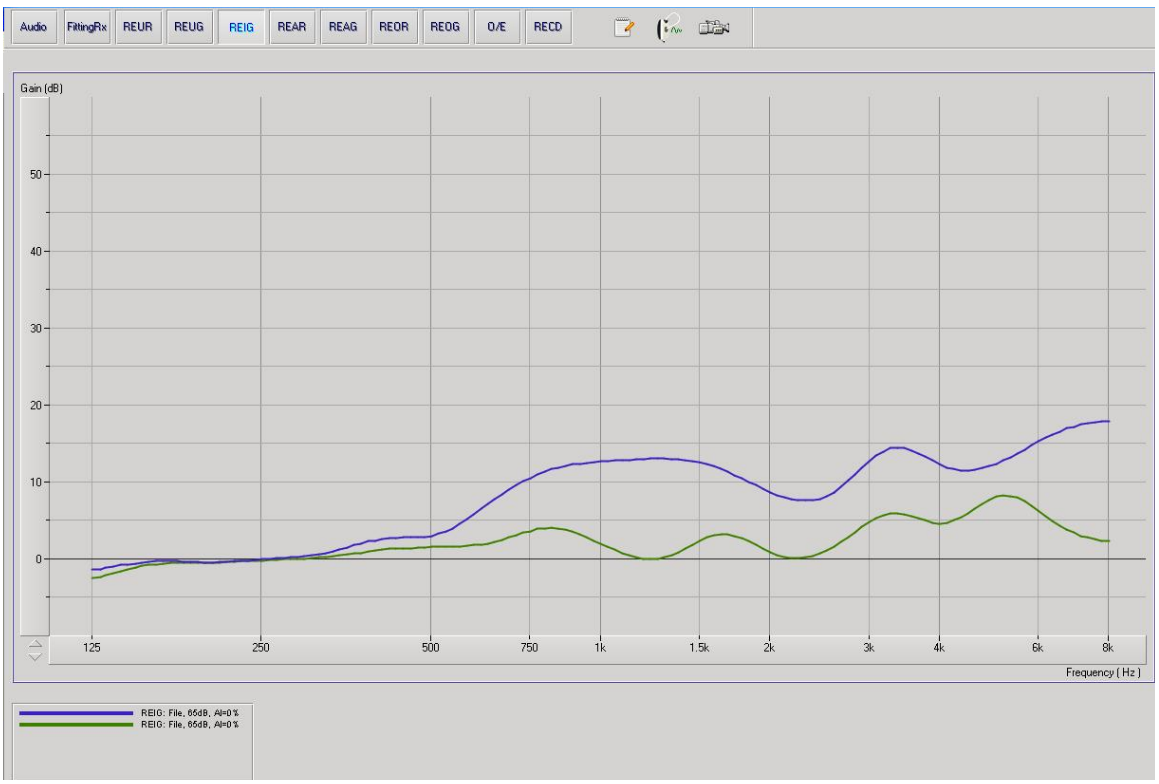

Figure 5.4: Insertion gain for subject 10. Only the left ear is plotted. Frequency is plotted on the x-axis, and insertion gain on the y-axis. NC insertion gain is plotted in blue, and WDRC insertion gain is plotted in green. More gain was being delivered overall with the NC.

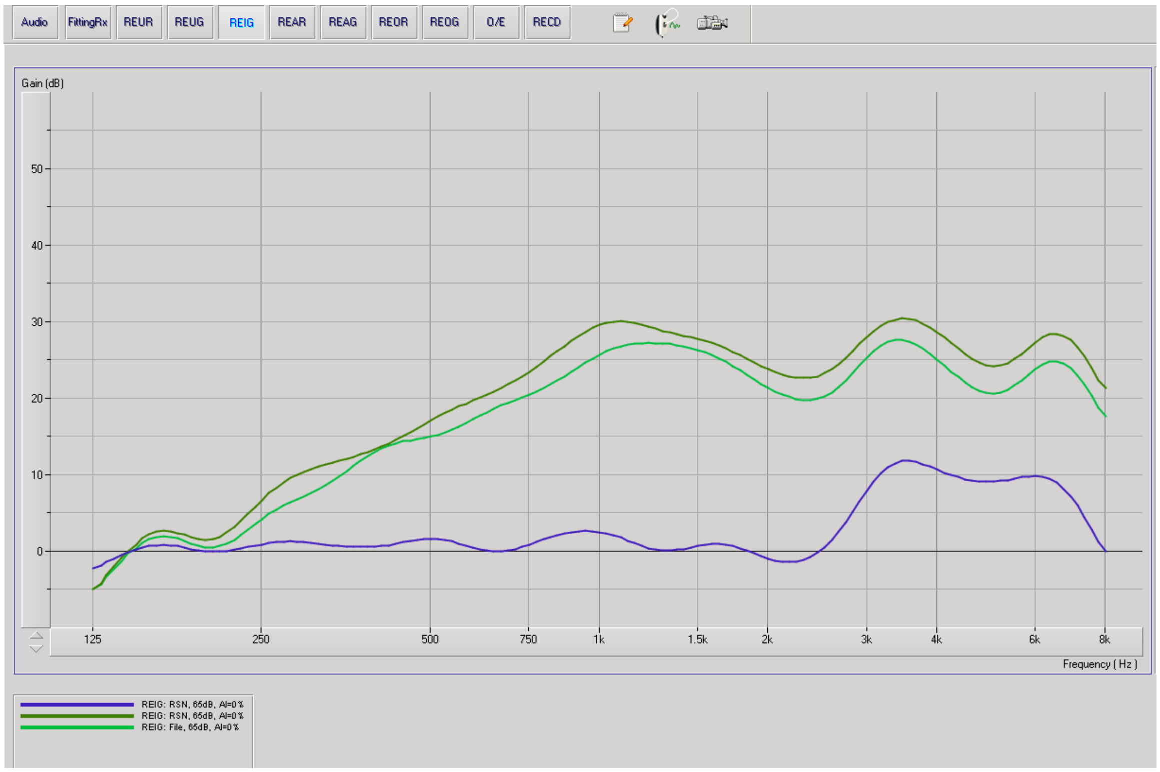

Figure 5.5: Insertion gain for subject 03. Only the left ear is plotted. Frequency is plotted on the x-axis, and insertion gain on the y-axis. NC insertion gain is plotted in green (one green curve for ICRA noise, another for speech-shaped noise), and WDRC insertion gain is plotted in blue. Much more gain was being delivered overall with the NC, and there was asymmetric gain between the ears for the NC (right ear is not shown).

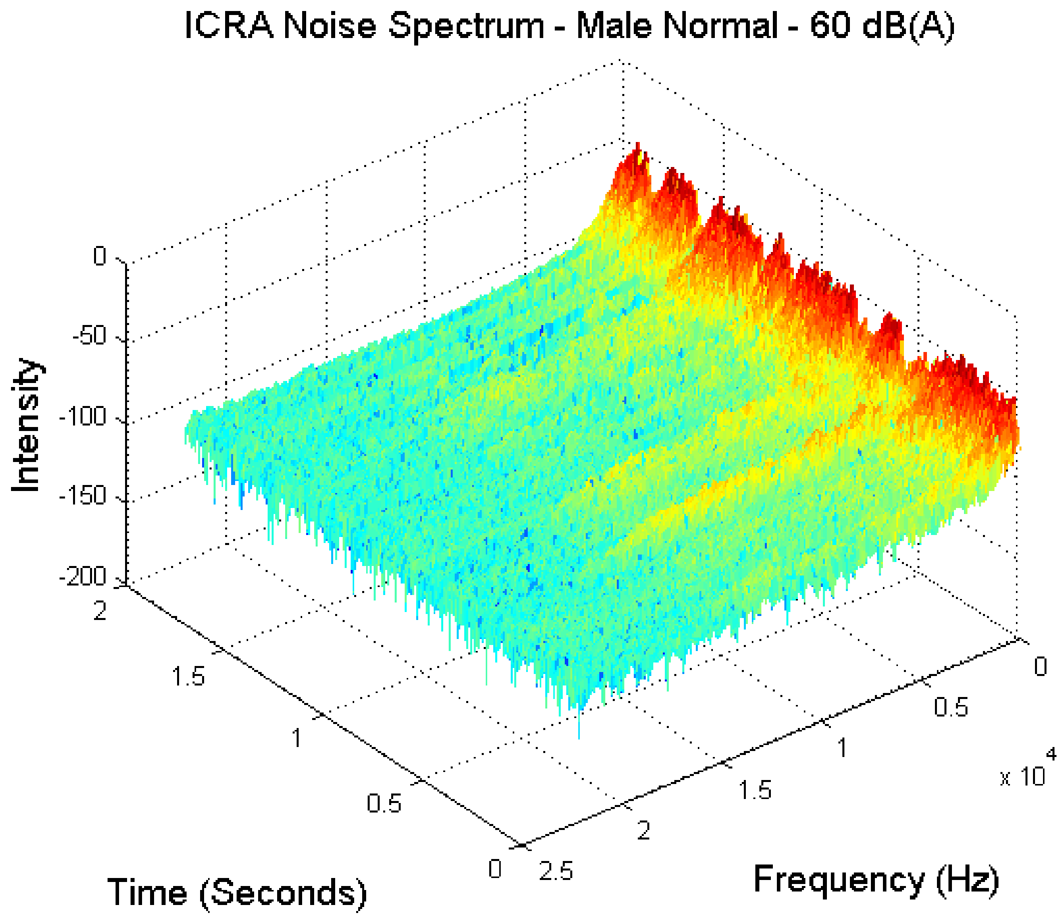

Figure 5.6: ICRA noise spectrum - 1 male speaker 60 dB(A). Similar to other speech shaped noise, most of the acoustic energy for this noise is contained in the lower frequencies, and the noise has speech-like temporal variations as well.

channels as well (WDRC does not do this), the real ear measurements taken here may be misleading. The ICRA stimulus is shaped to have similar spectrotemporal characteristics of speech, but its short term spectrum does not exactly resemble that of speech. Thus, a more informative approach to validating the gain of the NC hearing aids might be to present real speech. However, the real ear measurements taken for these three participants' hearing aid fittings does provide evidence that the NC may be providing too much gain in the low frequencies for some non-speech stimuli and some fittings. In noise, which is often lower frequency, the result may be that the noise is overamplified, masking cues in the higher frequencies which are very important for speech intelligibility. The evidence is consistent with freeform feedback we received from participants throughout the course of the study, who reported that noise seems to be overamplified with the NC.

5.3 Procedures Modified from NCStudy1

NCStudy2 used exactly the same procedures and room setup as specified in NC-Study1, with some minor changes. This section summarizes all of these changes. If no changes are reported for a given task, it means its procedure was not modified.

5.3.1 Hearing In Noise Test (HINT)

The length of the HINT task was increased, in order to improve the accuracy of SRT measurements. Instead of presenting 2 lists of 10 sentences for each hearing aid type, 4 lists of 10 sentences were presented. 2 sets of lists were created (lists 2-5, and lists 6-9), and were used to estimate SRTs. The use of each set of lists was counterbalanced across hearing aid type, and counterbalanced for the order in which the sets of lists was presented across participants.

5.3.2 Word Recognition

A word recognition task was added, with the purpose of gauging unaided and aided speech understanding in quiet. Such an objective measure is used in the audiology profession in order to assess whether the hearing aids provide any real speech intelligibility benefit in quiet.

Two lists were used (Auditec NU6 word lists), and they were always presented in the same order (list 1 for unaided, list 2 for aided). List 1 had 49 words, and list 2 had 50 words. The experimenter sat in the testing room with the touchscreen on his/her lap, and initiated each trial by pressing a button on the touchscreen, while the subject was asked to face the center speaker. On each trial, a word was played from the loudspeaker directly in front of the participant (same room setup as the HINT), and the participant was tasked with repeating the word that they heard. The experimenter (who had no hearing loss) coded the responses of participants as correct, or incorrect. The words were calibrated to an overall level of 65 dB SPL.

5.3.3 hVd

Due to ceiling effects on this task for some participants in NCStudy1, the hVd task was made more difficult for NCStudy2. In NCStudy1, there was a quiet and a noise (+5 SNR) condition. In NCStudy2, the quiet condition was modified to include noise (+7 SNR, noise level of 60.5 dB(A)), and the noise condition was made more difficult (+2 SNR, noise level of 65.5 dB(A)). The rest of the procedure was identical to that used in NCStudy1.

5.3.4 APHAB

In NCStudy1, participants filled out Form A of the APHAB electronically, whereas participants in NCStudy2 filled out a paper version. The software only allowed print-outs of graphs, as opposed to access to the raw data, and thus the change was made to paper.

The APHAB was administered 3 times in NCStudy2. The first administration took place on the day participants were fit with their first set of hearing aids, and they were asked to fill out only the unaided portion of Form A immediately before they were fit with hearing aids. The second occasion where participants filled out the APHAB was during the second test session for the first set of hearing aids, where participants were asked to fill out only the aided portion of Form A. The last administration took place during the second test session for the second set of hearing aids, and again, participants were asked to only fill out the aided portion of Form A.

5.4 Results

Analyses for all experiments were performed in R (R Core Team, 2013). Where appropriate, the Huynh-Feldt correction for deviations from sphericity ($\tilde{\epsilon}$) was used to compute p values from F tests on within-subjects factors (Maxwell & Delaney, 2004). In data tables, degrees of freedom and sums of squares for both the numerator (effect of interest) and denominator (error) are reported, along with F and p values, and a measure of effect size, generalized eta squared (abbreviated *ges*).

5.4.1 Speech Intelligibility Results

Hearing In Noise Test (HINT)

Speech reception threshold (SRT) estimates for NCStudy2 were computed slightly differently than those for NCStudy1, due to the increased number of sentences presented for NCStudy2. Instead of averaging the levels of the last 16 sentences, the levels of the last 36 sentences were averaged to estimate SRTs. A within-subjects ANOVA with a 2-level factor (hearing aid) was performed, with SRT as the dependent measure. There was no significant effect of hearing aid ($F(1,10) = 1.71$, $p = 0.220$). Group means and standard errors are plotted in Figure 5.7, and individual SRT means and standard deviations are plotted in Figure 5.8. Individual differences in speech intelligibility in noise were far greater than any difference between the two hearing aids.

Word Recognition

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and condition (unaided or aided). The dependent measure was the percentage of words correctly repeated. The only significant effect was condition ($F(1,10) = 38.53$, $p < 0.001$), such that aided performance was superior to unaided performance. Thus, the use of WDRC or NC hearing aids improved speech intelligibility in quiet. Detailed statistical results may be seen in Table 5.1. Group means and standard errors are plotted in Figure 5.9, and individual means and standard deviations are plotted in Figure 5.10.

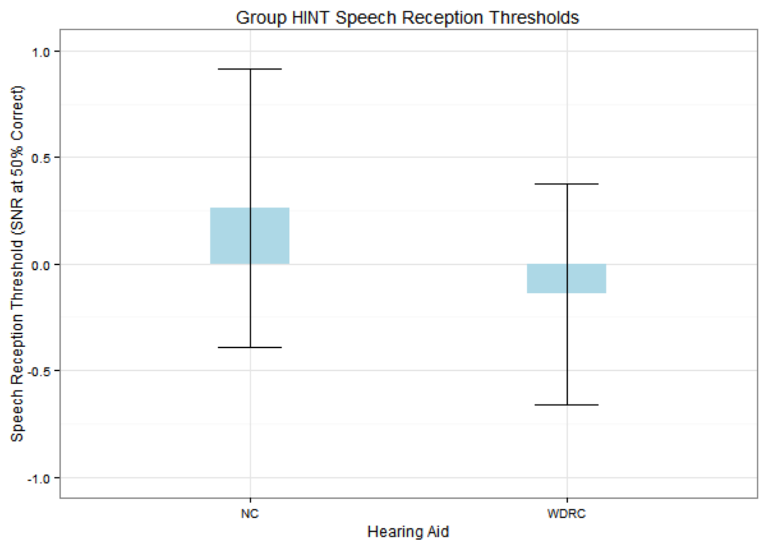

Figure 5.7: NCStudy2 group results on the HINT. Error bars represent ± 1 SEM. Speech reception thresholds were calculated by subtracting the noise level from the average presentation level of the last 36 sentences presented. There was no significant difference in performance between the NC and WDRC (F(1,10) = 1.71, p = 0.220).

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	26.7	177	1.51	0.248		0.005
cond	1	10	2480	644	38.53	< 0.001	***	0.298
HA:cond	1	10	0.7	195	0.04	0.849		0.000

Table 5.1: Word Recognition Within-Subjects ANOVA

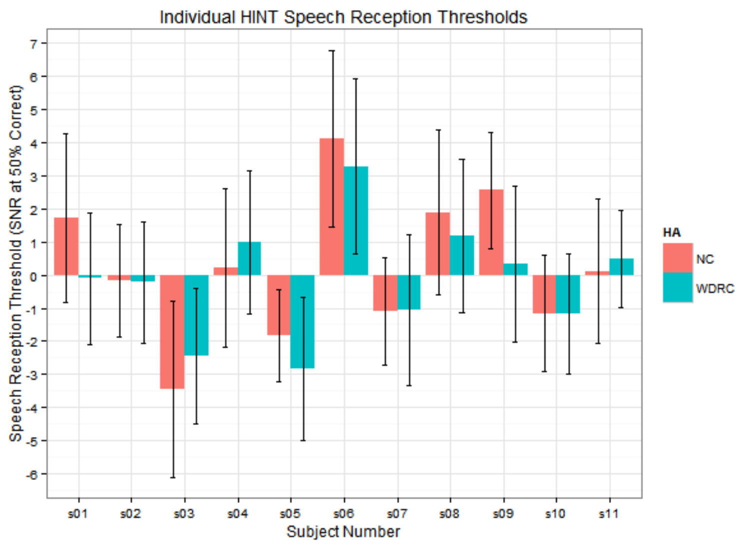

Figure 5.8: NCStudy2 individual results on the HINT. Error bars represent ± 1 SD. Speech reception thresholds were calculated by subtracting the noise level from the average presentation level of the last 36 sentences presented. Most participants performed approximately equally between NC and WDRC.

Figure 5.9: NCStudy2 group word recognition scores. Error bars represent ± 1 SEM. There was no significant difference in performance between NC and WDRC (F(1,10) = 1.51, p = 0.248), with each hearing aid conferring approximately a 15% increase in speech intelligibility performance in quiet.

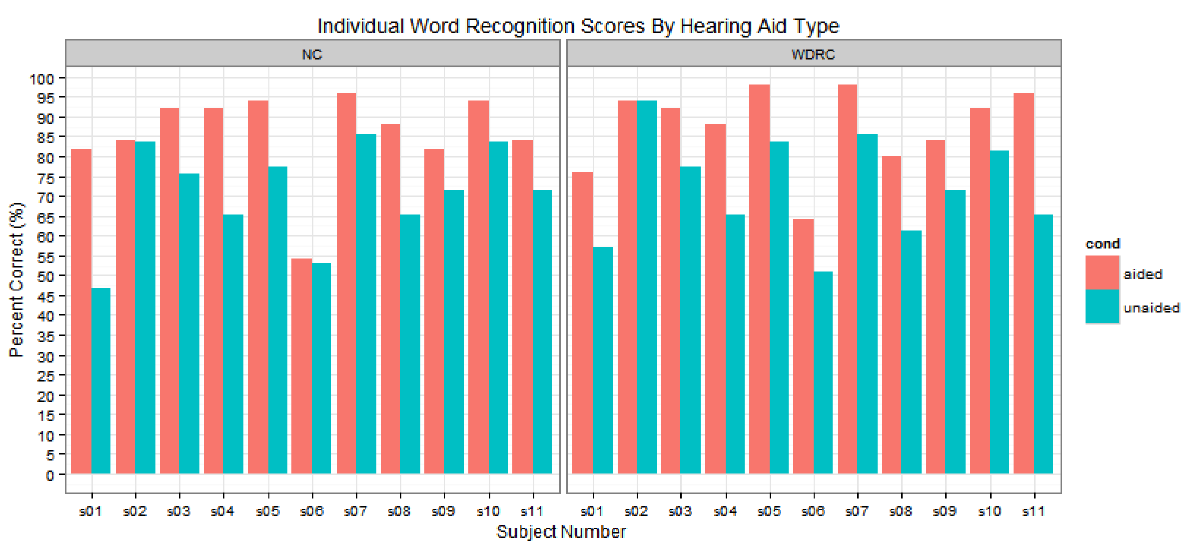

Figure 5.10: NCStudy2 individual word recognition scores. Most subjects received roughly the same amount of benefit from NC and WDRC.

Consonant Vowel Consonant (CVC)

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and condition (+7 SNR or +2 SNR). The dependent measure was the percentage of hVd tokens correctly identified. The only significant effect was condition ($F(1,10) = 26.87$, $p < 0.001$), such that the +2 SNR condition was more difficult than the +7 SNR condition. Detailed statistical results may be seen in Table 5.2. Group means and standard errors are plotted in Figure 5.11, and individual means and standard deviations are plotted in Figure 5.12.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	11.7	219.5	0.54	0.481		0.006
cond	1	10	523.8	195.0	26.87	< 0.001	***	0.199
HA:cond	1	10	23.0	88.4	2.60	0.138		0.011

Table 5.2: hVd Within-Subjects ANOVA

Although there were no significant differences between hearing aids on the overall percentage of hVd tokens correctly identified, there were minor consistent differences between the hearing aids on individual phonemes. Figure 5.13 plots a confusion matrix both for NC and WDRC averaged over both noise conditions, to illustrate the difficulty of individual phonemes in the experiment. Figure 5.14 plots the difference between the NC and WDRC confusion matrices for both the +7 SNR and +2 SNR conditions.

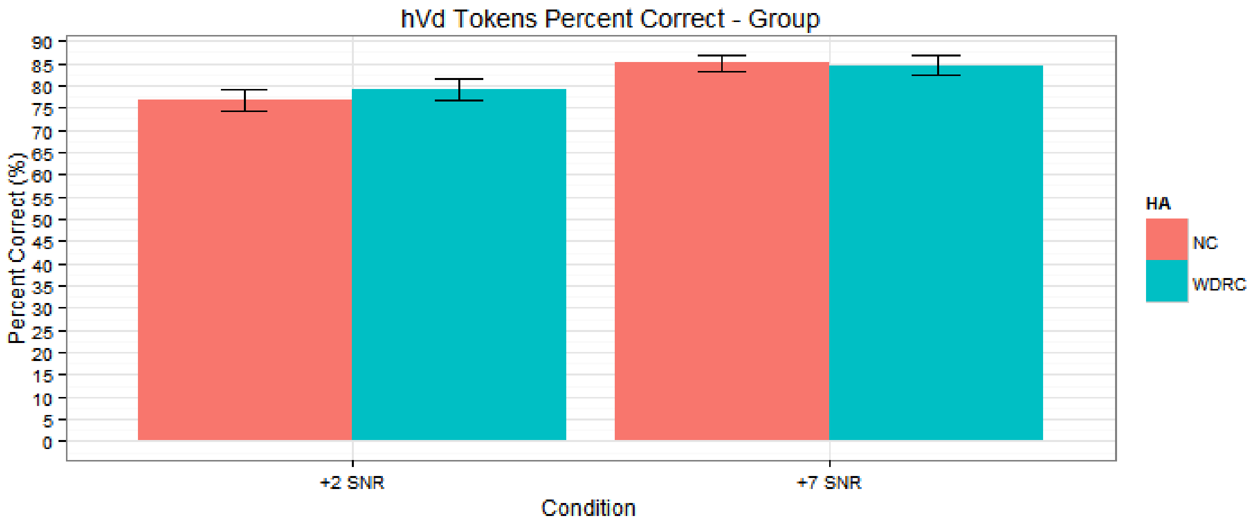

Figure 5.11: NCStudy2 hVd token percent correct - group. Error bars represent ± 1 SEM. There was no significant difference in performance between NC and WDRC on the percentage of hVd tokens correctly identified ($F(1,10) = 0.54$, $p = 0.481$).

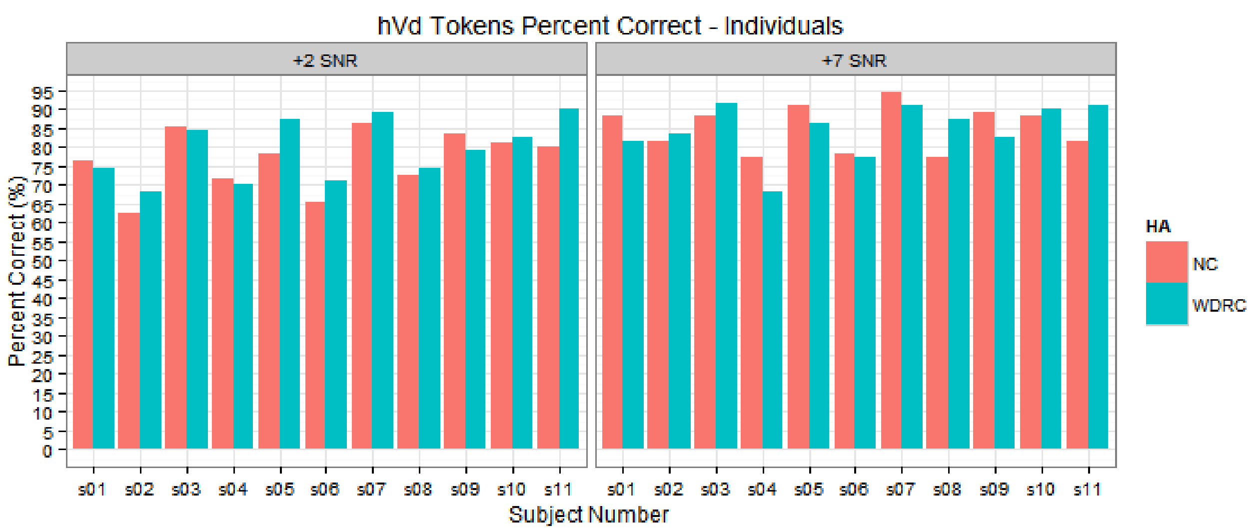

Figure 5.12: NCStudy2 hVd token percent correct - individuals. Some participants correctly identified more hVd tokens with WDRC, while others performed better with NC.

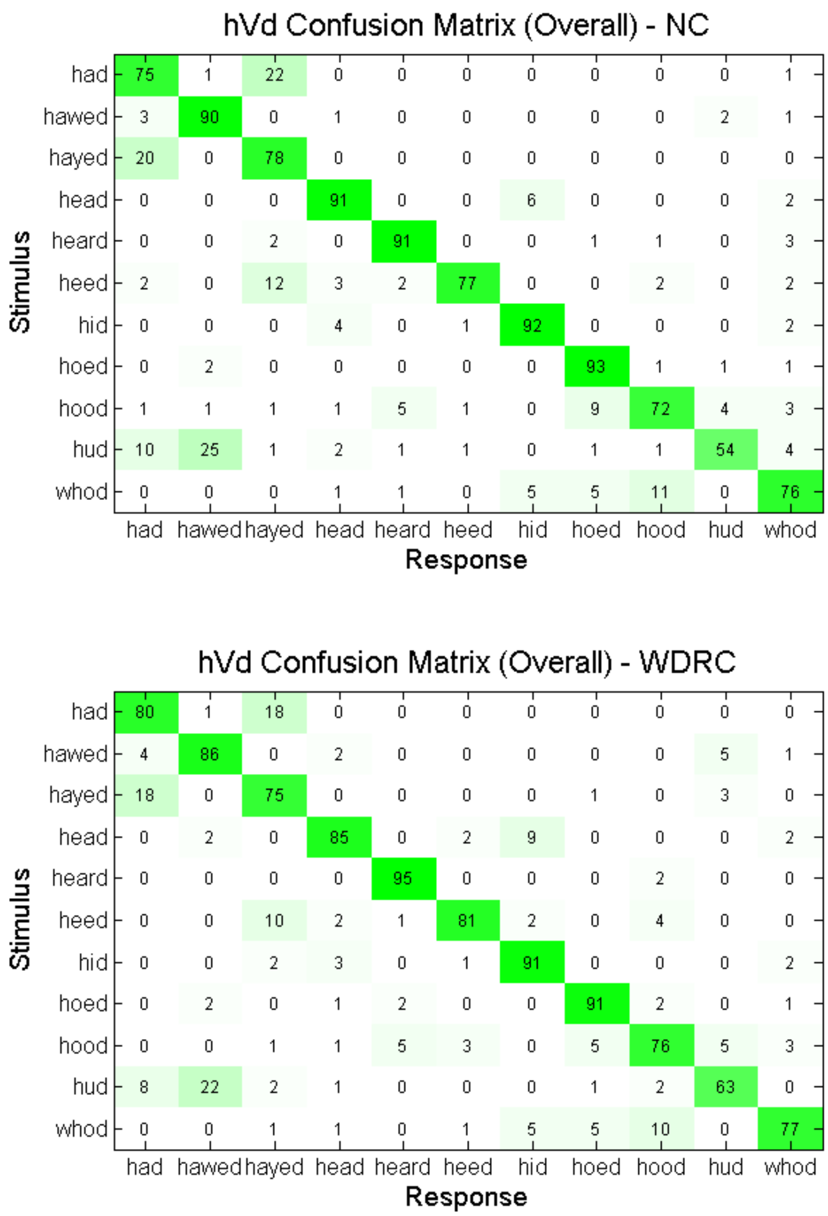

Figure 5.13: hVd confusion matrices. Data shown are for the noise and quiet conditions amalgamated together. *Top*: hVd confusion matrix for NC. *Bottom*: hVd confusion matrix for WDRC. For many of the syllables, performance is at near ceiling. Few differences exist between NC and WDRC; the syllables where WDRC and NC differ the most are /head/, and /hud/.

Vowel Consonant Vowel (VCV)

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and condition (quiet or noise). The dependent measure was the percentage of VCV tokens correctly identified. There were two significant main effects; a main effect of condition ($F(1,10) = 477.44$, $p < 0.001$), such that the noise condition was much more difficult than the quiet condition, and hearing aid ($F(1,10) = 10.55$, $p = 0.009$), such that WDRC outperformed the NC. Detailed statistical results may be seen in Table 5.3. Group means and standard errors are plotted in Figure 5.15, and individual means and standard deviations are plotted in Figure 5.16.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	289	274	10.55	0.009	**	0.088
cond	1	10	13800	289	477.44	< 0.001	***	0.821
HA:cond	1	10	7	137	0.51	0.493		0.002

Table 5.3: VCV Within-Subjects ANOVA

There were consistent differences between the hearing aids on individual phonemes. Figure 5.17 plots a confusion matrix both for NC and WDRC averaged over both conditions, to illustrate the difficulty of individual phonemes in the experiment. Figure 5.18 plots the difference between the NC and WDRC confusion matrices for both the quiet and noise conditions.

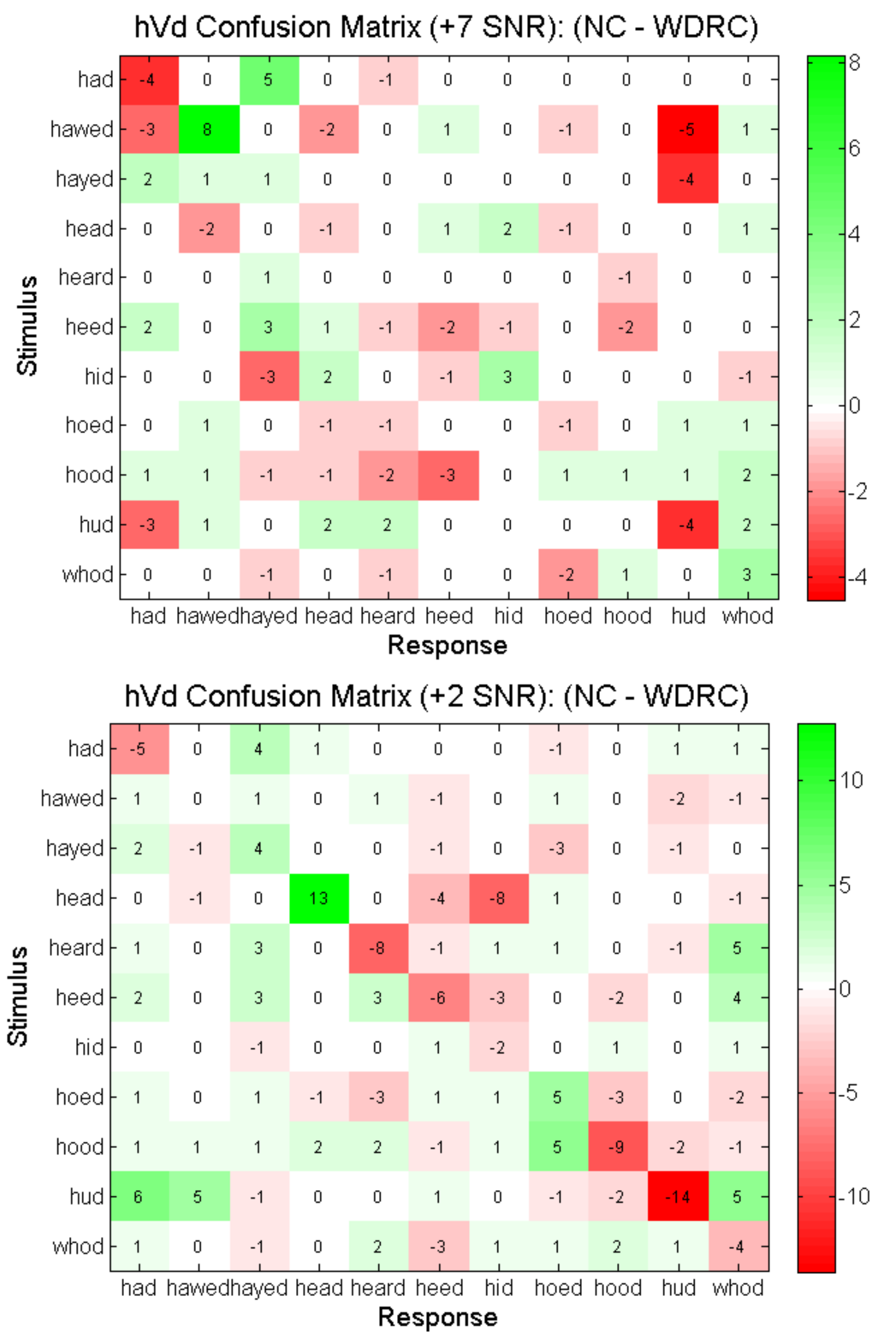

Figure 5.14: NCStudy2 difference between hVd confusion matrices (NC - WDRC). *Top:* Difference between hVd confusion matrices for the +7 SNR condition. *Bottom:* Difference between hVd confusion matrices for the +2 SNR condition. Each tile represents NC % response - WDRC % response. Differences between NC and WDRC are accentuated in the noisier condition; in noise, NC restores /head/ better than WDRC, while WDRC restores /hud/ better.

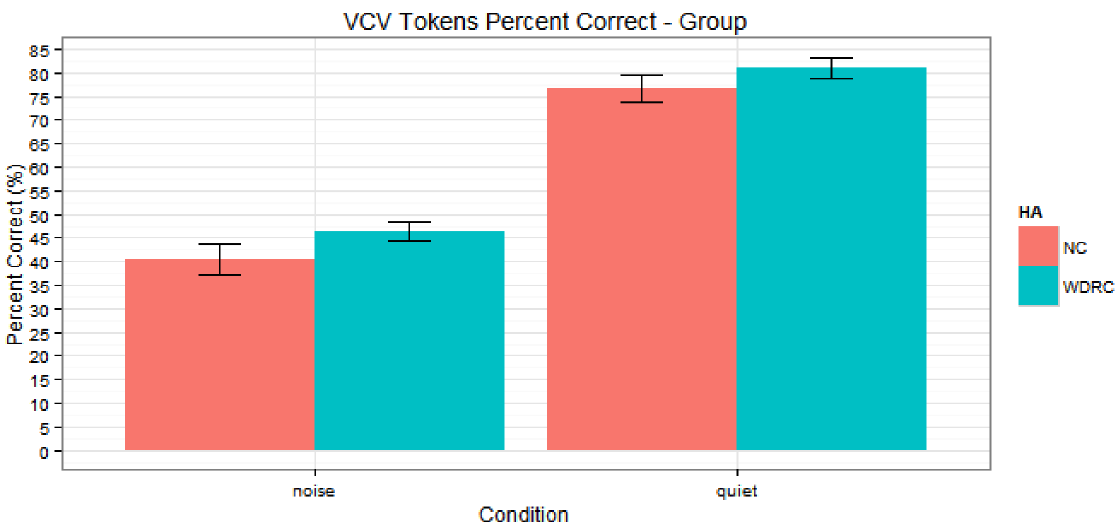

Figure 5.15: NCStudy2 VCV token percent correct - group. Error bars represent ± 1 SEM. There was a significant difference in performance between NC and WDRC on the percentage of VCV tokens correctly identified (F(1,10) = 10.55, p = 0.009).

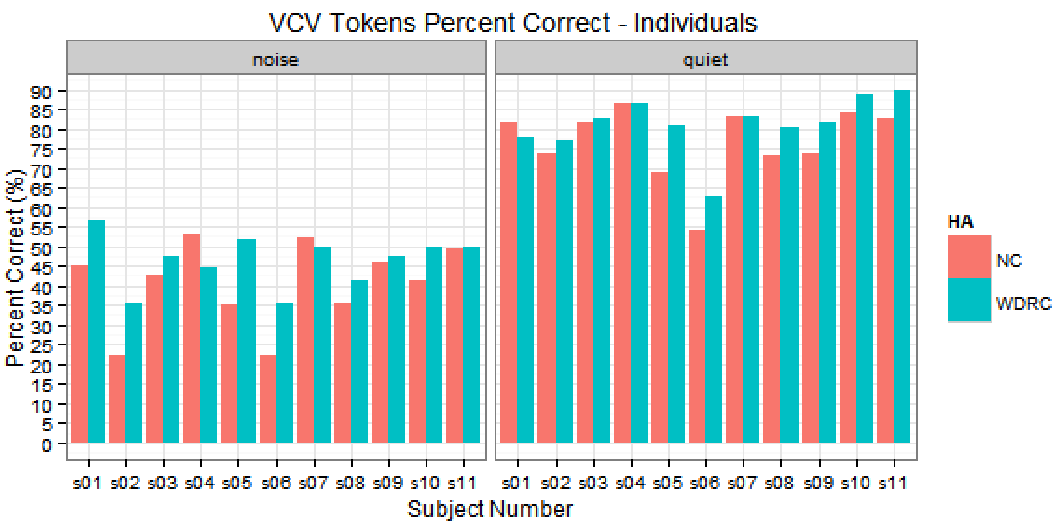

Figure 5.16: NCStudy2 VCV token percent correct - individuals. The main effect of hearing aid was not simply driven by a few participants performing better with WDRC hearing aids. There were only 2 participants out of 11 who identified more tokens correctly with the NC in noise, and only 1 out of 11 in quiet.

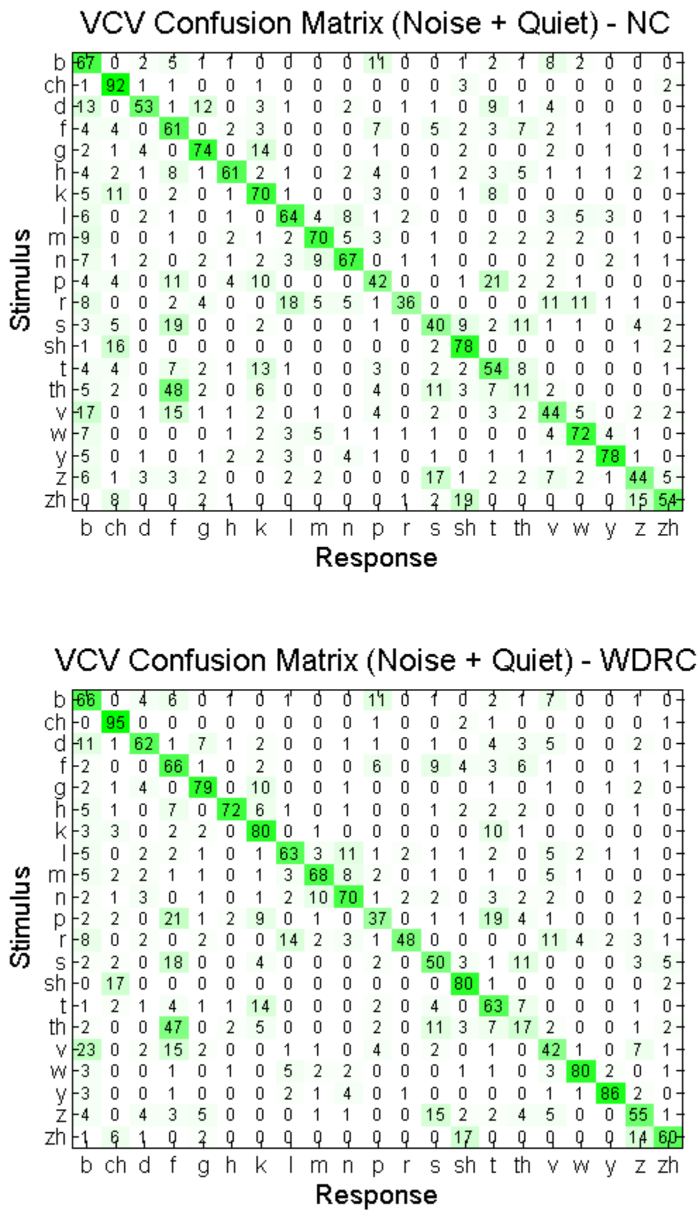

Figure 5.17: NCStudy2 VCV confusion matrices. Data shown are for the noise and quiet conditions amalgamated together. *Top*: VCV confusion matrix for NC. *Bottom*: VCV confusion matrix for WDRC.

5.4.2 Music Perception Results

Mistuned Harmonic

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and tone (200 Hz or 600 Hz). The dependent measure was the mistuned harmonic threshold. The only significant effect was a main effect of tone ($F(1,10) = 5.00$, $p = 0.049$), such that the higher complex tone (600 Hz) was more difficult than the lower complex tone (200 Hz). Detailed statistical results may be seen in Table 5.4. Group means and standard errors are plotted in Figure 5.19, and individual means and standard deviations are plotted in Figure 5.20.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	21	139	1.53	0.245		0.019
tone	1	10	229	459	5.00	0.049	*	0.170
HA:tone	1	10	1	165	0.03	0.859		0.000

Table 5.4: Mistuned Harmonic Within-Subjects Factorial ANOVA

Timbre Perception

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and harmonic (fundamental or second harmonic). The dependent measure was the harmonic intensity threshold. The only significant effect was a main effect of harmonic ($F(1,10) = 47.34$, $p < 0.001$), such that participants were more sensitive to changes in intensity of the fundamental than the second harmonic. Detailed statistical results may be seen in Table 5.5. Group means and standard errors are plotted in Figure 5.21, and individual means and standard deviations are plotted in Figure 5.22.

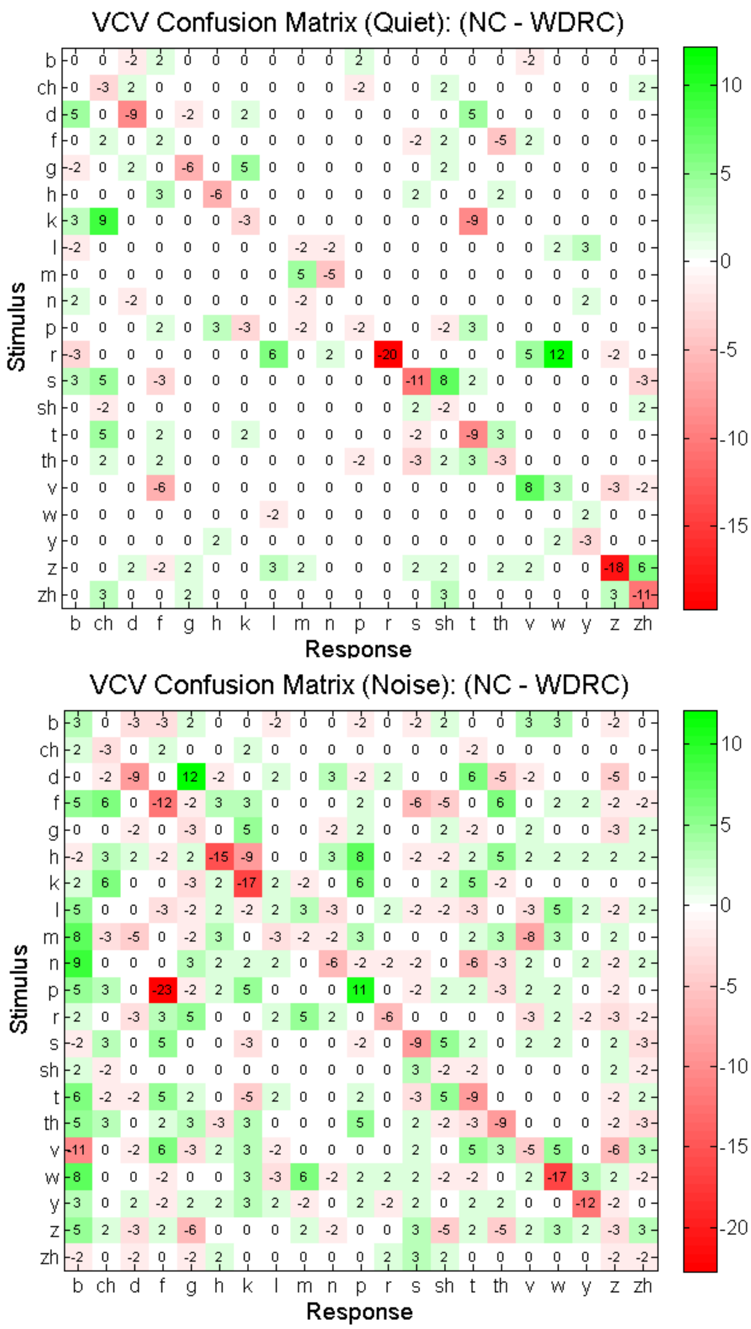

Figure 5.18: NCStudy2 difference between VCV confusion matrices (NC - WDRC). *Top:* Difference between VCV confusion matrices for the quiet condition. *Bottom:* Difference between VCV confusion matrices for the noise condition. Most of the consonants were restored better with WDRC.

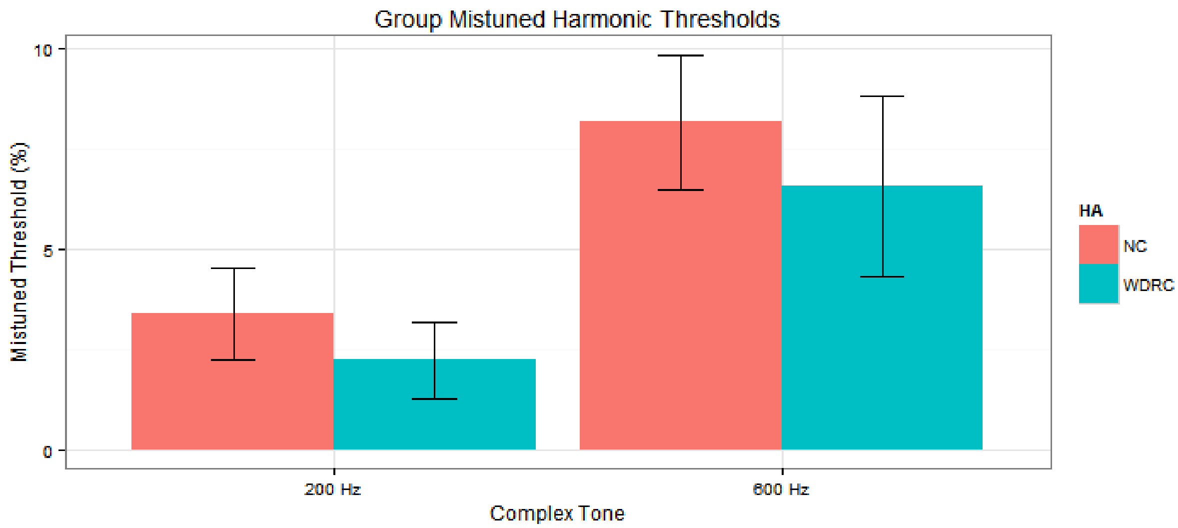

Figure 5.19: NCStudy2 mistuned harmonic - group results. Error bars represent ± 1 SEM. There was no significant difference between NC and WDRC ($F(1,10) = 1.53$, $p = 0.245$), but the higher frequency complex tone was significantly more difficult ($F(1,10) = 5.00$, $p = 0.049$).

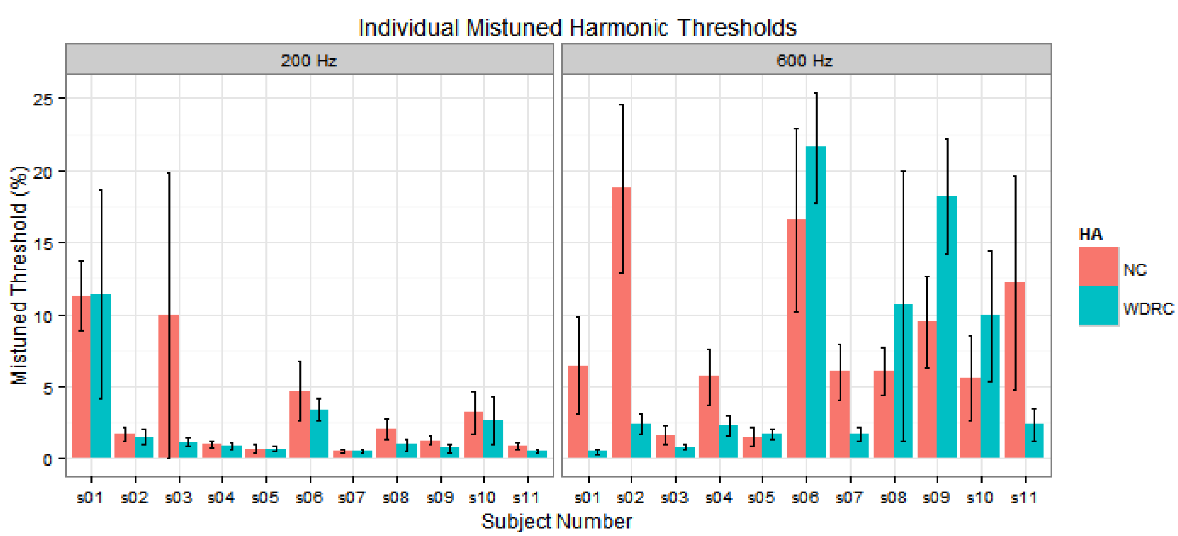

Figure 5.20: NCStudy2 mistuned harmonic - individual results. Error bars represent ± 1 SD. There are huge individual differences on this task.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	59	1822	0.32	0.583		0.008
harm	1	10	6299	1330	47.34	< 0.001	***	0.456
HA:harm	1	10	57	730	0.78	0.399		0.007

Table 5.5: Timbre Perception Within-Subjects Factorial ANOVA

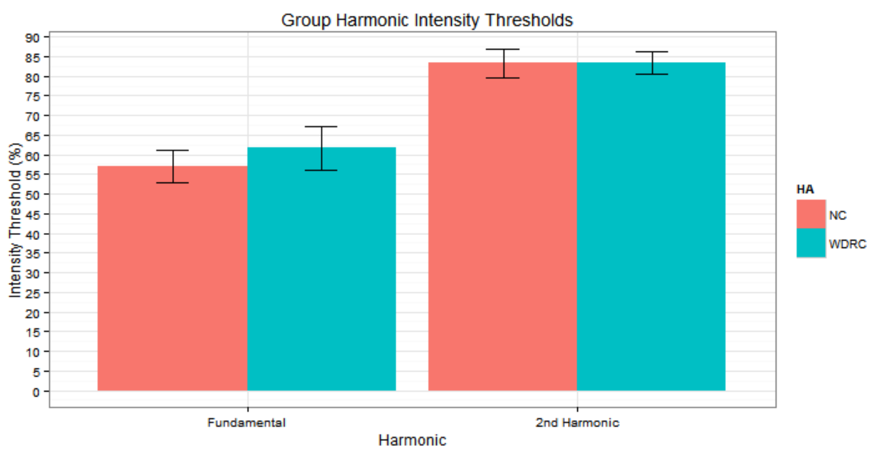

Figure 5.21: NCStudy2 timbre perception - group thresholds. Error bars represent $\pm$ 1 SEM. There was no significant difference between NC and WDRC ($F(1,10) = 0.32$, $p = 0.583$), but the fundamental was easier to discriminate than the second harmonic ($F(1,10) = 47.34$, $p < 0.001$).

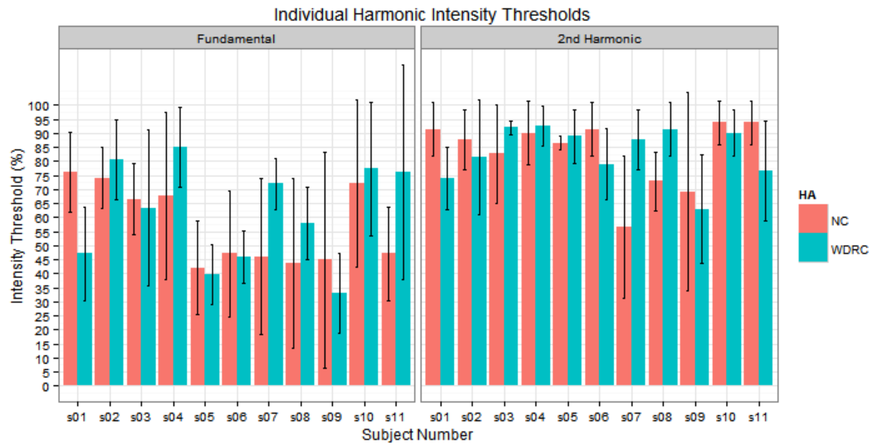

Figure 5.22: NCStudy2 timbre perception - individual thresholds. Error bars represent ± 1 SD.

Gap Detection

A within-subjects ANOVA was performed, with hearing aid (WDRC or NC) as the only factor. The dependent measure was the gap threshold. There was a main effect of hearing aid ($F(1,10) = 18.1$, $p = 0.002$), such that participants could detect a smaller silent gap in a sound while wearing WDRC. Detailed statistical results may be seen in Table 5.6. Group means and standard errors are plotted in Figure 5.23, and individual means and standard deviations are plotted in Figure 5.24.

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	566	314	18.1	0.002	**	0.343

Table 5.6: Gap Detection Within-Subjects ANOVA

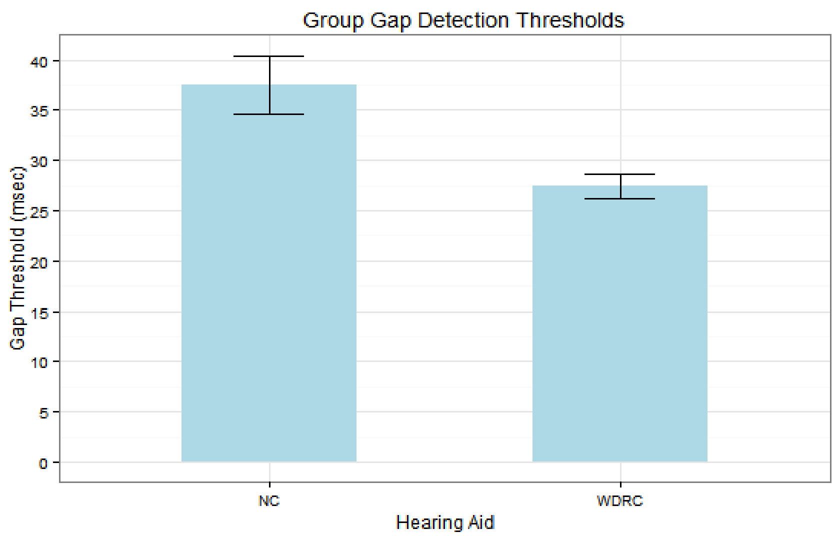

Figure 5.23: NCStudy2 gap detection - group thresholds. Error bars represent ± 1 SEM. Performance on the task was significantly better with WDRC than NC, as determined with a within-subjects ANOVA (F(1,10) = 18.1, p = 0.002).

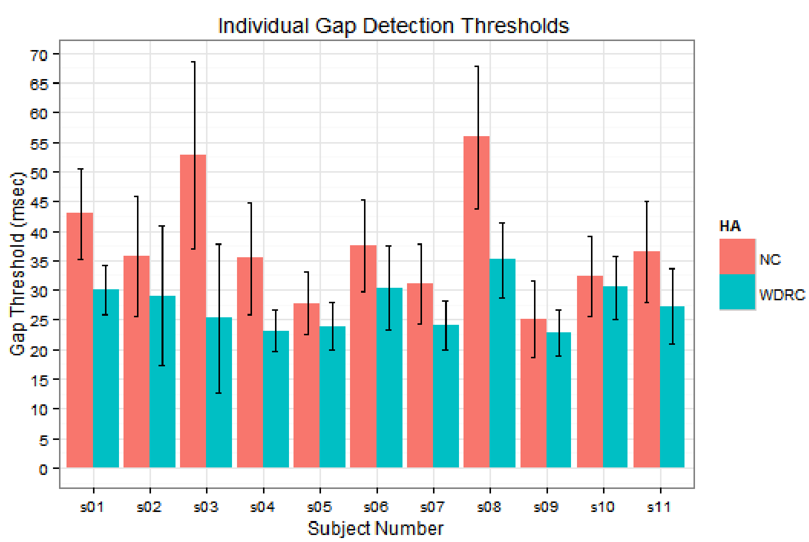

Figure 5.24: NCStudy2 gap detection - individual thresholds. Error bars represent ± 1 SD. All 11 participants performed better on the task with WDRC.

5.4.3 Sound Localization Results

A within-subjects ANOVA was performed, with three within-subjects factors, hearing aid (WDRC or NC), stimulus (low freq, high freq, or phone), and angle (0 to 90° in 15° increments). The dependent measure was the error (|Stimulus° - Response°|). After correcting for deviations from sphericity with the Huynh-Feldt correction factor ($\tilde{\epsilon}$), the only significant effect was a main effect of stimulus ($F(2,20) = 121.64$, $p < 0.001$, $\tilde{\epsilon} = 0.643$), such that the high frequency stimulus was much more difficult than the phone or low frequency stimuli. There was also a trending main effect of angle ($F(6,60) = 2.75$, $p = 0.098$, $\tilde{\epsilon} = 0.287$), and a trending interaction of stimulus and angle ($F(12,120) = 2.57$, $p = 0.093$, $\tilde{\epsilon} = 0.188$). Detailed statistical results may be seen in Table 5.7. Group means and standard errors are plotted, as a function of stimulus in Figure 5.25, as a function of angle in Figure 5.26, and combined in Figure 5.27. Figure 5.28 also shows the distribution of responses on each condition for both NC and WDRC.

5.4.4 Subjective Measure Results

APHAB

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and APHAB subscale (Ease of Communication, Reverberation, Background Noise, Aversiveness of Sounds). The dependent measure was the average score from 1-7 on each subscale. Group means and standard errors are plotted as a function of subscale in Figure 5.29.

After correcting for deviations from sphericity with the Huynh-Feldt correction factor ($\tilde{\epsilon}$), all main effects and the interaction were significant (see Table 5.8). Thus,

ANOVA								
Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	69	1621	0.42	0.530		0.002
stim	2	20	28876	2374	121.64	< 0.001	***	0.469
angle	6	60	2321	8450	2.75	0.020	*	0.066
HA:stim	2	20	160	2271	0.70	0.506		0.005
HA:angle	6	60	95	1492	0.64	0.700		0.003
stim:angle	12	120	1961	7617	2.57	0.005	**	0.057
HA:stim:angle	12	120	494	3127	1.58	0.107		0.015

Mauchly's Test for Sphericity			
Effect	W	p	p<.05
stim	0.347	0.009	**
angle	< 0.001	< 0.001	***
HA:stim	0.242	0.002	**
HA:angle	0.031	0.170	
stim:angle	< 0.001	< 0.001	***

Sphericity Corrections						
Effect	GGe	p[GG]	p[GG]<.05	HFe	p[HF]	p[HF]<.05
stim	0.605	< 0.001	***	0.643	< 0.001	***
angle	0.251	0.107		0.287	0.098	.
HA:stim	0.569	0.437		0.593	0.442	
HA:angle	0.466	0.587		0.664	0.639	
stim:angle	0.154	0.106		0.188	0.093	.
HA:stim:angle	0.267	0.212		0.409	0.185	

Table 5.7: Sound Localization Within-Subjects ANOVA

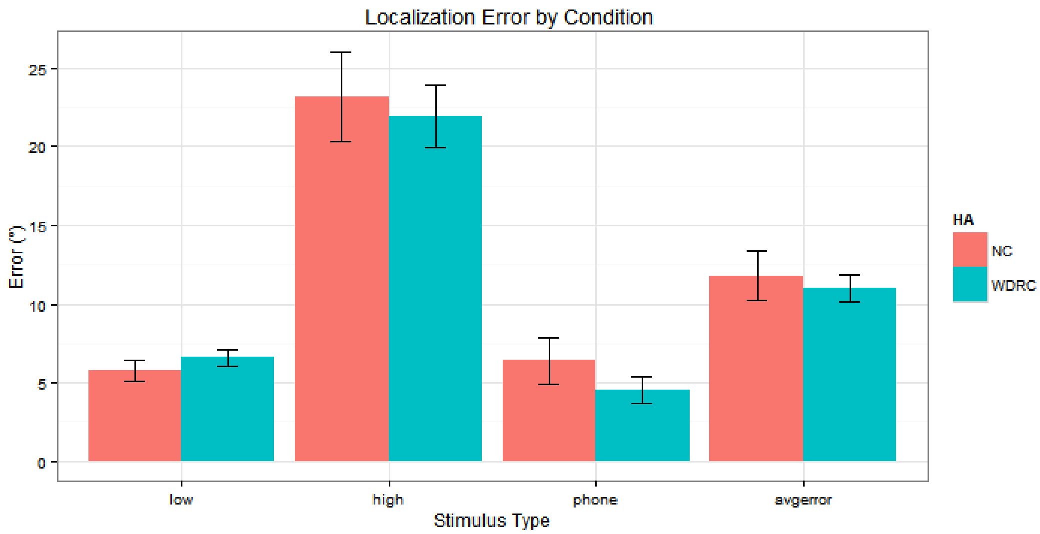

Figure 5.25: NCStudy2 sound localization results - stimulus type. Error bars represent ± 1 SEM. There was a main effect of stimulus type ($F(2,20) = 121.64$, $p < 0.001$, $\tilde{\epsilon} = 0.643$), such that the higher frequency sound was more difficult to localize.

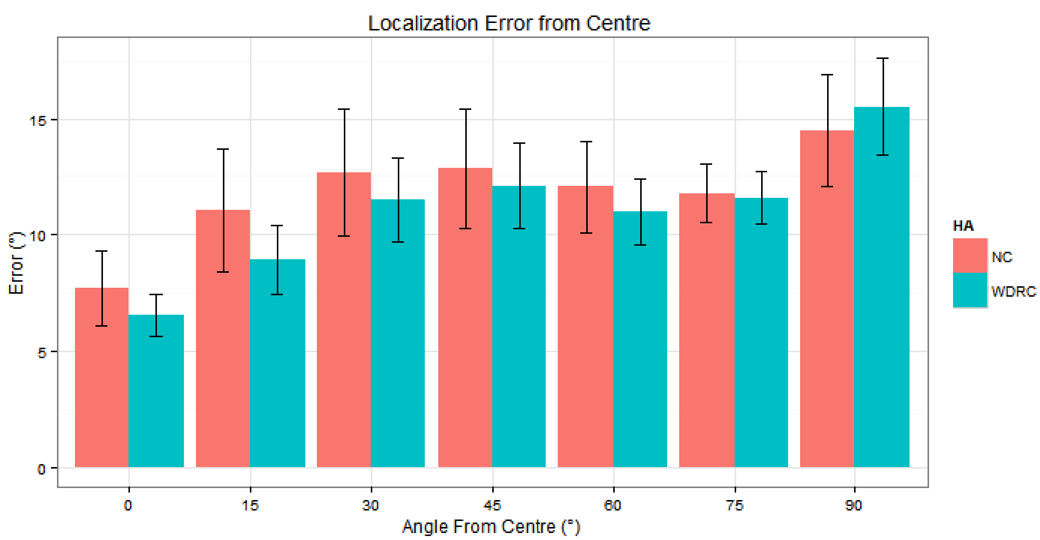

Figure 5.26: NCStudy2 sound localization results - angle. Error bars represent ± 1 SEM. Generally speaking, the further the stimulus is presented from centre, the greater the error.

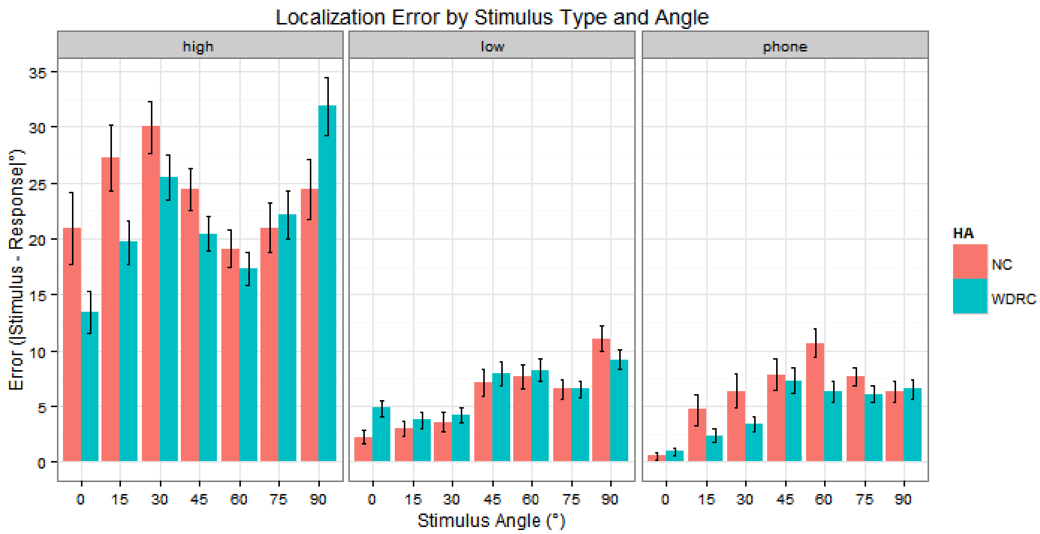

Figure 5.27: NCStudy2 sound localization results - stimulus type and angle. Error bars represent ± 1 SEM.

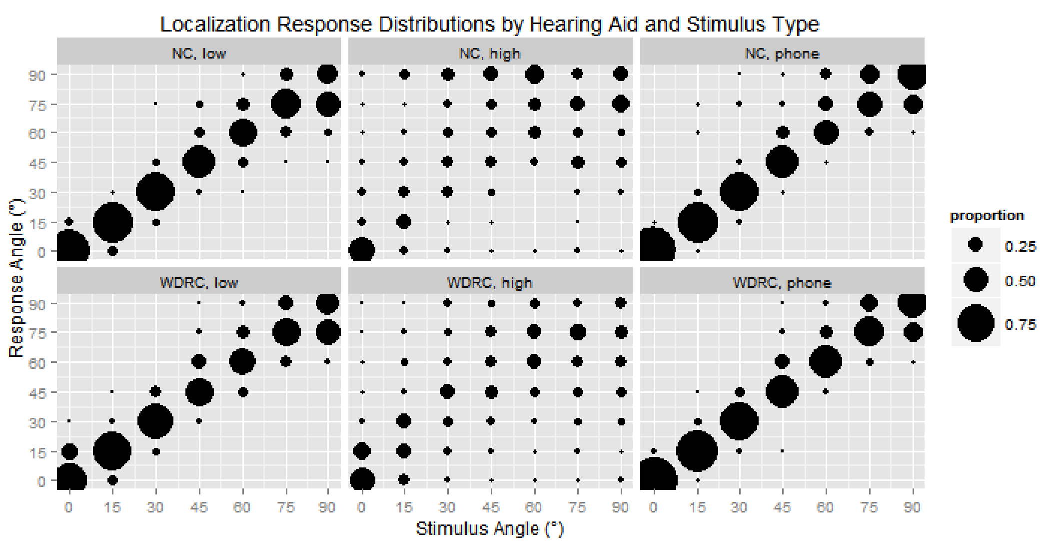

Figure 5.28: NCStudy2 sound localization response distributions. Circle area is proportional to number of responses at a particular stimulus, response configuration.

simple main effects of hearing aid at each level of APHAB subscale were investigated (see Table 5.9 for detailed results). There was a significant simple main effect of hearing aid for the Aversiveness of Sounds subscale ($F(1,10) = 16.20$, $p = 0.002$), such that participants expressed more aversion to sounds while wearing the NC than they did for WDRC, and there was a trending simple main effect of hearing aid for Background Noise ($F(1,10) = 3.61$, $p = 0.087$), such that participants expressed more problems with background noise while wearing the NC. The trending main effect of background noise should be interpreted with caution, as a correction for multiple comparisons was not been applied, and it is only a trend. It is of the author's opinion that it is worth mentioning, since it is potentially useful for interpretation of the results on the objective tests.

The Aversiveness of Sounds subscale contains items related to overly loud/uncomfortable alarms or bells, traffic noises, running water, construction work, fire engines, and screeching tires.

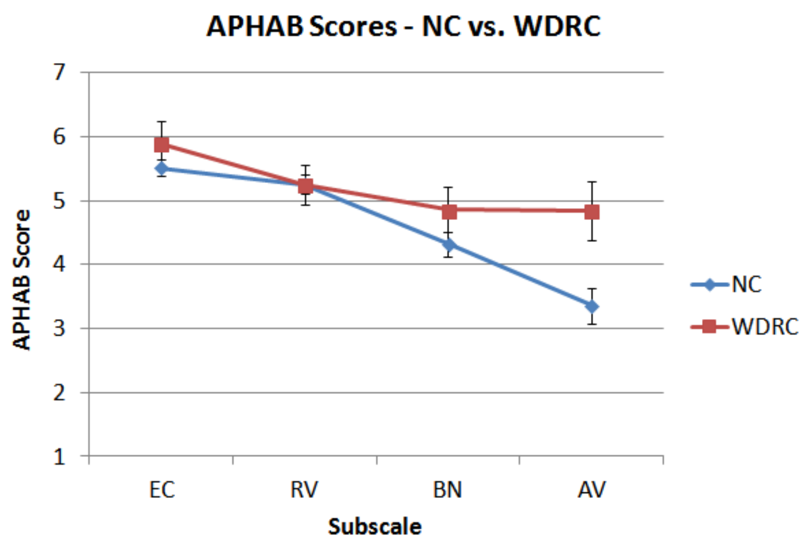

Figure 5.29: NCStudy2 APHAB results. APHAB subscale abbreviations are plotted on the x-axis, corresponding to: Ease of Communication, Reverberation, Background Noise, and Aversiveness of Sounds, respectively. Lower APHAB scores correspond to more problems. Error bars represent ± 1 SEM. A within-subjects ANOVA indicated that there was an interaction between HA and APHAB subscale ($F(3,30) = 6.74$, $p = 0.005$, $\tilde{\epsilon} = 0.721$). Simple main effects of HA tested for each APHAB subscale showed that the NC was associated with significantly greater aversion to sounds ($F(1,10) = 16.20$, $p = 0.002$).

ANOVA								
Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	7.9	10.6	7.44	0.021	*	0.093
subscale	3	30	33.3	28.9	11.52	< 0.001	***	0.302
HA:subscale	3	30	6.7	9.9	6.74	0.001	**	0.080

Mauchly's Test for Sphericity			
Effect	W	p	p<.05
subscale	0.055	< 0.001	***
HA:subscale	0.329	0.087	

Sphericity Corrections						
Effect	GGe	p[GG]	p[GG]<.05	HFe	p[HF]	p[HF]<.05
subscale	0.423	0.003	**	0.457	0.003	**
HA:subscale	0.598	0.008	**	0.721	0.005	**

Table 5.8: APHAB Within-Subjects Factorial Anova

Subscale	DFn	DFd	SSn	SSd	F	p	p<.05	ges
EC	1	10	0.79	4.70	1.68	0.224		0.0507
RV	1	10	0.00	3.91	0.00	0.965		0.000
BN	1	10	1.58	4.38	3.61	0.087	.	0.081
AV	1	10	12.20	7.50	16.20	0.002	**	0.283

Table 5.9: Simple Main Effect of HA for each APHAB Subscale

Sound Quality

A within-subjects ANOVA was performed, with two within-subjects factors, hearing aid (WDRC or NC), and questionnaire subtype (music or overall). The dependent measure was a composite score, averaged over all items in each condition for the questionnaire. The music quality composite score was an average of 4 items, whereas the overall quality composite score was an average of 10 items. The only significant effect was a main effect of questionnaire subtype ($F(1,10) = 13.23$, $p = 0.005$), which is not very interesting. Detailed statistical results may be seen in Table 5.10. Group means and standard errors are plotted in Figure 5.30.

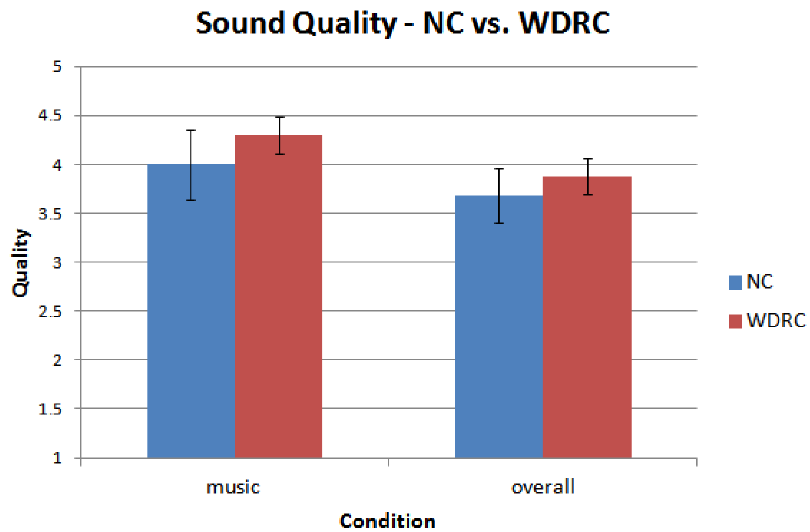

Figure 5.30: NCStudy2 sound quality results. Music quality was assessed on a scale of 1-5 on various dimensions (naturalness, pleasantness, fullness, clarity). For each person and set of hearing aids, a composite score was computed by averaging a participant's scores over all dimensions. Overall quality was computed by averaging scores over several listening situations, each measured on a scale of 1-5, ranging from unpleasant to pleasant. Error bars represent ± 1 SEM. The listening situations included speech in quiet with one person, speech with several talkers, car radio, telephone, television, small and large music performances, speech/music at the cinema, and favorite recordings of the participant's favorite music. There were no significant differences between NC and WDRC on sound quality ($F(1,10) = 0.50$, $p = 0.496$).

Structured Diary

To properly analyze the data from the structured diary, since there were three different dependent measures, the data were first split into three separate subsets: one for the loudness of target speech, another for the clarity of speech, and last, for the loudness of background noise. A within-subjects ANOVA was performed for

Effect	DFn	DFd	SSn	SSd	F	p	p<.05	ges
HA	1	10	0.68	13.64	0.50	0.496		0.022
cond	1	10	1.49	1.12	13.23	0.005	**	0.047
HA:cond	1	10	0.02	2.27	0.11	0.751		0.001

Table 5.10: Sound Quality Within-Subjects ANOVA

each subset of data, with two within-subjects factors, hearing aid (WDRC or NC), and listening situation (conversing in a quiet living room, conversing on a car/bus, having a meal with 2-3 people, conversing with a noisy group, listening to someone at a distance, and television at home). The subset of data for clarity was largely uninteresting, and is ignored here. There were close to significant interactions between hearing aid and listening situation for both the loudness of target speech data ($F(5,30) = 2.12$, $p = 0.13$, $\tilde{\epsilon} = 0.620$) and the loudness of background noise data ($F(5,30) = 2.19$, $p = 0.09$, $\tilde{\epsilon} = 0.89$). Once again, no corrections for multiple comparisons were made here, so the results must be interpreted with caution. Figure 5.31 depicts results on each listening situation for the target loudness data, and Figure 5.32 depicts results on each listening situation for the background noise data.

In Figure 5.31, although no statistical analyses were performed, the loudness of target speech for the NC generally seems to louder, and closer to the optimal score (3) than WDRC, particularly in quiet, while watching television, and while having a meal with 2-3 people.

In Figure 5.32, although no statistical analyses were performed, it is quite clear that the greatest difference in background noise ratings between the NC and WDRC is with a group of people in a noisy situation. Participants tend to report that the loudness of background noise is too loud with the NC for this particular situation, while for the other situations there are no differences between the NC and WDRC.

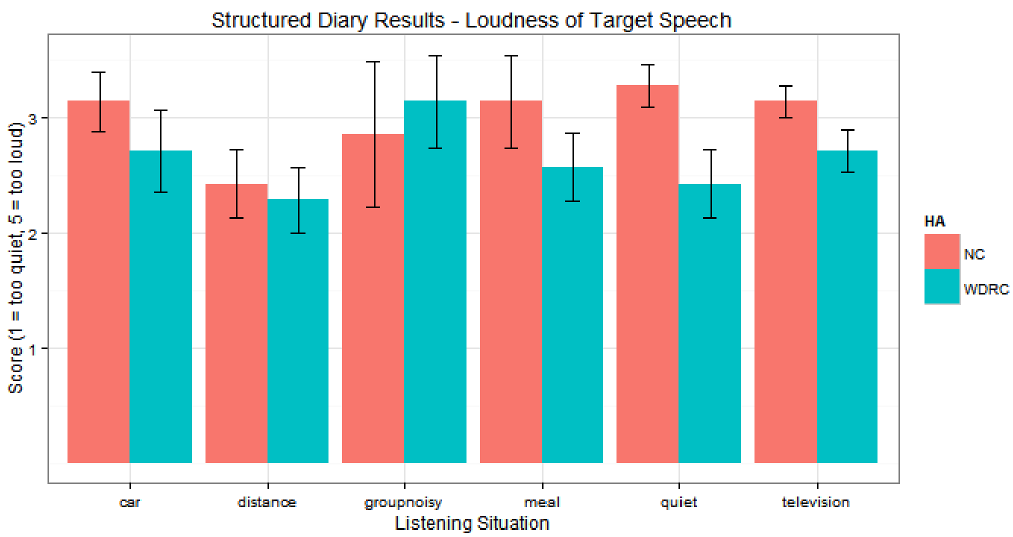

Figure 5.31: NCStudy2 structured diary results - loudness of target speech. Error bars represent ± 1 SEM. The loudness of target speech for the NC generally seems to louder, and closer to the optimal score (3) than WDRC, particularly in quiet, while watching television, and while having a meal with 2-3 people.

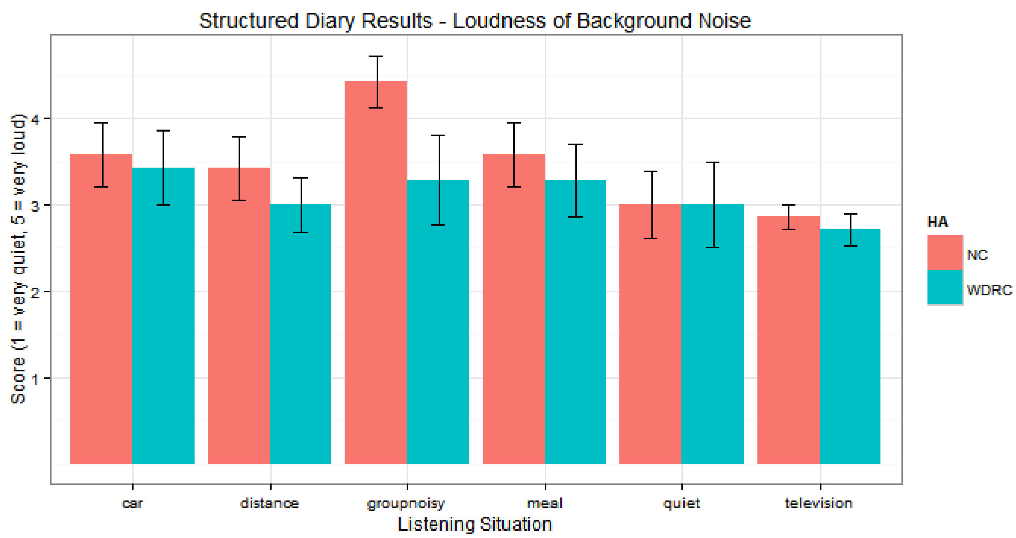

Figure 5.32: NCStudy2 structured diary results - loudness of background noise. Error bars represent ± 1 SEM. Participants tend to report that the loudness of background noise is too loud with the NC while speaking with a group of people in a noisy situation, while for the other situations there are no differences between the NC and WDRC.

Feedback Form

Table 5.11 tabulates results from the feedback form. On some questions, participants expressed no preference for one hearing aid over the other; in these cases, the observation was ignored when summing the amount of participants preferring WDRC or the NC. Overall, 7 participants preferred WDRC, and 4 participants preferred the NC. The main reasons for not liking the NC was that it was reported to be too loud relative to the WDRC, it produced uncomfortable sounds (most likely attributable to entrainment), and two participants complained that one side was louder than the other. The main reasons for not liking the WDRC, was frequent feedback sounds, and a perceived lack of amplification.

Situation	NC Count	WDRC Count
Music	5	6
Speech in quiet	4.5	6.5
Speech in noise	3.5	7.5
Television	5	6
Telephone	3	8
Overall	4	7

Table 5.11: Hearing Aid Preference for Different Listening Situations

5.5 Discussion

The discussion of NCStudy2 proceeds in similar fashion to the discussion for NCStudy1. First, potential limitations to the study design are outlined, followed by a discussion of results on the various tasks, so that the reader may interpret the results while keeping the study limitations in mind.

5.5.1 Limitations

Unfortunately, despite being a much more controlled study than NCStudy1, there were still a few factors in the study that we were unable to control for, which may affect the interpretation of the results.

Features

First, noise reduction and feedback cancellation had to be enabled on the test hearing aids, in order to ensure participants' acceptance of the hearing aids in their everyday listening environments. Although the implementation of these features were, for the most part, identical between WDRC and NC, the constraint that these features must be identically implemented does not necessarily prevent them from contaminating the results. The reason is that noise reduction and feedback cancellation could interact with the amplification algorithm, and for an ideal comparison of two amplification algorithms, all features on the hearing aids ought to be disabled. During the design of the study, this issue was anticipated, and a solution was proposed to attempt to correct for this study design flaw. The solution proposed the creation of 2 programs on participants' hearing aids (program 1: features enabled, and program 2: features disabled). The intention was for program 1 to be the everyday listening program, and for program 2 to be used strictly in the testing environment. The solution was sound, in theory, but after programming the hearing aids, it was determined that program 2 would be unusable, at least for the WDRC hearing aids, because there was an inordinate amount of feedback. As a result, participants were told to use only program 1 on their hearing aids, and all participants complied.

In particular, the mistuned harmonic and timbre perception tasks were likely highly affected by the interaction between the amplification algorithm and the feedback canceller. Some participants reported an additional ringing or an additional tone during the presentation of tonal stimuli while wearing the NC hearing aids (entrainment), and, as a result, found it difficult to perform to the best of their ability on these tasks. By contrast, there were no entrainment-like symptoms reported for WDRC by any of the participants in the study. The reason for entrainment occurring in the NC, but not the WDRC, could be due to the increased gain in the NC hearing aids, or slight differences in the implementation of the feedback canceller. No statistically significant differences were observed between hearing aids on the mistuned harmonic and timbre perception tasks, but perhaps if entrainment was eliminated, or if features were entirely disabled and the proper care was taken to eliminate any source of feedback, this question could be addressed properly. There was no effect of entrainment on the other tasks (the problem only occurs in response to sufficiently tonal inputs), so results from the other tasks may be interpreted knowing there was little or no effect of the feedback canceller on these tasks.

Gain

Second, there appeared to be quite different amounts of gain between the NC and WDRC hearing aids, according to real ear measurements that were collected (Figures 5.3, 5.4, and 5.5). In any comparison of hearing aid amplification algorithms, the gain is expected to differ between the algorithms, but perhaps not to the extent that the gain differed between hearing aids in the current study. The WDRC hearing aids, though they were fit according to the DSL [i/o]* prescription, were given less gain than the prescribed target. The WDRC software required the specialist to input the hearing aid experience of participants on a scale of 1 to 4 (1 = new hearing aid user, 4 = long time hearing aid user), and the specialist naturally selected a value of 1, and during the adjustment session was sometimes able to increase the value to 2. Generally, in hearing aid fittings, it takes a while to adapt to the increased volume, so patients are often initially under-fit, and gain is gradually increased when patients ask for it, or when the specialist suggests it (Brooks, 1973). By contrast, the NC gain is prescribed independent of any parameter specifying a user's hearing aid experience, which partly explains why more gain was given for the NC overall than the WDRC hearing aids.

Adjustments

The software used to adjust the hearing aids also differs between hearing aid manufacturers, so when comparing two hearing aid technologies in the context of a study that has hearing aid adjustments, the efficacy of the adjustment depends on the clinician's familiarity with the software, and clinician's skill with one hearing aid technology versus the other.

5.5.2 Conclusions from NCStudy2

Speech Intelligibility

A realistic speech in noise test did not show any significant differences in speech intelligibility between the two hearing aids, as well as a test of word recognition in quiet. Vowel perception was also about the same between hearing aids. However, the NC hearing aids performed more poorly on consonant recognition.

Consonants that were especially worse with the NC were /d/, /h/ (mainly noise),

/k/ (noise only), /r/ (mainly quiet), /s/, /t/, /w/ (noise only), and /z/ (mainly quiet). These consonants fall into the categories of stop consonants (d, k, t), fricatives (h, s, z), and a couple of glides (r, w). In order to better understand why consonant recognition was worse for these consonants, an overview of the acoustic features of these groups of consonants must first be presented.

The acoustic energy for stop consonants is relatively wideband, although /t/, and less so, /k/, have a lot of high frequency energy (Hamill & Price, 2008). Some other important characteristics of stop consonants are the stop gap, which is the length of time between the stop consonant and the next vowel, and the voice onset time, the time between the beginning of phoneme production and when the vocal folds start to vibrate. The stop gap and voice onset times are useful for differentiating between voiced and unvoiced consonants. Formant transition contours are also very important in the classification of stop consonants, in that stop consonants differ with respect to how a consonant's spectrum transitions into the subsequent vowel formants. For instance, for the words /bet/, /det/, and /get/, /b/ has an ascending second formant transition, /d/ has a flat second formant transition, while /g/ has a descending second formant transition.

Fricatives are long in duration relative to the other consonants, and contain even more high frequency energy than stop consonants. /s/ and /z/ contain the most high frequency energy out of all the English consonants.

Glides are very vowel-like, in that they mainly contain low frequency energy. Glides also tend to be longer in duration, but /r/ is an exception to the rule. Similar to the stop consonants, the formant transitions may be used as a cue for characterizing different glides.

Out of the consonants which the NC had trouble with, some common themes emerge. One is that the consonants mostly had high frequency energy, and formant transitions are especially important for these consonants to help with recognition. Therefore, it may be that the NC is not providing enough amplification at very high frequencies, or that the formant transitions are not being restored adequately. Another possibility is that the prior or subsequent vowel is overamplified in the context of the VCV experiment, resulting in greater amounts of perceptual forward or backward masking of the consonant sound.

To investigate these hypotheses further, a sensible experiment would involve isolating each of these features in some stimulus, and experimentally manipulating these features one by one, in order to determine which of these features are distorting consonant perception for the NC hearing aid.

Music Perception

One hearing aid did not result in superior frequency resolution, as measured by the mistuned harmonic task, nor was timbre perception more accurate with one hearing aid over another, unlike the trending result of favorable timbre perception for the NC in NCStudy1. In NCStudy1, there were no complaints of entrainment during the procedure for either set of hearing aids, but in NCStudy2, there were complaints of entrainment, only for the NC hearing aids. Thus, task performance could have been affected by entrainment for the NC, and for this reason it is difficult to interpret these results.

The most surprising result was that gap detection thresholds were consistently lower with WDRC than the NC hearing aids, indicating that temporal resolution is superior with WDRC. There may be many reasons for this discrepancy in performance, but two reasons quickly come to mind. One, is that compression is known to affect the detection of gaps in narrow bands of noise (Glasberg & Moore, 1992; Moore et al., 2001). Specifically, when compression is applied to a narrow band noise stimulus, such as that used in the gap detection task employed in this thesis, one is able to detect smaller gaps in a noise band (Glasberg & Moore, 1992). So, perhaps the difference in performance between the NC and WDRC on this task is due to less compression given by the NC, or a different type of compression used altogether (slow-acting, fast-acting). The second possibility which could account for the worse performance for the NC hearing aids, is a difference in the amount of gain applied between the two hearing aid types. Supposing that the NC gave more gain outside of the 1000 Hz region (where the noise band was centred), particularly just below the 1000 Hz region, then the noise band centred at 1000 Hz could be masked disproportionately more with the NC.

To evaluate the compression hypothesis, measures should be taken to quantify how much and what type of compression is being applied by the NC hearing aid. To test the gain hypothesis, a gap detection experiment with noise bands centred at different frequencies could be designed, taking care to measure the amount of gain given to each participant in the experiment.

Sound Localization

There was no significant main effect or interaction involving the hearing aid factor for the sound localization experiment. However, looking at the localization error by stimulus type and angle graph (Figure 5.27), there did appear to be some differences between the NC and WDRC for high frequency stimuli presented close to centre (at 0°, 15°, and 30° angles), with the NC performing worse. One potential reason for this discrepancy is that a couple of the NC hearing aid fittings gave asymmetrical gain between the two ears, perhaps distorting ILD cues for these participants, which are important for the localization of high frequencies. To test this hypothesis, localization performance should be investigated on an individual basis. One other possible reason is that the NC did not provide enough high frequency gain, reducing performance for the NC in the high frequency condition. This hypothesis is supported by the fact that localization performance for the phone stimulus (which contains both high and low frequencies) also tended to be worse for the NC, yet localization performance for the low frequency stimulus was roughly equal.

Subjective Measures

Several differences were observed between the two hearing aid types for the subjective measures collected in the current study. The APHAB results indicated that background noise was overly loud with the NC compared to the WDRC, and certain sounds were especially loud (screeching tires, alarms/bells, running water, traffic noises, construction work, fire engines). These findings were corroborated by evidence obtained from the structured diary used in the study. The structured diary results indicated that background noise was too loud for the NC, but only when speaking with a group of people in a noisy situation. One advantage that the NC had was that target speech tended to be closer to an optimal volume, as reported in the structured diary. Overall sound quality and music quality was not reported to be better with one hearing aid type versus the other.

The results from the subjective measures suggest that the NC may not be providing enough compression, or that there is too much gain in the low frequencies and not enough in the high frequencies. Background noise, which tends to have greater energy in the lower frequencies, might be overamplified with the NC.

Overall

Taken together, the results from the whole study suggest that noise may be over-amplified and under-compressed with the NC. A lack of compression, over-amplification of the low frequencies and potential under-amplification of the high frequencies are likely contributing to worse consonant perception, and worse temporal resolution for the NC hearing aid.

Chapter 6

Conclusions

6.1 Summary of Experiments

This thesis reports results from two separate hearing aid studies, comparing two different hearing aid technologies: WDRC, a standard amplification algorithm used in the hearing aid industry, and the NC, a novel hearing aid which uses a computational model of the auditory periphery to attempt to restore normal auditory nerve output. The first study recruited participants with prior hearing aid experience, who already owned a set of WDRC hearing aids. The second study recruited participants with no prior hearing aid experience, was double-blind, and used WDRC and NC hearing aids which were identical in virtually all respects except the amplification algorithm programmed onto the hearing aids. The aim was to compare the two hearing aid technologies on as many different auditory domains as possible (speech intelligibility, music perception, sound localization, and subjective measures).

The aim of the first study was to determine whether an implementation of the NC was comparable to WDRC hearing aids currently out on the market. Due to many factors which were outside of the experimenter's control (lack of real ear measurement tools for gain verification, different features programmed on the WDRC and NC hearing aids, the age of the hearing instruments tested, the quantity and quality of hearing aid adjustments that participants received, hearing aid adaptation effects, and a lack of access to participants' audiometric history) the results of this first study are not easily interpretable, and so results for this study were considered preliminary. One hearing aid type did not confer an advantage in speech in noise, however, the NC seemed not to restore particular consonants as well as WDRC. Timbre perception tended to be superior for NC, such that finer changes to musical timbre were more easily detected while participants wore NC hearing aids. Overall hearing aid preference was in favor of the NC.

The aim of the second study was to control for as many factors as possible which potentially influenced the results obtained from the first study. The NC and WDRC hearing aids looked identical, had the same hardware and very similar implementation of hearing aid features (DNR, FC), and the experimenter and participants did not know which hearing aid type the participants were wearing at any given instant during the study. There was no difference in speech in noise performance between the hearing aids for a realistic speech in noise test, and both hearing aids offered about the same improvement in word recognition in quiet (approximately 15%). Similar to the first study, the NC did not restore particular consonants (fricatives, stop consonants, glides) as well as WDRC. There were no differences in music perception ability or music quality while wearing the NC or WDRC, and music quality was not greater for one hearing aid over the other. Temporal resolution was superior for the WDRC hearing aids as determined by a gap detection task. Subjective measures indicated that background noise was reported to be too loud with the NC, particularly while conversing in a noisy group situation, and there was greater aversion to sounds with the NC.

6.2 Contributions to Research

Throughout the course of this research program, many obstacles were encountered, and most of them were surmounted. It is very challenging to conduct completely controlled hearing aid studies, because it is difficult to isolate only one or two variables to manipulate, which is crucial for any experimental design. In the context of hearing aid research, such variables include the hearing aid amplification algorithm, the hardware and electrical components used on selected test hearing aids, the large array of features programmed on modern hearing aids (FC, DNR, directional microphones, automatic programs, volume control, etc.), and the inexact science of performing hearing aid adjustments. For any future clinical trials of hearing aids, one may consult this thesis for guidance on how to design and conduct a proper study of hearing aid technologies.

In addition, the current thesis applied music perception methods to hearing aid research, for which there is limited scientific literature on the topic. It was found that musical stimuli (complex tones, musical instruments) can produce entrainment artifacts, which are added tones by the feedback canceller in response to what it detects as feedback. To eliminate entrainment artifacts in the context of a music perception experiment, one strategy is to submerge the musical stimuli in low-level background noise. When programming these experiments, care must be taken to set the background noise to an appropriate level so as not to mask the musical stimuli, while at the same time ensuring that there are no entrainment artifacts. A more effective strategy is to turn the feedback cancellation system off when testing music perception abilities, but then real feedback becomes a possibility.

The most important contribution of the current research program was perhaps its identification of auditory domains that the NC could improve upon, and its offering of suggestions for why the NC may be under-performing in these auditory domains (consult the Discussion in Chapter 5 for more details).

6.3 Future Research

It is unclear whether the differences between WDRC and the NC on the tasks reported above are due to real differences between the NC and WDRC amplification algorithms, or whether programming these algorithms on a hearing aid causes the NC to under-perform. To answer such a question, different versions of WDRC algorithms should be obtained in software, and compared to the NC algorithm in software, using the stimuli from this study as input.

Another very interesting avenue of research could be to use electroencephalography to investigate to what extent the neural code is restored back to normal at various stages along the auditory pathway. One such technique, called complex auditory brain responses (cABRs), allows one to do just this, as the cABR physically resembles the evoking auditory stimulus (Anderson & Kraus, 2010).

Appendix A

Sound Quality Questionnaire

1. How clear does music sound with the hearing aids you are *currently* wearing?
   1 (unclear) 2 (moderately unclear) 3 (neither clear nor unclear) 4 (moderately clear) 5 (clear)

2. How pleasant does music sound with the hearing aids you are *currently* wearing?
   1 (unpleasant) 2 (moderately unpleasant) 3 (average) 4 (moderately pleasant) 5 (pleasant)

3. How natural does music sound with the hearing aids you are *currently* wearing?
   1 (unnatural) 2 (moderately unnatural) 3 (average) 4 (moderately natural) 5 (natural)

4. How full does music sound with the hearing aids you are *currently* wearing?
   1 (thin) 2 (moderately thin) 3 (neither thin nor full) 4 (moderately full) 5 (full)

For the following conditions, rate how pleasant the listening experience is while wearing the hearing aids you are *currently* wearing:
1 = unpleasant
2 = moderately unpleasant
3 = average
4 = moderately pleasant
5 = pleasant

1. Listening to a small live performance (solo/small group).

2. Listening to a large live performance (concert/orchestra).

3. Listening to TV programs.
4. Listening on the telephone.
5. Listening to another talker in quiet.
6. Listening to a car radio.
7. Listening to another talker amongst competing talkers.
8. Listening to speech in an auditorium/gymnasium.
9. Listening to speech/music at the cinema.
10. Listening to your favourite recordings of your favourite music.

References

Abel, S. M. (1972). Duration discrimination of noise and tone bursts. *The Journal of the Acoustical Society of America*, *51*, 1219-1223.

Aibara, R., Welsh, J., Puria, S., Goode, R. (2001). Human middle-ear sound transfer function and cochlear input impedance. *Hearing Research*, *152(1)*, 100-109.

Anderson, S., Kraus, N. (2010). Objective neural indices of speech-in-noise perception. *Trends in Amplification*, *14(2)*, 73-83.

Anshel, A., Kipper, D. (1988). The influence of group singing on trust and cooperation. *Journal of Music Therapy*, *25(3)*, 145-155.

Arthur, R., Pfeiffer, R., Suga, N. (1971). Properties of two-tone inhibition in primary auditory neurones. *The Journal of Physiology*, *212(3)*, 593-609.

Becker, S., Bruce, I. C. (2002, March). Neural coding in the auditory periphery: Insights from physiology and modeling lead to a novel hearing compensation algorithm. In *Workshop on neural information coding.* Les Houches France.

Bench, K. A., J., Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13(3)*, 108-112.

Bentler, R., Chiou, L. (2006). Digital noise reduction: An overview. *Trends in Amplification*, *10(2)*, 67-82.

Boike, K. T., Souza, P. E. (2000). Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification. *Journal of Speech, Language and Hearing Research*, *43(2)*, 456-468.

Bondy, J., Becker, S., Bruce, I., Trainor, L., Haykin, S. (2004). A novel signal-processing strategy for hearing-aid design: neurocompensation. *Signal Processing*, *84(7)*, 1239-1253.

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

British Society of Audiology. (2004). *Recommended procedure for pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels.*

Bronkhorst, A. W., Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *83*, 1508-1516.

Brooks, D. (1973). Gain requirements of hearing aid users. *Scandinavian Audiology*, *2(4)*, 199-205.

Brooks, D. (1996). The time course of adaptation to hearing aid use. *British Journal of Audiology*, *30(1)*, 55-62.

Bruce, I. C., Sachs, M. B., Young, E. D. (2003). An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses. *The Journal of the Acoustical Society of America*, *113*, 369-388.

Byrne, D., Dillon, H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, *7(4)*, 257-265.

Byrne, D., Parkinson, A., Newall, P. (1990). Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired. *Ear and Hearing*, *11(1)*, 40-49.

Carney, L. (1994). Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. *Hearing Research*, *76(1)*, 31-44.

Chasin, M., Russo, F. A. (2004). Hearing aids and music. *Trends in Amplification*, *8(2)*, 35-47.

Chen, Z., Becker, S., Bondy, J., Bruce, I. C., Haykin, S. (2005). A novel model-based hearing compensation design using a gradient-free optimization method. *Neural computation*, *17(12)*, 2648-2671.

Church, R. M., Getty, D. J., Lerner, N. D. (1976). Duration discrimination by rats. *Journal of Experimental Psychology. Animal Behavior Processes*, *2(4)*, 303-312.

Clarkson, M. G., Clifton, R. K., Perris, E. E. (1988). Infant timbre perception: Discrimination of spectral envelopes. *Perception and Psychophysics*, *43(1)*, 15-20.

Cooke, M., Scharenborg, O. (2008, September). The interspeech 2008 consonant challenge. In *Proceedings of the interspeech* (p. 1765-1768).

Cornelisse, L. E., Seewald, R. C., Jamieson, D. G. (1995). The input/output formula: A theoretical approach to the fitting of personal amplification devices. *The Journal of the Acoustical Society of America*, *97*, 1854-1864.

Cox, R., Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing*, *16(2)*, 176-186.

Cox, R., Gilmore, C. (1990). Development of the profile of hearing aid performance (PHAP). *Journal of Speech, Language and Hearing Research*, *33(2)*, 343-357.

Cox, R., Johnson, J., Alexander, G. (2012). Implications of high-frequency cochlear dead regions for fitting hearing aids to adults with mild to moderately severe hearing loss. *Ear and Hearing*, *33(5)*, 573-587.

Davies-Venn, E., Souza, P., Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology*, *18(8)*, 688-699.

Deroche, M., Culling, J. (2011). Narrow noise band detection in a complex masker: Masking level difference due to harmonicity. *Hearing Research*, *282(1)*, 225-235.

Dillon, H. (1999). Nal-nl1: A new procedure for fitting non-linear hearing aids. *The Hearing Journal*, *52(4)*, 10-12.

Emiroglu, S., Kollmeier, B. (2008). Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. *Brain Research*, *1220*, 199-207.

Ferguson, S. H., Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*, 259-271.

Fitzgibbons, P. J., Wightman, F. L. (1982). Gap detection in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *72*, 761-765.

Gfeller, K., Lansing, C. R. (1991). Melodic, rhythmic, and timbral perception of adult cochlear implant users. *Journal of Speech, Language and Hearing Research*, *34(4)*, 916-920.

Gfeller, K., Witt, S., Mehr, M. A., Woodworth, G., Knutson, J. (2002). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Annals of Otology, Rhinology and Laryngology*, *111*, 349-356.

Giraudi-Perry, D. M., Salvi, R. J., Henderson, D. (1982). Gap detection in hearing-impaired chinchillas. *The Journal of the Acoustical Society of America*, *72*, 1387-1393.

Glasberg, B. R., Moore, B. C. (1992). Effects of envelope fluctuations on gap detection. *Hearing Research*, *64(1)*, 81-92.

Glasberg, B. R., Moore, B. C., Bacon, S. P. (1987). Gap detection and masking in hearing-impaired and normal-hearing subjects. *The Journal of the Acoustical Society of America*, *81*, 1546-1556.

Glista, D., Scollie, S., Bagatto, M., Seewald, R., Parsa, V., Johnson, A. (2009). Evaluation of nonlinear frequency compression: Clinical outcomes. *International Journal of Audiology*, *48(9)*, 632-644.

Goines, L., Hagler, L. (2007). Noise pollution: A modern plague. *Southern Medical Journal*, *100(3)*, 287-294.

Goto, M., Hashiguchi, H., Nishimura, T., Oka, R. (2008, October). Rwc music database: Music genre database and musical instrument sound database. In *Proceedings ISMIR* (Vol. 3, p. 229-230).

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, *61*, 1270-1277.

Guinan Jr., J. J. (2012). How are inner hair cells stimulated? Evidence for multiple mechanical drives. *Hearing Research*, *292*, 35-50.

Hamill, T., Price, L. L. (2008). *The Hearing Sciences.* Plural Publishing, San Diego, CA.

Hartmann, W. M. (1996). Pitch, periodicity, and auditory organization. *The Journal* *of the Acoustical Society of America*, *100*, 3491-3502.

Hawkins, D. B., Cook, J. A. (2003). Hearing aid software predictive gain values: How accurate are they? *The Hearing Journal*, *56(7)*, 26-28.

Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K. (1995). Acoustic characteristics of american english vowels. *The Journal of the Acoustical Society of America*, *97*, 3099-3111.

Hopkins, K., Moore, B., Stone, M. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *The Journal of the Acoustical Society of America*, *123*, 1140.

Humes, L. E., Christensen, L., Thomas, T., Bess, F. H., Hedley-Williams, A., Bentler, R. (1999). A comparison of the aided performance and benefit provided by a linear and a two-channel wide dynamic range compression hearing aid. *Journal of Speech, Language and Hearing Research*, *42(1)*, 65.

Jenstad, L. M., Seewald, R. C., Cornelisse, L. E., Shantz, J. (1999). Comparison of linear gain and wide dynamic range compression hearing aid circuits: Aided speech perception measures. *Ear and Hearing*, *20(2)*, 117-126.

Keidser, G., Dillon, H. R., Flax, M., Ching, T., Brewer, S. (2011). The nal-nl2 prescription procedure. *Audiology Research*, *1(1)*, e24.

Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., Convery, E. (2006). The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers. *International Journal of Audiology*, *45(10)*, 563-579.

Kiefte, M., Enright, T., Marshall, L. (2010). The role of formant amplitude in the perception of /i/ and /u/. *The Journal of the Acoustical Society of America*, *127*, 2611-2621.

Kiefte, M., Kluender, K. R. (2005). The relative importance of spectral tilt in monophthongs and diphthongs. *The Journal of the Acoustical Society of America*, *117*, 1395-1404.

Knight, W., Rickard, N. (2001). Relaxing music prevents stress-induced increases in subjective anxiety, systolic blood pressure, and heart rate in healthy males and females. *Journal of Music Therapy*, *38(4)*, 254-272.

Kochkin, S. (2009). MarkeTrak VIII: 25-year trends in the hearing health market. *Hearing Review*, *16(11)*, 12-31.

Kochkin, S. (2010). Marketrak viii: Consumer satisfaction with hearing aids is slowly increasing. *Hearing Review*, *17(4)*, 12-34.

Kochkin, S., Beck, D., Christensen, L., Compton-Conley, C., Fligor, B., Kricos, P., ... Turner, R. (2010). MarkeTrak VIII: The impact of the hearing healthcare professional on hearing aid user success. *The Hearing Journal*, *63(1)*, 19-20.

Larson, V. D., Williams, D. W., Henderson, W. G., Luethke, L. E., Beck, L. B., Noffsinger, D., ... Rappaport, B. (2000). Efficacy of 3 commonly used hearing aid circuits. *JAMA: The Journal of the American Medical Association*, *284(14)*, 1806-1813.

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

Lin, F. R. (2011). Hearing loss and cognition among older adults in the united states. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *66(10)*, 1131-1136.

Lobarinas, E., Salvi, R., Ding, D. (2013). Insensitivity of the audiogram to carboplatin induced inner hair cell loss in chinchillas. *Hearing Research*, *302*, 113-120.

Looi, V., McDermott, H., McKay, C., Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear and Hearing*, *29(3)*, 421-434.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, *103(49)*, 18866-18869.

Maxwell, S., Delaney, H. (2004). *Designing experiments and analyzing data: A model comparison perspective*. Routledge.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58(3)*, 177-192.

McCreery, R., Venediktov, R., Coleman, J., Leech, H. (2012). An evidence-based systematic review of directional microphones and digital noise reduction hearing aids in school-age children with hearing loss. *American Journal of Audiology*, *21(2)*, 295-312.

McDermott, H. (2004). Music perception with cochlear implants: A review. *Trends in Amplification*, *8(2)*, 49-82.

Mehrgardt, S., Mellert, V. (1977). Transformation characteristics of the external human ear. *The Journal of the Acoustical Society of America*, *61*, 1567-1576.

Middlebrooks, J. C., Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, *42(1)*, 135-159.

Miller, R., Schilling, J., Franck, K., Young, E. (1997). Effects of acoustic trauma on the representation of the vowel [bold/e/] in cat auditory nerve fibers. *The Journal of the Acoustical Society of America*, *101*, 3602-3616.

Moore, B. (2007). *Cochlear hearing loss: physiological, psychological and technical issues*. Wiley-Interscience.

Moore, B., Glasberg, B. (1988). Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. *The Journal of the Acoustical Society of America*, *83*, 1093-1101.

Moore, B., Glasberg, B. R. (1998). Use of a loudness model for hearing-aid fitting. i. linear hearing aids. *British Journal of Audiology*, *32(5)*, 317-335.

Moore, B., Glasberg, B. R., Alcántara, J. I., Launer, S., Kuehnel, V. (2001). Effects of slow- and fast-acting compression on the detection of gaps in narrow bands of noise. *British Journal of Audiology*, *35(6)*, 365-374.

Moore, B., Marriage, J., Alcántara, J., Glasberg, B. R. (2005). Comparison of two adaptive procedures for fitting a multi-channel compression hearing aid: Comparación de dos procesos de adaptación de auxiliares auditivos con compresión multicanal. *International Journal of Audiology*, *44(6)*, 345-357.

Nelson, P. B., Thomas, S. D. (1997). Gap detection as a function of stimulus loudness for listeners with and without hearing loss. *Journal of Speech, Language and Hearing Research*, *40(6)*, 1387-1394.

Nilsson, M., Soli, S. D., Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*, 1085-1099.

Noble, W., Byrne, D. (1990). A comparison of different binaural hearing aid systems for sound localization in the horizontal and vertical planes. *British Journal of Audiology*, *24(5)*, 335-346.

Noreña, A. J. (2011). An integrative model of tinnitus based on a central gain controlling neural sensitivity. *Neuroscience & Biobehavioral Reviews*, *35(5)*, 1089-1109.

Oxenham, A., Bacon, S. (2003). Cochlear compression: perceptual measures and implications for normal and impaired hearing. *Ear and Hearing*, *24(5)*, 352-366.

Parsa, V. (2006). Acoustic feedback and its reduction through digital signal processing. *The Hearing Journal*, *59(11)*, 16-18.

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

Pickles, J. (2008). *An introduction to the physiology of hearing.* Emerald Group Publishing Limited, 3rd edition.

Plomp, R. (1983). Perception of speech as a modulated signal. In *Proceedings of the tenth international congress of phonetic sciences* (p. 29-40). Dordrecht, Foris.

Plomp, R., Steeneken, H. J. M. (1969). Effect of phase on the timbre of complex tones. *The Journal of the Acoustical Society of America*, *46*, 409-421.

Potter, R. K., Steinberg, J. C. (1950). Toward the specification of speech. *The Journal of the Acoustical Society of America*, *22*, 807-820.

Qiu, C., Salvi, R., Ding, D., Burkard, R. (2000). Inner hair cell loss leads to enhanced response amplitudes in auditory cortex of unanesthetized chinchillas: evidence for increased system gain. *Hearing Research*, *139(1-2)*, 153-171.

R Core Team. (2013). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org

Rahne, T., Böhme, L., Götze, G. (2011). Timbre discrimination in cochlear implant users and normal hearing subjects using cross-faded synthetic tones. *Journal of Neuroscience Methods*, *199(2)*, 290-295.

Robles, L., Ruggero, M., Rich, N. (1986). Basilar membrane mechanics at the base of the chinchilla cochlea. i. input-output functions, tuning curves, and response phases. *The Journal of the Acoustical Society of America*, *80*, 1364-1374.

Rose, J., Brugge, J., Anderson, D., Hind, J. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of Neurophysiology*, *30(4)*, 769-793.

Ruggero, M., Robles, L., Rich, N. (1992). Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression. *Journal of Neurophysiology*, *68(4)*, 1087-1099.

Saarikallio, S., Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music*, *35(1)*, 88-109.

Samson, S., Zatorre, R. J. (1994). Contribution of the right temporal lobe to musical timbre discrimination. *Neuropsychologia*, *32(2)*, 231-240.

Samson, S., Zatorre, R. J., Ramsay, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*, *125(3)*, 511-523.

Sarampalis, A., Kalluri, S., Edwards, B., Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language and Hearing Research*, *52(5)*, 1230.

Schellenberg, E. (2004). Music and cognitive abilities. *Current Directions in Psychological Science*, *14(6)*, 317-320.

Schneider, B. A., Hamstra, S. J. (1999). Gap detection thresholds as a function of tonal duration for younger and older listeners. *The Journal of the Acoustical Society of America*, *106*, 371.

Schwartz, O., Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, *4(8)*, 819-825.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., ... Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends in Amplification*, *9(4)*, 159-197.

Shailer, M. J., Moore, B. C. (1983). Gap detection as a function of frequency, bandwidth, and level. *The Journal of the Acoustical Society of America*, *74*, 467-473.

Shi, L., Carney, L., Doherty, K. (2006). Correction of the peripheral spatiotemporal response pattern: A potential new signal-processing strategy. *Journal of Speech, Language and Hearing Research*, *49(4)*, 848.

Simpson, A., Hersbach, A., McDermott, H. (2005). Improvements in speech perception with an experimental nonlinear frequency compression hearing device. *International Journal of Audiology*, *44(5)*, 281-292.

Smith, N. A., Trainor, L. J., Shore, D. I. (2006). The development of temporal resolution: between-channel gap detection in infants and adults. *Journal of Speech, Language and Hearing Research*, *49(5)*, 1104-1113.

Tambs, K. (2004). Moderate effects of hearing loss on mental health and subjective well-being: results from the nord-trøndelag hearing loss study. *Psychosomatic*

Medicine, *66(5)*, 776-782.

Trautwein, P., Hofstetter, P., Wang, J., Salvi, R., Nostrant, A. (1996). Selective inner hair cell loss does not alter distortion product otoacoustic emissions. *Hearing Research*, *66(5)*, 71-82.

Trehub, S. E., Endman, M. W., Thorpe, L. A. (1990). Infants' perception of timbre: Classification of complex tones by spectral structure. *Journal of Experimental Child Psychology*, *49(2)*, 300-313.

Unnikrishnan, K. P., Venugopal, K. P. (1994). Alopex: A correlation-based learning algorithm for feedforward and recurrent neural networks. *Neural Computation*, *6(3)*, 469-490.

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., Wouters, J. (2006). Horizontal localization with bilateral hearing aids: without is better than with. *The Journal of the Acoustical Society of America*, *119*, 515-526.

Vickers, D., Moore, B., Baer, T. (2001). Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies. *The Journal of the Acoustical Society of America*, *110*, 1164-1175.

Walden, B. E., Grant, K. W., Cord, M. T. (2001). Effects of amplification and speechreading on consonant recognition by persons with impaired hearing. *Ear and Hearing*, *22(4)*, 333-341.

Wang, J., Powers, N., Hofstetter, P., Trautwein, P., Ding, D., Salvi, R. (1997). Effects of selective inner hair cell loss on auditory nerve fiber threshold, tuning and spontaneous and driven discharge rate. *Hearing Research*, *107(1-2)*, 67-82.

Weinstein, B., Ventry, I. (1982). Hearing impairment and social isolation in the elderly. *Journal of Speech, Language and Hearing Research*, *25(4)*, 593-599.

Wiener, F., Ross, D. (1946). The pressure distribution in the auditory canal in a progressive sound field. *The Journal of the Acoustical Society of America*, *18*, 401. Retrieved from http://link.aip.org/link/?JASMAN/18/401/1