Neural Correlates of Phonetic Learning in Adult Listeners with Cochlear Implants

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Sharon Elizabeth Miller

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Yang Zhang, Ph.D., Peggy B. Nelson, Ph.D.

August 2015

Acknowledgements

First and foremost, I am grateful to my primary advisor, Yang Zhang, for his unwavering support, wisdom, and patience. I consider it an extreme privilege to call Yang my mentor. To my co-advisor, Peggy Nelson, I offer sincere gratitude for the guidance and support you have provided me throughout my graduate studies.

I owe immeasurable thanks to my other committee members, Bert Schlauch and Andrew Oxenham, who offered valuable insights on my research, and whose generosity and support is deeply appreciated.

This dissertation would not have been possible without the technical assistance and support of Edward Carney. To my friend and colleague, Heather Kreft, thank you for all of your help with subject recruitment and unfailing support.

A special thanks to my labmates, Tess Koerner, Luodi Yu, Erin Diamond, and Liz Tusler-Meyer who have helped me in too many ways to count. Additional thanks to my colleagues and friends for their help along the way: Aparna Rao, Ben Munson, Yingjiu Nie, Jane Carlstrom, Sarah Angerman, Arlene Carney, Melanie Gregan, Liz Anderson, and Yu-Wen Chen.

I am deeply indebted to my research participants for their selfless dedication and commitment to improving outcomes for persons with hearing loss. I am also grateful for funding provided by The Capita Foundation and the Doctoral Dissertation Fellowship that made this work possible.

Finally, I am profoundly grateful for my family and dear friends who have never stopped believing in me and without whom this dissertation simply would not exist.

Dedication

This dissertation is dedicated to my parents for their endless love and support and for encouraging me to never stop asking questions.

Abstract

Speech perception is a product of an individual's linguistic experience. Postlingually deafened cochlear implant (CI) recipients, persons who acquired speech and language with normal acoustic hearing, need to learn to remap degraded electric inputs provided by the implant to previously learned language patterns. The mechanisms underlying the perceptual remapping and whether formal auditory training can promote phonetic learning in CI users remain unclear.

This dissertation used behavioral and auditory event-related potential (ERP) methods to examine phonetic learning of the difficult /ba/-/da/ and /wa/-/ja/ speech contrasts in adult CI recipients. Behavioral and neural measures were collected before and after high variability identification training. Behavioral experiments employed identification and discrimination tasks, and the ERP experiments used an oddball paradigm to elicit the mismatch negativity (MMN) response associated with preattentive phonetic categorization.

The results indicated substantial neural plasticity for phonetic learning in adult postlingually deafened CI listeners can be induced by high variability identification training. The training protocol significantly improved perception of naturally produced speech in postlingually deafened CI recipients, and listeners generalized their learning to unfamiliar talkers. Fine scale behavioral and neural measures suggest enhanced phonetic categorization skills supported the observed improvements in phonetic perception. These findings have potential clinical implications related to the aural rehabilitation process following receipt of a cochlear implant device.

Table of Contents

a. List of Tables

List of Figures

Chapter I: Introduction

I. Overview

Speech perception is a fundamental component of language acquisition and successful communication for people with or without hearing loss. It involves a perceptual mapping process from the acoustic speech signal from the speaker to the meaningful linguistic representations of phonemes, syllables, words, and sentences in the listener’s mind. This dissertation project uses both behavioral and electrophysiological measures to better understand how listening experience affects the neural coding of speech in adult CI patients. A deeper understanding of the cortical mechanisms underlying speech perception with a cochlear implant (CI) and how formal auditory training alters the neural coding of electric speech has important clinical implications for persons with severe-to-profound hearing loss. Linking changes in behavior with changes in the brain will allow for the development of improved rehabilitation protocols that exploit the mechanisms of cortical plasticity in CI patients.

The dissertation project consists of three studies. The first study examines the validity and reliability of the event-related potential (ERP) approach and signal processing techniques to remove CI-related artifacts for assessing the neural encoding of speech for listeners with CIs. The second study investigates whether perceptual learning of speech at the segmental level in CI patients can be induced by a training protocol designed to promote perceptual grouping of similar stimuli and phonetic categorization. The final study of this dissertation examines the fine scale behavioral and neural correlates underlying improvements in speech perception in CI listeners after formal auditory training. Together, these three studies will shed light on the neural substrates and challenges underlying language processing and speech learning with a CI for various users.

This chapter will provide an introduction to speech perception with a CI and how it differs from normal, acoustic hearing. Basic questions of how CI recipients adapt to the new electric inputs will be introduced and evidence in support of formal auditory training will be highlighted. Electrophysiological approaches to assessing speech perception in CI recipients will also be discussed. An overview outlining the hypotheses and design of three experiments in the project is presented at the end of this chapter.

II. Background

A. The cochlear implant device

Cochlear implants (CIs) are neural prostheses that provide hearing sensation via electrical stimulation of the auditory nerve. It is considered one of the groundbreaking biomedical achievements in the last three decades (NIH fact sheet; http://report.nih.gov/nihfactsheets/ViewFactSheet.aspx?csid=83). Roughly 188,000 patients around the globe have received a CI device. In the United States, over 41,000 adults and 26,000 children have been implanted (nih.gov). The candidacy criteria to receive a CI differ for children and adults. According to the U.S. Food and Drug Administration, children 12 months and older are eligible to receive an implant if they have bilateral, profound sensorineural hearing loss and limited benefit, or lack of progress in auditory skill development, with appropriate binaural hearing aids over a three to six month period. Adults with prelingual or postlingual, bilateral severe-to-profound hearing loss and limited sentence recognition who receive little benefit from traditional hearing aid amplification also meet the candidacy criteria to receive an implant (fda.gov). In order to better understand how the CI operates in these individuals, a brief description of how the auditory system encodes sound will be provided (See Pickles, 1988;Gelfand, 2004 for more details).

Encoding sound in the auditory system

In normal acoustic hearing, the outer, middle, and inner ear (Fig 1.1a) perform a series of transformations that convert acoustic pressure variations in the environment into neural signals. The outer ear collects sound waves that are transmitted down the ear canal, causing vibration of the tympanic membrane. The vibration of the tympanic membrane sets the ossicles into motion, transforming the acoustic pressure variations into mechanical energy. The mechanical vibrations are transformed into electrical impulses in the inner ear when the ossicles impinge against the oval window of the fluid-filled cochlea, setting up a traveling wave that displaces the basilar membrane and causes a shearing motion of hair cells along the basilar membrane. The deflection of a hair cell triggers a neural impulse that is conducted via the auditory nerve through subcortical processing stations to the brain.

Within the cochlea, the frequency of incoming sounds is coded by both place and temporal mechanisms. The place mechanism operates via the mechanical properties of the cochlea and the tonotopic organization of the basilar membrane. The basilar membrane is narrow and stiff at the base and wide and floppy at the apex. The stiffness gradient causes high frequency sounds to maximally displace the basal end of the cochlea, while low frequencies cause maximal vibration amplitudes at the apical end of the cochlea. The corresponding hair cells on the basilar membrane are tonotopically organized, with hair cells on the basal part of the membrane being most sensitive to high frequencies, and the hair cells on the apical end most sensitive to low frequencies. The spiral ganglion nerve cells adjacent to the hair cells preserve the tonotopic organization of the basilar membrane. Frequency is also encoded via a temporal mechanism in the cochlea due to the synchronous auditory nerve firing patterns to the period of an acoustic waveform. At frequencies below 4000-5000Hz, the timing of auditory nerve firing is more likely to occur at one phase of the sinusoid cycle than another, a property known as phase locking. Due to this phase locking, the inter-spike intervals of the firing pattern occur at integer multiples of the waveform period, and these time intervals can be used to derive frequency information about the evoking tone.

Damage to the hair cells of the inner ear means the auditory system can no longer transform acoustic energy into neural impulses, resulting in hearing loss. Damage contained to the outer hair cells is consistent with hearing thresholds of 60 dB HL or better, whereas more widespread destruction that involves the outer and inner hair cells is consistent with hearing thresholds poorer than 60 dB HL (Van Tasell, 1993). Hair cells can be damaged by disease, ototoxic drugs, noise exposure, and a variety of other etiologies. When hearing loss reaches severe-to-profound levels such that the listener cannot benefit from sound amplification, a CI device may be prescribed. The CI device is designed to bypass the damaged cochlea and electrically stimulate the auditory nerve fibers directly.

Description of the CI device

Electric hearing with a CI device markedly differs from normal acoustic hearing. While there are multiple manufacturers of CI devices, the basic external and internal surgically implanted components of the device are the roughly same (Fig. 1.2). The external components consist of a microphone (Fig. 1.2A), speech processor (Fig. 1.2A) and head coil with a radio frequency transmission link (Fig 1.2B). The internal components consist of a receiver (Fig 1.2C) and an electrode array (Fig. 1.2D).

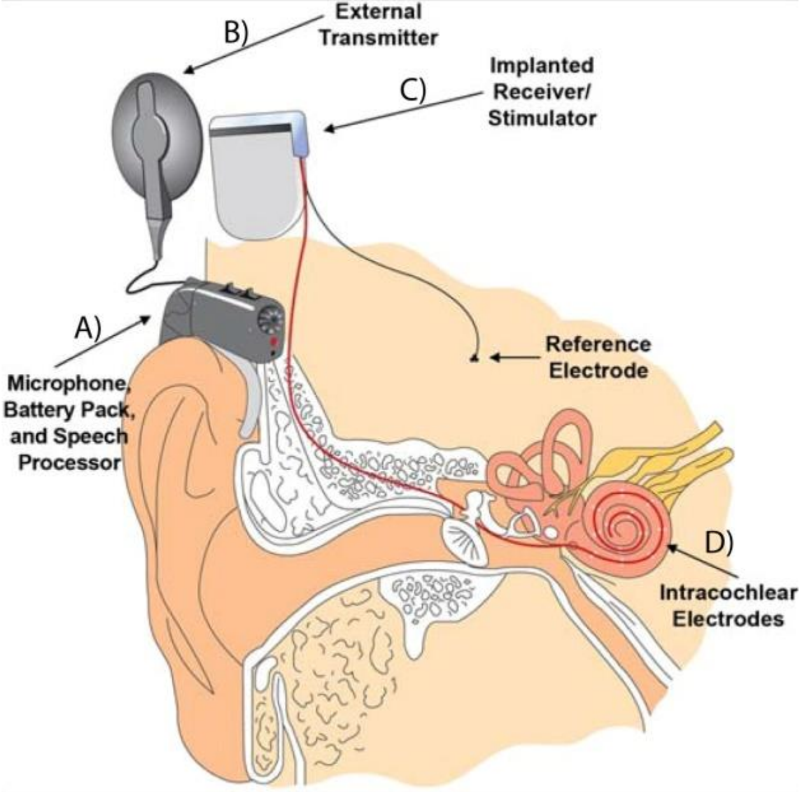


Fig. 1.1. Diagram of the external and internal components of the typical cochlear implant. A) Ear level microphone and speech processor. B) Head coil with radio frequency transmission link. C) Internal receiver. D) Electrode array. (Figure © MED-EL, reprinted with permission).

The external microphone collects acoustic signals from the environment and converts them to electric signals that are sent to the speech processor worn by the patient. The speech processor filters the incoming signal into multiple frequency bands and extracts the low frequency temporal envelope from each channel. Current pulses with amplitudes that are determined by the extracted envelope from each channel are then generated and transmitted to the internally implanted electrodes that correspond to the normal tonotopic distribution of the basilar membrane. Thus, the relative amplitudes of the current pulses delivered to each electrode reflect the spectral content of the incoming signal (see Loizou, 1998 for a review). The number and spacing of electrodes on an array varies by manufacturer and ranges anywhere from 8-22 channels. The CI device operates on the assumption that there are enough surviving spiral ganglion cells near the electrodes that, when stimulated, will fire and send signals to the brain.

The speech processor and electrode array of the CI are designed to mimic the function of a normal hearing cochlea; however, the place mechanism that operates in the normal functioning cochlea to encode frequency is severely limited by the CI device. The normal cochlea is sharply tuned where small changes in frequency cause large changes in the amplitude of vibration along the basilar membrane. In contrast, while the signal processor can filter incoming signals precisely, due to electrical current spread and the number of surviving spiral ganglion cells, the spatial selectivity of the CI is limited, severely reducing spectral resolution (Rubinstein, 2004). In addition, the speech processor of the CI only provides slow envelope cues and removes the temporal fine structure. Thus, the CI provides the brain with a highly impoverished and unnatural signal relative to the normal cochlea.

B. Speech perception with a cochlear implant

The speech signal

For speech to be intelligible with a CI device, the signal processing schemes have to encode information in the speech signal that is known to be perceptually important. Before considering speech perception with a CI, a brief review of the important acoustic information that contributes to consonant and vowel perception is necessary (See Borden, Harris, & Raphael, 2003; Johnson, 2003 for more details). The source-filter theory describes speech production as the output of a two stage process where a periodic or aperiodic sound source is generated and then independently filtered or shaped by the vocal tract (Fant, 1960; Flanagan, 1965). The frequency response of the vocal tract is determined by the position of the speech articulators during production. The speech signal that is ultimately produced can be divided into consonant and vowel categories that have distinct acoustic features.

The periodic vibration of the vocal folds serves as the source for all vowels. This periodic source is filtered through a relatively open vocal tract which produces a frequency response having a number of resonant peaks, or formants (e.g. F1, F2, and F3). The first two vowel formants (F1 and F2) are related to the position of the tongue and jaw during production. F1 is associated with vowel height or openness during production and F2 is associated with the front/back-ness of articulators. Access to the first two formant frequencies alone can produce accurate vowel discrimination of nearly 70% (Hillenbrand, Getty, Clark, & Wheeler, 1995; G. E. Peterson, 1959).

Consonants have either a periodic or aperiodic source that is filtered through a constricted vocal tract. Consonant production can be described by the place of articulation, manner of articulation, and voicing. Place of articulation describes where in the vocal tract the constriction occurs. Places of articulation include the following: bilabial, labiodental, dental, alveolar, palatoalveolar, retroflex, palatal, velar, and glottal. Manner of articulation describes the degree of constriction, the laterality of the tongue, and the airflow through nasal cavities. Manners of articulation include stops, nasals, fricatives, affricates, and approximants. Voicing refers to whether the consonant is produced with a voiced or voiceless source. Different places and manners of articulation have different acoustic realities and are typically classified based on spectral properties. Acoustic cues to place of articulation include, but are not limited to, the formant structure, direction of formant transitions into adjacent vowels, spectral burst, and frication spectrum. Different manners of articulation are cued by the presence of silence or frication noise, noise onset time, duration of formant transitions, and formant structure, among others.

Because CIs severely degrade the spectral content of speech signals, it is important to outline the temporal properties associated with vowel and consonant production as well. Rosen (1992) outlined an acoustic framework for describing the temporal properties of the speech signal that consists of the three following classifications: envelope, periodicity, and fine structure. The temporal envelope is defined as amplitude fluctuations occurring at rates of 2-50 Hz that provide segmental cues to manner of articulation, vowel identity, and the presence of voicing. Periodicity can be divided into subclasses of aperiodic and periodic sounds. Periodic sounds have amplitude variations that occur at rates of 50-500 Hz and aperiodic sounds have temporal fluctuations ranging from about 2000 Hz to 5000-10,000 Hz. Periodicity provides segmental cues to manner of articulation. Temporal fine structure cues refer to the wave shape within a single period or within a short interval of an aperiodic sound. Fine structure cues typically fluctuate at rates of 600-10,000 Hz and provide cues to place of articulation, vowel identity, voicing, and manner of articulation.

Acoustic limitations of CI signal processing

Current CI processing schemes limit the acoustic cues of speech that are available to the listener in a variety of ways. CIs provide only sparse spectral cues and are limited by the number of independent channels or electrodes. Cochlear implants typically provide only 8-22 frequency bands, but these bands are further limited by the lack of independence across channels due to electric current spread (Friesen, Shannon, Baskent, & Wang, 2001). In addition, the average insertion depth of the typical long electrode implant typically corresponds to an acoustic lower frequency limit of 1000 Hz. This mismatch means it is possible that a 500 Hz input, for example, could drive and electrode stimulating auditory nerves tuned to the 1000 Hz region. Finally, implants do not provide any temporal fine structure information because they convey mainly envelope fluctuations. The use of a fixed pulse rate further reduces any temporal fine structure information available to the CI user as well.

Despite acoustic limitations, CI listeners are capable of using the acoustic cues provided by the implant to understand speech (Dorman, Loizou, Spahr, & Maloff, 2002). The type of spectral and temporal information provided by the device explains some of the variability in speech understanding performance across users. CIs transmit mainly envelope cues, which as discussed previously, provide segmental information about the manner of articulation (Rosen, 1992). Accordingly, because the envelopes of different classes of consonants differ reliably, CI patients make relatively few errors on manner of articulation judgments (Dorman, et al., 2002). Using periodicity cues, CI listeners are usually able to determine whether a sound is voiced or voiceless as well (Dorman et al., 1990). Because temporal fine structure cues are not well transmitted and only course spectral cues are available through the implant, CI listeners struggle the most on vowel identification and consonant place of articulation judgments (Dorman, et al., 1990). For example, the /ba/-/da/ contrast differs in place of articulation and is cued by the frequency spectrum of the release burst of the initial consonant and the dynamic formant transitions to the following vowel (Hazan & Rosen, 1991). If the rapidly changing formants fall within the same acoustic filter band of the speech processor, the frequency change might not be present in the resultant pulses that are delivered to the implanted electrodes, limiting perception.

While certain phonetic contrasts are inherently more difficult based on implant signal processing, by using the available acoustic cues, some experienced CI recipients can achieve sentence recognition scores above 90% (Spahr & Dorman, 2004). However, when initially fit with an implant, speech perception performance can be quite low, and the time course of adaptation to electric hearing among individual CI users can vary dramatically (Krueger et al., 2008). Some have even found that most of the improvements in speech perception observed in adult CI users occur in the first 3-6 months of use, with minimal gains occurring after 6 months of being implanted, with some listeners obtaining very little benefit from the implant (Dorman & Loizou, 1997). That the adult CI user can adapt to the impoverished signals, and see improvements in speech understanding with time, suggests that cortical plasticity for remapping the degraded electric inputs to previously learned phonetic categories exists. However, some CI recipients receive little to no benefit from their devices, indicating passive adaptation to the CI device is not sufficient for some users. According to Moore and Shannon (2009), the variability in speech perception performance in CI users likely occurs because of three main factors: limitations in signal processing, differences in neural survival, and cortical plasticity. Research aimed at increasing spectral resolution by using tripolar stimulation (Bierer, 2007; Bierer, Bierer, & Middlebrooks, 2010; Bierer & Faulkner, 2010) and adding temporal fine structure cues to implant signal processors (Qi et al., 2012) are a few examples of device-related improvements that could potentially improve speech perception in future CI patients. For adult CI users currently struggling with his or her device, targeting central plasticity through active learning remains a viable option for improving speech understanding.

C. Formal auditory training in listeners with CIs

While of potential clinical benefit, the inclusion of formal auditory training in the adult CI rehabilitation process has been limited by the small number of controlled studies assessing efficacy and effectiveness of the treatment. The few training studies that have been completed, though, suggest that speech perception abilities in CI listeners are plastic and can be modified with training.

Fu, Galvin, Wang and Nogaki (2005) trained adult CI listeners to discriminate and identify naturally produced monosyllabic words and non-words and measured the effects of training on consonant and vowel recognition. Discrimination tasks presented listeners with three words, two of which were identical, and had the listener choose which one was different. Identification training involved presenting listeners with monosyllabic words and having them label the medial vowel or the initial or final consonant. After one month of training, vowel recognition significantly improved by 16% and consonant recognition by 13%. Results were not reported by individual phoneme, so it is unclear what speech sounds improved the most with training.

Stacey et al. (2010) also evaluated the effectiveness of word and sentence training strategies in CI users. A 2AFC procedure was used for the word training task where listeners were presented with naturally produced monosyllabic and multisyllabic words and asked to choose between two options. For the sentence training task, a sentence was presented acoustically followed by a screen with six key words. Listeners had to identify the three key words that had been presented in the sentence. Training was completed at home over the course of three weeks. Consonant recognition, vowel recognition, and sentence perception were measured before and after training, but significant improvements were only observed for consonant recognition. Again, consonant recognition scores were not reported by individual phoneme, so whether performance improvements were equivalent across phonemes is unclear.

Ingvalson, Lee, Fiebig, and Wong (2013) trained CI listeners to identify consonant and vowel stimuli using word, phrase, and sentence stimuli in the presence of background noise. Performance on the HINT sentence test and the QuickSIN was measured at pre-and posttest intervals after 4 sessions of training. After training, performance on the HINT test significantly improved at favorable SNRs and there was also a significant reduction in the SNR loss on the QuickSIN. Oba et al.(2011) also trained listeners in the presence of background noise but used nonspeech digit stimuli. Listeners were presented with three digits in a random order in the presence of speech babble and were asked to label the digits by clicking on the computer screen. Recognition of digits, HINT and IEEE sentences in speech babble and steady state noise was measured before and after training. The results suggest that digit recognition in babble significantly improved with training as did sentence recognition in steady noise for the two sentence tests.

As evidenced by previous studies, adult CI listeners have substantial cortical plasticity for different aspects of speech learning, even after many years of experience with their devices. However, the training methodologies differed dramatically across previous studies, so it is unclear what perceptual mechanisms were targeted by the training programs. A better understanding of the mechanisms supporting speech understanding improvements would lead to more effective rehabilitation protocols for individual CI recipients. Previous developmental (Kuhl, Conboy, Padden, Nelson, &

Pruitt, 2005; Kuhl et al., 2006; Tsao, Liu, & Kuhl, 2004) and cross-linguistic (Kuhl et al., 2008; Y. Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005) research suggests phonetic categorization skills are strongly correlated with later speech perception and production skills, making them an attractive target for training in CI users. Exclusively targeting phonetic categorization skills through training to improve speech perception in CI users has not been previously examined and warrants further study. The role of phonetic categorization in speech perception will be reviewed in the following section.

D. Speech perception and phonetic category acquisition

Speech perception is a highly complex task that involves the perceptual mapping of the variable spectral and temporal properties of the speech signal to internal mental representations of phonemes or linguistic units (Holt & Lotto, 2010). The phonetic category can be thought of as the basic perceptual unit of speech that has meaning and is distributed in acoustic space. For example, the phonetic category of /b/ will have a range of acceptable productions that are acceptable for category membership. Phonetic inventories differ across languages, and developmental research suggests that the mapping of the variable acoustic input to phonetic categories depends on native language experience in infancy (Kuhl, et al., 2008). At birth infants have the ability to discriminate any phonetic contrast (Kuhl, et al., 2006; Kuhl, Tsao, Liu, Zhang, & De Boer, 2001). However, by the age of 12 months, infants show an increased sensitivity to native contrasts and decreased ability to discriminate nonnative phonetic contrasts (Werker & Tees, 1984). It is thought that this process reflects a strong neural commitment to the statistical properties of the infant’s native language and a decrease in neural plasticity for learning other languages (Kuhl, et al., 2008). The warping of perceptual space that occurs with phonetic learning suggests speech perception is not based exclusively on the acoustic properties of speech, and developmental studies provide evidence that the perceptual parsing of the acoustic space into phonetic categories affects later language skills. For example, phonetic learning skills as early as 6 months of age are correlated with later speech comprehension and vocabulary and syntax skills (Kuhl, et al., 2005; Kuhl, et al., 2006; Tsao, et al., 2004). As the infant begins to map the perceptual space into phonetic categories, a competition between native and nonnative phonetic patterns emerges. Infants that display better native discrimination at 7 months of age perform better on later language assessments compared to infants that have better nonnative phonetic discrimination at the same age (Fenson et al., 2000). Whether the relationship between early language learning and phonetic perception emerges due to general auditory bottom-up processes (Benasich & Tallal, 2002; Holt & Lotto, 2010; Visto, Cranford, & Scudder, 1996) or due to higher order statistical, phonetic and phonotactic properties of one's native language (Kuhl, et al., 2005) is a subject of much debate. Regardless of the mechanism, phonetic perception plays an important role in later language learning.

Cross-linguistic behavioral studies also provide evidence that the perceptual warping associated with early phonetic category learning has long term effects on perception and ability to learn a second language. For example, the Japanese language does not use the /r/-/l/ contrast to distinguish meaning (Ladefoged & Maddieson, 1996) and adult Japanese listeners perform only at chance when discriminating between productions of American English /r/ and /l/, even though they have the ability to discriminate frequencies that cue the difference between the two sounds (Miyawaki et al., 1975) . While training can improve /r/-/l/ perception in adult Japanese listeners (Pisoni & Lively, 1995), performance usually remains below that of native speakers of American English speakers (Flege, Takagi, & Mann, 1995). It is hypothesized that neural commitments to the statistical properties of a first language creates perceptual interference and reduces the ability to learn a second language. For example, Iverson et al. (2003) documented that when identifying /r/ and /l/, Japanese speakers exhibit a hypersensitivity to an F2 cue that is unimportant for categorization of the contrast in English. The authors concluded that overreliance on an irrelevant F2 cue might interfere with phonetic learning of /r/-/l/ due to an inability to attend to the important F3 cue that distinguishes the two sounds. Thus, according to the authors, it is not that all /r/ and /l/ productions sound the same to Japanese listeners. Rather, they have formed category boundaries that are irrelevant for /r/-/l/ categorization. This finding implies that to successfully learn a nonnative contrast, it is likely that selective attention to previously ignored important acoustic-dimensions must be re-allocated.

Categorical Perception of Speech

The subtle perceptual sensitivities and warping that occurs during phonetic learning has been probed using the phenomenon of categorical perception (CP) of speech. Categorical perception of speech was first reported by Liberman, Harris, Hoffman, and Griffith (1957). Early speech perception experiments used continua varying in equivalent acoustic steps from one phoneme to another in order to look for acoustic cues that predicted perception. In one study, Liberman et al. (1957) created, a 14 step synthetic consonant-vowel continuum varying in equivalent F2 transition values at each step that had endpoints reliably labeled as /b/ and /g/. It was discovered that when asked to label stimuli 1 to 14 along the continuum, perception abruptly shifted from /b/ to /d/ around stimulus 4 and from /d/ to /g/ around stimulus 9. The points where these abrupt shifts occurred were considered the phonetic categorization boundaries. A second hallmark of categorical perception observed by the authors was that listeners displayed heightened sensitivity for discriminating pairs of stimuli from the continuum that crossed a phonetic boundary relative to stimulus pairs differing by the same acoustic amount but from the same phonetic category. Because listeners were only able to discriminate sounds they had labeled as different, the authors concluded phonemic categories influence the ability to make perceptual discriminations. While this strong conclusion that speech is special has been challenged by findings showing listeners can be trained to discriminate within phoneme category differences (Carney, 1977), the categorical perception paradigm can be used to examine how speech perception is shaped by native language experience and the location of phonetic boundaries for a given contrast in a listener.

Categorical perception paradigm and phonetic learning

The categorical perception paradigm has been used to examine the effects of native language experience and phonetic learning in a variety of populations. In a developmental study, Zlatin and Koenigsknecht (1976) examined categorical perception of synthetic ‘bees’-‘peas’ and ‘dime’-‘time’ continua in 2 year-olds, 6 year-olds and adults. The results indicated that by 2 years of age, children perceive speech sounds differing in voice onset time (VOT) categorically and have similar phoneme boundaries to 6 year-old children and adults. The youngest children did have the widest areas surrounding the phoneme boundaries, suggesting they needed larger acoustic differences in VOT to categorize the contrasts.

Cross-linguistic studies using identification and discrimination tasks also highlight the role of linguistic experience on categorical perception. Numerous studies with Japanese listeners have demonstrated that perception of the nonnative /r/-/l/ contrast is difficult and inaccurate and is characterized by poor labeling and chance discrimination of synthetic stimuli on categorical perception tasks (Miyawaki, et al., 1975; Yamada & Tohkura, 1992), especially for listeners with limited exposure to English (MacKain, Best, & Strange, 1981). However, with training, performance on identification and discrimination tasks has been shown to become more categorical (Y. Zhang et al., 2009). Thus while the theories underlying categorical perception differ, the CP paradigm can be used to assess the tuning of perceptual space.

E. High variability training and phonetic learning

Previous developmental and cross-linguistic studies provide strong evidence that native first language experience shapes phonetic categorization and later language learning. Some have theorized that the difficulties adults face when acquiring a nonnative contrast stem from the existence of a biologic 'critical period' for language acquisition whereby, after puberty, cortical plasticity for language learning is diminished (Lenneberg, 1967). However, more recent evidence suggests that changes in plasticity for phonetic learning are more linear (Flege, Munro, & MacKay, 1995), and that it is not equally hard to acquire all nonnative contrasts. Instead, nonnative contrasts that are the most similar to contrasts that exist in one's native language are more difficult and contrasts that are the most dissimilar to existing phonetic categories are easier to acquire (Best, 1994). Successful acquisition of a nonnative contrast is defined as efficiency and accuracy of categorization, transfer of learning to untrained stimuli and talkers, improved perception and production of the contrast, and stability of category learning over time (Pisoni & Lively, 1995; Y. Zhang & Wang, 2007). The degree and nature of brain plasticity for perceptual learning of a nonnative contrast in adulthood is highlighted by the success and failures of certain phonetic training protocols and the theories motivating the training tasks.

The mapping of the variable speech signal to internal phonetic categories is a complex task that must accommodate the acoustic realties of coarticulation and talker variability. Multiple theories exist as to how phonetic categorization and spoken word recognition occur, but they can be broadly classified into bottom-up, top-down, and interactive approaches. Bottom-up or analytic approaches to speech perception assume that perception of the continuous speech signal proceeds in a serial fashion where phonemes are recovered from the speech waveform and then parsed into words. In bottom-up approaches, phonetic analysis and phoneme identification occurs without any lexical feedback (Oden & Massaro, 1978). Top-down non-analytic approaches to speech perception assume phonetic analysis and categorization are influenced by long term memory and lexical knowledge. These theories assume perception depends on more than a singular abstract representation of a word or phoneme in the mental lexicon and proceeds when a word is first recognized and then broken into its constituent phonemes (see Pisoni & Levi, 2007 for a review) . Other hybrid models posit that speech perception is an interactive process based on bottom-up sensory analysis and stored, top-down lexical knowledge with feedforward and feedback processes (e.g. McClellend & Elman, 1986). Training protocols for learning nonnative phonetic contrasts are based on these theories of phoneme perception and, thus, typically employ tasks based on improving bottom-up or top-down skills. Different training programs have produced varying amount of gain in phonetic category acquisition. Pisoni and colleagues conducted a series of studies that will be reviewed in the following section that concluded top-down approaches that incorporate talker variability in the training tasks produce the most robust and stable nonnative phonetic categories.

The role of talker variability in speech perception was highlighted in early work by Pisoni and colleagues that demonstrated that intelligibility of spoken words in noise was faster and more accurate when a single talker generated the word lists as opposed to multiple talkers (Mullennix, Pisoni, & Martin, 1989). The greater intelligibility for a single talker suggested that acoustic and linguistic properties of speech were not perceived independently and led to further studies that documented that listeners encode detailed episodic information about a talker's voice (Palmeri, Goldinger, & Pisoni, 1993). Nygard, Sommers, and Pisoni (1994) later determined that talker familiarity facilitates recognition of novel words in noise relative to listeners trained with the same talkers but tested for generalization using unfamiliar talkers. Based on these results, Pisoni and colleagues suggested that the neural representation of words proceeds based on linguistic and structural acoustic properties (Pisoni & Lively, 1995). The relationship between talker-specific characteristics and word recognition suggests that speech perception is based on more than the existence of invariant abstract phoneme representations. This finding led to Pisoni and colleagues to examine the role of variability in the perceptual learning of nonnative phonetic contrasts and whether incorporating variability into training would facilitate acquisition of stable phonetic categories.

The experiments examining the role of stimulus variability in phonetic learning performed by Pisoni and colleagues were based on the hypothesis that to acquire a nonnative contrast, attention needs to be directed to the important critical cues for categorization. They predicted that by exposing listeners to stimulus variability and a wide array of possible exemplars, changes in selective attention to important categorization cues be would facilitated. In an initial study, Logan, Lively and Pisoni (1991) trained Japanese listeners to perceive the nonnative English /r/-/l/ contrast using five unique talkers. Listeners completed 15 days of two-alternative forced-choice identification training of naturally produced speech. The training materials consisted of naturally produced English words with /r/ and /l/ in different phonologic positions. Perception of minimal pairs of naturally produced words contrasted by /r/ and /l/ spoken by a familiar talker (a talker used during training) and an unfamiliar talker (a talker not used during training) was measured at pretest and posttest intervals. After training, performance significantly improved by similar degrees from pretest to posttest for both talkers and the greatest gains were seen for /r/-/l/ final phonetic environments. To compare these results with listeners trained without talker variability, Lively, Logan, and Pisoni (1993) replicated the identical experiment but used only one talker during training.

The generalization posttest to an unfamiliar talker highlighted the limitations of single-talker training. Unlike the previous study with multiple talkers used during training, listeners' mean accuracy for phoneme identification for the unfamiliar talker was unchanged from pretest to posttest intervals. The authors concluded that the use of multiple talkers during training maximizes the diversity of cues that listeners can use when acquiring a phonetic category which encourages generalization to new talkers (Lively, et al., 1993). To determine whether phonetic identification gains observed with high variability training were stable, Lively et al. (1994) replicated the original multiple talker training study but tested generalization of learning after 15 days of training and again 3 months after the study ended. Their results suggest that mean accuracy of /r/-/l/ identification significantly improved and that these gains remained stable three months after completing the training. These results suggest that incorporating talker variability in training improves long-term retention of phonetic learning.

In a more recent study, Iverson et al. (2005) investigated the effectiveness of four different training protocols for acquiring the nonnative /r/-/l/ contrast in Japanese listeners. The first training protocol was High Variability Phonetic Training (HVPT) (Logan, et al., 1991), where listeners were presented with natural productions of words spoken by multiple talkers with /r/ and /l/ in different phonetic environments. The second protocol used the Perceptual Fading technique (Jamieson & Morosan, 1989) where listeners were trained to attend to important acoustic cues for accurate categorization. Using this protocol, Japanese listeners were presented with the same stimuli used in the HVPT protocol, but at the beginning of training, the F3 difference that contrasts /r/-/l/

was exaggerated and then reduced at later training sessions. The third protocol was All Enhanced where the training stimuli always had an enhanced F3. Because Japanese listeners are hypersensitive to the unimportant F2 cue (Iverson, et al., 2003), the final training protocol was Secondary Cue Variability. In this training program, variability in the F2 cue was removed at the beginning of training so listeners could better attend to the important F3 cue. At later stages of training, variability in the F2 cue was added back to the training stimuli. Subjects were randomly assigned to one of the training protocols and were tested on identification of natural speech before and after 10 sessions of training. The results indicated that all four techniques produced highly significant gains in identification of natural speech, and there were no differences among techniques. As the signal processing needed to manipulate cues for the all the protocols except the HVPT was extensive, the authors concluded that using natural, unaltered speech is likely the best choice for training nonnative language acquisition.

F. Phonetic categorization with a cochlear implant

The postlingually deafened adult acquired phonetic categories and established boundaries between categories with normal hearing prior to receiving a CI device. In order to perceive speech, the postlingually deafened CI user must learn how the new degraded electrical patterns of activity provided by the implant map onto their previously learned phonetic categories. To date, only a few studies have examined phonetic categorization in CI recipients. Using a synthetic continuum of stimuli and the CP paradigm, Iverson (2003) documented that the locations of the identification boundaries and sensitivity peaks differ for CI listeners compared to listeners with NH for the /da/-

/ta/ contrast which is cued by differences in voice onset time. In adverse listening conditions, Munson and Nelson (2005) found listeners with CIs tended to have shallower identification function slopes and poorer endpoint accuracy than listeners with normal hearing for phonetic contrasts differing by a dynamic spectral cue. A longitudinal study of phonetic categorization by Lane et al. (2007) measured categorization of two speech contrasts immediately after receiving the CI and one year post implantation. Over the course of one year, identification slopes became steeper, suggesting more categorical-like perception, but goodness rating functions remained consistently poorer than normal hearing listeners. Taken together, adult CI users seem to display diminished phonetic categorization skills relative to listeners with normal hearing.

It is likely that device related variables that limit spectral resolution account for some of the diminished phonetic categorization skills observed in previous studies. However, recent evidence suggests that, similar to Japanese listeners acquiring a nonnative contrast, improper cue-weighting of available spectral cues could also explain the identification and discrimination results (Moberly et al., 2014; Winn & Litovsky, 2015). Moberly et al. (2014) examined spectral and amplitude cue weighting by adult postlingually deafened CI users relative to listeners with NH for the /ba/-/wa/ contrast and compared weighting strategies to word recognition and spectral and amplitude structure sensitivities. The results indicated that the listeners with CIs who used a spectral weighting strategy most similar to listeners with NH displayed superior word recognition performance relative to listeners who weighted the readily available amplitude cue more heavily. Sensitivity to spectral structure was somewhat related to use of the spectral weighting cue in CI listeners, but better sensitivity to spectral structure was not correlated with better word recognition performance. Sensitivity to amplitude structure cues did not predict usage of the cue. Thus, despite degradation by the CI device, the acoustic information necessary to identify the speech sounds was still accessible to the CI listeners, but some listeners were unable to use the available cues for proper categorization of the contrast. Whether high variability training known to promote acquisition of stable and robust nonnative contrasts can also promote phonetic learning and improve phonetic categorization in CI users is unknown and will be the focus of this dissertation.

G. Behavioral vs. Neurophysiologic Approach

Behavioral training studies of phonetic learning demonstrate that the central nervous system is plastic, or capable of change as a result of sensory experience in adulthood. Experience-dependent changes in behavior induced by training are typically monitored using the classical tests of accuracy and reaction time. This behavioral approach is based on the information-processing theory that assumes perception involves stages of processing (Atkinson & Shiffrin, 1971). The accuracy and time course between stimulus presentation and a subject's overt response is thought to reflect the complexity and efficiency of the mental information processing. With phonetic training, it is assumed reaction times would get faster and accuracy higher for a given categorization task. For example, Lively et al. (1994) documented that after training, not only did Japanese listeners' identification accuracy of /r/-/l/ significantly improve, but also reaction times to label the stimuli decreased by 600ms as well. The behavioral approach has yielded significant data regarding the effects of training on phonetic processing. However, behavioral data cannot inform us about the neural plasticity underlying observed changes in speech processing or how the functional neuroanatomy of language representation changes with training in the human brain. A better understanding of how speech percepts and improved phonetic categorization skills are represented in the central auditory system would lead to better rehabilitation protocols for CI recipients.

Neurophysiological and neuroimaging methods have recently emerged as powerful tools for monitoring the physiological changes that occur in the brain as a result of training, making it possible to examine the neural mechanisms underlying observed behavioral changes. Electroencephalography (EEG), Magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET) are the most commonly used methods for monitoring brain activities relates to language processing in the brain. Each method has strengths and weaknesses for studying linguistic processing due to the spatial-temporal resolution tradeoff across techniques. The fMRI and PET techniques, which measure hemodynamic blood flow and metabolic changes in brain activities, have superior spatial resolution and poor temporal resolution; EEG and MEG have exquisite temporal resolution along with modest spatial resolution. Because EEG and MEG have temporal resolution on the order of milliseconds and are noninvasive, they are useful tools for studying the neural processing of dynamic speech stimuli in a variety of patient populations.

The EEG technique uses electrodes placed on the scalp, and MEG uses super-conducting quantum interference devices (SQUIDs) located above the scalp to measure time varying electrical current and magnetic fields generated by postsynaptic cortical activity. Using an event-related paradigm, EEG (or MEG) responses time-locked to repeated stimulus presentations can be averaged to produce an auditory event-related potential (ERP) waveform (ERF for MEG). The ERP waveform (Fig. 1.3) consists of a series of positive and negative peaks described by amplitude and latency measures. The peak latency of ERP components reflects the neural travel time of a stimulus through the auditory pathway and the peak amplitude reflects the strength of neural processing of the physical or psychological properties of the stimulus.

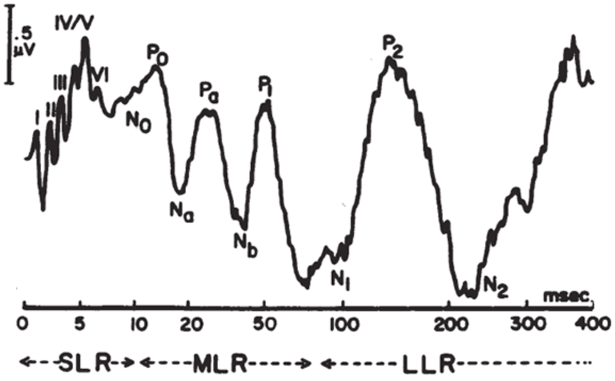


Figure 1.2. ERP waveform depicting the early (SLR), middle (MLR), and late (LLR) peak components. Positive polarity plotted up. (Figure from http://www.asha.org/policy/RP1987-00024/#r107).

ERP peak components are generally classified as early (first 10 ms), middle (10-80 ms), and late (80 ms and later) responses based on the response latency after stimulation. Generators of the early responses are thought to be the cochlea, auditory nerve, and low midbrain structures (Luck, 2005). The Auditory Brainstem Response (ABR) that is frequently used to assess hearing thresholds in clinical audiology practice is classified as an early response. The middle latency responses are thought have thalamocortical generators, and the late potentials originate in the auditory cortex. Because the late auditory ERP response can be elicited by a variety of auditory stimuli and paradigms and has been closely linked to perceptual processes such as detection and discrimination (see Martin et al., 1998 for a review) , it a potentially useful tool for assessing mechanisms of linguistic processing and phonetic learning in CI recipients.

The late cortical potentials can be broadly classified as exogenous, or obligatory, or endogenous responses. In adult listeners, auditory stimulation by short duration stimuli over repeated trials elicits a notable series of late obligatory peaks referred to as the P1-N1-P2 complex. The P1-N1-P2 complex can be passively recorded, meaning the listener does not have to attend or respond to the stimuli in order to measure the response. The P1-N1-P2 complex can be used to estimate behavioral thresholds (within 5-10 dB) (H. Davis, Hirsh, Shelnutt, & Bowers, 1967), but the ABR which is highly replicable, and can be recorded in sleeping and drowsy patients, is much more readily used clinically to evaluate the integrity of the auditory pathway (Stapells, Gravel, & Martin, 1995). However, obligatory ERP responses, including the P1-N1-P2 complex can be used to assess supra-threshold auditory skills, such as speech perception. When elicited in response to the onset of a sound, it provides information about the neural encoding of the acoustic properties of the signal that allows for behavioral detection. Previous research has determined that the peaks in the P1-N1-P2 complex reflect the neural encoding of the critical acoustic features that define consonant and vowel categories such as voice onset time (Digeser, Wohlberedt, & Hoppe, 2009; Sharma, Marsh, & Dorman, 2000; Zaehle, Jancke, & Meyer, 2007), place of articulation (Tavabi, Obleser, Dobel, & Pantev, 2007)

(Tavabi et al. 2007) and manner of articulation (Hari, 1991; Y. Zhang, et al., 2005). A brief description of the P1, N1, and P2 peaks that make up the complex are reviewed in the following section.

The P1 component is the first positive peak in the P1-N1-P2 complex and it typically occurs roughly 50ms after stimulus onset in adult listeners with NH. Neural generators of the auditory P1 are thought to include the primary auditory cortex, hippocampus, planum temporale, and lateral temporal regions (Huotilainen et al., 1998; Reite, Teale, Zimmerman, Davis, & Whalen, 1988). In adults, the P1 peak amplitude is usually relatively small, but this is the opposite trend in young children (Ceponiene, Rinne, & Naatanen, 2002). The N1, also known as N100, occurs around 100ms after stimulus onset or an abrupt change in the auditory signal and is thought to reflect stimulus encoding and the formation of that stimulus trace in sensory memory (Näätänen & Picton, 1987). Stimulus dependent changes in N1 in terms of amplitude, latency and cortical source estimation have been widely studied, suggesting that N1 reflects stimulus feature encoding (see Näätänen & Picton, 1987 for a review). Neural generators of the N1 include bilateral primary and secondary auditory cortex, thus, the N1 response is typically greatest when measured at the midline electrode Cz (Näätänen & Picton, 1987). P2 is not as well understood as N1, but it follows N1 and is elicited about 180ms after stimulus onset. P2 likely has multiple generators located in multiple auditory areas that include primary and secondary auditory cortices and the reticular activating system (Crowley & Colrain, 2004). P2 is thought to represent a stimulus classification process.

The P1-N1-P2 complex is typically elicited using short duration tones, clicks, or speech tokens, but it can also be elicited by longer duration stimuli. The P1-N1-P2 can also be elicited by an acoustic change within a longer duration stimulus, such as a consonant-vowel transition in a speech token (Hari, 1991; Ostroff, Martin, & Boothroyd, 1998). The resulting waveform elicited by these complex stimuli has a double peaked response with multiple P1-N1-P2 complexes. Some authors have termed this double peaked response the Acoustic Change Complex (ACC) (Ostroff, et al., 1998). The ACC can be elicited by a change in frequency, amplitude, or periodicity with a given stimulus. It is thought that the presence of the ACC indicates that the brain has detected a change within a speech sound, and when it is present, the neural capacity to discriminate the sounds exists (Martin, et al., 2008).

The Mismatch Negativity (MMN) is another type of late response recorded in the absence of attention and was discovered by Näätänen et al. (1978). The MMN is a frontocentral response with bilateral generators evoked using trains of frequent standard stimuli that are intermittently interrupted by infrequent deviant stimuli. The MMN is visualized as a negative displacement in the difference waveform obtained by subtracting the ERP response elicited by the deviant stimulus from that elicited by the standard. The MMN is generated when a standard stimulus builds up a sensory memory trace that is then violated by the deviant stimulus. The MMN peaks around 100-300ms after the onset of a discriminable change in an auditory speech or nonspeech stimulus. The presence of the MMN corresponds well to behavioral discrimination thresholds of changes in frequency, intensity, duration, stimulus pattern, and sound category (see Näätänen et al., 2007 for a review). Thus, the MMN can be said to be a marker of neural sensitivity.

As they relate to speech perception, the presence of the P1-N1-P2 complex suggests that a speech stimulus has been encoded at the level of the auditory cortex, the ACC indicates that a change within a speech stimulus has been detected, and the presence of the MMN indicates that a change within a train of stimuli has been processed at the pre-attentive level. These obligatory potentials can be recorded in the absence of sustained attention, and thus hold great potential benefit in assessing speech perception skills in clinical populations that cannot provide accurate behavioral responses. The importance of ERPs as a clinical tool for CI patients is highlighted by work done by Sharma and colleagues (Sharma, Dorman, Spahr, & Todd, 2002; Sharma, Dorman, & Spahr, 2002) who used ERPs to monitor neural changes to speech in pediatric CI recipients. The studies determined that the earlier the children were implanted, the more likely the P1 latency to speech was to be in the normal range and the better their speech and language outcomes were. This was vital, objective evidence that was used to advocate for lowering the acceptable age of implantation under the law. Likewise, ERPs have the potential to be used to monitor phonetic categorization skills in young CI recipients who cannot complete difficult CP tasks. The ability to objectively measure whether children are setting up appropriate phonetic boundaries would have important clinical and educational policy implications.

Phonetic categorization and the MMN

Neurophysiological research has produced numerous studies demonstrating that linguistic training can change central auditory processing. Much of the training work has focused on the MMN component of the ERP response to measure how experience modifies neuronal activities (see Näätänen et al., 2007 for a review). Specifically, the MMN response has been used to probe whether phonetic categorization reflects a learning-induced selective retuning of attention. In an MEG study, Näätänen et al. (1997) used the MMF to examine whether language-specific phoneme traces exist in the brain. In the study, native Finnish speakers were presented with standard strings of a native vowel interrupted by either another native vowel or a nonnative vowel that differed by F2 only. The MMF elicited by the contrastive vowel exemplar was significantly larger than the MMF evoked by the nonnative vowel, even though the acoustic difference was greater for the nonnative vowel. Source localization suggested that the MMF enhancement to the vowel originated in the left auditory cortex, and it was accompanied by bilateral auditory cortex activation to the acoustic change. Based on these results, Näätänen and colleagues suggested two parallel processes likely contribute to the mismatch response. The first component is a bilateral acoustic change detection process involving acoustic memory traces of the standards. The second component is a left hemisphere phoneme specific process based on native language experience. This theory is supported by additional studies that have found an enhanced MMN response to pairs of standard and deviant stimuli from distinct phoneme categories relative to the MMN evoked by a pair of stimuli differing by an acoustically equivalent amount but from the same phonetic category (Kraus, McGee, Carrell, & Sharma, 1995; Näätänen, et al., 1997;

Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000; Tremblay, Kraus, Carrell, & McGee, 1997; Y. Zhang, et al., 2005; Y. Zhang & Wang, 2007). These language specific MMN data suggest that native language experience shapes not only higher order categorization skills, but also lower level perceptual processes (Y. Zhang & Wang, 2007).

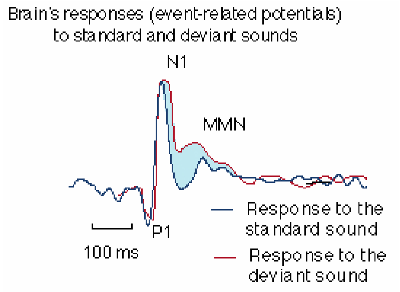


Figure 1.3. Example of the MMN response (shaded blue area) showing the ERP waveforms to the standard (blue) and deviant (red) stimuli. Negative polarity plotted up. (Figure from http://emcap.iua.upf.edu/BasicsMMN.html)

H. Phonetic learning and the MMN

The existence of language-specific memory traces suggests that linguistic experience shapes the neuronal architecture of the brain. Whether the cortex can be physically changed as a result of training has been the subject of multiple studies. Because the MMN is a marker of neural sensitivity**,** it has been extensively used to examine phonetic learning. Tremblay et al. (1997) trained native speakers of American English to discriminate a nonnative voice onset time (VOT) contrast that is phonemic to Hindi speakers, /ba/ syllables with either a -20ms or a -10ms VOT cue. After five days of training, behavioral discrimination of the contrast significantly improved, and behavioral improvement coincided with a significant increase in the duration and area of the mismatch response to the nonnative contrast as well. Source localization suggested that the enhanced MMN to the nonnative contrast was more pronounced in the left cortex.

In an MEG study, Zhang et al. (Y. Zhang, et al., 2009) collected neurophysiological responses before and after training native Japanese speakers to categorize the nonnative /r/-/l/ contrast. The training protocol used in the study was specifically designed to exploit the basic principles of infant-directed speech that are thought to support phonetic category acquisition (Y. Zhang, et al., 2005). Specifically, the training protocol incorporated adaptive exaggeration of the F3 cue that distinguishes /r/ and /l/, phonologic and talker variability, visible articulation cues, adaptive difficulty, and self-initiated selection. Measures of neural sensitivity using the MMF to the nonnative contrast were collected before and after training. Behavioral pre-posttests included phonetic identification and discrimination of synthetic and naturally produced /ra/-/la/ stimuli. After training, perception of naturally produced stimuli significantly increased and subjects generalized to unfamiliar talkers not used during training. Transfer of learning to unfamiliar talkers was associated with significant changes in identification of the synthetic stimuli near the phonetic boundary for /r/ and /l/ as well as significantly improved discrimination for an across phoneme category pair of stimuli. An enhanced MMF to an across phoneme category stimulus pair was observed in the left cortex after training as well. Thus, while the Japanese listeners did not obtain native native-like performance of the /r/-/l/ contrast, their perception of naturally produced stimuli significantly improved and this improvement was associated with behavioral and neural correlates of enhanced phonetic categorization. The results of the study provide strong evidence that enriched linguistic exposure can induce plasticity for phonetic learning.

Previous behavioral and neurophysiological training studies document that substantial cortical plasticity exists in adulthood for phonetic learning and that the mechanism of improvement appears to be improved phonetic categorization. This brain-behavior experimental paradigm has not been previously used to examine phonetic learning in postlingually deafened CI recipients.

I. Challenges in using the MMN paradigm in CI Users

Because of the presence of a magnet in the internal components of the CI device, fMRI and MEG are contraindicated for CI recipients, making EEG the safest and least invasive method to examine the cortical processing in these patients. Previous studies have collected EEG responses from CI recipients (Friesen & Tremblay, 2006; Kelly, Purdy, & Thorne, 2005; Kraus et al., 1993; Ponton et al., 1996; Ponton et al., 2000; Ponton, Moore, & Eggermont, 1999; Sharma, Dorman, & Spahr, 2002; F. Zhang et al., 2011); however, there are significant challenges when using ERPs to assess the brain mechanisms of speech perception in CI users.

When stimulated, the CI produces electrical artifacts that can contaminate the EEG signal recorded on the scalp (Gilley et al., 2006; Sharma, Dorman, Spahr, et al., 2002) (Fig 1.4).

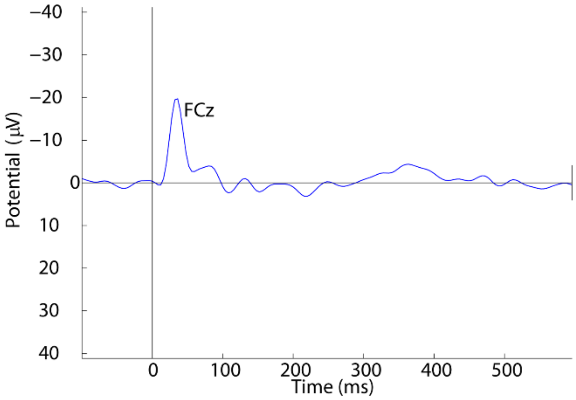


Figure 1.4. Averaged ERP waveform to a 1000Hz tone in an individual CI subject displaying large device-related artifact at the FCz electrode site. Negative is plotted up for the waveform.

Before using the MMN to examine phonetic learning in CI recipients, a reliable method for removing the artifact needs to be tested. Previous EEG studies have reported varying degree of artifact in their EEG data collected from CI recipients. Gilley et al. (2006) have described the typical CI related artifact in detail. CI related artifact usually consists of a positive pedestal that occurs with the stimulus onset, followed by a large negative overshoot and ringing in the amplifier filters. The speech processor can introduce a low level noise floor to the entire stimulus recording as well. The device related artifact is time-locked to the stimulus presentation, lasts for the duration of the stimulus, and can be 5-10 times larger than the average ERP components. If not dealt with properly, the artifact can mask the MMN, or worse, potentially be misinterpreted as a biologic response.

Previous EEG studies that evoked the MMN in CI recipients have dealt with the CI related artifact in different manners, but a lack of consensus as to what is considered artifact and the best way to remove it exists in the literature. In an early study, Kraus et al. (1993) examined the relationship between the MMN to a /da/-/ta/ speech contrast and speech perception abilities in adults with CIs who were classified as either ‘good’ or ‘poor’ CI users based on word recognition abilities. The authors reported no artifact in their EEG data, possibly because of the use of the MMN and subtraction technique; if the artifact was consistent for the standard and deviant stimuli it would be subtracted out of the data. The results indicated that the MMN was reliably recorded from the ‘good’ CI users and absent from the ‘poor’ user. The authors suggested that the MMN served as an objective measure of the neurophysiologic activity underlying the speech perception abilities in their group of CI recipients.

Lonka et al. (2004) also recorded the MMN in adult CI patients. The study tracked auditory discrimination skills in adult CI recipients immediately after receiving the CI and at 1 year and 3 year intervals after receiving the CI. At each time interval, the MMN to a vowel contrast was evoked in each CI listener as well. The authors reported that the EEG data were contaminated by large electrical artifacts for the first 5 months after receiving the CI. After 5 months, the CI artifact vanished from the EEG data recorded from these subjects. The study reported that as vowel discrimination improved, the amplitude to the MMN became larger as well, suggesting that the MMN reflects the cortical plasticity associated with improved phonetic processing.

In other EEG studies with CI users, Friesen and Tremblay (2006) successfully evoked the ACC to /si/-/shi/ stimuli and Brown et al. (2008) evoked the electrical ACC to changes in electrode stimulation. These studies used different filter settings to remove stimulus related artifacts. Friesen and Tremblay (2006) filtered the EEG data from 1.0-

20.0 Hz and Brown et al. (2008) used a bandpass filter from 1.0-1000 Hz. Sharma and colleagues utilize an optimized differential reference technique to remove artifact from their EEG data (Sharma, Dorman, Spahr, et al., 2002; Sharma, Dorman, & Spahr, 2002), but this approach only works for single channel recordings. When investigating phonetic processing and the effects of linguistic experience on cortical responses, the use of multiple channels to examine hemispheric effects is important.

A brief review of some of the EEG studies with CI recipients highlights the lack of uniformity in recording procedures and methods for removing the variable CI device-related artifact. Of importance, though, the studies demonstrate that the MMN can be used to assess phonetic processing in CI users and is an appropriate measure of neural sensitivity in CI users. However, before using the MMN paradigm to assess phonetic learning in CI recipients, it is vital to assess the validity of the EEG method and artifact removal technique in our own lab. The first study in this dissertation is aimed at assessing the validity and reliability of a more recent approach to removing CI artifact from EEG data. Newer evidence suggests that the Independent Component Analysis (ICA) method to statistically decompose the EEG activity recorded at the scalp into maximally independent sources (Delorme et al., 2011; Gilley, et al., 2006; Makeig, Debener, Onton, & Delorme, 2004) is a successful approach to separating artifact and noise from the EEG signal. Previous studies have used the ICA technique to separate CI device related artifact and biologic components of the EEG response (Gilley, et al., 2006; Viola et al., 2012; Viola, Thorne, Bleeck, Eyles, & Debener, 2011; F. Zhang, et al., 2011), but the methods have not been independently validated by an outside lab.

III. **Summary**

Adult postlingually deafened CI users display diminished phonetic categorization skills. Because previous research suggests that phonetic categorization skills are correlated with later language skills (Kuhl, et al., 2008; Kuhl, et al., 2005; Kuhl, et al., 2006; Tsao, et al., 2004; Y. Zhang, et al., 2005), it is important to determine whether formal training protocols known to aid in category acquisition (Pisoni & Lively, 1995) can induce plasticity for phonetic learning in CI listeners as well. The goal of this dissertation is to use both behavioral and electrophysiological measures to better understand the mechanisms underlying phonetic learning in CI recipients. The experiments included in this dissertation that aim to examine the cortical mechanisms underlying speech perception with a CI are as follows:

A. Experiment 1:

Before the MMN paradigm can be used to assess the neural correlates of phonetic learning in CI listeners, a method for removing CI device related artifact needs to be validated. The purpose of the first study in this dissertation was to validate our method of removing device artifact from our EEG data using the ICA method. To accomplish this aim, we collected EEG data from ten CI recipients and used manual and semi-automatic approaches to remove the device-related artifact from the ICA matrix. The reconstructed ERP waveforms using each artifact removal approach were then compared. To assess the reliability of our EEG method, the same CI subjects returned to the lab one year later, the experiment was repeated, and the ERP waveforms were compared across sessions. We hypothesized that the two approaches would yield similar results at each session, confirming the validity of the ICA method for dealing with CI artifact in EEG data.

B. Experiment 2:

Experiment 2 explored whether high variability identification training known to promote robust category formation could be exploited to improve perception of two phonetic contrasts in postlingually deafened CI users. To test this aim, a pretest-intervention-posttest paradigm was implemented. Nine CI recipients completed multiple sessions of high variability training in the laboratory over the course of two weeks. Perception of speech produced by familiar and unfamiliar talkers was measured before and after training in these listeners. A control group of untrained CI was also included who completed identical pre-and posttests over the same time course as the trainees. We hypothesized that perception of the two phonetic contrasts would improve for both familiar and unfamiliar talkers due to more abstract higher-order phoneme category learning.

C. Experiment 3:

Experiment 3 enrolled the same trained CI listeners that participated in Experiment 2. This experiment aimed to examine the underlying mechanisms of cortical plasticity that support phonetic learning by monitoring fine scale behavioral and neural correlates of phonetic categorization before and after the high variability training used in Experiment 2. To address these aims, behavioral and neural measures were collected before and after training. Behaviorally, phonetic identification and discrimination of the two speech contrasts were measured at pretest and posttest intervals. The electrophysiologic MMN response elicited by pairs of stimuli that crossed a phonetic boundary or were from the same phonetic category was measured before and after training in each subject as well. It was hypothesized that the training would improve phonetic perception in adult CI users due to enhanced phonetic categorization.

Chapter 2: Removing CI Artifact from EEG Data

Sections I-V are reprinted with permission from
Miller, S.E. and Zhang, Y. (2014). Validation of the Cochlear Implant Artifact
Correctiontool for auditory electrophysiology. *Neuroscience Letters*, 577, 51-55.

I. Introduction

Auditory event-related potentials (ERPs) are a valuable tool for hearing scientists as well as clinical audiologists to assess the neural encoding of speech and nonspeech sounds in different populations, including cochlear implant (CI) users (Martin, et al., 2008; Näätänen, 2003). Due to the delivery of the audio signal in the CI device via direct electric stimulation of the auditory nerve, other neurophysiological and neuroimaging tools such as functional Magnetic Resonance Imaging and magnetoencephalography are not feasible for the typical CI user (Pantev, Dinnesen, Ross, Wollbrink, & Knief, 2006). However, recording and analyzing ERP data from CI users faces a great challenge. The responses are usually contaminated by large device-related electrical artifacts upon auditory stimulation (Debener, Hine, Bleeck, & Eyles, 2008; Gilley, et al., 2006; Mc Laughlin, Lopez Valdes, Reilly, & Zeng, 2013; Viola, et al., 2012; Viola, et al., 2011). These artifacts are often 5-10 times greater than the cortical evoked responses and are time–locked to the stimulus presentation in each trial, which can obscure the cortical ERP components such as the N1-P2 complex (Gilley, et al., 2006) that are frequently used to examine the neural processing of speech sounds in terms of response latency, amplitude, topography, or cortical source localization.

Several methods have been developed to attenuate the CI-related electrical artifact in the EEG data, which require special stimulus choice or experimental setup. As the CI artifacts last for at least the duration of the evoking auditory stimulus, one approach has been to use very short stimuli (less than 50 ms). This approach could restrict the CI artifact to an early time window and minimize the contamination of later cortical responses of interest (Firszt, Chambers, Kraus, & Reeder, 2002), but it also precludes the use of more ecologically valid speech stimuli with longer durations. Others have implemented an optimized differential reference technique (Sharma, Dorman, Spahr, et al., 2002; Sharma, Dorman, & Spahr, 2002), where the reference electrode is placed on the scalp surface on a distant position from the recording electrode along the isopotential contour line of the artifact that passes the two points in space. While optimized reference placement has been shown to work well in single channel recordings, it is subject-dependent and cannot be used in multichannel studies (Gilley, et al., 2006). Another method to attenuate CI artifact takes advantage of an experimental design where the physical CI artifact remains constant for the trials of interest in two conditions and the analysis of cognitive responses requires the subtraction of the ERPs of the two conditions (Friesen & Picton, 2010). This method also has inherent limitations due to the experimental paradigm choice and the possible inequalities in the non-CI-related biological or environmental noises that are present in the trials.

A different approach without limitations on stimulus duration or experimental paradigm is to use independent component analysis (ICA). ICA statistically decomposes the EEG activity recorded at the scalp into maximally independent sources, and the resulting ICA matrix is a series of components that represent the underlying structure of the data. In theory, each component represents the activation of one independent contributing source to the averaged ERP, and thus artifactual components can be identified and linearly subtracted from the ICA matrix (Makeig, et al., 2004). The majority of publications that have used the ICA approach relied on manual inspection to remove components representing CI artifact from EEG data (Debener, et al., 2008; Gilley, et al., 2006; F. Zhang, Benson, & Fu, 2013; F. Zhang, et al., 2011). This manual process is laborious, subjective, and requires expert knowledge. Recently, Viola et al. (2012) developed the CI Artifact Correction (CIAC) algorithm which is an objective, semi-automatic approach to cluster and remove CI artifacts. The CIAC algorithm follows previous successful attempts to use ICA to remove CI artifacts from ERP data and apply source modeling to the reconstructed data (Debener, et al., 2008). The CIAC algorithm relies on both spatial and temporal properties of the independent components (ICs) at the subject group level. In their paper, Viola et al. (Viola, et al., 2012) validated the method by comparing the evoked responses reconstructed using a manual versus CIAC approach and found no significant differences in N1 or P2 latencies or N1-P2 peak-to-peak amplitudes across methods. As yet, there has been no independent validation of this CIAC approach from another lab. The purpose of the present study was to provide validation of the CIAC method using a unique EEG system and CI users with different devices. A secondary aim was to assess the stability of the evoked responses in CI users and to assess the reliability of both the CIAC algorithm and the manual approach by comparing ERPs from the same CI users at a second session, one year after the initial test.

II. Method

A. Participants

Ten (two males) right-handed, post-lingually deafened adult CI listeners, ranging in age from 45-75 years-old (Mean 58.9 years) participated in the study. All participants were native speakers of American English and reported no history of cognitive impairment. Six of the ten CI users were bilaterally implanted, and all CI users had at least 6 months experience with their devices. Table 2.1 displays the different subject and device profiles. Informed consent was obtained in compliance with the institutional Human Research Protection Program at the University of Minnesota.

Gender	Age (yrs)	CI use (yrs)	CI side	Etiology	Duration HL prior to implant (yrs)	Speech Processor	Speech Strategy
F	58.8	10.2	Right	Unknown	8	Harmony	HiRes-P with Fidelity120
F	61.2	3.7	Right*	Otoscerlosis	13	Harmony	HiRes-S with Fidelity120
F	53.3	8.6	Left*	Unknown	11	Harmony	HiRes-S with Fidelity120
M	45.6	15.8	Right	Maternal rubella	<1	Freedom	SPEAK
F	64.2	0.7	Left*	Familial; prog. SNHL	27	Harmony	HiRes-P with Fidelity120
M	75.3	22.8	Right*	Hereditary; prog. SNHL	4	ESPrit 3G	SPEAK
F	65.0	11	Left	Familial; prog. SNHL	7	Harmony	HiRes-S with Fidelity120
F	54.2	2.8	Right*	High fever	unknown	Harmony	HiRes-S with Fidelity120
F	56.0	1.3	Left	Unknown	<1	Harmony	HiRes-P with Fidelity120
F	55.6	2.2	Right*	Meniere's Disease	2.5	Harmony	HiRes-S with Fidelity120

Table 2.1: Subject and CI device characteristics. * Indicates bilateral CI user. The subjects completed two separate testing sessions, and the subject ages reflect age at the initial data collection session.

B. Experimental Protocol

Six of the ten subjects completed two testing sessions with approximately one year between sessions (mean 339.3 days between sessions). The remaining four subjects also completed two testing sessions, with four weeks between sessions. The identical experimental setup, stimuli, and procedures were used in each session

A 60 ms 1000 Hz pure tone stimulus with a 10 ms rise and fall time sampled at 44.1 kHz was used to evoke the ERPs. The stimulus was presented at 50 dB SL relative to the subject's threshold to the 1000 Hz tone. Participants were seated in a comfortable chair in an electrically and acoustically treated room (ETS-Lindgren Acoustic Systems). Stimuli were presented in a free sound field using EEVoke software (ANT Inc., the Netherlands) via bilateral loudspeakers (M-audio BX8a). The loudspeakers were placed at approximately 60 degree azimuth angle to each participant. The study utilized a passive listening design with 150 homogenous presentations of the stimulus. During the experiment, participants watched a muted movie of their choice on a 20-inch LCD TV located approximately 2.5 m from the listener. The interstimulus interval (offset-to-onset ISI) was randomized between 900-1000 ms.

C. EEG Data Acquisition and ERP Data Processing

For each session, continuous EEG activity was recorded using the Advanced Neuro Technology EEG system and a 64 channel Waveguard Cap (Rao, Zhang, & Miller, 2010). Participants wore only one implant during the recording sessions, and the electrodes on the cap located near the device were deactivated during data acquisition. Bilateral users selected their better ear for the experiment. The EEG data were bandpass filtered (0.016-200 Hz) and digitized using a sampling rate of 512 Hz. The Ag/AgCl electrodes on the cap were arranged in the standard 10-20 system with additional intermediate positions. The ground electrode was located at the AFz position. To reduce electrical artifact, the mastoid electrode contralateral to the user's CI was used as the reference electrode during recordings. The average electrode impedance was below 10 kOhm.

ERP averaging was performed offline with a common average reference using the EEGLAB toolbox (Delorme & Makeig, 2004) in MATLAB (Mathworks). The ERP epoch contained a 500 ms recording window and a 100 ms pre-stimulus baseline. Data were band pass filtered from 0.5 to 40 Hz using FIR filters (MATLAB) and then down-sampled to 500 Hz. A blind source separation ICA algorithm was applied to the data and equivalent current dipole modeling was performed on the resultant components using the DIPFIT plugin in EEGLAB (Delorme, et al., 2011). The initial number of components in the ICA matrix reflected the number of electrodes used during the recordings for each subject. Electrodes deactivated during the recordings were interpolated after performing the ICA and artifacts representing EOG and EMG activity were removed prior to averaging.

D. Removal of ICs reflecting CI artifact

Components representing CI artifact were removed from each subject's ICA matrix both manually and using the semi-automatic CIAC algorithm in separate analyses. Manual rejection of components reflecting CI artifacts was performed on each subject's data using criteria in accordance with Gilley et al. (Gilley, et al., 2006). Independent components were defined as CI artifacts and manually removed if 1) the activation occurred at the onset of the stimulus, 2) the activation occurred at the offset of the stimulus, 3) the duration of the activation was constant throughout the duration of the stimulus, and 4) scalp projections of the activation revealed centroid patterns on the side of the implant. Components were removed individually in a step-by-step fashion from the ICA matrix until only non-artifactual components remained. On average, fifteen components were manually removed from each subject's ICA matrix in each session (Table 2.2).

	CIAC Method		Manual Method	
Subject	Session 1	Session 2	Session 1	Session 2
F1	12	24	19	26
F2	12	13	21	18
F3	17	5	14	14
F4	13	12	15	17
F5	2	4	6	5
F6	13	7	13	11
F7	13	15	4	11
F8	7	7	6	12
M1	18	17	19	16
M2	9	8	25	24

Table 2.2: Number of components identified as artifact and removed from the ICA matrix for individual subjects across test sessions and methods.

Components reflecting CI artifact from the same original datasets were also identified using the semi-automatic CIAC algorithm. The CIAC algorithm uses both spatial and temporal criteria to separate cortical and artifactual components in the ICA matrix (see Viola et al. (Viola, et al., 2012) for complete description of the algorithm). Prior to running the CIAC algorithm, the user is required to first model equivalent dipoles to the ICs. Spatially, ICs with cortical sources will be dipolar and have a lower residual variance between the actual IC topography and the modeled projection for the equivalent dipole. Conversely, components representing CI artifacts will not be dipolar and will have higher residual variance. The CIAC algorithm requires the user to input a predefined threshold for the residual variance, and ICs having a residual variance greater than this value are defined as artifact. The CIAC algorithm also requires the user input the duration of the evoking auditory stimulus and to define the time window of the expected auditory ERP peaks (e.g. N1-P2). In the temporal domain, components reflecting CI artifact have the largest activations at the onset and offset of the stimulus, whereas ICs reflecting cortical activity have the greatest activations in the time window of the N1-P2 responses. Based on this pattern, the CIAC algorithm computes the ratio of the RMS amplitude of the temporal derivative of an IC in the artifact onset/offset window and the RMS amplitude for the time window where the cortical responses are expected. The IC with the largest ratio is then chosen as a template for that CI user and the topographical map of that IC is correlated with the other topographical maps of the other ICs for that user. ICs with a ratio larger than a pre-defined value or having a correlation with the template above a predefined value are defined as artifact. Using the recommended parameters, (Viola, et al., 2012), the present study utilized a threshold of residual variance of 20%, a threshold derivative of 2.5, and a threshold of correlation of 0.9. In the present study, the evoking stimulus was 60ms in duration, and the time window of the expected responses was 80-250 ms. On average, the CIAC algorithm removed eleven components from each subject’s ICA matrix across sessions (Table 2).

E. ERP and Statistical Analysis

ERP peak extraction was performed at the Cz electrode site. After removal of CI artifacts, peak amplitudes and latencies for the N1 and P2 components elicited by the 1000 Hz tone were extracted from the averaged waveforms of each subject from the two separate sessions. Based on the grand mean ERP waveforms, the following latency ranges were used in extracting the peaks elicited by the 1000 Hz tone: N1: 80-120 ms and P2: 175-300 ms.

To compare the ERP responses reconstructed using the manual and CIAC approaches and whether the responses were stable over time, a repeated-measures analysis-of-variance (ANOVA) was performed using the data from the six subjects who completed testing over the course of one year. The main effects of CI artifact removal strategy from the ICA matrix (manual vs. CIAC) and session (first vs. second) on peak amplitudes and latencies from the individual data were assessed. To assess the association between the two approaches, the coefficient of determination (R^2) was computed for the correlation between peak N1 and P2 latencies and amplitudes for the manual and CIAC methods. Test-retest reliability of each method was also assessed by computing R^2 for the correlation between peak N1 and P2 latencies and amplitudes across sessions.

To further determine the reliability of the methods and quantify the noise present within and across test sessions, the odd and even trials from ten individual subject data sets were analyzed separately using the manual and CIAC approaches. R^2 was computed for the correlations between peak N1 and P2 amplitudes within and across sessions for the odd-even split averaged data sets. Considering the recommended number of trials necessary for adult auditory N1 and P2 responses (Luck, 2005), data from the two separate test sessions were combined to generate the split-half ERP data for the odd- and even-numbered trials. Where applicable, Bonferroni or Greenhouse-Geisser corrections were applied to all reported *p* values.

III. Results

Clear N1-P2 components for the 1000 Hz tone were observed at electrode Cz for both the manual and CIAC approaches in both test sessions (see Fig.2.1 for grand average data, Fig. 2.2 for a representative individual subject).

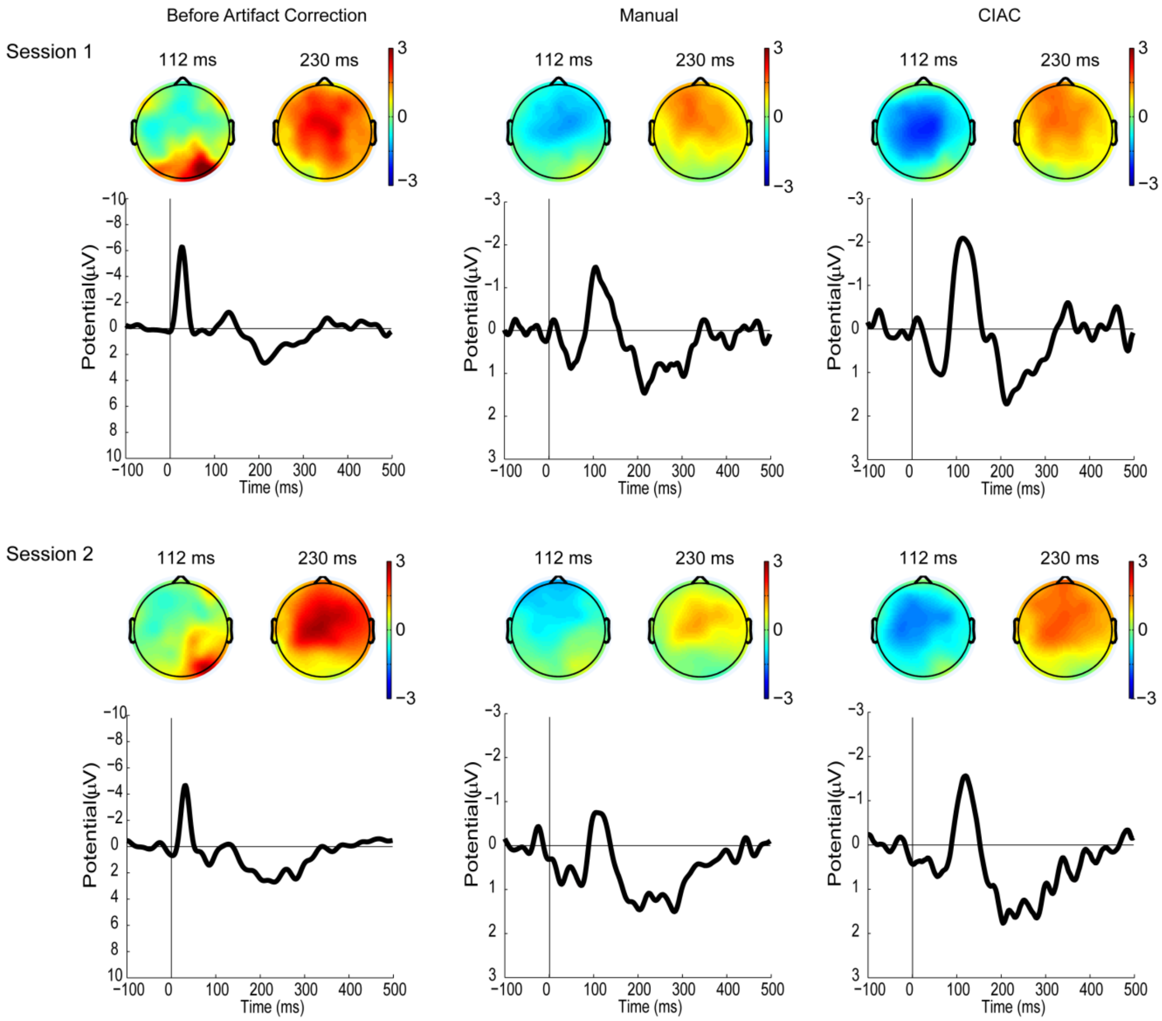


Figure 2.1 Grand mean averaged waveforms plotted without artifact correction and using the manual and CIAC algorithm at the Cz electrode site. Common average reference is used and negative is plotted up for the waveforms. The scalp topographies at latencies corresponding to N1 and P2 are plotted above the averaged waveforms.

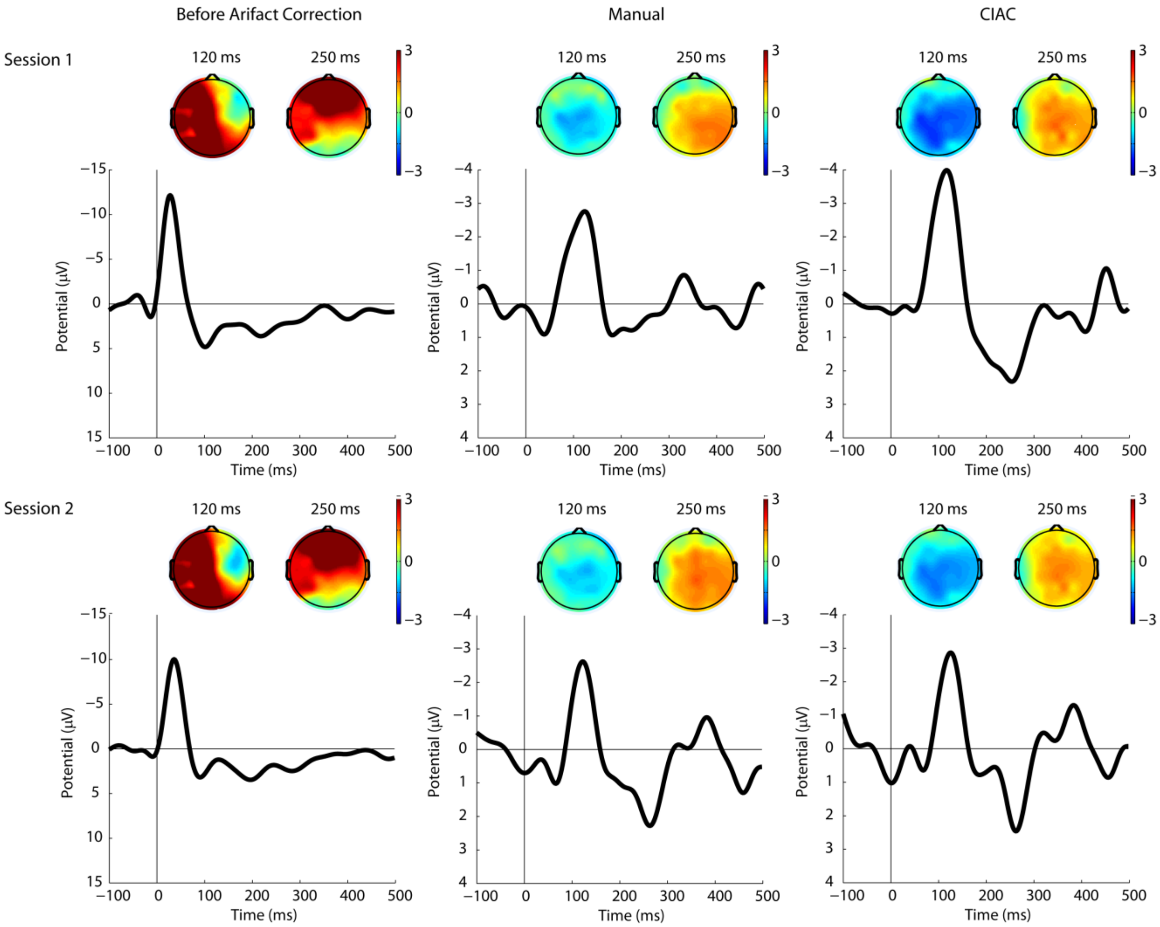
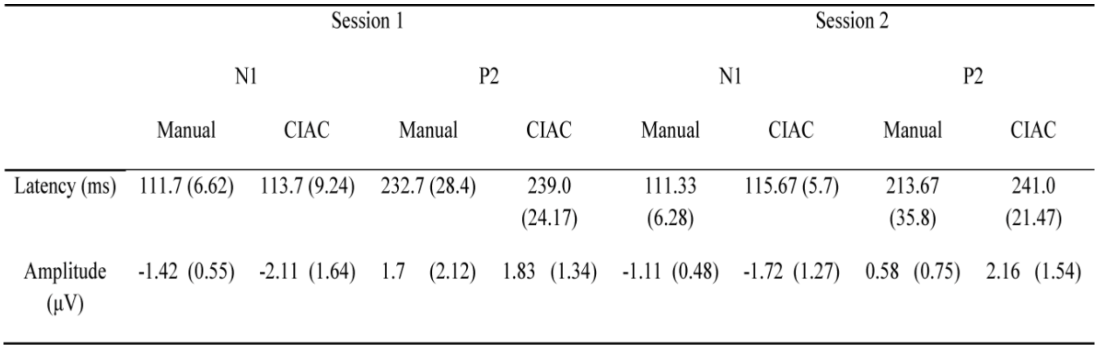


Figure 2.2 Averaged waveforms from an individual subject plotted without artifact correction and using manual removal and the CIAC algorithm at the Cz electrode site. Common average reference is used and negative is plotted up for the waveforms. Note the difference in scale on the vertical axis for the uncorrected waveforms.

The grand mean peak amplitude and latency values and standard deviations from both analyses are summarized in Table 2.3.



	Session 1				Session 2			
	N1		P2		N1		P2	
	Manual	CIAC	Manual	CIAC	Manual	CIAC	Manual	CIAC
Latency (ms)	111.7 (6.62)	113.7 (9.24)	232.7 (28.4)	239.0 (24.17)	111.33 (6.28)	115.67 (5.7)	213.67 (35.8)	241.0 (21.47)
Amplitude (μV)	-1.42 (0.55)	-2.11 (1.64)	1.7 (2.12)	1.83 (1.34)	-1.11 (0.48)	-1.72 (1.27)	0.58 (0.75)	2.16 (1.54)

Table 2.3: Mean peak latency and amplitude values of the individual subjects at electrode Cz. Standard deviations in parentheses.

Repeated-measures ANOVAs indicated no significant main effect of CI artifact removal strategy on N1 latencies (*F*(1,5) =2.4, *p*=0.18), N1 amplitudes (*F(*1,5)=2.8, p=0.15), P2 latencies (*F*(1,5) =5.0, *p*=0.8) and P2 amplitudes (*F*(1,5) = 3.04, *p*=0.14). These results suggest that the two artifact removal approaches yielded statistically equivalent results. The main effect of session was also not significant for N1 latencies (*F*(1,5) =0.17, *p*=0.7), N1 amplitudes (*F*(1,5) =2.31, *p*=0.19), P2 latencies (*F*(1,5) =1.16, *p*=0.33), or P2 amplitudes (*F*(1,5) =0.89, *p* = 0.39). These results indicate that the ERP peak responses were stable across over the course of approximately one year. There were no significant interactions between the main effects of removal strategy and session for N1 or P2 peak responses, suggesting the reconstructed ERP waveforms using either the CIAC or manual artifact removal strategy were relatively stable over the two sessions.

Peak N1 amplitudes (R^2 = 0.94, *p*=0.005), N1 latencies (R^2 = 0.71, *p*=0.02), and P2 amplitudes (R^2 = 0.84, *p*=0.002) were significantly correlated between the manual and

CIAC approaches. Peak P2 latencies were also correlated across methods, but with marginally significant results (R^2 = 0.61, p=0.06). These findings suggest that there is a high degree of association between the averaged ERP responses from manual and CIAC methods.

Test-retest reliability measurements indicated that for the CIAC method, N1 peak latencies (R^2 = 0.71, p=0.02) P2 peak latencies (R^2 = 0.77, p=0.009) and P2 peak amplitudes (R^2 = 0.89, p=0.0004) were significantly correlated across sessions. For the manual ICA rejection approach, N1 peak latencies (R^2 = 0.84, p=0.009) and N1 peak amplitudes (R^2 = 0.67, p=0.03) were significantly correlated across sessions. P2 amplitudes were also correlated by a marginally significant amount (R^2 = 0.6, p=0.06). These results suggest that both methods produce stable results over multiple test sessions.

Correlational analysis of the odd-even split average data showed significant effects for the N1 and P2 responses. For N1 peak amplitudes, the odd and even trials were significantly correlated within a test session for the CIAC method (R^2 = 0.66, p=0.04) and marginally correlated for the manual approach (R^2 = 0.56, p=0.09). For P2 peak amplitudes, the odd and even trials were significantly correlated for the CIAC (R^2 = 0.72, p=0.01) and manual (R^2 = 0.94, p=0.0003) approaches as well. These results indicate that both methods have a high degree of within session and across session reliability.

IV. Discussion

This study provides an independent validation of the CIAC algorithm using a different set of CI subjects and EEG equipment. Furthermore, our results suggest that N1-P2 peak responses remain stable in adult CI users over the course of one year and have a high degree of within session reliability when using either a manual or CIAC approach to remove artifact from EEG data. Examination of the topographical maps at the N1 and P2 latencies (Fig. 2.1) also suggests that the manual and CIAC methods produced similar scalp activations across sessions as well, further validating the use of the CIAC. Similar across-session stability of the N1-P2 complex responses was previously reported in normal hearing listeners and considered to be an important reliability indicator for potential clinical applications (Tremblay, Kraus, McGee, Ponton, & Otis, 2001).

The results suggest that EEG studies using CI subjects need to give serious thought on how to deal with the electrical artifacts produced by the devices. Our data provide evidence that EEG data can be corrupted by CI artifact, even when using a short-duration stimulus. We used a 60 ms 1000 Hz tone to examine N1-P2 peaks in CI users which occurred at 112 ms and 230 ms respectively. The onset of the CI artifact occurred at 26 ms which is much earlier than N1-P2, but examination of the topographical maps and waveforms (Fig. 2.1) clearly shows that N1 can be severely compromised by the preceding artifact. In theory, if the artifact lasted for only 60ms poststimulus, these later peaks should not be affected. But examination of the ERP waveforms without CI artifact removal shows that the assumption can be problematic. The CI artifact sustains over a time window beyond the cessation of the auditory stimulus.

To remove the large CI device-related artifact from our EEG data, we utilized ICA and both a manual and semi-automatic CIAC removal strategy. Our results indicated the peak N1 and P2 responses were significantly correlated across methods, indicating the two approaches yielded similar results. To assess the stability of the evoked responses and the reliability of the two methods, EEG data were collected in the same CI subjects using the identical evoking stimulus at two separate sessions, with approximately one year between sessions. Our results indicate that manual removal of the artifacts from the ICA matrix and use of the CIAC algorithm produced similar ERP waveform and topographical results across sessions, suggesting both are appropriate methods for analyzing ERP data in CI users. The split-averaged data analysis also suggests that both methods also have a similar degree of within session reliability as well.

While both the manual and CIAC approaches are valid methods for removing CI artifact from EEG data, they each have their own strengths and weaknesses. Of the two approaches, the CIAC is a more objective and reliable method for removing artifacts and eliminates the laborious manual identification of the components representing CI artifact. Examination of our individual ERP data indicated that the manual approach tended to be more conservative, with a greater number of components identified as artifact and removed from the ICA matrix compared to the CIAC approach in each subject. This difference likely resulted in the slightly greater peak N1 and P2 peak amplitudes (Table 3) found when using the CIAC approach; however, the differences across methods were not statistically significant.

V. Conclusion

EEG/ERP studies using CI subjects need to address how electrical artifact from the CI devices affects their data. The CIAC semi-automatic algorithm that uses spatial and temporal characteristics of ICs to remove CI artifact from EEG data is a viable and efficient method when using a short duration stimulus in a multichannel recording system. The stability of the N1-P2 responses in our data analysis across recording sessions with one-year time span indicates that CIAC is a reliable tool for scientific research and potential clinical applications.

Chapter 3: Phonetic identification training in postlingually deafened cochlear implant listeners

Miller, S.E. Zhang, Y., and Nelson, P.B. (2015). Efficacy of high-variability phonetic identification training in postlingually deafened cochlear implant listeners. *Journal of Speech Language Hearing Research,* submitted.

I. Introduction

Cochlear implants are neural prostheses that have the potential to improve speech perception in postlingually deafened adults, but significant variability in patient outcomes remains (see Shannon, 2002 for a review) . After implantation, postlingually deafened adults who acquired speech and language with normal acoustic hearing prior to the onset of deafness, need to learn how the neural activation patterns provided by electric hearing map onto previously learned phonemic language patterns (Boothroyd, 2010; Svirsky et al., 2001). The mechanisms that support this perceptual remapping and whether targeted auditory training can promote this process remain unclear. The present study investigated whether a form of high variability identification training, known to promote perceptual grouping of similar stimuli (Goldstone, 1998; Pisoni & Lively, 1995), enhanced phoneme perception in adult postlingually deafened cochlear implant (CI) users.

Previous research has documented that formal auditory training can improve consonant recognition (Fu, et al., 2005; Stacey, et al., 2010), vowel recognition (Dawson & Clark, 1997; Fu, et al., 2005), and sentence perception (Ingvalson, et al., 2013; Oba, et al., 2011) in adult cochlear implant users (see Fu and Galvin, 2007, Fu and Galvin, 2008, Ingvalson and Wong, 2013, or Henshaw and Fergusen, 2013 for reviews). These results are encouraging and have important clinical implications because they suggest even long-term CI users' speech perception abilities are plastic and can improve over time. The training materials and protocols varied dramatically across previous studies, but most studies trained at the word and sentence level (Fu, et al., 2005; Ingvalson, et al., 2013; Stacey, et al., 2010), with some also including a form of phonetic contrast training that encouraged listeners to attend to small acoustic differences, such as formant transitions and voice onset times across minimal pairs of monosyllabic words (Fu, et al., 2005; Fu & Galvin, 2007). The present study adopted a different, more linguistically simple training approach and investigated whether basic phonetic identification training alone can improve phoneme recognition in postlingually deafened adult CI users.

Language acquisition and cross-linguistic research provide evidence in support of training at the basic phonetic level. Developmental research has documented that better native phonetic discrimination at a young age is strongly correlated with later language skills (Kuhl, et al., 2005; Kuhl, et al., 2006; Tsao, et al., 2004). Cross-linguistic research also suggests that early phonetic learning plays a pivotal role in the ability to learn a nonnative contrasts later in life (Kuhl, et al., 2008; Y. Zhang, et al., 2005). Adults' success when acquiring a nonnative phonetic contrast typically remains below that of native speakers, but there is not a complete loss of perceptual sensitivity (McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Pisoni & Lively, 1995; Y. Zhang & Wang, 2007), and certain training methods are known to promote more stable and robust phonetic category acquisition. When learning a nonnative phonetic contrast, identification training that incorporates talker and phonologic context variability has been shown to produce the largest behavioral gains in adults, as evidenced by accuracy and efficiency of categorization, transfer of learning and long term retention (Pisoni & Lively, 1995). Unlike discrimination training where listeners might be responding to stimulus differences, identification training requires a listener to identify a single stimulus on every trial, forcing a higher normalization process towards category level response (Pisoni & Lively, 1995). For example, Strange and Dittmann (1984) found that being able to discriminate relevant acoustic dimensions of the /r/-/l/ contrast did not transfer to robust /r/-/l/ perception in Japanese listeners. Conversely, Lively et al. (1993) used high variability, multiple talker identification training to teach Japanese listeners to perceive the non-native /r/-/l/ contrast. Their results indicated that not only did identification of /r/-/l/ improve, but the listeners also generalized to unfamiliar talkers and phonologic contexts, indicating they had abstracted robust mental representations of the /r/-/l/ categories. Forcing category level responses during high variability identification training is thought to enhance attention to between category phonetic differences and reduce attention to within category stimulus level differences (Pisoni & Lively, 1995), thereby encouraging listeners to group perceptually similar stimuli into the same phonetic category.

The present study explored whether the attention weighting mechanisms known to promote robust category formation in developmental and cross linguistic studies could be exploited to improve perception of the /ba/-/da/ and /wa/-/ja/ contrasts in postlingually deafened CI users. Multiple talker phonetic identification training was employed and phoneme perception was measured before and after training. Pre-post phonetic testing included unfamiliar talkers (talkers not used during training) and familiar talkers (talkers used during training) to assess whether the phonetic identification training promoted robust category formation and generalization of learning. We hypothesized that perception of the two phonetic contrasts would improve for both familiar and unfamiliar talkers due to more abstract higher-order phoneme category learning. To verify the efficacy of the speech training paradigm, we also examined test-and-retest scores in a different group of CI users who did not receive the training.

II. Materials and methods

A. Subjects

Fourteen right-handed, postlingually deafened adult cochlear implant users participated in the training study (ages 45.6-75.3 years of age, mean 60.8 years). Nine of the listeners were in the experimental training group (mean age 58.2 years, 8 years of CI use) and the other five listeners were in a control group (mean age 65.5 years, 7 years of CI use). The untrained listeners, strictly speaking, should be considered a pseudo-control group due to the heterogeneity of subject and device characteristics and non-randomized assignment to the groups described below. All participants were native speakers of American English and reported no history of cognitive impairment. Nine of the fourteen CI users were bilaterally implanted, and all CI users had at least six months experience with their devices. Table 3.1 displays the different subject and device profiles. Informed consent was obtained in compliance with the institutional Human Research Protection Program at the University of Minnesota.

Sex	Age (yrs)	CI use (yrs)	CI side	Etiology	Duration HL prior to implant (yrs)	Speech Processor	Speech Strategy
F	58.8	10.2	Right	Unknown	8	Harmony	HiRes-P with Fidelity120
F	61.2	3.7	Right*	Otoscerlosis	13	Harmony	HiRes-S with Fidelity120
F	54.2	0.9	Right*	Progressive SNHL; Mondinis	27	Harmony	HiRes-S with Fidelity120
F	64.2	0.7	Left*	Familial; prog. SNHL	27	Harmony	HiRes-P with Fidelity120
F	65.0	11	Left	Familial; prog. SNHL	7	Harmony	HiRes-S with Fidelity120
F	54.2	2.8	Right*	High fever	Unknown	Harmony	HiRes-S with Fidelity120
F	45	3.0	Right	Measles	35	Harmony	HiRes-P with Fidelity120
M	45.6	15.8	Right	Maternal rubella	<1	Freedom	SPEAK
M	75.3	22.8	Right*	Hereditary; prog. SNHL	4	ESPrit 3G	SPEAK
F	53.3	8.6	Left*	Unknown	11	Harmony	HiRes-S with Fidelity120
M	75	5.1	Left*	Trauma	25	Harmony	HiRes-S with Fidelity120
M	68	7.0	Right*	Unknown	3	Harmony	HiRes-P with Fidelity120
F	56.0	1.3	Left	Unknown	<1	Harmony	HiRes-P with Fidelity120
F	75.0	12.5	Left*	Otosclerosis	22	ESPrit 3G	SPEAK

Table 3.1: Subject and CI device characteristics. * indicates bilateral CI user. The dotted line separates the trained (above) and control (below) listeners.

B. Experimental Design

The current study used a pretest-intervention-posttest design to assess the effects of multi-talker identification training on perception of the /ba/-/da/ and /wa/-/ja/ contrasts. The nine CI listeners in the training group completed pretest measures of phoneme perception, followed by two weeks of auditory training, followed by posttest measures of phoneme perception. The CI listeners in the pseudo-control group did not undergo training and completed the identical pretest and posttest measures of phoneme perception over the same time course as the trainees. Assignment to the experimental and pseudo-control groups was based on pretest phoneme perception scores and subject availability. Subjects with average pretest phoneme identification scores below 70% correct who were able to commit to multiple lab visits were enrolled in the training group. Two high performing subjects with phoneme identification scores near ceiling were enrolled in the pseudo-control group to assess procedural learning. Three additional low performing CI subjects who could not commit to the training protocol were also enrolled in the pseudo-control group in order to make comparisons across groups.

The pre-posttest and training sessions took place in the laboratory inside a double-walled sound-attenuated booth (ETS-Lindgren Acoustic Systems). The speech materials used in the pre-posttest and training sessions were recorded from eleven talkers (six males) into a Sennheiser high-fidelity microphone in a carpeted, double-walled sound booth (ETS-Lindgren Acoustic Systems) and digitally recorded to disk (44.1 kHz). All speech materials were equated for root mean square (RMS) intensity level (Sony Sound Forge) and were presented in the free field using E-prime (Psychology Software Tools, Inc) via bilateral loudspeakers (M-audio BX8a). The loudspeakers were placed at approximately 45 degree azimuth angle to each participant. The materials were presented at 50 dB SL relative to the subject's threshold to a 1000 Hz tone. The same presentation level was used for a listener's pre-posttest and training sessions.

C. Pretest and posttest stimuli and procedures

The pre-posttest sessions used naturally produced /ba/, /da/, /wa/, and /ja/ stimuli recorded from four native speakers of American English (two males, two females). The speech contrasts were chosen because they vary based on dynamic spectral cues which can be subject to misperception in adult CI users (Munson & Nelson, 2005). One of the female talkers was familiar to the trainees because the talker was also included in the training program. The other three talkers used in the pre-posttest sessions were classified as unfamiliar and not used during training in order to assess transfer of learning in the trained subjects.

Behavioral identification of the test stimuli was measured at pre-posttest intervals for all CI listeners. The forced choice identification tests presented listeners with ten trials of the /ba/, /da/, /wa/, and /ja/ stimuli from each of the four talkers (160 total stimulus presentations). Listeners indicated their responses by clicking on a screen with orthographic labels of the stimuli from a given contrast (‘ba’ or ‘da’; ‘wa’ or ‘ya’). All possible identification responses were taken into account and a bias-free estimate of perceptual sensitivity (d') was computed for each contrast (Macmillan & Creelman, 2004).

D. Training stimuli and protocol

The training stimuli consisted of naturally produced /ba/, /da/, /wa/, and /ja/ productions recorded from eight native speakers of American English (four males, four females). A custom, computer-based training program was designed, and subjects completed four two hour sessions of training over the course of two weeks in the laboratory. Unlike discrimination training which encourages listeners to attend to small, within category differences (Carney, 1977), identification training is more naturalistic and encourages listeners to attend to higher, more abstract category-level differences across stimuli (Pisoni & Lively, 1995). Training was self-directed and included a four alternative forced choice task. For each trial, listeners were presented with a screen displaying four icons representing the different speech tokens (/ba/, /da/, /wa/, and /ja/) along with a photographic facial image of the talker. Trainees were instructed to click on an iconic button to hear a stimulus presentation of their choice. After listening to the selected stimulus presentation, the next trial was initiated. Training was implemented in blocks, with each block consisting of 160 trials. To begin, only two unique talkers (one female and one male) were included in a training block. After completing a training block, trainees took a short identification quiz of 16 tokens from each talker (two productions of each syllable). Adaptive scaffolding was incorporated in the training (Y. Zhang, et al., 2009), and if quiz performance exceeded 90% correct, two additional talkers (one female and one male) were added to subsequent training blocks until eight talkers were included in a training block. Subjects repeated a given training block until achieving 90% correct on the quiz. Training ended once the listener obtained 90% correct phoneme identification of the eight talkers. Subjects completed the training sessions at their own pace, and every subject finished the training in four sessions (approximately eight hours in total).

E. Statistical analysis

Effects of *test session* (pretest and posttest) and *stimulus identity* (/ba/, /da/, /wa/, and /ja/) on percent correct identification were assessed using a repeated-measures analysis-of-variance (ANOVA) in SYSTAT (Version 10.2). The categorical factors of *group* (trained versus pseudo-control) and *talker* (Male 1, Male 2, Female 1, and Female 2) were included in the ANOVA model to examine training and talker intelligibility effects. Where applicable, Bonferroni or Greenhouse-Geisser corrections were applied to the reported *p* values. Post-hoc t-tests were also conducted to further understand how each factor of interest contributed to significant interaction effects in the ANOVA test.

III. Results

A. Effects of training

Percent correct identification of the /ba/-/da/ and /wa/-/ja/ contrasts from each of the four talkers in the pre-post test sessions was calculated for the trained and control subjects (Fig. 3.1). Consistent with previous speech perception studies on cochlear implant users, we found very large intersubject variability at pre- and post- tests for both speech contrasts (Fig. 3.3). Nevertheless, there were significant training-induced changes in the pre- and post- test results, which were not observed in the controls.

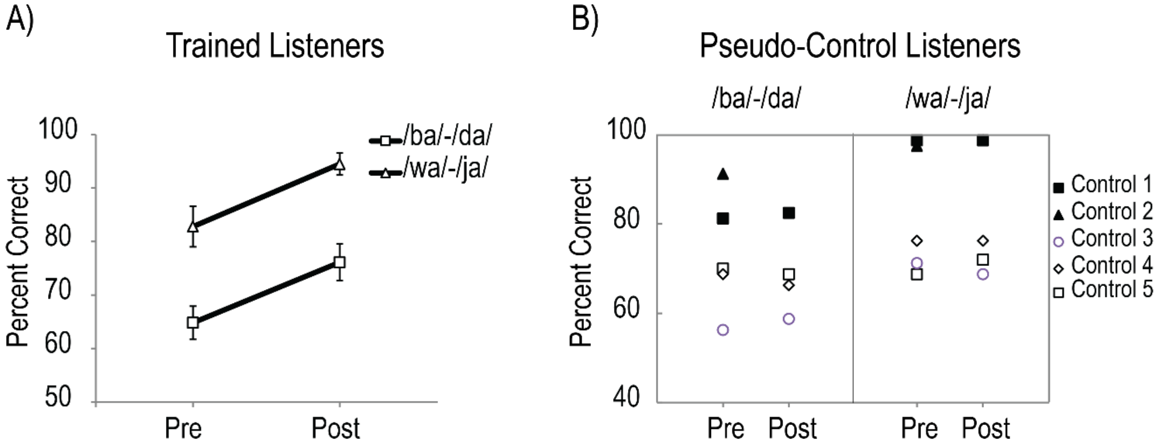


Figure 3.1 Average percent correct phoneme identification (error bars ± 1SE of the mean) of the /ba/-/da/ and /wa/-/ja/ contrasts across pre-posttest sessions for the A) trained listeners and B) pseudo-control listeners. The individual data are plotted for the pseudo-control listeners. Low performing controls, classified as < 70% correct on a pretest phoneme identification task, plotted with open symbols.

The repeated-measures ANOVA indicated a significant *group* x *test session* interaction ($F(1,51) = 3.97$, $p < 0.05$). Post hoc tests indicated that the multiple talker training program significantly increased average phoneme identification scores from the pretest to the posttest sessions in the trained listeners ($F(1,35) = 22.28$, $p < 0.01$) (Fig.3.1A), but not in the pseudo-control group ($F(1,19) = 0.57$, $p > 0.05$) (Fig. 3.1B). To test whether this lack of significance in the pseudo-control group was driven by the inclusion of high performing control listeners who had a small margin for improvement, a univariate post-hoc ANOVA that included only the low performing controls was performed. The results indicated that phoneme identification performance was not significantly different from pretest to posttest for the low performing controls ($F(1,11) = 2.16$, $p > 0.05$) (Fig 3.1B), suggesting the significant improvement observed in the trained group was not likely due to procedural learning.

A significant *test session* x *stimulus identity* interaction was observed (F(3,153) =4.4, p< 0.01), suggesting percent correct improvement from pre-to posttest was not equivalent for the phonemes. Post hoc analysis of the trained group indicated that significant improvements after training were confined to /ba/ ($t(35) = -2.4$, $p < 0.05$) and /wa/ ($t(35) = -3.62$, $p < 0.01$). Trainees improved their identification of /ja/, but the improvement was only marginally significant ($t(35) = -1.93$, $p = 0.06$). The between subjects factor of *talker* was not significant ($F(3,51) = 1.98$, $p > 0.05$), indicating that, on average, the four talkers used in the study were equally intelligible. However, there was a significant *talker* x *stimulus identity* interaction ($F(9,153) = 10.1$, $p < 0.01$), suggesting that intelligibility differed across the four stimuli for a given talker.

B. Transfer of learning

To determine whether the training related gains in phoneme identification were confined to the familiar talker (the talker used in the pre-post test and training sessions), or whether trainees generalized to unfamiliar talkers not used during training, a second repeated-measures ANOVA that included only the trained subjects' data was performed. The analysis included *talker familiarity* (familiar versus unfamiliar) as a categorical variable; the within subjects factors were identical to the initial ANOVA model. Where applicable, Bonferroni or Greenhouse-Geisser corrections were applied to the reported *p* values.

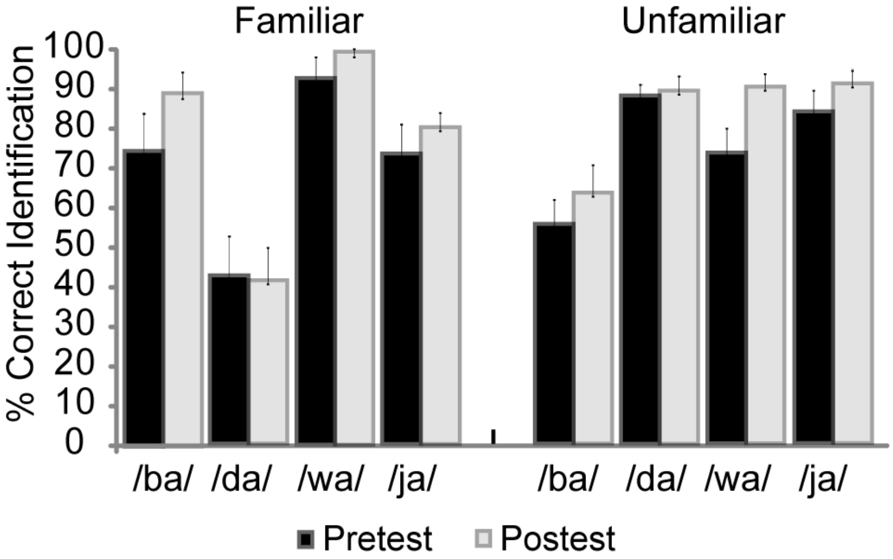


Figure 3.2 Pre-post identification scores for the speech sounds in the trained listeners sorted by familiar and unfamiliar talkers

Despite the existence of large intersubject variability, significant transfer of learning was found in the trainees (Fig. 3.2). The *talker familiarity* factor was not significant ($F(1,34) = 0.77$, $p > 0.05$), indicating that the observed training gains were not confined to the familiar talker alone. There was a significant *stimulus identity* x *talker familiarity* interaction ($F(3,102) = 18.18$, $p < 0.01$) indicating different amounts of identification improvement for a given syllable across familiar and unfamiliar talkers. Post hoc analysis of this significant interaction indicated that percent correct /wa/ identification improved significantly more for the unfamiliar talker than the familiar talker. Identification of /ba/, /da/, and /ja/ was not significantly different for either the unfamiliar or familiar talker.

IV. Discussion

The purpose of the present study was to examine whether multiple talker phonetic training can improve perception of the /ba/-/da/ and /wa/-/ja/ speech contrasts in postlingually deafened adult CI users. Our results indicated that perception of the contrasts significantly improved for both familiar and unfamiliar talkers by an average of 11.5%, consistent with more robust category formation in the trained CI listeners. The implications of our findings and comparison to previous results will be discussed.

A. Phonetic learning in CI users

Significant improvements in phonetic identification after training were observed for /ba/, /wa/, and /ja/ in the present study, but a significant degree of variability in the amount of learning across trainees existed (Fig. 3.3).

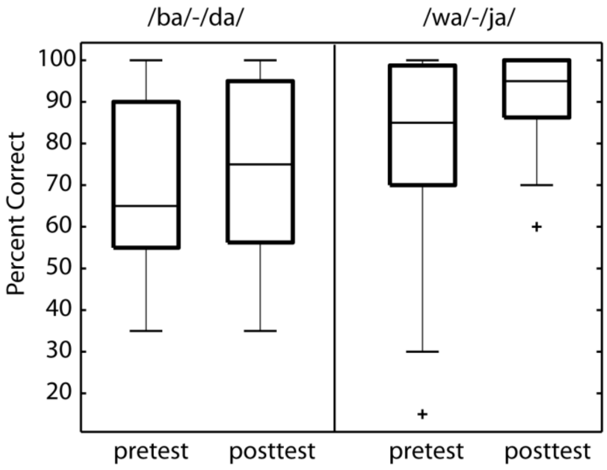


Figure 3.3 Box plot of the trained subjects' pre-and posttest identification scores for the two speech contrasts. The edges of the boxes represent the 25th and 75th percentiles of the distribution with the median denoted by the horizontal line within the box. The whiskers extend to the most extreme individual data points not considered outliers. + denotes outliers that are greater than 1.5x interquartile range.

The CI listeners were all postlingually deafened, meaning they had acquired language with a normal auditory system prior to implantation. In the present study, the trained listeners' average identification of the speech contrasts improved, but even after training, some subjects still experienced significant difficulty perceiving the speech contrasts in quiet. For example, for the the /ba/-/da/ contrast, changes in d' for the trainees ranged from 0.1 to 2.8, and for /wa/-/ja/ contrast, changes in d' ranged from 0 to 3.6. What limits the ability to relearn the speech sounds for some listeners remains unknown, but it is possible that device characteristics, stimulus properties, listener strategy, or a combination of these factors play a role.

A given speech sound has multiple acoustic cues that converge on the same phoneme or language pattern. The transformed electrical signal provided by the implant could give rise to redundant but degraded spectral and temporal cues present that no longer conform to the mental representation of the speech categories established via prior learning, which would limit performance. In the present study, before and after training, overall listener performance was superior for /wa/ compared to /ba/, with pretest scores ranging from 52.5% to 100% and posttest scores ranging from 75% to 100% for /wa/. Spectrally, /ba/ and /wa/ are similar and likely stimulate the same implant electrodes; however, the sounds differ in manner of articulation, with /wa/ having longer formant transition and amplitude rise times. It is possible that the difference in performance noted across /ba/ and /wa/ stimuli in the present study was due to CI device coding of the relationship between the important formant and amplitude cues across the two stimuli.

A secondary explanation for the limited learning noted with some CI subjects is related to listener strategy. It is possible that despite spectral degradation by the CI device, the acoustic information necessary to identify the speech sounds was still accessible to the CI listeners, but they were unable to use it for proper categorization. A recent study by Moberly et al. (2014) examined phonemic cue weighting of the /ba/-/wa/ contrast in adult postlingually deafened CI users and found that superior word recognition performance was related to the ability to use the same weighting cues as normal hearing listeners, independent of spectral discrimination abilities. Their results suggest that being able to discriminate acoustic cues is not enough to produce optimal word recognition, and instead attention to important cues is what matters. It is possible some of our less successful CI listeners had adopted non-ideal linguistic listening strategies that reduced learning. Similarly, some of our CI listeners' neural commitment to previously learned language patterns (Y. Zhang, et al., 2009; Y. Zhang, et al., 2005) might be too strong to be overcome with training. As the average age of our CI users was 60.8 years, the adult brain at such an advanced age might not have substantial plasticity to adequately deal with the degraded inputs.

Finally, it is possible that a combination of input characteristics and listener strategy limited phonetic learning in the current study. It is well established that normal hearing listeners benefit from talker familiarity during word recognition (Bradlow, Nygaard, & Pisoni, 1999; Nygaard & Pisoni, 1998; Nygaard, et al., 1994). It is thought that NH listeners perform some form of talker normalization to deal with the acoustic variability present in spoken language, and speech perception is enhanced when listeners have previously been exposed to a talker because demands of talker normalization have been reduced (Pisoni & Lively, 1995). Cochlear implants provide limited spectral information and due to constant pulse rates, users rely mainly on temporal cues for pitch perception. It is possible that the degraded acoustic inputs provided by the implant limit perceptual normalization across talkers because of the way phoneme identity and talker specific information are coded by the device. We observed an interaction between talker and stimulus identity in the present study, suggesting that the CI users were sensitive to some form of talker specific characteristics, but the specific cues that contributed to this effect are unknown. Acoustic analysis of the stimuli from the different talkers in the present suggests that duration cues might be contributing to this effect, though, as the talkers with the shortest productions tended to be misperceived to a greater degree.

B. Integration with previous speech training results

Fu and colleagues (2005) previously documented that five weeks of intensive phonetic contrast training of monosyllabic words (targeting attention to medial vowels, for example) significantly improved overall consonant and vowel perception in adult CI users. The present study trained listeners at an even more linguistically basic phonetic level and measured perception of two speech contrasts before and after training. Our results indicated that identification improved by an average of 11.5% for each of the two contrasts. This degree of improvement is similar to that found by Fu et al. (2005) who documented a 13.5% improvement in consonant discrimination, but the authors did not report results by individual consonant, so it is unclear if they found similar improvements for the /b/-/d/ and /w/-/j/ contrasts. Stacey and Summerfield (2008) previously used noise-vocoded speech and trained normal hearing listeners using a synthetic phonetic discrimination task and found no significant improvements in consonant recognition. While it is difficult to compare our results with CI listeners to NH participants listening to CI simulated speech, it is possible that the greater behavioral gains we documented were due to our use of identification training with natural speech that promoted higher order category learning (Pisoni & Lively, 1995). It is important to note that the present study differed from previous work in the total amount of time spent on training. We trained listeners on two phonetic contrasts for approximately eight hours, whereas Stacey and Summerfield (2008) trained 11 phonetic contrasts for a shorter period of time (nine 20 minute sessions) and Fu et al. (2005) trained listeners for close to five weeks. It is possible that time spent on training and the total amount of contrasts trained led to the observed differences across studies.

C. Clinical Implications and Future Directions

The present study found phonetic identification training improved phoneme perception in experienced adult CI listeners. Even though there was extreme heterogeneity in the trained and control groups, these data add to the existing training literature and provide preliminary support for the inclusion of formal auditory training in the CI rehabilitation protocol. Feedback from several of the older CI subjects in our study suggested the self-directed training protocol was easy to follow and did not induce high levels of stress or frustration. The level of difficulty and degree of subject engagement are important factors to consider when designing a training protocol, and future clinical studies will need to examine the roles age and device experience play in adherence to a training program.

The present study examined only two phonetic contrasts differing in place of articulation, and future work should aim to include a variety of different contrasts. Further research is also needed to examine the limits of phonetic learning in cochlear implant users and how phonetic categorization is related to speech recognition at word and sentence levels. It remains unclear how CI users deal with the excessive talker variability in spoken language, and more studies are needed to examine the relationship between intelligibility and acoustic characteristics of different talkers and explain the inter-subject variability in terms of how individual CI users may perceptually weigh the relative importance of different acoustic cues for speech discrimination and categorization at the segmental level and beyond (Fogerty, 2011; Moberly et al., 2014; Winn & Litovsky, 2015). Future work should also examine developmental effects of phonetic learning in pediatric CI users as well as prelingually deafened adult CI users who developed language with abnormal auditory input. Finally, it will be important to determine how perceptual weighting schemes and language learning/remapping influence neural coding of speech in CI users.

V. Conclusions

The present study found that phonetic identification training with multiple talkers improved phonetic perception in postlingually deafened CI users and listeners generalized their learning to unfamiliar talkers. This pattern of results is consistent with enhanced phonemic categorization of the trained speech sounds. Significant individual variability in the training group warrants further study to examine the sources of this variability and to assess limits of phonetic learning in postlingually deafened CI users.

Chapter 4: Neural correlates of phonetic learning in cochlear implant listeners

Miller, S.E. Zhang, Y., and Nelson, P.B. (2015). Neural correlates of phonetic learning in postlingually deafened cochlear implant listeners. *Ear and Hearing,* submitted.

I. Introduction

Cochlear implants (CI) are neural prostheses for people with severe-to-profound hearing loss who cannot benefit from sound amplification. CIs provide sound awareness and can often improve speech perception in postlingually deafened adults, persons who lost their hearing after acquiring speech and language with normal hearing (see Shannon 2002 for a review). While CI technology has continued to improve, a great degree of variability in speech performance outcomes still exists among individual adult CI users. For the postlingually deafened adult, the best CI device would ideally be able to faithfully reproduce speech sound patterns they previously experienced with acoustic hearing. However, current CI signal processing extract only amplitude envelope cues and discard the temporal fine structure of speech. Thus, the postlingually deafened adult needs to adapt to these new acoustic cues and relearn how these spectrally degraded inputs map onto their previously learned phonemic categories. It remains unclear what perceptual and neurophysiological mechanisms support this phonetic categorization in adult postlingually deafened CI users.

Mapping the acoustic input to phonetic categories is specific to one's native language experience and is thought to reflect a strong neural commitment to the statistical properties of language learned in infancy (Kuhl et al. 2008). In this speech perception model, the perceptual space is warped categorically, whereby differences across phoneme categories (e.g. /b/ vs. /d/) become exaggerated and within phoneme category differences are perceptually diminished (e.g. two different productions of /d/) (Feldman, Griffiths, & Morgan, 2009; Iverson, et al., 2003; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). Evidence of this perceptual warping comes from tasks involving identification and discrimination of phonemes along continua that vary in equivalent physical steps from one phoneme to another (e.g. a continuum from /ba/ to /da/ that varies by changing the F2 formant transition by a discrete amount in each step). In these tasks, listeners are either asked to label a stimulus from the continuum or whether pairs of stimuli from the continuum are the same or different. One hallmark of the warping associated with categorical perception includes an abrupt change in perception from one phoneme to the other along the continuum, evidenced by a sigmoidal identification function with a steep slope, as opposed to a gradual change in perception across the different steps. The other hallmark of categorical perception is heightened discrimination of stimuli pairs that cross an identification boundary and poor discrimination for stimuli pairs that differ by the same acoustic amount but are on the same side of the identification boundary.

To date, only a few studies have examined phonetic categorization in CI users. Iverson (2003) investigated phonetic categorization of a synthetic /da/-/ta/ continuum in postlingually deafened CI users and found that CI users tended to have identification boundaries and sensitivity peaks at different points along the continuum than listeners with normal hearing. In addition, the identification boundary widths in the CI users were wider and had shallower slopes than listeners with normal hearing, which is expected for listening in adverse conditions. Munson and Nelson (2005) investigated phonetic identification in adult CI users of four synthetic speech continua in quiet and noise: /wa/-/ja/, /i/-/u/, /ra/-/la/, and /say/-/stay/. Overall, their results suggest that CI users tended to have shallower slopes of their identification functions and identified endpoint stimuli less accurately than listeners with normal hearing. This pattern of results was exaggerated for the condition of listening in noise, especially for contrasts differing by a dynamic spectral cue. Lane et al. (2007) examined phoneme category structure longitudinally in adult CI users and measured phonetic identification and discrimination of two speech continua shortly after receiving the CI and one year post implantation. The results suggest that identification slopes became steeper with time, indicating more categorical-like perception, but goodness rating functions were poorer than normal hearing listeners soon after receiving the implant and did not improve over the course of one year. Taken together, previous behavioral results suggest that adult CI users seem to display diminished phonetic categorization relative to normal hearing peers, even after years of experience with their devices.

Accurate phonetic categorization is important to consider in CI users. In particular, it is of clinical importance to examine whether formal intervention can improve phonetic categorization and speech perception outcomes in persons using CIs. Both developmental (Kuhl, et al., 2005; Kuhl, et al., 2006; Tsao, et al., 2004) and cross-linguistic (Kuhl, et al., 2008; Y. Zhang, et al., 2005) research suggests that phonetic categorization skills are correlated with later speech perception and production skills. Prior cross-linguistic studies have shown strong evidence that phonetic categorization can be improved by formal training, which is associated with enhanced neural discriminatory sensitivity to speech sounds. For example, high variability phoneme identification training can improve the accuracy and efficiency of phonetic categorization of the non-native /r/-/l/ contrast in Japanese speakers (Lively, et al., 1993) and this phonetic learning is retained over time (Lively, et al., 1994). High variability identification training requires a listener to label a stimulus on every trial and incorporates talker and phonologic context variability in the training protocol. It is thought that forcing category level responses during identification training aids in the perceptual warping process that enhances attention to between category phonetic differences and reduces attention to within category stimulus level differences, which can be more effective than training protocols using a discrimination task (Pisoni & Lively, 1995). It remains to be tested whether this type of formal intensive identification training aimed at improving phonetic categorization would work for adult CI users.

Previous research with adult CI users has shown that formal auditory training can improve a variety of speech perception measures (see Fu and Galvin, 2007, Fu and Galvin, 2008 or Henshaw and Fergusen, 2013 for reviews) including consonant recognition (Fu, et al., 2005; Stacey, et al., 2010), vowel recognition (Dawson & Clark, 1997; Fu, et al., 2005), and sentence perception (Ingvalson, et al., 2013; Oba, et al., 2011) among others. These training results are clinically important because they suggest that there should be sufficient cortical plasticity available in this population for phonetic learning. But the mechanisms underlying the training-induced improvements are unknown, and it is also unclear whether they are related to improved phonetic categorization.

The notion that phonetic categorization reflects a learning-induced selective retuning of attention is supported by previous neurophysiologic studies of phonetic learning utilizing the cortical Mismatch Negativity (MMN) component of the auditory event-related potential (ERP) response. The MMN is a preattentive cortical response with a fronto-central topographical distribution that is elicited by presenting trains of standard stimuli that are intermittently interrupted by a deviant stimulus with a noticeable change (see Näätänen et al., 2007 for a review). The MMN waveform is manifested as an increased negativity to the deviant stimulus and is best seen by subtracting the response waveform to the standard from that of the deviant. Amplitude, latency, and duration measures can be quantified to denote the presence and strength of the MMN.

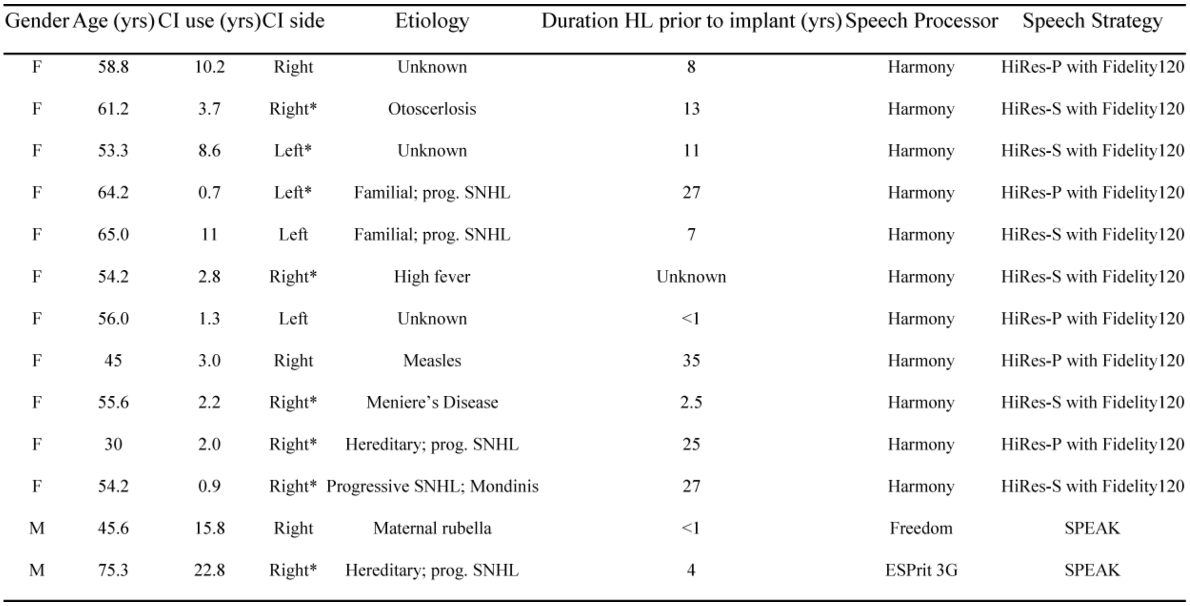
Prior studies of the MMN and phonetic categorization have found that the MMN response elicited by pairs of standard and deviant stimuli from distinct phoneme categories is enhanced relative to the MMN evoked by a pair of stimuli differing by an acoustically equivalent amount but from the same phonetic category (Kraus, et al., 1995; Näätänen, et al., 1997; Rivera-Gaxiola, et al., 2000; Tremblay, et al., 1997; Y. Zhang, et al., 2005; Y. Zhang & Wang, 2007). In a previous study of phonetic learning, Zhang et al. (2009) measured the mismatch field (MMF) in Japanese speakers acquiring the nonnative /r/ and /l/ categories before and after high variability identification training. The results suggested that even though the Japanese listeners did not achieve native-like performance, their perception of naturally produced /r/ and /l/ significantly improved, and this improvement was correlated with an enhanced mismatch response in the left hemisphere to the across category stimuli as well as steeper identification function slopes behaviorally. These brain and behavior data suggest that enhanced phonetic categorization likely supported the observed improvements in /r/-/l/ perception for the Japanese listeners in the study.

The present training study was motivated by the success of previous cross-linguistic studies. We aimed to examine the underlying mechanisms of brain plasticity related to phonetic learning by monitoring fine scale behavioral and neural correlates of phonetic categorization before and after high variability identification training in the same trained adult CI users that participated in Experiment 2. The central questions of the present study were twofold: 1) Whether formal identification training could improve phonetic categorization of two difficult speech contrasts in adult postlingually deafened CI users, and 2) Whether improvements in phonetic perception after training would be correlated with neural markers associated with phonetic learning. To answer these questions, behavioral phonetic identification and discrimination of the contrasts were measured before and after training in each CI user. Likewise, the electrophysiologic MMN response elicited by an across phoneme category pair and a within phoneme category pair were examined before and after training in each subject as well. We hypothesized that high variability identification training would improve phonetic identification of speech in adult CI users and that observed improvements would be related to enhanced phonetic categorization. Enhanced phonetic categorization might also be associated with the development of a sharper phonetic boundary to a contrast and a larger MMN to the across phoneme category pair relative to an acoustically equivalent within phoneme category pair after training. Our experimental design essentially followed that of a previous phonetic training study (Zhang et al., 2009). To test the robustness of training effects, we tested the learning of two sets of speech contrasts in the adult CI users. To verify transfer of learning, we used well-controlled synthetic speech and non-speech stimuli for the pre- and post- tests and the test material were not part of the training experience. To investigate domain specificity of phonetic learning, we used control stimuli of non-speech sounds that tracked the critical formant transition patterns in the speech contrasts.

II. Materials and Methods

A. Subjects

Thirteen right-handed, postlingually deafened adult cochlear implant users (11 females) participated in the study (ages 30.1-75.3 years, mean 55.3 years). Nine of the thirteen listeners were in the experimental group and received training and the other four listeners were high performing subjects that did not undergo training. The high performing listeners that did not receive training were included to assess effects of procedural learning and whether test-retest ERP responses would show changes in the absence of training. All participants were native speakers of American English and reported no history of cognitive impairment. Eight of the thirteen CI users were bilaterally implanted, and all listeners had at least six months experience with their devices. Table 4.1 displays the different device and subject profiles. Informed consent was obtained in compliance with the institutional Human Research Protection Program at the University of Minnesota.



Gender	Age (yrs)	CI use (yrs)	CI side	Etiology	Duration HL prior to implant (yrs)	Speech Processor	Speech Strategy
F	58.8	10.2	Right	Unknown	8	Harmony	HiRes-P with Fidelity120
F	61.2	3.7	Right*	Otoscerlosis	13	Harmony	HiRes-S with Fidelity120
F	53.3	8.6	Left*	Unknown	11	Harmony	HiRes-S with Fidelity120
F	64.2	0.7	Left*	Familial; prog. SNHL	27	Harmony	HiRes-P with Fidelity120
F	65.0	11	Left	Familial; prog. SNHL	7	Harmony	HiRes-S with Fidelity120
F	54.2	2.8	Right*	High fever	Unknown	Harmony	HiRes-S with Fidelity120
F	56.0	1.3	Left	Unknown	<1	Harmony	HiRes-P with Fidelity120
F	45	3.0	Right	Measles	35	Harmony	HiRes-P with Fidelity120
F	55.6	2.2	Right*	Meniere's Disease	2.5	Harmony	HiRes-S with Fidelity120
F	30	2.0	Right*	Hereditary; prog. SNHL	25	Harmony	HiRes-P with Fidelity120
F	54.2	0.9	Right*	Progressive SNHL; Mondinis	27	Harmony	HiRes-S with Fidelity120
M	45.6	15.8	Right	Maternal rubella	<1	Freedom	SPEAK
M	75.3	22.8	Right*	Hereditary; prog. SNHL	4	ESPrit 3G	SPEAK

Table 4.1: Subject and CI device characteristics. * Indicates bilateral CI user.

B. Training stimuli and protocol

The nine trainees completed the training. The training stimuli consisted of naturally produced /ba/, /da/, /wa/, and /ja/ recordings from eight native speakers of American English (four males). The training stimuli were recorded using a Sennheiser high-fidelity microphone in a sound treated booth (ETS-Lindgren Acoustic Systems) and digitally recorded to disk (44.1 kHz sample rate). All training materials were equated for root mean square (RMS) intensity level (Sony Sound Forge). The stimuli were presented in the free field via bilateral loudspeakers (M-audio BX8a) placed at approximately 45 degree azimuth angle to each listener. Stimuli were presented at 50 dB SL relative to the subject's threshold to a 1000 Hz tone. The same presentation level was used across training sessions for a given listener.

Subjects completed all training sessions in the laboratory, and the training package was implemented using E-Prime (Psychology Software Tools, Inc) on a Dell PC inside an acoustically shielded booth (ETS-Lindgren Acoustic Systems). The training program incorporated aspects of high variability identification training previously used in Zhang et al. (2009) that were shown to improve phonetic categorization. Training consisted of four two hour sessions, completed over the course of two weeks at the subject's own pace. Training was implemented in blocks, with each block consisting of 160 stimulus presentations. For each trial in a block, four icons representing /ba/, /da/, /wa/, and /ja/ and a photographic facial image of the talker (Fig.4.1) were displayed on a screen in front of the listener.

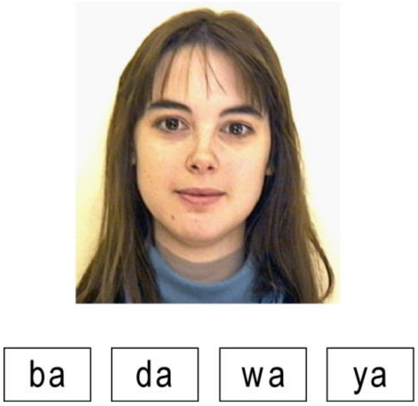


Fig. 4.1: Screenshot of the training program. Subjects clicked on one of the four icons representing /ba/, /da/, /wa/, and /ja/ and saw a photographic facial image of the talker on each training trial.

Training was self-directed and trainees were instructed to click on the iconic button of their choice to hear a stimulus presentation. Adaptive scaffolding was incorporated into the training by varying the number of unique talkers per block (Y. Zhang, et al., 2009). Training began with two talkers in a block (one female and one male), and after completing each block, trainees took an identification quiz. If listeners scored 90% correct or higher on the quiz, two additional talkers (one female and one male) were added to subsequent training blocks until eight talkers were included. Training ceased once the listener obtained 90% correct phoneme identification for the eight talkers.

C. Pre-post test stimuli

The behavioral and electrophysiologic pre-post tests utilized synthesized speech and nonspeech stimuli. The test speech stimuli consisted of two eleven step synthetic continua: /ba/-/da/ and /wa/-/ja/. The /ba/-/da/ and /wa/-/ja/ contrasts were chosen because they differ based on dynamic spectral cues which can be difficult for adult CI users to perceive (Munson & Nelson, 2005). The use of synthetic speech allowed us to create /ba/-/da/ and /wa/-/ja/ contrasts that differed in the length of the formant transition only (/wa/ and /ja/ have longer transitions than /ba/ and /da/), and enabled us to examine the role of temporal cues in phonetic perception. Using synthetic speech also ensured that the /ba/ and /wa/ and /da/ and /ja/ stimulated the same electrodes of a listener's implant. The nonspeech stimuli were sinewave correlates of the /ba/-/da/ and /wa/-/ja/ continua and were included to assess auditory versus phonetic processing. The /ba/-/da/ and /wa/-/ja/ speech continua were created using the HLsyn speech synthesizer (Sensimetrics, Inc, Massachusetts, MA). The nonspeech stimuli were created in MATLAB (Mathworks). The test stimuli were all equated for root mean square (RMS) intensity level and sampled at 22,050 Hz (Sony Sound Forge).

The first speech continuum was an 11 step /ba/-/da/ continuum. Each stimulus in the continuum had a duration of 170 ms. The initial formant transitions were 60 ms long and the steady-state vowel portion was 110 ms. The F0 was held constant at 110 Hz across stimuli. The steady-state /ɑ/ portion of the stimuli had an F1 value of 720 Hz, an

F2 value of 1240 Hz, and an F3 value of 2500 Hz. The F1 onset for the /ba/ endpoint was 400 Hz, the F2 was 900 Hz and the F3 was 2580 Hz. For the /da/ endpoint, the F1 onset was 400 Hz, F2 was 1700 Hz, and F3 was 2580 Hz. During the transition period of the stimulus from 0-60 ms, the formant values were interpolated from the onset values to the steady state formant values. Intermediate stimuli on the continuum were created by varying only the F2 onset value of each stimulus in equal linear steps from 900 to 1700 Hz; F1 and F3 values were identical for all stimuli in the continuum.

The second speech continuum was an 11 step /wa/-/ja/ continuum. Each stimulus on the continuum was 170 ms in length. The initial formant transitions were 80 ms long and the steady state vowel was 90 ms long. The F0 was held constant at 110 Hz across stimuli. The steady-state /a/ portion of the stimuli was identical to that used in the /ba/-/da/ continuum. The F1, F2, and F3 onsets for the /wa/ endpoint were identical to the /ba/ stimulus. For the /ja/ endpoint, the F1, F2, and F3 values were the same as the /da/ endpoint as well. During the transition period of the stimulus from 0-80 ms, the formant values were interpolated from the onset values to the steady state formant values. Identical to the /ba/-/da/ continuum, intermediate stimuli on the continuum were created by varying the F2 onset value of each stimulus in equal linear steps from 900 to 1700 Hz. F1 and F3 values were the same for all stimuli in the continuum.

The nonspeech test stimuli were sinewave analogs of the /ba/-/da/ and /wa/-/ja/ speech continua. The nonspeech stimuli were 170 ms in duration and consisted of three pure tone glides concatenated to three steady-state pure tones. The frequency values used in the nonspeech stimuli were identical to the formant transitions and steady state formants used in the synthesized /ba/, /da/, /wa/, and /ja/ speech stimuli described previously. The duration of the formant transitions and steady state portions of the nonspeech stimuli were identical to the speech continua as well. Intermediate stimuli on the nonspeech continua were created by varying the onset values of each stimulus in equal linear steps in the same manner as the speech stimuli.

D. Pre-post procedures

Pre-post behavioral procedures

The trained subjects completed the same pre-post behavioral tests under identical experimental conditions before and after training. The untrained listeners completed the pre-post behavioral tests over the same time course as the trained listeners. All pre-post test sessions took place inside a double-walled sound attenuated booth (ETS-Lindgren Acoustic Systems), and test stimuli were presented in the free field using E-prime (Psychology Software Tools, Inc) via bilateral loudspeakers (M-audio BX8a). Identical to the training procedure, the loudspeakers were placed at approximately 45 degree azimuth angle to each participant, and the stimuli were presented at 50 dB SL relative to the subject's threshold to a 1000 Hz tone. The presentation level was held constant for a listener's pretest and posttest sessions. Unlike the training sessions, no visual cues were provided.

The synthetic speech continua were tested using identification and AX discrimination tasks. The nonspeech continua were tested using only an AX discrimination task as pilot data suggested that the nonspeech stimuli were not speechlike and could not be labeled. In total, listeners completed eight behavioral test blocks (five discrimination and three identification tests). The order of tests was counterbalanced across listeners and the presentation order of the stimuli within a test block was randomized. Listeners completed practice sessions with the endpoint stimuli prior to beginning each identification or discrimination test.

The identification tests presented listeners with twenty trials of each stimulus along the speech continuum. Listeners indicated their responses by clicking on a screen with orthographic labels of the two endpoint stimuli from a given continuum (‘ba’ or ‘da’; ‘wa’ or ‘ya’). The discrimination tasks had subjects decide whether two stimuli from the continuum were the same or different. Responses were indicated by clicking on boxes with orthographic ‘same’ or ‘different’ labels. The pairs of stimuli were presented in both directions twenty times each with an ISI of 250 ms. False positive rates were assessed by presenting listeners with an equal number of control trials where the same stimulus from the continuum was presented twice.

Pre-post electrophysiological procedures

EEG data were collected from trainees before and after training. The untrained group completed the same pre-post tests over the same time course as the trained listeners. A passive listening double oddball paradigm (Näätänen, et al., 2007; Xi, Zhang, Shu, Zhang, & Li, 2010; Y. Zhang, et al., 2009) was used to evoke the MMN responses. The double oddball paradigm differs from the classic single oddball paradigm in that the MMN is elicited using two different deviant stimuli in the same block. By using the double oddball paradigm, we could compare the listener’s MMN response evoked by a within phoneme category deviant to that evoked by an across phoneme category deviant, allowing us to examine phoneme categorization within the same subject.

In the present study, MMN responses were elicited using pairs of stimuli from the /ba/-/da/ and /wa/-/ja/ speech and nonspeech continua. Both within and across phoneme category pairs with equivalent acoustic intervals from each continuum were selected to examine whether the MMN response could reflect phonetic learning effects. For the /ba/-/da/ speech and nonspeech analogs, stimuli pairs 5-9 and 1-5 were used as the across and within category pairs respectively. In the double oddball paradigm for the /ba/-/da/ speech and nonspeech analogs, stimulus 5 was the standard stimulus, stimulus 1 was the within category deviant, and stimulus 9 was the across category deviant. For the /wa/-/ja/ speech and nonspeech analogs, stimuli pairs 3-7 and 7-11 were used as the across and within category pairs. Stimulus 7 was the standard stimulus, stimulus 11 was the within category deviant, and stimulus 3 was the across category deviant. These pairs were selected based on pilot data that showed them to be consistently across or within category for a group of CI listeners. More importantly, using each CI listener's MAP, it was verified that the within and across category pairs of stimuli stimulated an equal number of implant electrodes for each subject enrolled in the study. Stimulus pairs were presented in four consecutive test blocks (/ba/-da/ speech, /ba/-/da/ nonspeech analogs, /wa/-/ja/ speech, and /wa/-/ja/ nonspeech analogs), and the test blocks were counterbalanced across listeners. In each test block, the standard to deviant ratio was 85:15. To control for the acoustic differences between the standard and deviant stimuli, in each test block 150 presentations of the deviant stimuli were also presented alone (Kraus, et al., 1993). The interstimulus interval was randomized between 900-1000 ms. Participants watched a muted movie of their choice on a 20-inch LCD TV located approximately 2.5 m from the listener and were instructed to ignore the stimuli during the recording session.

E. EEG Data Acquisition and ERP Data Analysis

Continuous EEG activity was recorded from each listener using the Advanced Neuro Technology EEG system and a 64 channel Waveguard Cap (Miller & Zhang, 2014; Rao, et al., 2010). The Ag/AgCl electrodes on the cap were arranged in the standard 10-20 system with additional intermediate positions, and the ground electrode was located at the AFz position. For each recording session, listeners wore only one implant and the electrodes on the cap located near the device were deactivated for data acquisition. Bilateral users were instructed to select their better ear for the experiment and the identical setup was used for both recording sessions. To reduce electrical artifact, the mastoid electrode contralateral to the user's CI was used as the reference electrode during recordings. The EEG data were bandpass-filtered from 0.016-200 Hz and digitized at 512 Hz. The average electrode impedance was kept below 10 kOhm.

ERP averaging was performed offline with a common average reference using the EEGLAB toolbox (Delorme & Makeig, 2004) in MATLAB (Mathworks). The ERP epoch contained a 700 ms recording window and a 100 ms pre-stimulus baseline. Data were bandpass filtered from 0.5 to 20 Hz using FIR filters (MATLAB) and then down-sampled to 500 Hz.

Due to the presence of CI related artifacts in the EEG signal, a blind source separation ICA algorithm was applied to the data, with the initial number of components in the ICA matrix reflecting the number of electrodes used during the recordings for each subject (Miller & Zhang, 2014). Electrodes that had been deactivated during the recordings were interpolated after performing the ICA. Artifacts representing EOG and EMG activity were removed prior to averaging.

Independent components representing CI artifacts were identified and manually removed from each subject's ICA matrix using temporal and spatial criteria outlined by Gilley et al. (2006) and previously used in Miller and Zhang (2014). Components were defined as CI artifacts and subsequently removed if the artifact activation occurred at either stimulus onset or offset or if the duration of the activation was the same as the duration of the stimulus. In addition, component scalp projections indicating centroid, non-dipolar patterns on the side of the implant were also labeled as artifact and removed. Components were removed in a step-by-step fashion from the ICA matrix until only non-artifactual components remained. Pairwise *t*-tests confirmed that across listeners there was no significant difference in the number of components used to derive the ERP waveforms across the two test sessions for the speech ($t(17) = -1.03$, $p > 0.05$) or nonspeech conditions ($t(17) = -1.4$, $p > 0.05$).

F. Behavioral and MMN data analysis

Behavioral data analysis

Individual behavioral identification data were first converted to percentages and then subjected to a probit analysis. After fitting the probit function to the binary identification responses, the slope, boundary, and boundary width of the function were analyzed. The slope of the probit function reflects how well the endpoint stimuli in a binary response task are differentiated, with a steeper slope indicating a sharper boundary between the identification function endpoints and, presumably, more categorical-like perception. The boundary location represents the 50% identification point on the function and ranged from intervals 1 to 11, with 1 representing /ba/ on the /ba/-/da/ continuum (/wa/ on the /wa/-/ja/ continuum) and 11 /da/ (/ja/ on the /wa/-/ja/ continuum). The boundary width reflects the distance of interval separation from 25% to 75% identification of /ba/ (or /wa/) on the identification function. Individual discrimination data were analyzed for overall percent correct. The task took into account all possible response categories and a bias-free estimate of perceptual sensitivity (d') was computed (Macmillan & Creelman, 2004). Paired Student's *t*- tests were also performed separately on the /ba/-/da/ and /wa/-/ja/ contrasts to assess the effects of training (pretest vs. posttest) on the individual identification and discrimination data for the trained and untrained listeners.

MMN data analysis

To control for differences in the number of stimulus presentations and inherent acoustic differences between the standard and deviant stimuli, the MMN was calculated using standard and deviant responses to the identical stimulus (Kraus, et al., 1993; Y. Zhang, et al., 2009). The MMN waveform was derived by subtracting the averaged deviant waveform from the averaged standard waveform. Based on the grand mean MMN waveforms, peak MMN amplitudes and latencies were extracted from the difference waveform from each subject in the latency range of 140 to 270 ms from the

FCz and Cz electrodes. The average MMN amplitude from a fixed time window 20 ms before and after the peak MMN amplitude was also extracted for each subject.

Effects of training (pretest vs. posttest), speech continuum (/ba/-/da/ vs. /wa/-/ja/), acoustic-phonetic identity (speech vs. nonspeech), and phonetic category (within vs. across phoneme category) on peak MMN latencies, amplitudes, and MMN amplitude window from the individual data were assessed using repeated-measures analysis-of-variance (ANOVA). The grouping factor of trained vs. untrained listeners was included in the ANOVA model. Post-hoc repeated-measures univariate ANOVAs were performed on all significant main effects. Where applicable, Bonferroni or Greenhouse-Geisser corrections were applied to the reported p values. Brain-behavior correlates were assessed using a Pearson Correlation analysis.

III. Results

A. Behavioral effects of training

Discrimination scores (controlling for the false positive rate) for the stimulus pairs on the /ba/-/da/ continuum showed evidence that the identification training served to direct CI users' attention to the important F2 transition cue for categorizing the /ba/ vs. /da/ speech stimuli (Fig. 4.2A). Average percent correct discrimination for the across category 5-9 stimulus pair significantly increased with training from 64% to 71% for the trained listeners ($t(8) = 2.3$, $p < 0.05$). Average percent correct discrimination for the within category 1-5 and 7-11 stimulus pairs on the /ba/-/da/ continuum did not significantly change with training ($p > 0.05$) (Fig. 4.2A). For the nonspeech acoustic correlates of the /ba/-/da/ continuum, average percent correct discrimination for the across category 5-9 pair did not show any significant changes or enhancement at the phonetic boundary with training ($t(8) = -1.66$, $p = 0.13$) (Fig. 4.2C). Likewise, the nonspeech acoustic correlates of the 1-5 and 7-11 within category pairs on the /ba/-/da/ continuum also did not significantly change with training ($p > 0.05$) (Fig. 4.2C).

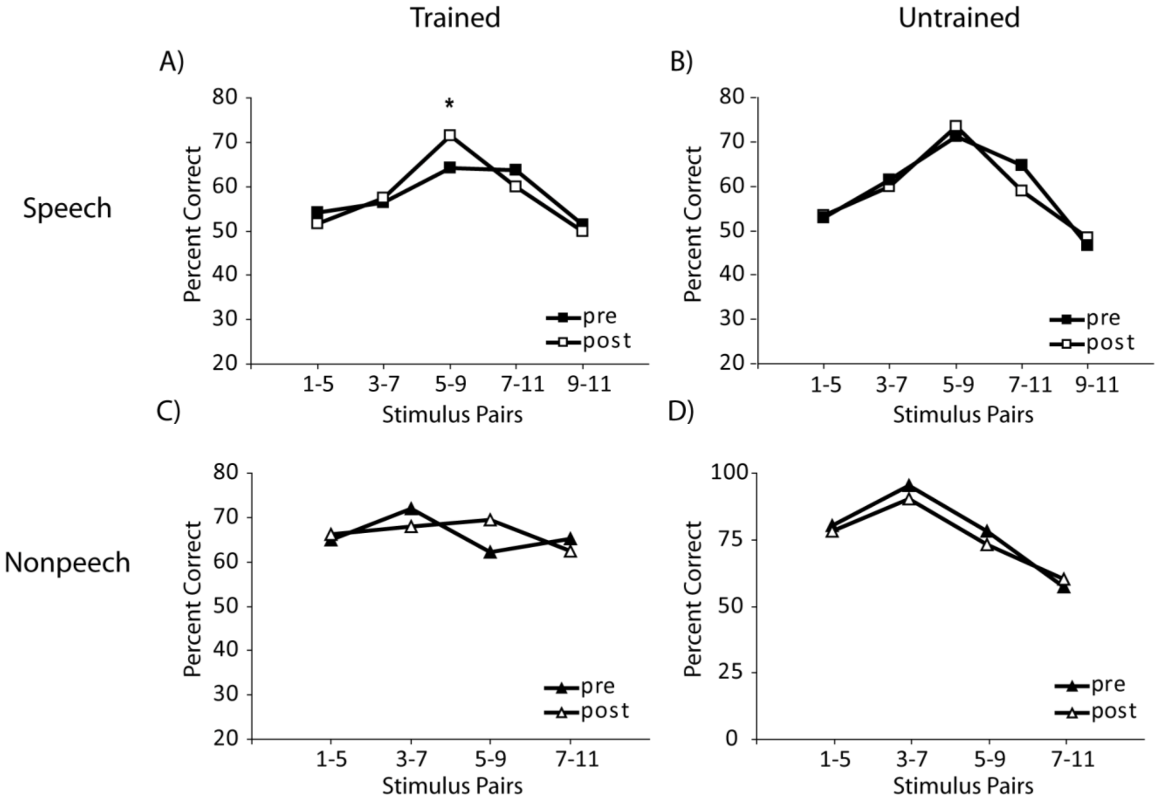


Fig. 4.2: Behavioral discrimination results. A) Discrimination scores for the synthetic /ba/-/da/ speech stimuli for the trained CI listeners. B) Discrimination scores for the synthetic /ba/-/da/ speech stimuli for the untrained CI listeners. C) Discrimination scores for the synthetic /ba/-/da/ nonspeech stimuli for the trained CI listeners. D) Discrimination scores for the synthetic /ba/-/da/ nonspeech stimuli for the untrained CI listeners. Stimulus pair 5-9 represents the cross-category pair [*p<0.05].

There were no significant effects of training at the group level observed for the slope (*t*(8) = -0.47, *p* = 0.65) or boundary width (*t*(8) = 1.27, *p* = 0.24) of the /ba/-/da/ identification functions from pretest to posttest (Fig. 4.3A). Training shifted the CI users' phonetic boundary on the /ba/-/da/ continuum, but failed to reach significance (*t*(8) = -1.8, *p* = 0.05, one tailed).

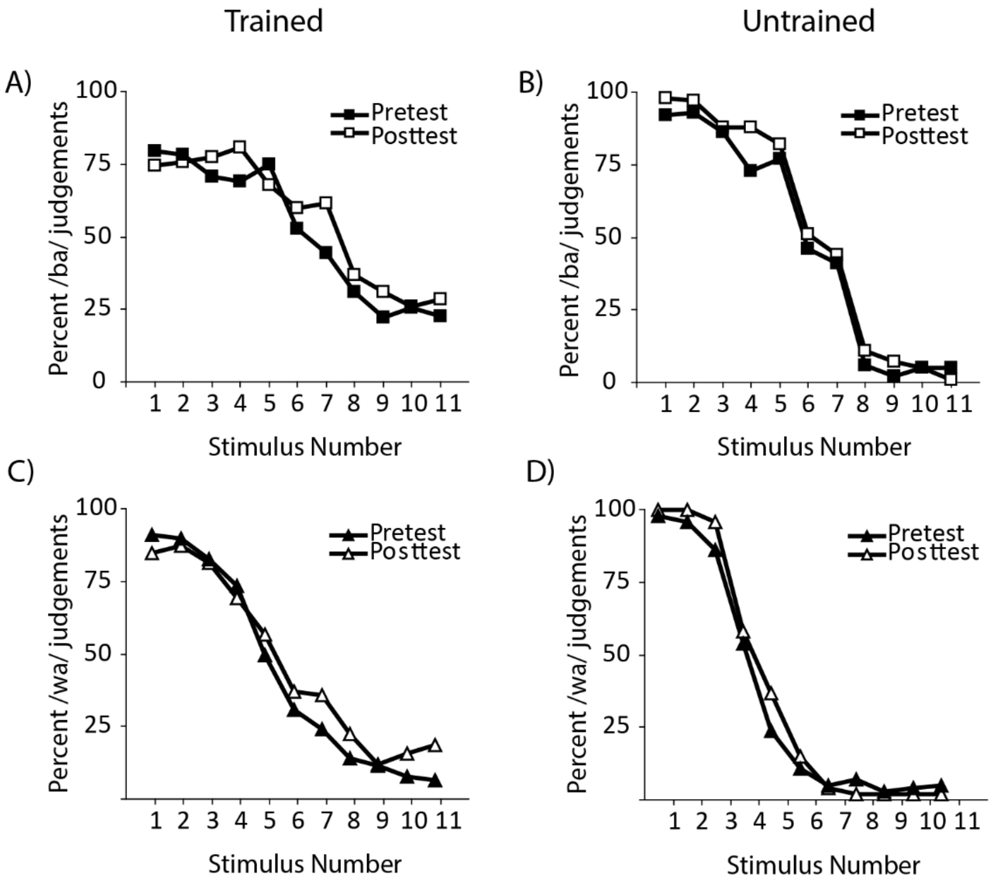


Fig. 4.3: Pre-posttest identification scores for the synthetic /ba/-/da/ (top panel) and /wa/-/ja/ (bottom panel) speech stimuli. A) Identification scores for the trained listeners for the /ba/-/da/ continuum. B) A) Identification scores for the untrained listeners for the /ba/-/da/ continuum. C) Identification scores for the trained listeners for the /wa/-/ja/ continuum. D) Identification scores for the untrained listeners for the /wa/-/ja/ continuum.

Discrimination scores for the /wa/-/ja/ synthetic speech continuum did not significantly change from pretest to posttest for any of the stimulus pairs on the continuum ($p > 0.05$) (Fig. 4.4A). Likewise, training did not significantly change percent correct discrimination for any of the across or within category pairs for the nonspeech acoustic correlates ($p > 0.05$) (Fig. 4.4C). No significant effects of training were observed for the slope ($t(8) = 0.42$, $p = 0.68$), boundary ($t(8) = -0.47$, $p = 0.65$) or boundary width ($t(8) = -0.55$, $p = 0.59$) of the /wa/-/ja/ of the identification functions from pretest to posttest (Fig 4.3C).

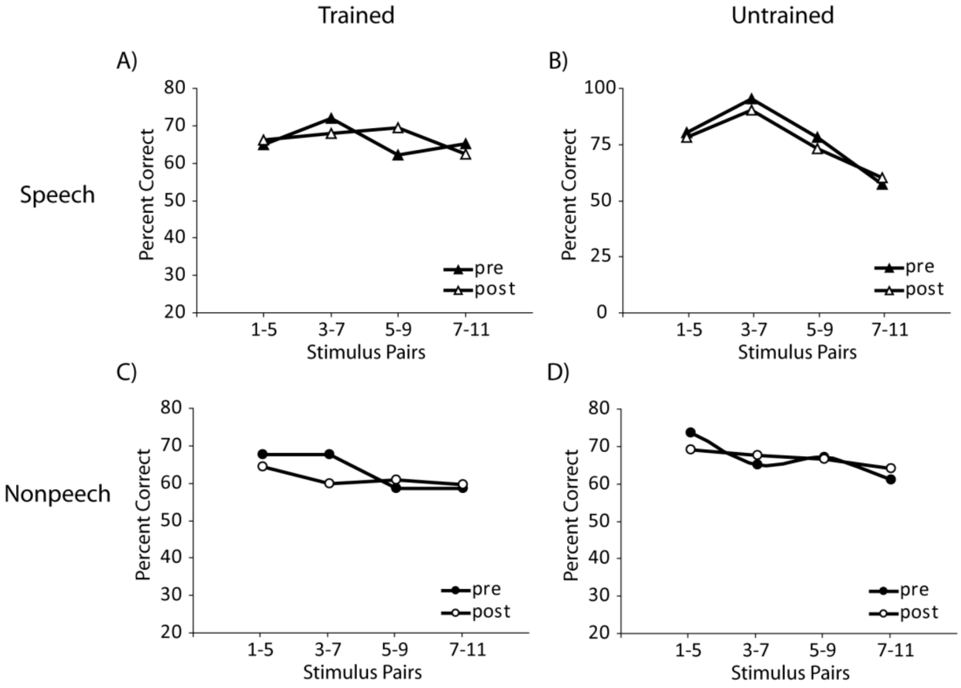


Fig. 4.4: Behavioral discrimination results. A) Discrimination scores for the synthetic /wa/-/ja/ speech stimuli for the trained CI listeners. B) Discrimination scores for the synthetic /wa/-/ja/ speech stimuli for the untrained CI listeners. C) Discrimination scores for the synthetic /wa/-/ja/ nonspeech stimuli for the trained CI listeners. D) Discrimination scores for the synthetic /wa/-/ja/ nonspeech stimuli for the untrained CI listeners. Stimulus pair 3-7 represents the cross-category pair [*p<0.05].

As expected, the high performing CI listeners who did not undergo training showed no significant changes in slope, boundary, or boundary width from pretest to posttest for either the /ba/-/da/ (Fig. 4.3B) or /wa/-/ja/ (Fig. 4.3D) contrasts ($p > 0.05$) . The high performing listeners did not show any significant changes to their discrimination scores from pretest to posttest for either the /ba/-/da/ (Fig. 4.2) or /wa/-/ja/ (Fig. 4.4) speech or nonspeech contrasts ($p > 0.05$). In comparison with the trainee group, the high performing listeners exhibited significantly steeper slopes for the /ba/-/da/ ($F(1,12) = 7.53$, $p = 0.02$) and /wa/-/ja/ ($F(1,12) = 6.26$, $p = 0.03$) contrasts at both the pretest and posttest sessions (Fig. 4.3). The two groups of CI listeners did not significantly differ on boundary location or boundary width at the pretest or posttest for either speech contrast ($p > 0.05$).

B. ERP data

Clear standard and deviant ERP response waveforms were observed across all speech and nonspeech conditions. The grand mean peak MMN latency, and average MMN amplitude over the fixed time window of the difference waveforms used in the statistical analysis for the trained are summarized in Table 4.2 and the untrained listeners in Table 4.3.

	Amplitude Window (uV)				Latency (ms)			
	/ba/-/da/		/wa-/ja/		/ba/-/da/		/wa-/ja/	
Speech	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Across	-0.34	-1.36	-0.38	-0.93	207.11	217.33	232.44	208.44
	(0.37)	(1.3)	(0.21)	(0.58)	(57.69)	(37.7)	(48.17)	(31.3)
Within	-0.75	-0.488	-0.43	-0.66	198.44	240.67	218.44	211.33
	(0.45)	(0.64)	(0.72)	(0.78)	(54.61)	(16.8)	(36.55)	(44.29)
Nonspeech	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Across	-0.39	-0.86	-0.97	-0.74	178.22	227.33	198.67	213.11
	(0.48)	(1.24)	(0.52)	(0.87)	(43.5)	(39.29)	(42.25)	(41.69)
Within	-0.72	-0.76	-0.91	-0.29	202.89	200.67	209.56	205.78
	(0.92)	(0.85)	(0.92)	(0.73)	(53.14)	(48.49)	(34.9)	(38.7)

Table 4.2: Average pretest and posttest amplitude window and latency values (± standard deviation) of the MMN response for the trained CI listeners used in the statistical analysis.

	Amplitude Window (uV)				Latency (ms)			
	/ba/-/da/		/wa-/ja/		/ba/-/da/		/wa-/ja/	
Speech	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Across	-0.44	-0.5	-0.65	-0.66	197.23	218.62	227.69	215.85
	(0.69)	(0.95)	(0.39)	(0.12)	(51.96)	(36.03)	(50.33)	(28.5)
Within	-0.06	-1.11	-0.99	-0.51	198.0	236.62	215.38	213.23
	(1.26)	(0.29)	(0.5)	(0.5)	(48.69)	(21.99)	(38.36)	(39.89)
Nonspeech	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Across	-1.34	-0.67	-0.57	-0.75	193.08	220.31	203.53	206.46
	(1.57)	(0.64)	(0.51)	(0.39)	(49.13)	(49.54)	(38.92)	(45.88)
Within	-0.69	-0.49	-0.77	-0.79	214.15	206.92	202.46	202.31
	(0.48)	(0.89)	(1.08)	(0.94)	(51.18)	(51.58)	(41.38)	(37.47)

Table 4.3: Average pretest and posttest amplitude window and latency values (± standard deviation) of the MMN response for the untrained CI listeners used in the statistical analysis.

C. Training effects in the MMN analysis

Repeated-measures ANOVA for MMN amplitude window (average MMN amplitude in the time window 20 ms before and after peak MMN amplitude) indicated a significant four-way interaction between the factors of *speech continuum* (/ba/-/da/ vs. /wa/-/ja/), *acoustic-phonetic identity* (speech vs. nonspeech), *training* (pretest vs. posttest), and *group* (trained vs. untrained) ($F(1,11) = 6.06$, $p = 0.03$). In addition, there was also a significant four-way interaction between the factors of *speech continuum*, *training*, *phonetic category* (across vs. within phoneme category), and *group* ($F(1,11) = 9.08$, $p = 0.01$).

Post-hoc univariate ANOVAs and paired Students *t*-tests were performed separately on the trained and untrained listeners to investigate factors that contributed to the significant interactions. For the trained listeners, the /ba/-/da/ across category speech stimuli evoked significantly larger MMN amplitude window responses after training ($t(8) = 2.69$, $p = 0.03$) (Fig. 4.5A), but the evoked responses to the within category speech stimuli from the /ba/-/da/ continuum did not change from pre-to posttest ($t(8) = -0.2$, $p = 0.85$) (Fig. 4.5B).

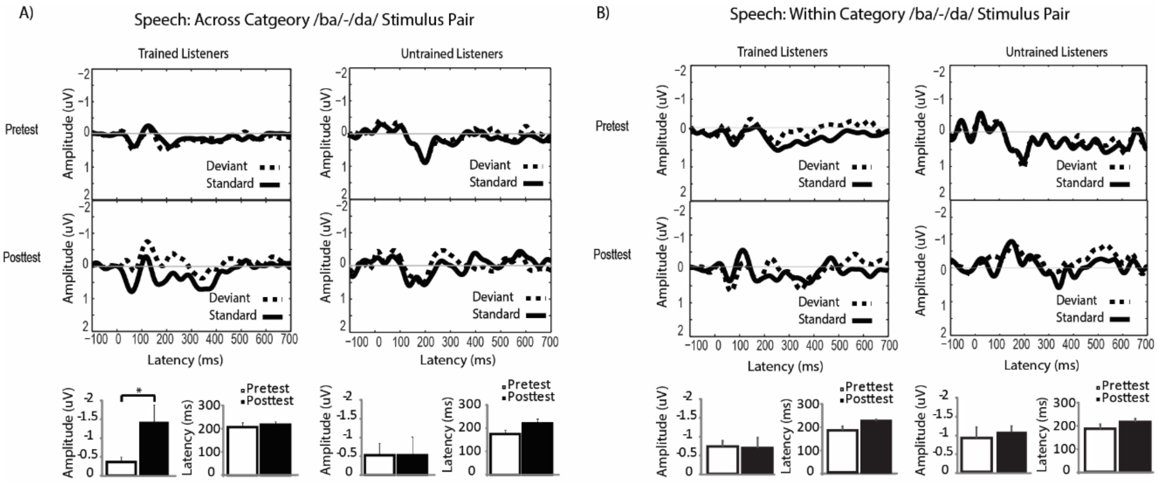


Fig. 4.5: Grand mean mismatch negativity (MMN) responses from the pretest and posttest sessions elicited by the /ba/-/da/ A) Across category and B) Within category speech stimuli pairs. Negative polarity plotted up for the ERP signal. Bar graphs display the grand mean MMN amplitude window and peak latency values for the trained and untrained listeners across test sessions [*$p<0.05$].

The trained listeners' MMN amplitude window responses to the across category /wa/-/ja/ speech stimuli also significantly increased from pretest to posttest ($t(8) = 2.22$, $p = 0.04$) (Fig.4.6A), but there was no change to the MMN waveforms evoked by the within category /wa/-/ja/ speech stimuli after training (*t*(8) = = 0.32, *p* = 0.56) (Fig. 4.6B).

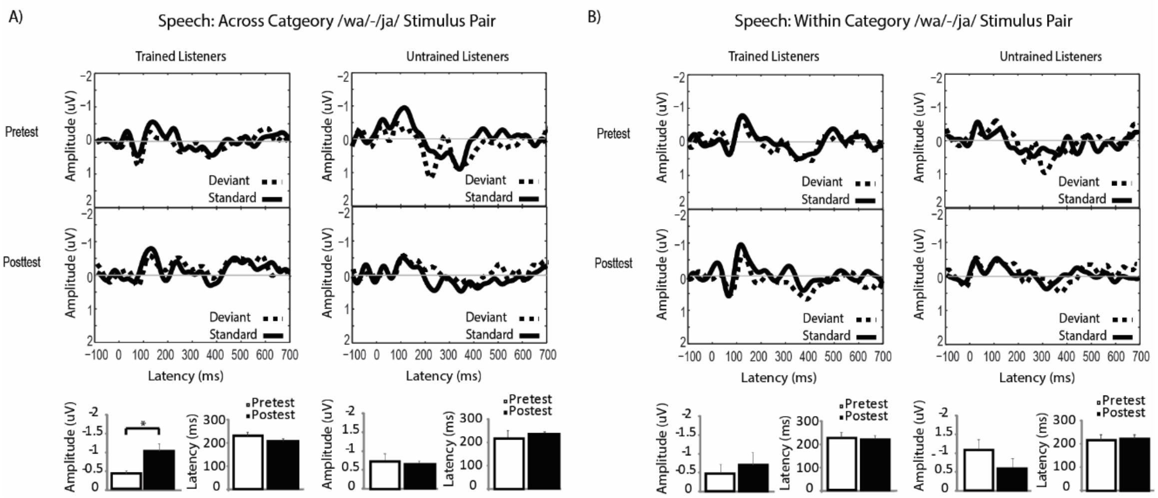


Fig.4.6: Grand mean mismatch negativity (MMN) responses from the pretest and posttest sessions elicited by the /wa/-/ja/ A) Across Category and B) Within Category speech stimuli pairs for the trained and untrained listeners. Stimulus pair 7-11 was used to evoke the within phoneme category and stimulus pair 3-7 the across phoneme category MMN response. Negative polarity plotted up for the ERP signal. Bar graphs display the grand mean MMN amplitude window and peak latency values for the trained and untrained listeners across test sessions [*p<0.05]

There were no significant changes from pretest to posttest for MMN amplitude window for the across category nonspeech correlates for /ba/-/da/ (*t*(8),*p* = 0.31) (Fig. 4.7A) or /wa/-/ja/ (*t*(8) = -0.78, *p* = 0.46) (Fig. 4.8A) for the trained listeners.

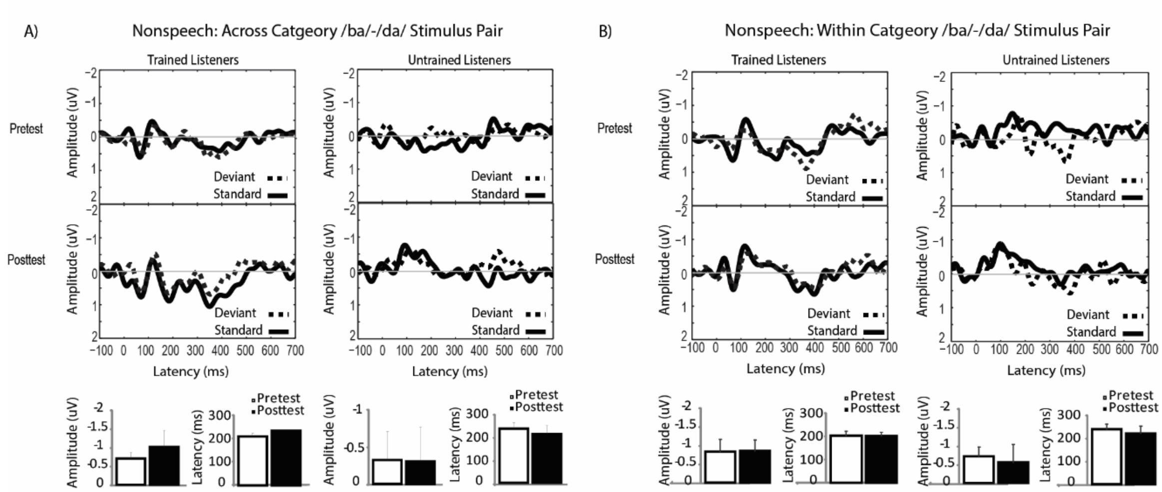


Fig. 4.7: Grand mean mismatch negativity (MMN) responses from the pretest and posttest sessions elicited by the /ba/-/da/ A) Across category and B) Within category nonspeech stimuli pairs. Negative polarity plotted up for the ERP signal. Bar graphs display the grand mean MMN amplitude window and peak latency values for the trained and untrained listeners across test sessions [*p<0.05].

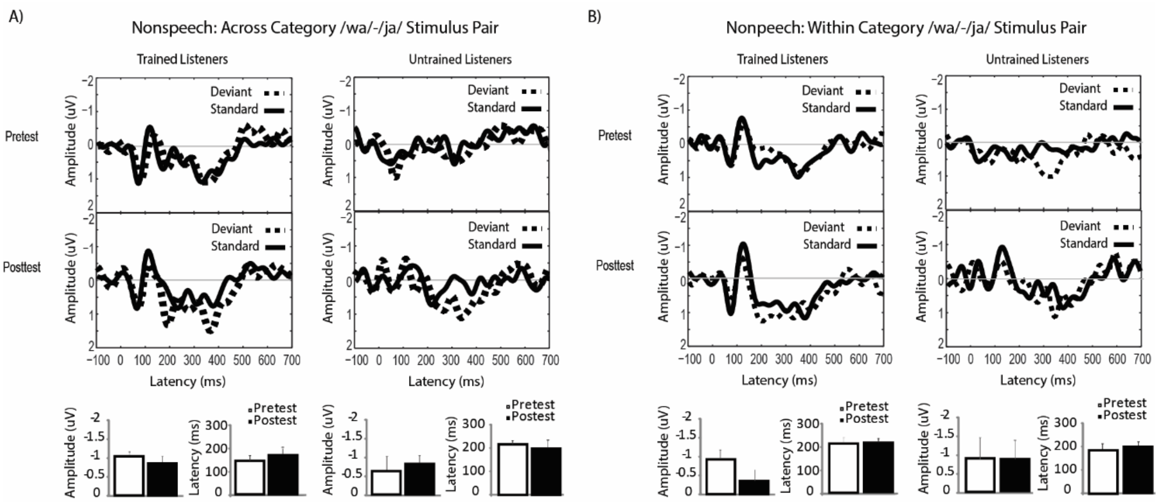


Fig.4.8: Grand mean mismatch negativity (MMN) responses from the pretest and posttest sessions elicited by the /wa/-/ja/ A) Across Category and B) Within Category nonspeech stimuli pairs for the trained and untrained listeners. Stimulus pair 7-11 was used to evoke the within phoneme category and stimulus pair 3-7 the across phoneme category MMN response. Negative polarity plotted up for the ERP signal. Bar graphs display the grand mean MMN amplitude window and peak latency values for the trained and untrained listeners across test sessions [*p<0.05]

Likewise, no significant changes in MMN amplitude window were observed for within category nonspeech correlates for /ba/-/da/ (*t*(8) = = 0.13, *p* = 0.91) (Fig. 4.7B) or /wa/-/ja/ *t*(8) = -1.3, *p* = 0.23) (Fig. 4.8B). For the untrained listeners, there were no significant changes in MMN amplitude window across the two test sessions for the within or across category speech or nonspeech correlates for either /ba/-/da/ or /wa/-/ja/.

Repeated-measures ANOVA for MMN latencies indicated that there were no significant main effects or interactions for the *training* (pretest vs. posttest), *speech continuum* (/ba/-/da/ vs. /wa/-/ja/), *acoustic-phonetic identity* (speech vs. nonspeech), or *phonetic category* (within vs. across phoneme category) variables for the trained or untrained listeners.

D. Training effects in the brain-behavior correlations

A high degree of variability in the individual CI user training data was observed, allowing for brain-behavior correlations to be examined. The MMN amplitude window evoked by the across category /ba/-/da/ speech stimuli (stimulus pair 5-9) increased from pretest to posttest for eight out of the nine trainees. A significant negative correlation was found between the change in the /ba/-/da/ across category MMN amplitude window response and the change in the steepness of slope of the individual /ba/-/da/ identification functions after training (Pearson's r = -0.809, *p* = 0.0085) (Fig. 4.9A). Thus, the CI users that showed the greatest enhancement in across category MMN amplitude also exhibited the sharpest category boundaries for the /ba/-/da/ contrast after training. The CI user that exhibited the largest change in MMN amplitude and change in the steepness of slope for the /ba/-/da/ contrast is a statistical outlier; however, removing the subject from the analysis only marginally changed the correlation (Pearson's r = -0.63, p = 0.09). Similarly, the MMN amplitude window evoked by the across category /wa/-/ja/ stimulus pair, increased from pretest to posttest for eight out of the nine trainees. A marginally significant negative correlation was also found between the change in the /wa/-/ja/ across category MMN amplitude window response and the change in the slope of the individual /wa/-/ja/ identification functions after training (Pearson's r = -0.62, *p* = 0.07) (Fig. 4.9B). Taken together, the brain-behavioral correlations suggest that the MMN reflects the training induced changes in behavior associated with improved phonetic categorization.

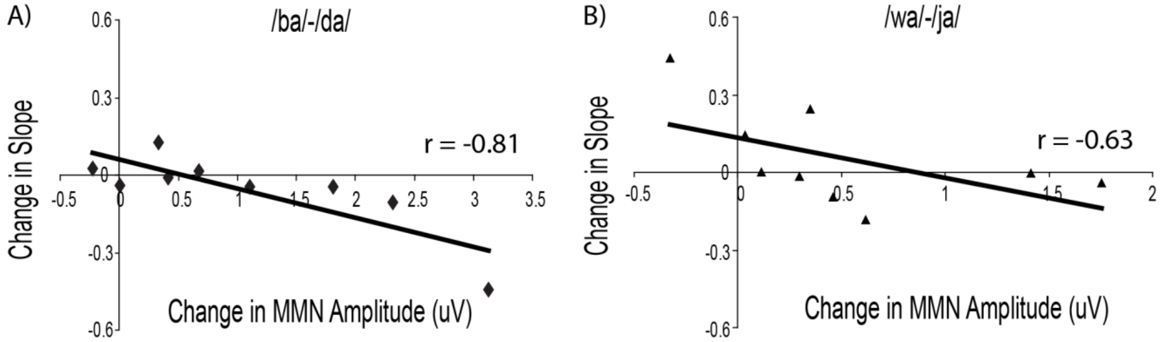


Fig. 4.9: Brain-behavior correlations of training induced plasticity for phonetic learning. A) Negative correlation between changes in the slopes of the individual identification functions and across phoneme category MMN amplitudes for the /ba/-/da/ speech stimuli. B) Negative correlation between changes in the slopes of the individual identification functions and across phoneme category MMN amplitudes for the /wa/-/ja/ speech stimuli.

IV. Discussion

As CI users tend to exhibit poorer phonetic categorization than listeners with normal hearing, the present study aimed to determine if high variability identification training could improve categorization of two speech contrasts. We were further interested in testing whether the electrophysiological MMN response would faithfully reflect changes in behavioral sensitivity. We hypothesized that the high-variability training protocol would sharpen phonetic boundaries and enhance the MMN response to an across phoneme category pair relative to an acoustically equivalent within phoneme category pair. Our brain and behavior data largely support our hypotheses and suggest identification training strengthened phonetic category development in adult postlingually deafened CI users. The detailed results and implications of the study and future directions will be discussed here.

A. Phonetic learning in adult CI users

In the present study, eight hours of high variability identification training improved phonetic categorization of speech contrasts differing by dynamic formant transitions in adult CI listeners. However, the individual trainees were highly variable in their ability to categorize the /ba/-/da/ and /wa/-/ja/ contrasts. Training significantly enhanced behavioral sensitivity to the across category stimulus pair for the /ba/-/da/ contrast. This change in perceptual sensitivity was accompanied by marginal changes in labeling of the stimuli along the /ba/-/da/ continuum, whereby, for the trained subjects, the slope of the /ba/-/da/ identification function became steeper with training, and the boundary width became smaller in six out of the nine trainees. These results collectively suggest that training succeeded in directing the CI users’ attention to the important F2 cue that categorizes /ba/ and /da/ and that training sharpened the phonetic boundary between the two phonemes.

Training did not appear to significantly alter behavioral identification or discrimination of the synthetic /wa/-/ja/ stimuli. However, in six of the nine trained CI listeners, the slope of the identification function became more negative, and for seven of nine trainees, the boundary width became smaller. These results suggest responses became more categorical-like and less variable with training for the /wa/-/ja/ contrast.

When comparing phonetic categorization of /ba/-/da/ vs. /wa/-/ja/, the trained CI users exhibited significantly sharper phonetic boundaries and perceived the /wa/-/ja/ contrast more categorically than the /ba/-/da/ contrast at both the pretest and posttest sessions, suggesting that the /wa/-/ja/ contrast was more salient than the /ba/-/da/ contrast for most listeners. The two contrasts contained the same spectral information and differed only in the length of the formant transition, meaning they should have stimulated the same electrodes of the implant. Thus, the more accurate categorization of the /wa/-/ja/ contrast suggests that simply having access to spectral structure does not result in equivalent phonetic categorization ability. Instead, the duration of the spectral cues appears to be important as well. This salience of duration cues during a phonetic categorization task is consistent with previous work by Winn et al. (2012) who found CI users tend to weight the duration cue more heavily than the formant transition, formant structure or voicing cues when categorizing /s/ and /z/. Further work will need to be done to determine the role of temporal cues and whether they can be exploited to enhance phonetic learning in adult CI users.

B. Neural Markers of Phonetic Learning

A fundamental goal of neuroimaging research is to link changes in behavior to changes in the brain in order to better understand the neural circuits that support language (Y. Zhang & Wang, 2007). Our behavioral and EEG training data suggest that high variability identification training altered perceptual sensitivity in adult CI users and these learning-induced changes were reflected by the MMN response. For both speech contrasts, the amplitude of the MMN response to the across phoneme category stimulus pair significantly increased with training, but the MMN response elicited by an acoustically equivalent within phoneme category pair remained unchanged. Because we did not observe any changes in the MMN response across test sessions for the untrained CI listeners, the observed changes were likely due to our training program and not repeated exposure to the speech sounds. Of importance, with training we found significant correlations between behavioral changes in the sharpness of the phonetic boundaries and changes in the enhancement of the MMN evoked by pairs of stimuli crossing the phonetic boundary for each contrast. Whether changes in behavioral discrimination to stimuli that cross a phonetic boundary also correlate with changes in the MMN response were not tested, but will be examined in a future analysis. The training-induced changes in MMN appear to be related to phonetic, rather than acoustic, learning, as the MMNs for the acoustic control stimuli were unchanged with training. However, it should be noted that the MMN was present to the nonspeech correlates at the pretest and posttest, suggesting that the training selectively affected speech learning without compromising acoustic sensitivity to nonspeech stimuli with similar spectral properties to the speech (Miyawaki, et al., 1975). This pattern of results is consistent with Näätänen's model of the MMN (Näätänen, et al., 1997; Näätänen, et al., 2007) that suggests two separate processes contribute to the mismatch response to speech: a left-hemisphere dominant component related to linguistic processing and a bilateral component related to acoustic processing. Our training may have encouraged CI users to attend to the important F2 transition cue for categorizing the two speech contrasts. Thus, it is likely the training-induced MMN enhancement we observed for the speech stimuli was related to the parallel detection of both the acoustic F2 change and the phonetic change.

C. Success and limitations of training

Our training program improved phonetic categorization of the two speech contrasts in the trained CI users, but significant individual variability in training-related phonetic learning was observed. The training program used in the present study was designed to enhance attention to the important F2 transition cue that contrasts /ba/-/da/ and /wa/-/ja/. However, it is possible that some CI users might be able to utilize a different acoustic cue in the naturally produced speech stimuli in the training sessions to categorize the speech contrasts, limiting the success of our training program for some listeners. Winn and Litovsky (2015) recently examined the perceptual weighting of the formant transition and spectral tilt cues used to categorize the /ba/-/da/ contrast in CI users. Their results suggest that, unlike NH listeners, CI users tend to weight the spectral tilt, the balance of high versus low frequency information, more heavily than the formant transition cue when categorizing /ba/ vs. /da/. However, they did find that the CI users that utilized formant transition perceptual weighting strategies most similar to NH listeners when categorizing the /ba/-/da/ contrast exhibited the best word recognition scores. Similarly, Moberly et al. (2014) documented that CI users that utilized perceptual weighting strategies most similar to NH listeners when categorizing the /ba/-/wa/ contrast exhibited superior word recognition performance as well. It is possible that the listeners in our study that exhibited the greatest training-related gains were using a different perceptual listening strategy to categorize the speech sounds than the CI users that showed only small behavioral changes. Thus, it may be beneficial for future training studies to examine what protocols are most successful in helping CI users retune their weighting strategies in order to maximize improvements in phonetic categorization. Future work will also need to determine if the training protocol, including the duration as well as the number of training sessions used for each individual CI user, will need to vary based on the listening strategy that a given CI user has adopted.

All our CI listeners in the trainee group completed the training sessions in the laboratory which allowed tight control over the training protocol. However, this approach may have limited phonetic learning in some CI users due to the artificial nature of the task. Language acquisition is inherently multimodal (Kuhl, 2000), but the training stimuli were presented without visual articulation cues. In typical face-to-face speech perception, both auditory and visual cues are available, and the visual signal can provide additional redundant information that can help overcome auditory degradations of speech. When learning a nonnative speech contrast, audiovisual training has been previously shown to be beneficial for improving both the perception (Hardison, 2003) and production (Hazan, Sennema, Iba, & Faulkner, 2005) of the phonemes. It is also well established that listeners with hearing loss rely on the visual signal for place of articulation cues (Grant, Walden, & Seitz, 1998), so it may be likely that audiovisual stimuli presentations during training (Zhang et al., 2009) could strengthen the sensorimotor process involved in producing and perceiving speech, improving stored phonemic representations as well.

D. Clinical implications and Future Directions

The results of the present training study are of potential clinical importance for CI users. Here we report both behavioral and neural sensitivity measures of phonetic categorization in adult postlingually deafened CI users. Our results suggest that our training protocol was successful in inducing neural plasticity related to phonetic learning in adult CI users through enriched linguistic exposure, even in users who had significant experience with their devices. Previous behavioral studies have shown that training can improve a variety of speech perception measures in CI users (Fu, et al., 2005; Fu & Galvin, 2007; Stacey, et al., 2010), and importantly, past EEG studies with CI listeners have also documented that the MMN response correlates well with behavioral measures such as spectral discrimination thresholds (Lopez-Valdes et al., 2013), timbre discrimination thresholds (Rahne, Plontke, & Wagner, 2014), and speech perception measures (Kelly, et al., 2005; Kraus, et al., 1993). Thus, by collecting behavioral and EEG data during training, we can shed light on the neural processing underlying plasticity-related changes in behavior, making them a potentially valuable clinical tool. For example, in the present study, we observed that the MMN was significantly enhanced with training for the /wa/-/ja/ contrast even though we saw only small changes in identification and discrimination of the contrast behaviorally. If only looking at the behavioral responses, it might have been assumed that the present training protocol was unsuccessful. However, Tremblay et al. (1998) previously documented that changes in

ERP responses could occur prior to observed behavioral improvements in speech discrimination and perception after auditory training. Thus, it would be clinically useful to monitor ERP responses, such as the MMN, during the CI rehabilitation process to determine whether cortical-level discriminative sensitivity to the speech contrasts has taken place at a pre-attentive stage and has the potential to be translated into behavioral discrimination. The ability to assess phonetic categorization objectively in the absence of attention has large clinical implications for pediatric CI users and other listeners that do not always provide reliable behavioral results.

While our results suggest that the MMN can be used to assess phonetic categorization and speech discrimination for individual CI users, the measurement of the MMN response is not currently part of the audiologic test battery. One of the main reasons the MMN has not been incorporated into clinical practice is related to the high inter-subject variability of the amplitude of the MMN response. Typically, the MMN is elicited with a single oddball paradigm, and the amplitude of the MMN response using this technique can be quite small or even absent at the individual level, despite stimuli being behaviorally discriminable. Because the scale of the response can differ dramatically across individuals for the same set of stimuli, the MMN is often statistically tested at the group level, which may not directly inform clinical assessment at the individual subject level. In addition, with the single oddball paradigm, the MMN response is simply deemed to be present or absent relative to the zero crossing of the response waveform for a given subject; when the response is present, there is no true reference within the same subject or a standardized scale to compare the MMN response.

The double oddball paradigm that we used in the present study could overcome the shortcomings of the single oddball design and has the potential to be used clinically to assess phonetic categorization in individual subjects. The double oddball paradigm uses two distinct deviants to elicit MMN responses and thus provides a baseline MMN response within the same subject to which comparisons can be made. In particular, phonetic categorization can be examined in an individual by evoking the MMN using both within and across phoneme category pairs of stimuli that differ by equivalent acoustic amounts, but have different phonologic status. The MMN evoked by the within phoneme category pair serves as a baseline comparison to the MMN evoked by the across category pair for each subject. A comparison or ratio of these two responses for an individual avoids the problem of MMN amplitude scale disparity across individuals.

We acknowledge that the purpose of our study was only to determine whether the MMN would reflect changes in behavioral sensitivity after training for a group of CI users, not to examine the clinical sensitivity of the MMN at the individual level. However, our individual subject data do suggest that our high performing CI users who exhibited superior phonetic categorization skills have larger across versus within phoneme category ratios than our lower performing CI subjects (Tables 4.2 and 4.3). Future work will need to focus on developing appropriate methods and norms for comparing the across versus within phoneme category MMN responses before being used clinically to assess auditory perception and phonetic categorization in both adults and children.

V. Conclusion

The present study demonstrates that high variability identification training can induce neural plasticity and enhance phonetic categorization of two speech contrasts differing by dynamic formant transitions in postlingually deafened adult CI users. Changes in behavioral sensitivity to the phonetic contrasts were reflected in the electrophysiologic data for the trained CI users as evidenced by the larger MMN responses to across phoneme category pairs of stimuli relative to acoustically equivalent within phoneme category pairs. The trained CI listeners exhibited significant variability in phonetic categorization skills and future training studies will need to examine what protocols produce the most robust category development for individual CI users. Future work will also need to focus on refining the double oddball paradigm for collecting MMN responses before it can be used clinically to assess phonetic learning in individual CI listeners.

Chapter 5: General Discussion

I. Summary

Cochlear implants provide only coarse spectral cues, and the postlingually deafened CI recipient needs to learn how these degraded auditory inputs map onto to previously acquired phonetic categories. The experiments included in this dissertation assessed whether targeted auditory training could induce plasticity for learning difficult phonetic contrasts in CI recipients. The mechanisms that support the perceptual learning of electric speech were examined using behavioral and electrophysiological measures. In order to use ERPs to examine the neural correlates of phonetic learning in CI recipients, a validation study of our method for removing CI device-related electrical artifact from EEG data was also completed. The results of these experiments suggest that high variability identification training improved perception of naturally produced speech and CI listeners generalized their learning to unfamiliar talkers. Fine grained behavioral and neural measures suggest that an attention-weighting mechanism related to enhanced phonetic categorization supported the improved phonemic perception. The theoretical implications of this work will be discussed.

II. Removing CI artifact from EEG data

Auditory event-related potentials (ERPs) can be used to assess the integrity of the central auditory pathway and have been successfully used to examine the neural coding of speech sounds in a variety of clinical populations (see Martin et al., 2008 for a review). Despite their audiological significance, ERPs have not been routinely incorporated into rehabilitation protocols for CI recipients, likely because of the presence of large CI device-related electric artifacts in the EEG signal. Experiment 1 aimed to validate the ICA method for removing CI device-related artifact from EEG data collected in our lab. The results indicated that the manual and semi-automated approaches that use spatial and temporal criteria to identify artifactual components in an ICA matrix can be reliably used to remove CI artifact from EEG data and reconstruct the ERP response. The results of Experiment 1 validated the use of EEG to answer our questions about the cortical mechanisms underlying phonetic learning in adult CI recipients.

The ICA technique was successful in minimizing CI device-related artifact from our EEG data, but there are some limitations to this approach. The most critical issue when using the ICA method is the correct identification and removal of artifactual components from the ICA matrix. The number of active electrodes, the type of stimulation (monopolar vs. bipolar), and the speech processing algorithms differed greatly across subjects in the present study. These differences likely contributed to the variability in the amplitude and scalp distribution of CI device-related artifact we observed across subjects, making manual identification and removal of the artifact laborious. The ICA method relies on a set of assumptions about the spatial and temporal characteristics of the CI artifact, and to reduce the subjective nature of manual inspection, we implemented strict criteria for removing artifactual components from our EEG data. However, the use of strict criteria and the variability in artifact we observed meant it was not always possible to remove the artifact when it was very small in amplitude as we ran the risk of removing cortical responses as well.

The ICA approach has been used to successfully attenuate CI device-related artifact from EEG data collected from adult (Gilley, et al., 2006; Viola, et al., 2012; Viola, et al., 2011) and pediatric (Bakhos et al., 2012) CI patients, but due to the laborious and subjective nature of artifact removal, more work is needed before being implemented clinically to assess the sensitivity and specificity of the technique. Future studies will also need to refine the automated selection of artifact and focus on understanding how individual CI speech processors and electrode arrays affect the type of artifact observed on the scalp.

III. Perceptual learning of speech

Experiment 2 documented that multiple talker phonetic identification training improved perception of naturally produced /ba/-/da/ and /wa/-/ja/ speech contrasts by an average of 11.5% in postlingually deafened adult CI users, and listeners generalized their learning to unfamiliar talkers. The success of the training protocol used in this study sheds light on the underlying perceptual mechanisms that support robust phonetic category acquisition in adult CI listeners.

In a comprehensive review of perceptual learning, Goldstone (1998) outlined the following four mechanisms of how learning occurs: attention-weighting, differentiation, imprinting, and unitization. Perceptual learning can be defined as any relatively long-lasting change in a perceptual system to a stimulus array that improves one's ability to respond to the environment following practice or experience with the stimulus array (Gibson, 1963; Goldstone, 1998). In his review, Goldstone (1998) highlighted the role of an attention-weighting mechanism in the perceptual learning of phonetic categories.

Attention-weighting refers to the increased attention paid to important perceptual dimensions and the decreased attention given to irrelevant dimensions. Through this perceptual learning mechanism, listeners can be trained to selectively direct attention toward important stimulus aspects that aid in categorization at different levels of information processing.

The categorical perception of native speech sounds (Liberman, et al., 1957; Studdert-Kennedy, et al., 1970) is a classic example of perceptual adaptation based on an attention-weighting mechanism where perception is heightened for stimuli that cross a phonetic boundary relative to stimuli from the same phonetic category. Whether this perceptual adaptation of phonetic categories is a domain general or domain specific process is the subject of much debate, but previous studies have documented that phonetic categorization is subject to learning. Evidence that adult listeners can reshape phonetic categories to accommodate variations in input comes from training studies examining category acquisition of nonnative contrasts (Lively, et al., 1993), accented speech (Bradlow & Bent, 2008), and degraded vocoded speech (M. H. Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005) (see Samuel & Kraljic, 2009 for a review) . The success of certain training protocols over others in promoting robust and stable phonetic categories highlights the role attention-weighting plays in phonetic learning. Identification training protocols that promote perceptual grouping of similar stimuli, rather than discrimination training that promotes perceptual distinctions between stimuli, have been shown to induce more plasticity for the perceptual learning of phonetic categories (Pisoni & Lively, 1995). In other words, identification protocols are designed to target an attention-weighting mechanism, as opposed to discrimination training that targets a differentiation mechanism, and produce greater gains in phonetic category acquisition. The results of Experiment 2 suggest that targeting an attention-weighting mechanism via high variability identification training can improve phoneme perception in experienced adult CI listeners. The trained CI listeners were able to generalize learning to unfamiliar talkers, suggesting they had abstracted higher order category information. Whether or not the phonetic learning we observed remained stable is unclear as listeners were not brought back to the lab for testing after an extended period of time. While the results of Experiment 2 alone cannot rule out that simple time-on-task or general improvements in focused attention accounted for the training-related improvements, this explanation would be unlikely as other previous work has shown training protocols that require sustained visual memory training have minimal effects on speech perception performance in adult CI users (Oba, Galvin, & Fu, 2013).

Experiment 2 trained CI listeners using a training protocol designed to target an attention-weighting mechanism and found significant improvements in phoneme perception. Experiment 3 empirically tested whether the improvement in the trained listeners' perception of naturally produced speech was due to a perceptual retuning of attention by monitoring fine scale behavioral and neural correlates of phonetic categorization before and after training. The EEG data from Experiment 3 showed an enhanced MMN response to an across phoneme category pair relative to an acoustically equivalent within phoneme category pair for both the /ba/-/da/ and /wa/-/ja/ contrasts. In addition, the behavioral results suggested training significantly enhanced behavioral sensitivity across the phonetic boundary for the /ba/-/da/ contrast, but only marginally altered identification and discrimination of the /wa/-/ja/ stimuli. Of importance, though, significant correlations between our behavioral and neural sensitivity measures of phonetic categorization were observed for both speech contrasts. Taken together, the brain and behavior data from Experiment 3 suggest that enhanced phonetic categorization skills supported the group level improvements in phonemic perception observed in Experiment 2. It is unlikely that these results can be explained as a change in basic auditory sensitivity, as opposed to changes in selective attention, as listeners' behavioral and neural measures of sensitivity for the nonspeech correlates did not change with training. The lack of change in the untrained listening group makes it unlikely that procedural learning could explain the results either.

It is important to emphasize that the improvements in phonetic categorization observed in the present study occurred at the group level. However, consistent with previous work examining CI patient outcomes (see N.R. Peterson, Pisoni, & Miyamoto, 2010 for a review), significant variability in the amount of phonetic learning was observed across individual CI users in the present study. While there is no single metric that has been shown to predict individual speech perception outcomes in CI recipients, duration of deafness and the number of surviving spiral ganglion cells, etiology of hearing loss, communication modality, cognitive factors, and years of CI are all likely sources of the variability in speech perception abilities among CI users (see Holden et al., 2013 for a review).

IV. Neural basis of perceptual learning of speech

Consistent with previous studies, Experiment 3 provides evidence that training induced changes in phonetic learning were reflected by the MMN, a neural marker of sensitivity (e.g. Y. Zhang et al., 2009). The exact nature of how neuronal ensembles in the brain were altered as a result of training is unknown, but several scenarios are possible. Training-related improvements could be represented in the nervous system by 1) an increase in neural firing or cortical recruitment, 2) sharper tuning of cells (or a decrease in the size of the neuronal ensemble), or 3) a change in temporal coherence or neural synchrony (Gilbert, Sigman, & Crist, 2001). Because EEG is a far field response measured at the scalp with limited spatial resolution, theories as to how the central nervous system was altered by training in the present study remain speculative at best. Whatever the mechanism, the EEG data and MMN responses were collected without having the listener attend to the stimuli, thus, it seems reasonable to assume that the listening experience altered some lower level perceptual process as well as higher-level categorization skills.

V. Success and limitations of training protocol and individual variability

At the group level, the training program improved phonetic categorization of difficult speech contrasts for a group of trained CI listeners. However, significant individual variability in phonetic learning was observed. Examination of individual behavioral and neural sensitivity measures results suggests that the more successful learners (individuals whose d' scores for naturally produced speech improved) appear to display better phonetic categorization skills (Figures 5.1 and 5.2). Figure 5.1 plots the behavioral identification and across category MMN for the /ba/-/da/ stimuli at pre-and posttest for an individual whose d' for the naturally produced /ba/-/da/ speech contrast improved by 0.8 with training (from 1.9 at pretest to 2.7 at posttest).

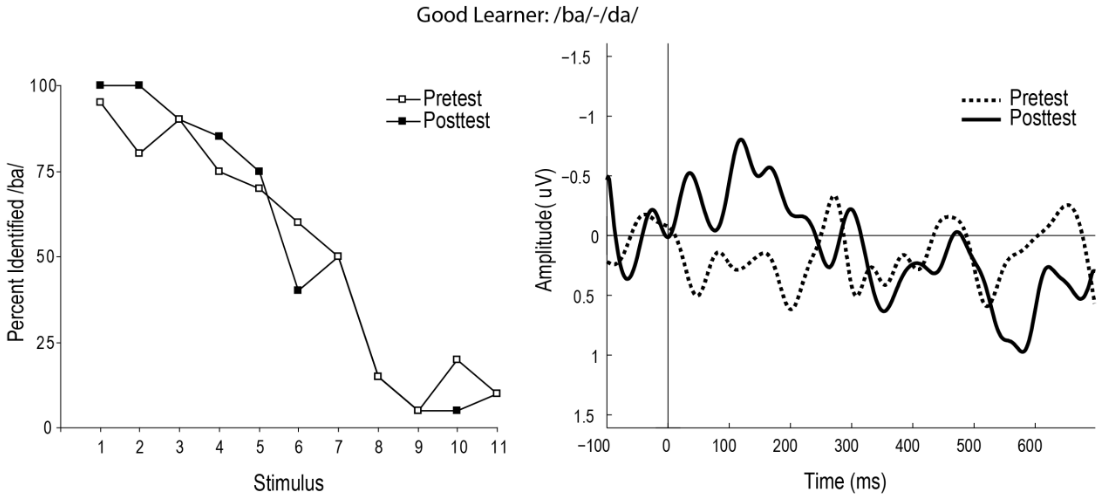


Fig. 5.1: Individual data for a subject whose d' for the naturally produced /ba/-/da/ contrast improved by 0.8. A) Behavioral identification of the synthetic /ba/-/da/ continuum at pre-and posttest intervals. B) MMN difference waveforms from pre- and posttests evoked by the across-category /ba/-/da/ stimulus pair.

In contrast, 5.2 plots the behavioral identification and across category MMN for the /ba/-/da/ stimuli at pre-and posttest for an individual who displayed minimal training effects (d' of 1.2 at pretest and 1.3 at posttest).

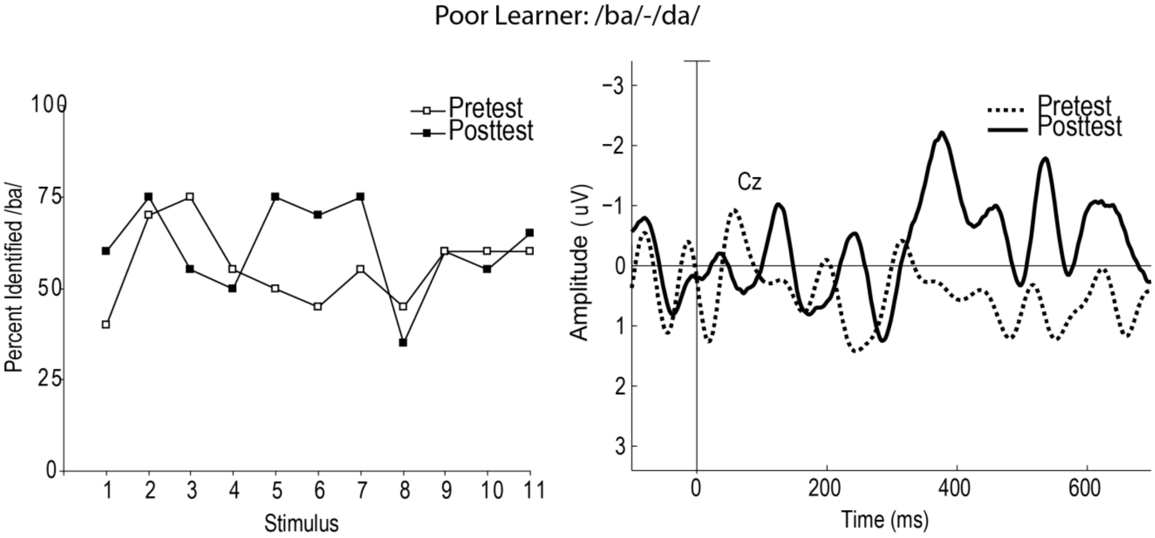


Fig. 5.2: Individual data for a subject whose d' for the naturally produced /ba/-/da/ contrast did not change with training. A) Behavioral identification of the synthetic /ba/-/da/ continuum at pre-and posttest intervals. B) MMN difference waveforms from pre-and posttests evoked by the across-category /ba/-/da/ stimulus pair.

What limits learning is unknown, and further work will be needed to assess the device and patient-centered sources of variability. Of note, the less successful learner (Fig 5.2) had a much longer duration of deafness before receiving the implant than the successful learner. Previous work has shown that duration of deafness and age of implantation can be correlated with later speech perception skills (Eggermont & Ponton, 2003) as well as auditory evoked potential peak latencies (Eggermont, Ponton, Don, Waring, & Kwong, 1997; Sharma, Dorman, Spahr, et al., 2002; Sharma, Dorman, & Spahr, 2002).

VI. Future Directions

The observed variability in patient outcomes for CI recipients highlights the need for more research regarding formal auditory training and perceptual learning of speech. That the majority of gain in speech perception experienced by adult CI listeners occurs within the first 3-6 months, with some patients receiving little to no benefit from their devices, makes it clinically imperative to determine whether active learning can improve outcomes. This dissertation provides evidence that training can alter the neural coding of speech, which has obvious clinical implications, but many important questions remain. Of primary importance, whether or not objective neurophysiologic measures can assess phonetic learning at the individual level in CI recipients remains to be determined. The double-oddball paradigm has the potential to be used clinically on individual patients, but significant research focusing on the development of age-appropriate norms is needed first. Future studies will also need to examine how device and patient-related variables affect learning and whether individualized training protocols need to be developed. For example, it is possible that the less successful learners in this study had stronger neural commitments to previously learned language patterns that required more training or a different training paradigm that included signal enhancement or visual articulation cues (Y. Zhang, et al., 2009). Phonetic learning of different speech contrasts and whether phonetic categorization skills generalize to improved word and sentence recognition remains to be determined as well.

VII. Conclusion

The series of studies included in this dissertation provide evidence that electroencephalograpy (EEG) can be used as a noninvasive and objective tool for measuring how listening experience alters neural processing of speech in adult CI listeners. In sum, there are two main findings of the present work. First, the results suggest that substantial neural plasticity for phonetic learning in adult CI recipients can be induced using high variability identification training. Second, for the trained listeners, changes in behavioral sensitivity to the phonetic contrasts were reflected in the electrophysiologic MMN response. The fine grained behavioral and neural measurements suggest the underlying mechanism for the improved phoneme perception we observed was due to enhanced phonetic categorization. These results have potentially important clinical implications related to the aural rehabilitation process following receipt of a CI device.

References

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Sci Am, 225*(2), 82-90.

Bakhos, D., Roux, S., Robier, A., Bonnet-Brilhault, F., Lescanne, E., & Bruneau, N. (2012). Minimization of cochlear implant artifact in cortical auditory evoked potentials in children. *Int J Pediatr Otorhinolaryngol, 76*(11), 1627-1632.

Benasich, A. A., & Tallal, P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. *Behav Brain Res, 136*(1), 31-49.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transitionfrom speech sounds to spoken words*. Cambridge, MA: MIT Press.

Bierer, J. A. (2007). Threshold and channel interaction in cochlear implant users: evaluation of the tripolar electrode configuration. *J Acoust Soc Am, 121*(3), 1642-1653.

Bierer, J. A., Bierer, S. M., & Middlebrooks, J. C. (2010). Partial tripolar cochlear implant stimulation: Spread of excitation and forward masking in the inferior colliculus. *Hear Res, 270*(1-2), 134-142.

Bierer, J. A., & Faulkner, K. F. (2010). Identifying cochlear implant channels with poor electrode-neuron interface: partial tripolar, single-channel thresholds and psychophysical tuning curves. *Ear Hear, 31*(2), 247-258.

Boothroyd, A. (2010). Adapting to changed hearing: the potential role of formal training. *J Am Acad Audiol, 21*(9), 601-611.

Borden, G. J., & Harris, K. S. (1980). *Speech science primer : physiology, acoustics, and perception of speech*. Baltimore: Williams & Wilkins.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707-729.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Percept Psychophys, 61*(2), 206-219.

Brown, C. J., Etler, C., He, S., O'Brien, S., Erenberg, S., Kim, J. R., et al. (2008). The electrically evoked auditory change complex: preliminary results from nucleus cochlear implant users. *Ear Hear, 29*(5), 704-717.

Carney, A. E. (1977). Noncategorical perception of stop consonants differing in VOT. *J Acoust Soc Am, 62*(4), 961-970.

Ceponiene, R., Rinne, T., & Naatanen, R. (2002). Maturation of cortical sound processing as indexed by event-related potentials. *Clin Neurophysiol, 113*(6), 870-882.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin Neurophysiol, 115*(4), 732-744.

Davis, H., Hirsh, S. K., Shelnutt, J., & Bowers, C. (1967). Further validation of evoked response audiometry (ERA). *J Speech Hear Res, 10*(4), 717-732.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen, 134*(2), 222-241.

Dawson, P. W., & Clark, G. M. (1997). Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant. *Ear Hear, 18*(6), 488-501.

Debener, S., Hine, J., Bleeck, S., & Eyles, J. (2008). Source localization of auditory evoked potentials after cochlear implantation. *Psychophysiology, 45*(1), 20-24.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21.

Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., et al. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Comput Intell Neurosci, 2011*, 130714.

Digeser, F. M., Wohlberedt, T., & Hoppe, U. (2009). Contribution of spectrotemporal features on auditory event-related potentials elicited by consonant-vowel syllables. *Ear Hear, 30*(6), 704-712.

Dorman, M. F., & Loizou, P. C. (1997). Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants. *Am J Otol, 18*(6 Suppl), S113-114.

Dorman, M. F., Loizou, P. C., Spahr, A. J., & Maloff, E. (2002). Factors that allow a high level of speech understanding by patients fit with cochlear implants. *Am J Audiol, 11*(2), 119-123.

Dorman, M. F., Soli, S., Dankowski, K., Smith, L. M., McCandless, G., & Parkin, J. (1990). Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant. *J Acoust Soc Am, 88*(5), 2074-2079.

Eggermont, J. J., & Ponton, C. W. (2003). Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: correlations with changes in structure and speech perception. *Acta Otolaryngol, 123*(2), 249-252.

Eggermont, J. J., Ponton, C. W., Don, M., Waring, M. D., & Kwong, B. (1997). Maturational delays in cortical evoked potentials in cochlear implant users. *Acta Otolaryngol, 117*(2), 161-163.

Fant, G. (1960). *Acoustic theory of speech production, with calculations based on X-ray studies of Russian articulations*. s'Gravenhage,: Mouton.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol Rev, 116*(4), 752-782.

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: Don't shoot the messenger. *Child Development, 71*(2), 323-328.

Firszt, J. B., Chambers, R. D., Kraus, & Reeder, R. M. (2002). Neurophysiology of cochlear implant users I: effects of stimulus current level and electrode site on the electrical ABR, MLR, and N1-P2 response. *Ear Hear, 23*(6), 502-515.

Flanagan, J. L. (1965). *Speech analysis, synthesis, and perception*. New York,: Academic Press.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *J Acoust Soc Am, 97*(5 Pt 1), 3125-3134.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Lang Speech, 38 (Pt 1)*, 25-55.

Friesen, L. M., & Picton, T. W. (2010). A method for removing cochlear implant artifact. *Hear Res, 259*(1-2), 95-106.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am, 110*(2), 1150-1163.

Friesen, L. M., & Tremblay, K. L. (2006). Acoustic change complexes recorded in adult cochlear implant listeners. *Ear Hear, 27*(6), 678-685.

Fu, Q. J., Galvin, J., Wang, X., & Nogaki, G. (2005). Moderate auditory training can improve speech performance of adult cochlear implant patients. *Acoustics Research Letters Online-Arlo, 6*(3), 106-111.

Fu, Q. J., & Galvin, J. J., 3rd. (2007). Perceptual learning and auditory training in cochlear implant recipients. *Trends Amplif, 11*(3), 193-205.

Fu, Q. J., & Galvin, J. J., 3rd. (2008). Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hear Res, 242*(1-2), 198-208.

Gelfand, S. A. (2004). *Hearing : an introduction to psychological and physiological acoustics* (4th ed.). New York: Marcel Dekker.

Gibson, E. J. (1963). Perceptual learning. *Annu Rev Psychol, 14*, 29-56.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron, 31*(5), 681-697.

Gilley, P. M., Sharma, A., Dorman, M., Finley, C. C., Panch, A. S., & Martin, K. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. *Clin Neurophysiol, 117*(8), 1772-1782.

Goldstone, R. L. (1998). Perceptual learning. *Annu Rev Psychol, 49*, 585-612.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *J Acoust Soc Am, 103*(5 Pt 1), 2677-2690.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24*(4), 495-522.

Hari, R. (1991). Activation of the human auditory cortex by speech sounds. *Acta Otolaryngol Suppl, 491*, 132-137; discussion 138.

Hazan, V., & Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. *Percept Psychophys, 49*(2), 187-200.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual perception and production of training on the consonants by Japanese learners of English. *Speech Communication, 47*(3), 360-378.

Henshaw, H., & Ferguson, M. A. (2013). Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence. *PLoS One, 8*(5), e62836.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J Acoust Soc Am, 97*(5 Pt 1), 3099-3111.

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., et al. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear, 34*(3), 342-360.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Atten Percept Psychophys, 72*(5), 1218-1227.

Huotilainen, M., Winkler, I., Alho, K., Escera, C., Virtanen, J., Ilmoniemi, R. J., et al. (1998). Combined mapping of human auditory EEG and MEG responses. *Electroencephalogr Clin Neurophysiol, 108*(4), 370-379.

Ingvalson, E. M., Lee, B., Fiebig, P., & Wong, P. C. (2013). The effects of short-term computerized speech-in-noise training on postlingually deafened adult cochlear implant recipients. *J Speech Lang Hear Res, 56*(1), 81-88.

Ingvalson, E. M., & Wong, P. C. (2013). Training to improve language outcomes in cochlear implant recipients. *Front Psychol, 4*, 263.

Iverson, P. (2003). Evaluating the function of phonetic perceptual phenomena within speech recognition: an examination of the perception of /d/-/t/ by adult cochlear implant users. *J Acoust Soc Am, 113*(2), 1056-1064.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. *J Acoust Soc Am, 118*(5), 3267-3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47-57.

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: a comparison of the prototype and perceptual fading techniques. *Can J Psychol, 43*(1), 88-96.

Johnson, K. (2003). *Acoustic and auditory phonetics* (2nd ed.). Malden, Mass.: Blackwell Pub.

Kelly, A. S., Purdy, S. C., & Thorne, P. R. (2005). Electrophysiological and speech perception measures of auditory processing in experienced adult cochlear implant users. *Clin Neurophysiol, 116*(6), 1235-1246.

Kraus, N., McGee, T., Carrell, T. D., & Sharma, A. (1995). Neurophysiologic bases of speech discrimination. *Ear Hear, 16*(1), 19-37.

Kraus, N., Micco, A. G., Koch, D. B., McGee, T., Carrell, T., Sharma, A., et al. (1993). The mismatch negativity cortical evoked potential elicited by speech in cochlear-implant users. *Hear Res, 65*(1-2), 118-124.

Krueger, B., Joseph, G., Rost, U., Strauss-Schier, A., Lenarz, T., & Buechner, A. (2008). Performance groups in adult cochlear implant users: speech perception results from 1984 until today. *Otol Neurotol, 29*(4), 509-512.

Kuhl, P. K. (2000). A new view of language acquisition. *Proc Natl Acad Sci U S A, 97*(22), 11850-11857.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos Trans R Soc Lond B Biol Sci, 363*(1493), 979-1000.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the “Critical Period. *Language Learning and Development, 1*, 237–264.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev Sci, 9*(2), F13-F21.

Kuhl, P. K., Tsao, F. M., Liu, H. M., Zhang, Y., & De Boer, B. (2001). Language/culture/mind/brain. Progress at the margins between disciplines. *Ann N Y Acad Sci, 935*, 136-174.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK ; Cambridge, Mass.: Blackwell.

Lane, H., Denny, M., Guenther, F. H., Hanson, H. M., Marrone, N., Matthies, M. L., et al. (2007). On the structure of phoneme categories in listeners with cochlear implants. *J Speech Lang Hear Res, 50*(1), 2-14.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York,: Wiley.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol, 54*(5), 358-368.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *J Acoust Soc Am, 94*(3 Pt 1), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *J Acoust Soc Am, 96*(4), 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *J Acoust Soc Am, 89*(2), 874-886.

Loizou, P. C. (1998). Mimicking the Human Ear. *IEEE Signal Processing, 15*(5), 101-130.

Lonka, E., Kujala, T., Lehtokoski, A., Johansson, R., Rimmanen, S., Alho, K., et al. (2004). Mismatch negativity brain response as an index of speech perception recovery in cochlear-implant recipients. *Audiol Neurootol, 9*(3), 160-162.

Lopez-Valdes, A., Mc Laughlin, M., Viani, L., Walshe, P., Smith, J., Zeng, F. G., et al. (2013). Auditory mismatch negativity in cochlear implant users: a window to spectral discrimination. *Conf Proc IEEE Eng Med Biol Soc, 2013*, 3555-3558.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass.: MIT Press.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369-390.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide* (2nd ed.). New York: Cambridge University Press.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn Sci, 8*(5), 204-210.

Martin, B. A., Tremblay, K. L., & Korczak, P. (2008). Speech evoked potentials: from the laboratory to the clinic. *Ear Hear, 29*(3), 285-313.

Mc Laughlin, M., Lopez Valdes, A., Reilly, R. B., & Zeng, F. G. (2013). Cochlear implant artifact attenuation in late auditory evoked potentials: a single channel approach. *Hear Res, 302*, 84-95.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogn Affect Behav Neurosci, 2*(2), 89-108.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cogn Psychol, 18*(1), 1-86.

Miller, S., & Zhang, Y. (2014). Validation of the cochlear implant artifact correction tool for auditory electrophysiology. *Neurosci Lett, 577C*, 51-55.

Miyawaki, K., Liberman, A. M., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., et al. (1975). Effect of Linguistic Experience - Discrimination of [R] and [L] by Native Speakers of Japanese and English. *Perception & Psychophysics 18*(5), 331-340.

Moberly, A. C., Lowenstein, J. H., Tarr, E., Caldwell-Tarr, A., Welling, D. B., Shahin, A. J., et al. (2014). Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? *J Speech Lang Hear Res, 57*(2), 566-582.

Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: awakening the deafened brain. *Nat Neurosci, 12*(6), 686-691.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *J Acoust Soc Am, 85*(1), 365-378.

Munson, B., & Nelson, P. B. (2005). Phonetic identification in quiet and in noise by listeners with cochlear implants. *J Acoust Soc Am, 118*(4), 2607-2617.

Näätänen, R. (2003). Mismatch negativity: clinical research and possible applications. *Int J Psychophysiol, 48*(2), 179-188.

Näätänen, R., Gaillard, A. W., & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol (Amst), 42*(4), 313-329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432-434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol, 118*(12), 2544-2590.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept Psychophys, 60*(3), 355-376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech Perception as a Talker-Contingent Process. *Psychol Sci, 5*(1), 42-46.

Oba, S. I., Fu, Q. J., & Galvin, J. J. (2011). Digit Training in Noise Can Improve Cochlear Implant Users' Speech Understanding in Noise. *Ear and Hearing, 32*(5), 573-581.

Oba, S. I., Galvin, J. J., 3rd, & Fu, Q. J. (2013). Minimal effects of visual memory training on auditory performance of adult cochlear implant users. *J Rehabil Res Dev, 50*(1), 99-110.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychol Rev, 85*(3), 172-191.

Ostroff, J. M., Martin, B. A., & Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear Hear, 19*(4), 290-297.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *J Exp Psychol Learn Mem Cogn, 19*(2), 309-328.

Pantev, C., Dinnesen, A., Ross, B., Wollbrink, A., & Knief, A. (2006). Dynamics of auditory plasticity after cochlear implantation: a longitudinal study. *Cereb Cortex, 16*(1), 31-36.

Peterson, G. E. (1959). Vowel formant measurements. *J Speech Hear Res, 2*(2), 173-183.

Peterson, N. R., Pisoni, D. B., & Miyamoto, R. T. (2010). Cochlear implants and spoken language processing abilities: review and assessment of the literature. *Restor Neurol Neurosci, 28*(2), 237-250.

Pickles, J. O. (2012). *An introduction to the physiology of hearing* (4th ed.). New York: Academic.

Pisoni, D., & Levi, S. V. (2007). Representations and representational specificity in speech perception and spoken word recognition. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.

Pisoni, D., & Lively, S. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 433-459): York Press.

Ponton, C. W., Don, M., Eggermont, J. J., Waring, M. D., Kwong, B., & Masuda, A. (1996). Auditory system plasticity in children after long periods of complete deafness. *Neuroreport, 8*(1), 61-65.

Ponton, C. W., Eggermont, J. J., Don, M., Waring, M. D., Kwong, B., Cunningham, J., et al. (2000). Maturation of the mismatch negativity: effects of profound deafness and cochlear implant use. *Audiol Neurootol, 5*(3-4), 167-185.

Ponton, C. W., Moore, J. K., & Eggermont, J. J. (1999). Prolonged deafness limits auditory system developmental plasticity: evidence from an evoked potentials study in children with cochlear implants. *Scand Audiol Suppl*, *51*, 13-22.

Qi, B., Krenmayr, A., Zhang, N., Dong, R., Chen, X., Schatzer, R., et al. (2012). Effects of temporal fine structure stimulation on Mandarin speech recognition in cochlear implant users. *Acta Otolaryngol*, *132*(11), 1183-1191.

Rahne, T., Plontke, S. K., & Wagner, L. (2014). Mismatch negativity (MMN) objectively reflects timbre discrimination thresholds in normal-hearing listeners and cochlear implant users. *Brain Res, 1586*, 143-151.

Rao, A., Zhang, Y., & Miller, S. (2010). Selective listening of concurrent auditory stimuli: An event-related potential study. *Hearing Research, 268*(1-2), 123-132.

Reite, M., Teale, P., Zimmerman, J., Davis, K., & Whalen, J. (1988). Source Location of a 50 Msec Latency Auditory Evoked Field Component. *Electroencephalography and Clinical Neurophysiology, 70*(6), 490-498.

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behav Brain Res, 111*(1-2), 13-23.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci, 336*(1278), 367-373.

Rubinstein, J. T. (2004). How cochlear implants encode speech. *Curr Opin Otolaryngol Head Neck Surg, 12*(5), 444-448.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Atten Percept Psychophys, 71*(6), 1207-1218.

Shannon, R. V. (2002). The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. *Am J Audiol, 11*(2), 124-127.

Sharma, A., Dorman, M., Spahr, A., & Todd, N. W. (2002). Early cochlear implantation in children allows normal development of central auditory pathways. *Ann Otol Rhinol Laryngol Suppl, 189*, 38-41.

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear Hear, 23*(6), 532-539.

Sharma, A., Marsh, C. M., & Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *J Acoust Soc Am, 108*(6), 3030-3035.

Spahr, A. J., & Dorman, M. F. (2004). Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices. *Arch Otolaryngol Head Neck Surg, 130*(5), 624-628.

Stacey, P. C., Raine, C. H., O'Donoghue, G. M., Tapper, L., Twomey, T., & Summerfield, A. Q. (2010). Effectiveness of computer-based auditory training for adult users of cochlear implants. *Int J Audiol, 49*(5), 347-356.

Stacey, P. C., & Summerfield, A. Q. (2008). Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally distorted speech. *J Speech Lang Hear Res, 51*(2), 526-538.

Stapells, D. R., Gravel, J. S., & Martin, B. A. (1995). Thresholds for auditory brain stem responses to tones in notched noise from infants and young children with normal hearing or sensorineural hearing loss. *Ear Hear, 16*(4), 361-371.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Percept Psychophys, 36*(2), 131-145.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Theoretical notes. Motor theory of speech perception: a reply to Lane's critical review. *Psychological Review, 77*(3), 234-249.

Svirsky, M. A., Silveira, A., Suarez, H., Neuburger, H., Lai, T. T., & Simmons, P. M. (2001). Auditory learning and adaptation after cochlear implantation: a preliminary study of discrimination and labeling of vowel sounds by cochlear implant users. *Acta Otolaryngol, 121*(2), 262-265.

Tavabi, K., Obleser, J., Dobel, C., & Pantev, C. (2007). Auditory evoked fields differentially encode speech features: an MEG investigation of the P50m and N100m time courses during syllable processing. *Eur J Neurosci, 25*(10), 3155-3162.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *J Acoust Soc Am, 102*(6), 3762-3773.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport, 9*(16), 3557-3560.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear, 22*(2), 79-90.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Dev, 75*(4), 1067-1084.

Van Tasell, D. J. (1993). Hearing loss, speech, and hearing aids. *J Speech Hear Res, 36*(2), 228-244.

Viola, F. C., De Vos, M., Hine, J., Sandmann, P., Bleeck, S., Eyles, J., et al. (2012). Semi-automatic attenuation of cochlear implant artifacts for the evaluation of late auditory evoked potentials. *Hear Res, 284*(1-2), 6-15.

Viola, F. C., Thorne, J. D., Bleeck, S., Eyles, J., & Debener, S. (2011). Uncovering auditory evoked potentials from cochlear implant users with independent component analysis. *Psychophysiology, 48*(11), 1470-1480.

Visto, J. C., Cranford, J. L., & Scudder, R. (1996). Dynamic temporal processing of nonspeech acoustic information by children with specific language impairment. *J Speech Hear Res, 39*(3), 510-517.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *J Acoust Soc Am, 75*(6), 1866-1878.

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: effects of spectral degradation and electric hearing. *J Acoust Soc Am, 131*(2), 1465-1479.

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *J Acoust Soc Am, 137*(3), 1430.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience, 170*(1), 223-231.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Percept Psychophys, 52*(4), 376-392.

Zaehle, T., Jancke, L., & Meyer, M. (2007). Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features. *Behav Brain Funct, 3*, 63.

Zhang, F., Benson, C., & Fu, Q. J. (2013). Cortical Encoding of Pitch Contour Changes in Cochlear Implant Users: A Mismatch Negativity Study. *Audiol Neurootol, 18*(5), 275-288.

Zhang, F., Hammer, T., Banks, H. L., Benson, C., Xiang, J., & Fu, Q. J. (2011). Mismatch negativity and adaptation measures of the late auditory evoked potential in cochlear implant users. *Hear Res, 275*(1-2), 17-29.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., et al. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *Neuroimage, 46*(1), 226-240.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *Neuroimage, 26*(3), 703-720.

Zhang, Y., & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition, 10*(2), 147–160.

Zlatin, M. A., & Koenigsknecht, R. A. (1976). Development of the voicing contrast: a comparison of voice onset time in stop preception and production. *J Speech Hear Res, 19*(1), 93-111.