Understanding Factors Contributing to Variability in Outcomes of Cochlear-Implant Users

A Dissertation
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Erin R. O'Neill

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

Advisor: Andrew J. Oxenham, PhD

June 2020

Acknowledgments

Five years ago, I sat across from Andrew Oxenham and expressed my desire to complete a PhD in Cognitive and Brain Sciences, having no previous research experience and a less than stellar college GPA. I am eternally thankful that Andrew saw my passion for knowledge and took a chance on me. Andrew was also open to me pursuing my own research ideas, even though several of them fell outside his typical wheelhouse. Though we approach cochlear-implant research from different angles and perspectives, this dissertation is a true testament to collaboration and mutual respect between us.

The opportunity to pursue the research topics covered in Chapter 6 would not have been possible without the guidance, mentoring, and resources of Peggy Nelson and technical support from John Basile. I am incredibly thankful to both of them for making that study possible.

I would also like to thank Heather Kreft and Morgan Parke for their support in the completion of this dissertation. Heather was a tremendous asset in regards to data collection, science and life wisdom, stress management, and much needed lunch outings on campus. Morgan was a long-term undergraduate research assistant who contributed countless hours to data collection and was instrumental in the development of the corpus of nonsense sentences (Chapter 4).

This dissertation would also not have been possible without the patience and dedication of my cochlear-implant participants and older normal-hearing listeners. I want to give a special thanks to my family and friends who were willing to endure hours of listening torture to help me complete my PhD.

Finally, I would like to thank my parents, my labmates, and Brian Hotchkiss for both listening to me air my frustrations and providing much needed encouragement during the course of my PhD.

Abstract

Cochlear implants (CIs) are a beneficial and often life-changing treatment for individuals with profound sensorineural hearing loss. However, despite advances in device design and surgical techniques, clinicians continue to see a very wide range of outcomes for individual CI users. Some CI users are able to converse successfully over the phone and in noisy environments, while others struggle to communicate effectively even in quiet, one-on-one conversations. Clinical differences between patients, such as the onset and duration of deafness, and anatomical factors related to the electrode-to-neuron interface, explain relatively little of this overall variability in speech perception. This dissertation addresses various perceptual, cognitive, and social factors, such as spectral resolution, working memory, intellectual efficiency, social engagement, and coping strategies, which may account for some of the individual differences in speech understanding found in the CI population. First, the amount of variability inherent to difficult auditory tasks was assessed by measuring the variance in the perception of degraded speech in the normal-hearing population. Associations between the perception of degraded speech, working memory, and intellectual efficiency were also explored in CI users and both young and age-matched normal-hearing controls. The listening strategy of "filling in the blanks" or leveraging context to understand words that are not heard or misheard was also explored by using novel sentences with and without semantic context. Finally, ecological momentary assessments were used to assess social and listening behavior of CI users outside of the lab setting, by having participants complete short surveys on their smartphones while engaging in normal daily activities. Results indicate that listening difficulty associated with degraded speech perception and working memory abilities account for some variance seen in outcomes of CI users. Strategies such as leveraging semantic context and using visual cues to supplement auditory information are also widely used in the CI population. Finally, greater levels of social engagement were associated with better speech perception outcomes in individual CI users. These new insights into cognitive and social factors influencing outcomes in CI users could be used by clinicians to tailor rehabilitation and manage expectations in individual patients.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

The cochlear implant (CI) is a unique auditory prosthesis that uses electrical pulses to stimulate the auditory nerve, providing a means to access sound for individuals with profound sensorineural hearing loss. Since its inception in the 1960s, the CI has undergone several changes and improvements, progressing from a single-channel to a multi-electrode device, initially providing no open-set speech recognition but, through improved processing strategies developed in the 1990s, now conveying enough sound information for many recipients to converse on the telephone (Zeng, Rebscher, Harrison, Sun, & Feng, 2008). In its modern form, the CI is manufactured by five different companies worldwide, with internally implanted electrode arrays ranging from 12 to 22 electrodes and external speech processors employing strategies such as continuous interleaved sampling, current steering, and noise reduction algorithms that together allow the average user to identify 60% of words in quiet.

The device itself uses an external microphone, located behind the ear or on the side of the head, to pick up sound which is then converted to a digital signal, processed and transmitted across the skull via a radio frequency which is received by an internal coil, implanted on the inside of the skull near the mastoid bone (see Figure 1.1). An internal stimulator then decodes the signal and converts it into electrical pulses, which are sent down wires and into electrodes threaded through the scala tympani of the cochlea. The electrodes, which are interspersed along the cochlea, ensure that electrical pulses filtered into different frequency ranges reach specific neural populations to mimic normal auditory frequency selectivity and stimulate the auditory nerve in such a way that the brain will interpret these pulses as organized sound.

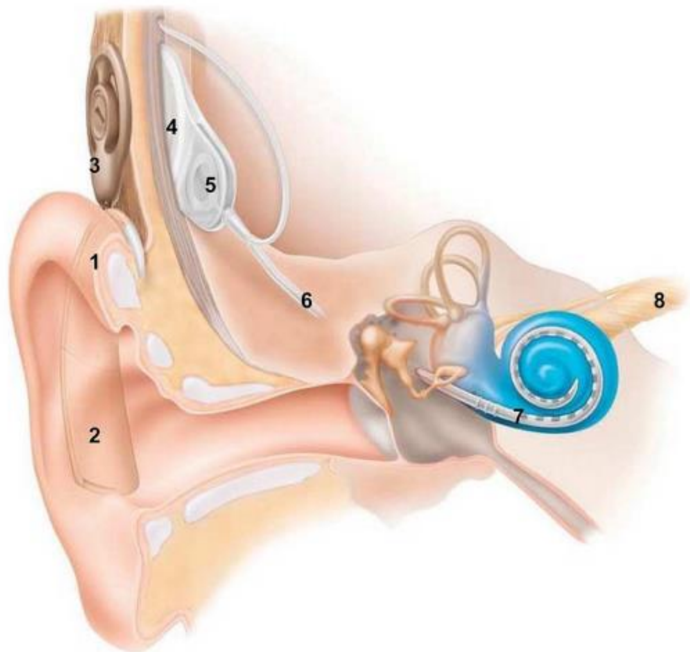

Figure 1.1

Diagram of a cochlear implant on a human patient. The numbers on the figure mark 1) the behind-the-ear external processor with ear hook, 2) the battery case, 3) the headpiece with magnet and coil, 4) the internal receiver with magnet and coil, 5) the simulator, 6) wires threaded into the cochlea, 7) the electrode array, and 8) the auditory nerve. Image from Zeng et al. (2008).

While the ingenuity and success of the CI in restoring a sense of hearing to those with profound hearing loss is undisputed, the variability in speech perception outcomes among recipients has puzzled scientists for some time (Hast, Schlücker, Digeser, Liebscher, & Hoppe, 2015). Many studies have investigated the possible connection between speech perception outcomes and peripheral factors, such as insertion depth of the electrode array and electrode-neural distance (Van Der Marel, Briaire, Verbist, Muurling, & Frijns, 2015); etiology of hearing loss, duration of deafness and age at implantation (Cosetti & Waltzman, 2012; Mahmoud & Ruckenstein, 2014); and the frequency selectivity and survival of neurons in the cochlea, in conjunction with the overall spectral resolution of electrical sound (Anderson, Nelson, Kreft, Nelson, & Oxenham, 2011; Choi et al., 2018; Henry & Turner, 2003; Won, Drennan, & Rubinstein, 2007). More cognitive factors may also be a source of variance in outcomes of CI users, with several studies exploring possible connections between working memory, general intelligence, and cognitive control with speech perception (Akeroyd, 2008; Arehart, Souza, Baca, & Kates, 2013; Hua, Johansson, Magnusson, Lyxell, & Ellis, 2017; Lunner, 2003; Moberly, Houston, Harris, Adunka, & Castellanos, 2017; Rönnberg et al., 2013). While the role of factors like duration of deafness and neural survival, in relation to hearing outcomes, have proven to be fairly clear cut, the contribution and interplay between many other factors, both peripheral and cognitive, that could be influencing performance in CI users remain highly debated. One goal of this chapter is to explore individual differences in the CI population and their normal-hearing (NH) peers on auditory tasks of equal difficulty and relate performance to peripheral, cognitive, psychosocial, and perceptual learning factors that may be influencing variability.

The remainder of this chapter is divided into three sections. The following section reviews peripheral factors that may contribute to variability in the speech perception of CI users, including surgical differences, etiology, duration of deafness, years of CI use, age at implantation and spectral resolution. The next section explores non-peripheral factors, such as working memory, general intelligence, cognitive control, listening strategies, psychological traits and social behaviors that may influence hearing outcomes in CI users. The final section describes our initial study of individual differences in both perceptual and cognitive tasks in CI users and NH listeners.

1.1 Review of Peripheral Factors Impacting Variability

1.1.1 Clinical and surgical differences within the CI population

One reason why it has been difficult for auditory scientists and clinical audiologists to pinpoint factors contributing most to the variability in outcomes among CI users is that the CI population itself is very heterogeneous. As of 2012, roughly 60% of the just under 100,000 people in the United States that have received CIs are adults and 40% are children ("Quick Statistics About Hearing," 2016). Within the adult population, the majority of recipients (85 to 90%) suffered hearing loss after acquiring spoken language (postlingually deafened), though some prelingually deafened individuals elect to receive CIs, as well (Holder, Reynolds, Sunderhaus, & Gifford, 2018). A key finding within the pediatric population, that has even influenced FDA regulations, is that children with severe-to-profound hearing loss from birth have better hearing outcomes with CIs the earlier they receive the device (Dorman & Wilson, 2004; Tobey et al., 2013). Similarly, among adult CI recipients, those who maintained normal hearing through adolescence perform better with CIs than adults who lost hearing at some point during childhood, before language acquisition was fully completed (Buckley & Tobey, 2011; Moon et al., 2014; Teoh, Pisoni, & Miyamoto, 2004). It is clear that access to sound during development of spoken language is important for good speech understanding later in life. Thus, children who receive CIs earlier in development and have access to auditory input for longer than those who receive CIs later in childhood, are more able to successfully decipher spoken language after implantation. By the same token, adults who acquire hearing loss midway through language development and receive CIs after a period of auditory deprivation critical to development perform more poorly than adults who lose hearing in their 30s or 40s. Just as age at implantation in early development has significant consequences for future CI performance, receiving a CI very late in life, when cognition starts to decline, can also make learning how to interpret sound in a new way more difficult (Beyea et al., 2016; Moberly et al., 2018).

While factors such as the duration of deafness or age at implantation are useful predictors at their extremes, the majority of CI users are postlingually deafened adults who receive CIs after childhood but before old age. Unfortunately, when the same factors are analyzed within this vast population of CI users, the results are muddled, with some studies asserting these factors to be predictive of performance and other studies finding no such correlations. A study by Blamey et al. (1996) tried to get to the bottom of these different findings by conducting a meta-analysis of 13 studies that analyzed duration of deafness, age at implantation, age at onset of hearing loss, etiology and duration of implant use as possible factors accounting for variability in performance within a sample 808 postlingually deafened CI users. Significant correlations between each of these factors and auditory performance are shown in the table below.

Table 1.1

Significant correlations between secondary factors and auditory performance reported in the literature. N is the number of CI users in the study; p-values indicate the degree of statistical significance; n.s. indicates that a statistical analysis was carried out, but no significant correlation was found. A blank indicates that the factor was not evaluated in the study. The * denotes a significant difference found between CI users who were deafened by meningitis and those with other aetiologies, but the p-value was not quoted. Data from Blamey et al. (1996).

Study	N	Aetiology	Age at implantation	Duration of deafness	Age at onset of deafness	Duration of implant use
Miller et al. (1986)	28		$p < 0.05$	$p < 0.05$		$p < 0.01$
Dowell et al. (1986)	40					$p < 0.01$
Dorman et al. (1989)	50	*	$p < 0.05$	$p < 0.01$		
Parkin et al. (1989)	20	n.s.	n.s.	n.s.	n.s.	
Shea et al. (1990)	20		n.s.	$p < 0.01$	n.s.	
Dorman et al. (1990)	27	n.s.	n.s.	n.s.		$p < 0.01$
Kileny et al. (1991)	10			$p < 0.05$	n.s.	
Blamey et al. (1992)	64	n.s.	$p < 0.05$	$p < 0.001$		
Gantz et al. (1993)	48	n.s.	$p < 0.05$	$p < 0.01$		
Shipp and Nedzelski (1994)	32		n.s.	$p < 0.0001$		
van Dijk et al. (1995)	38		n.s.	n.s.		
Summerfield and Marshall (1995)	119			$p < 0.01$		
Battmer et al. (1995)	132	$p < 0.01$	n.s.	$p < 0.01$	$p < 0.01$	

What is immediately apparent from this compilation of data is that the role each of the given factors play in accounting for variability in hearing outcomes is inconsistent across studies. For every factor measured, excluding duration of implant use, there is at least one study that found a significant correlation between the given factor and speech perception outcomes and at least one study that did not. In fact, there were more studies that found no correlation between either age at implantation or age at onset of deafness and hearing outcomes than studies that did find significant correlations between measures. Overall, the correlations found between age at implantation and speech perception were strongest in individuals implanted after the age of 60, in line with research related to the effects of cognitive decline on hearing outcomes with CIs. However, among middle-aged CI users, age of implantation was not predictive of performance. Similarly, the age at onset of deafness had little effect on performance for those who lost hearing before age 60. For those who lost hearing after the age of 60, there was a slight negative correlation, when comparing across studies.

The extent to which the etiology, or the cause of hearing loss, influences hearing outcomes within the adult CI population has also been a question explored in several studies, but to little avail. One issue inherent in drawing connections between CI performance and underlying disease processes is that a large proportion of individuals with profound hearing loss do not know the cause of their hearing loss. Of the 808 CI users analyzed in the Blamey et al. (1996) study, 439 individuals (54%) reported the cause of their hearing loss as unknown. In general, clinicians and medical researchers have a good understanding of the underlying causes of conductive hearing loss, but still struggle to account for the variety of factors that contribute to sensorineural hearing loss. Due to this lack of a specific diagnosis for many CI users, the correlations between etiology and hearing outcomes have historically been very weak. That being said, a few specific diagnosis have been associated with especially poor CI outcomes, including hearing loss caused by meningitis and bacterial labyrinthitis, whereas patients with Meniere’s disease have shown slightly better performance, on average. However, Blamey et al. (1996) found that these differences account for less than 2% of variance within the CI population.

All three studies (Dorman, Dankowski, McCandless, Parkin, & Smith, 1990; Dowell, Mecklenburg, & Clark, 1986; Miller et al., 1986) in the Blamey et al. meta-analysis that analyzed duration of CI experience as a factor contributing to variability found significant correlations with speech perception outcomes. While this is interesting, upon closer examination this positive relationship between experience and outcomes was only found to be significant in the first year after implantation, presumably reflecting users acclimatization to electrical sound and not their eventual mastery of the device. The factor with the biggest contribution to variance was duration of deafness, accounting for 13% of variability in outcomes. A steady negative correlation between longer periods of auditory deprivation and poorer hearing outcomes was significant within the CI population. Overall, Blamey et al. (1996) found that that the duration of deafness, age at onset of deafness, etiology and duration of implant experience together accounted for only 21% of variance within the postlingual adult CI population.

Fifteen years after the original Blamey et al. study, an updated meta-analysis of 2251 adult CI users implanted since 2003, found that the same four factors only accounted for 10% of variance within the modern postlingual CI population (Blamey et al., 2013). The factor accounting for the greatest amount of variance in the 1996 study, duration of deafness, explained much less variability in the more recently implanted population, since periods of auditory deprivation prior to implantation are shrinking among those with profound hearing loss. The recent trend to provide CIs to patients much more advanced in age also shifted age-related trends noted in the 1996 study, upwards in the 2013 study. The age at implantation and age at onset of hearing loss accounting for the largest amount of variance in CI users above the age of 70, rather than above the age of 60, with those between the ages of 60 and 70 no longer representing the oldest tier of implanted users. Unfortunately, after 15 years of medical research, 53% of patients still reported the cause of their hearing loss as unknown. Thus, etiology of hearing loss remains a poor predictor of CI hearing outcomes. Due to the relative loss of significance of these three factors in recent years, the duration of CI experience actually accounted for the most amount of variance in performance among CI users in the 2013 study. More recently implanted patients showed continued improvement in speech perception after the one-year mark and also more rapid improvement in the first few months after implantation. This new trend most likely reflects the changes in implantation criteria to include both older individuals and those with hearing loss that is less severe. While trends in the nature of variability have changed slightly in recent years, these clinical factors still account for a very small amount of variance in performance within the CI population.

In light of the lack of strength in predictive value of these factors, clinicians and researchers have explored surgical differences as a possible source of variability in CI hearing outcomes. Though CI users are viewed as a single patient population, comprised of individuals who elected to pursue the same treatment for hearing loss, the actual CI device characteristics vary widely across patients. Currently there are five CI manufacturers, including Cochlear Americas, Advanced Bionics, Med-El, Nurotron and Oticon Medical, each with a number of different external processors and internal electrode arrays. On top of these potential equipment differences between patients, a number of surgical factors vary from patient to patient, including the insertion depth of the electrode array, the relative distance between the array and the modiolus (central axis of the cochlea) and the scala within the cochlea (scala tympani or scala vestibuli) in which the array is inserted. A recent study by Lazard et al. (2012), in which authors re-analyzed data from the 2013 Blamey et al. study, found significant differences in speech perception in quiet for CI users using devices from different manufacturers, but the difference between the highest- and lowest-performing manufacturer was relatively small at only 14% . Each manufacturer was also not equally represented in the sample of CI users, which could make this difference in performance between manufacturers even smaller. The same study also found no significant difference in performance between patients who underwent two different surgical approaches (cochleostomy or round window) or had different electrode array insertion angles, depending on the array used and actual placement of the array in the cochlea. CI users with more active electrodes post-implantation did have significantly better speech perception in quiet than those with fewer electrodes turned on, but this difference was not significant when comparing speech perception in noise. Another study analyzing more detailed surgical factors in 203 CI users with Advanced Bionic devices found no correlations between the surgical insertion distance, the angular insertion depth, or the wrapping factor of the electrode array and speech perception outcomes (Van Der Marel et al., 2015).

One surgical factor that has been shown to be predictive of hearing outcomes in CI users is the number of electrodes inserted into the scala tympani (Finley et al., 2008). While CI surgeons do their best to insert the entire electrode array into the scala tympani, which is closest to auditory neurons within the cochlea, patient anatomy or difficulty during surgery can force some electrodes into the scala vestibuli. The figure below shows the variability in the number of electrodes inserted into the correct scala between patients.

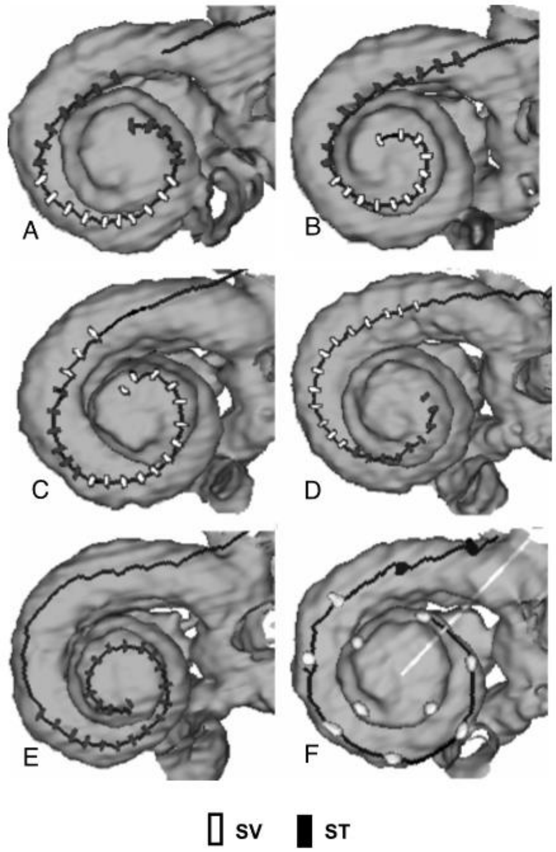

Figure 1.2

CT-based views of electrode position and insertion depth in six different ears. White rectangles represent electrodes inserted into the scala vestibuli and black rectangles represent electrodes inserted into the scala tympani. Figure from Finley et al. (2008).

Finley et al. found that more electrodes inserted into the scala tympani was associated with better speech perception outcomes. However, this study only looked at 14 CI users and thus it is debatable whether or not this correlation would hold for the greater CI population. Since surgical factors alone account for relatively little variance in performance between CI users, researchers have considered whether a combination of surgical factors and the survival and distribution of healthy neurons within the cochlea together might account for more of this variability. Unfortunately, the techniques currently available to directly count or measure the number and distribution of neurons within the cochlea is too invasive to be conducted in living humans. Thus, a number of non-invasive, behavioral methods have been developed in an attempt to measure approximate neural survival within the cochlea.

1.1.2 Measures of Spectral Resolution

In a healthy individual with normal hearing, the ability of the auditory system to perceive sound clearly is contingent upon the fidelity of thousands of hair cells and auditory neurons distributed evenly along the entire length of the cochlea. Different acoustic sound frequencies maximally stimulate different places along the basilar membrane, with lower frequencies activating primarily neurons located closer to the apex, or innermost part of the cochlea, and higher frequencies activating neurons located nearer to the base, or outermost part of the cochlea. The ability of a healthy cochlea to resolve different frequencies is what allows individuals with normal hearing to appreciate complex orchestral pieces and converse in a noisy restaurant with relative ease. Unfortunately, individuals with CIs do not experience the same level of precise frequency tuning or spectral resolution. While the mechanics of a healthy cochlea are precise in the conduction and transmission of acoustic signals, CIs bypass these damaged structures and deliver sound via electrical pulses that emit from individual electrodes and are received and interpreted by auditory neurons. As a result, the same sound processed by thousands of hair cells in a normal auditory system is processed by no more than 22 discrete electrodes in a CI. In addition, the spectral resolution of CI users is further degraded by the spread of electrical current within the cochlea. When the electrical pulses are emitted, the signal is rather broad, activating a wider population of neurons than would be excited for a given frequency in a normal auditory system. Since each individual CI user has a slightly different pattern of cochlear damage, electrode array insertion depth and distance between the array and the auditory neurons, researchers postulate that the combination of these factors may account for a significant amount of variance within the CI population. One of the most direct behavioral methods used to assess spectral resolution in CI users is measuring spatial tuning curves (STCs), which show to what extent neurons closer or further from a given electrode are activated when electrical pulses are sent to that electrode. STCs are measured using a forward masking paradigm. For each masker electrode, the level necessary to effectively mask the low-level probe at the test electrode is measured (Nelson, Donaldson, & Kreft, 2008). For example, if very low levels of masker on electrodes neighboring the target electrode effectively mask the probe tone, then tuning at that electrode is very broad, implying that the same population of auditory neurons is activated for a wide range of frequencies. On the other hand, if higher levels of masker are needed to mask the probe tone at the target electrode, then different populations of neurons, that have little overlap, are activated for different frequencies. Examples of broad and narrow STCs in CI users are shown below.

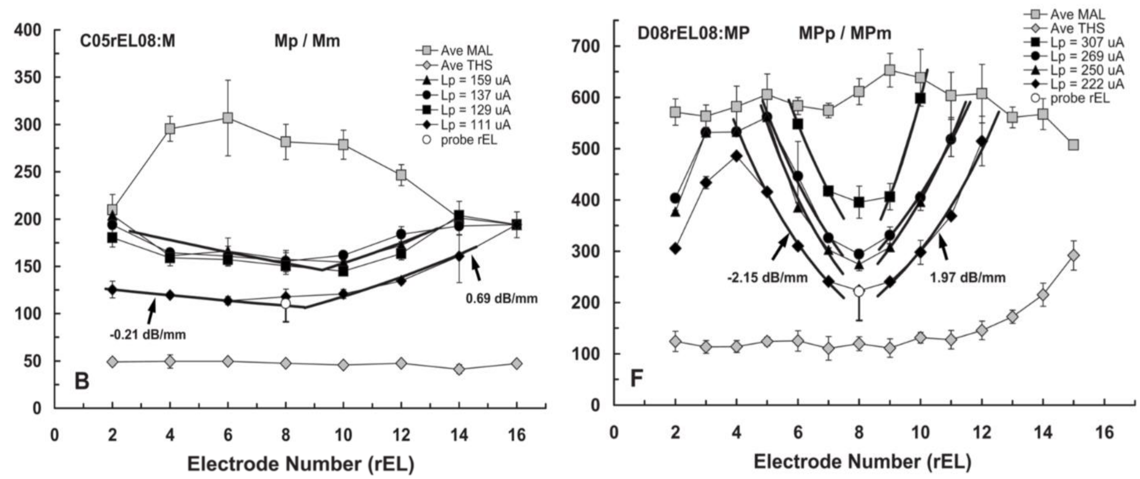

Figure 1.3

STCs for two Advanced Bionics CI users for probe electrode 8. Maximum acceptable loudness and minimum detectable thresholds are denoted by gray squares and diamonds, respectively. Masker levels required to mask the probe stimulus, at each masker electrode, are shown by black symbols, with each black symbol representing different probe levels tested. Images from Nelson et al. (2008).

Using STCs as a measure of spectral resolution is relatively effective and can provide detailed clues into the nature of neural survival spanning the length of the cochlea when completed for all electrodes in a given CI user. However, the procedure used to measure STCs is very time consuming, taking anywhere from 20 minutes to 5 hours per electrode, depending on the specifics of the technique. Measuring STCs at only a handful of electrodes is also hard to relate to speech perception, since good speech understanding depends on good spectral resolution along the entire length of the electrode array and not just in specific regions. In light of these drawbacks, alternative methods have been created and used by researchers in an attempt to determine spectral resolution along the length of the electrode array, but in a fraction of the time.

One such method that is widely used in CI research is called spectral ripple discrimination. In this method, a broadband Gaussian noise (e.g., 350-5600 Hz) is spectrally rippled, or modulated, at adaptively varying rates to create sinusoidal peaks and troughs in the noise that are closer or farther apart. This measure is typically administered as a three-interval, three-alternative forced-choice task where two of the intervals contain spectrally rippled noise with the same phase, and one interval contains a spectrally rippled noise with an 180-degree phase shift, or inverted ripple. CI users are to choose the stimulus that sounds different, in effect identifying the inverted ripple. After completing a number of runs, a ripple rate threshold is determined with greater ripples per octave corresponding to better spectral resolution and fewer ripples per octave corresponding to poorer spectral resolution (see Figure 1.4 below).

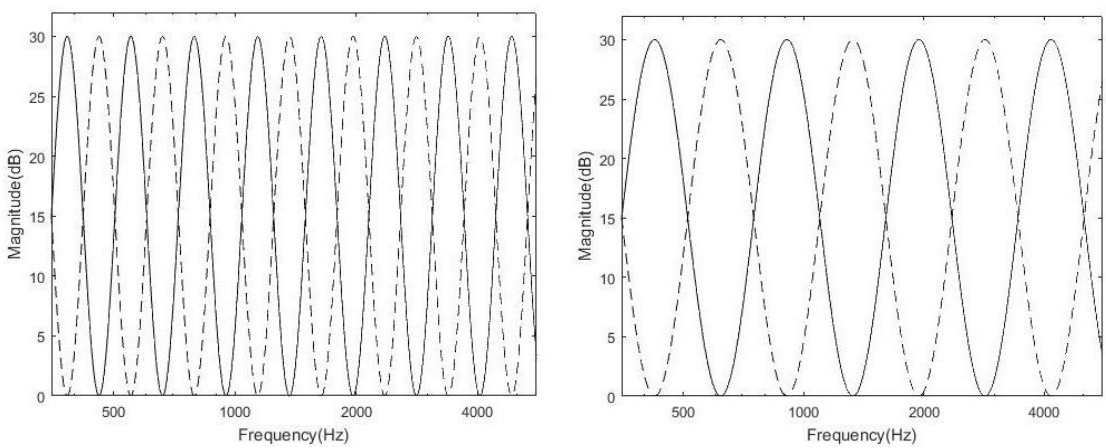

Figure 1.4

The left and right panels represent faster and slower ripple rates, respectively. The solid lines represent spectral ripples presented in two of the three intervals, and the dashed lines represent the inverted ripples to be identified by participants.

At higher ripple rates, where peaks of the inverted ripple are closer in frequency to those of the standard ripple, finer frequency tuning is needed to successfully discriminate between the two ripple phases. Thus, those CI users who can only discriminate between stimuli at slower rates likely have broader tuning or poorer neural survival. However, the connection between spectral resolution measured using STCs and ripple discrimination has been controversial. One difference between the measures that makes direct comparisons difficult is that the ripple discrimination stimulus is broadband, whereas STCs measure tuning at discrete locations in the cochlea. Anderson et al. (2011) investigated the similarity between these measures, finding that STCs measured at an electrode in the center of the array were not correlated with broadband ripple discrimination thresholds, but were significantly correlated with ripple thresholds measured using narrow bands corresponding to those used to measure STCs. Interestingly, when two outliers were removed from analysis, both correlations became highly significant. Anderson et al. (2011) reasoned, as have others (Azadpour & Mckay, 2012; McKay, Rickard, & Henshall, 2018), that these outliers may be using different cues to discriminate between ripples, such as spectral edges, or that they may have particularly uneven neural survival. Though the spectral ripples used in the ripple discrimination method are broadband, a CI user arguably only needs to detect a change in level at a single discrete location in the cochlea to successfully complete the task. Thus, a CI user with excellent neural survival in the apical end of the cochlea but with poor neural survival along the rest of the cochlea would likely have a very high (good) ripple discrimination threshold but poor speech perception, since good spectral resolution over a range of frequencies is needed to understand speech well.

Another widely used measure of spectral resolution is spectral ripple detection. Broadband spectral ripples are also utilized in this method but instead of adaptively varying the ripple rate, the amplitude of each ripple, or the ripple depth, is adaptively varied across trials (Anderson, Oxenham, Nelson, & Nelson, 2012; Bernstein & Green, 1988; Eddins & Bero, 2007; Litvak, Spahr, Saoji, & Fridman, 2007). This procedure is also typically a three-interval, three-alternative forced-choice task, but differs from the ripple discrimination measure in that two of the three intervals contain un-rippled or un-modulated broadband (e.g., 350-5600 Hz) Gaussian noise, and one interval contains spectrally rippled broadband noise. Across trials, the ripple rate is held constant but the depth of the modulation is varied, with larger modulation depths creating larger perceptual differences between the interval with modulated noise and the intervals with un-modulated noise. Similar to the ripple discrimination procedure, CI users completing the ripple detection measure are instructed to choose the interval that is different, in effect identifying the interval containing rippled noise. The amplitude of the modulated noise is adaptively varied until a ripple detection threshold is reached, indicating the smallest modulation depth that can be reliably detected, for a given ripple rate. Higher detection thresholds are thought to indicate poorer spectral resolution as fluctuations in the stimulus need to be more pronounced to correctly identify the modulated stimulus. Poorer spectral resolution, or broader frequency tuning, leads to a relative smoothing of the rippled stimulus, making it harder to detect amplitude fluctuations. Consequently, this measure should be informative for spectral resolution needed to understand speech as several studies have demonstrated a decrease in speech understanding when spectral contrasts are decreased (Bacon & Brandt, 1982; Van Tasell, Soli, Kirby, & Widin, 1987).

However, the ripple rate at which this connection to spectral resolution and speech perception should be most informative has been controversial. Theoretically, broader frequency tuning, seen in CI users, should most affect the resolution of ripples at higher rates, (1 and 2 ripples per octave) since even the poor spectral resolution of CIs is unlikely to affect the representation of very low modulation rates, such as 0.25 and 0.5 ripples per octave. In contrast, several studies have shown ripple detection thresholds at slower ripple rates to be most predictive of speech perception in CI users (Anderson et al., 2012; Litvak et al., 2007; Saoji, Litvak, Spahr, & Eddins, 2009a). This finding may indicate that speech perception is, at least partially, reliant on intensity resolution, which is the ability of CI users to compare intensity differences in a signal across frequency bands. Some researchers have explored the possible correlation between intensity resolution and speech perception (Anderson et al., 2012; McKay et al., 2018), but the relationship between intensity resolution, ripple detection at slow rates, and speech perception remains controversial. So while it is unlikely that ripple detection is a direct measure of spectral resolution in CI users, its high correlation with measures of speech perception still renders it a useful predictor of performance, despite the lack of understanding of the underlying mechanism actually being measured.

In addition to measures of spectral resolution, measures of temporal resolution, such as amplitude modulation detection, have also been found to correlate with speech perception in CI users (Fu, 2002). Since speech is a dynamic signal that changes in both spectral and temporal dimensions, some researchers have thought that maybe a combination of spectral and temporal (timing) resolution may be an even more informative predictor of CI performance (Won et al., 2015). In a 2015 study, Won et al. used a spectrotemporal modulation (STM) detection paradigm (Chi, Gao, Guyton, Ru, & Shamma, 1999; Elhilali, Chi, & Shamma, 2003) to explore this hypothesis. The method employed by Won et al. (2015) used a two-interval, two-alternative forced choice procedure in which one interval contained spectrotemporally modulated broadband (354-5656 Hz) Gaussian noise and the other interval contained un-modulated (unprocessed) Gaussian noise. CI users were instructed to choose the interval with the chirping or moving sound, in effect identifying the STM stimulus. The modulation depth of the STM stimuli was adaptively varied and a STM threshold was determined, with higher STM thresholds corresponding to poorer STM resolution. Three spectral ripple rates (0.5, 1 and 2) and two temporal rates (5 and 10 Hz) were tested. Examples of these stimuli are shown below.

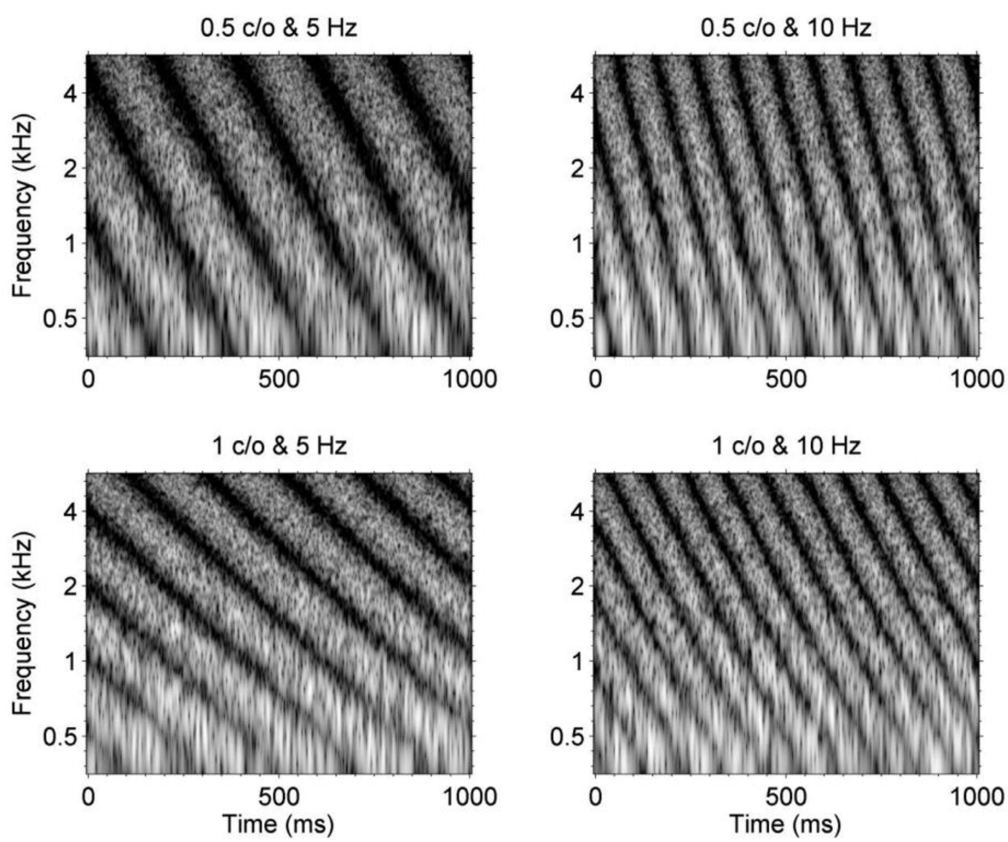

Figure 1.5

Spectrodograms of STM stimuli for four combinations of spectral and temporal rates. Figure from Won et al. (2015).

In agreement with other studies (Anderson et al., 2012; Litvak et al., 2007), Won et al. found STM stimuli with slower spectral ripple rates (0.5 and 1) to be most informative for predicting speech outcomes in CI users. Interestingly, this correlation was not affected by changes in the temporal rate of the STM stimuli, indicating that spectral resolution, or possibly the ability to detect slow broadband frequency changes, is more important for speech understanding than the ability to detect slow timing changes. This finding renders the need for and validity of the STM method, rather than strictly spectrally-focused methods for measuring spectral resolution in CI users somewhat debatable. However, all four methods used to measure frequency tuning and neural survival in CI users discussed in this section have been used to try and account for variability in performance of CI users, with more or less success depending on the parameters of each individual study.

1.1.3 Variability in Speech Perception Accounted for by Spectral Resolution

Measures of spectral resolution have been used in a number of different studies with some finding significant correlations with speech performance (Drennan, Anderson, Won, & Rubinstein, 2014; Gifford et al., 2018; Henry & Turner, 2003; Henry, Turner, & Behrens, 2005; Won et al., 2007, 2015) and others finding mixed results depending on the method of measuring spectral resolution and speech materials used (Anderson et al., 2011, 2012). The proportion of variance that different measures of spectral resolution have accounted for in speech perception for CI users ranges from 22 to 68% across studies. Some studies also find the proportion of variance accounted for by spectral resolution measures to be not significant. These differences in findings between studies are also likely due to small sample sizes, with most studies testing 15 to 30 CI users. Since measures of spectral resolution are rarely done in clinics, it is difficult for researchers to look at this relationship in a large sample, and control for other factors that may influence variability, such as duration of deafness. The largest study looking at this relationship was done recently by Gifford et al. (2018) in which ripple detection thresholds were compared with three different speech materials in 578 CI users. Gifford et al. found significant correlations between ripple detection thresholds when compared to words and sentences in quiet, as well as sentences in noise, for postlingual adult CI users. However, spectral resolution only accounted for 25 to 27% of variability in speech perception for postlingual adult CI users and the relationship did not reach significance for the group of pediatric CI users tested. In addition, the authors did not control for other factors that may influence variability, such as duration of deafness or age at implantation, which could make the variance solely accounted for by spectral resolution even smaller.

Results from other studies using smaller samples are difficult to compare, given the range of different speech materials used including, vowels and consonants in a closed set, words in either an open or closed set, and full sentences. These speech materials have also been presented either in quiet, multi-talker babble, or speech-shaped Gaussian noise. Furthermore, performance on these speech materials has been compared to different measures of spectral resolution, including STCs, spectral ripple discrimination, spectral ripple detection and STM detection. A few informative trends can be extracted when considering these studies together. First, correlations tend to be strongest when comparing the perception of relatively short speech spondees, such as vowels or words, in quiet with spectral ripple discrimination thresholds (Drennan et al., 2014; Henry & Turner, 2003; Henry et al., 2005; Won et al., 2007). This makes sense, since good frequency resolution in a narrow region of the cochlea would allow a CI user to successfully complete both tasks. A second trend found in the data is that the correlation between speech perception and spectral resolution tends to decrease or fail to reach significance when longer speech materials, like full sentences, in noise are used (Anderson et al., 2011, 2012). Finally, significant correlations between spectral resolution and speech perception also decrease when other factors, such as age or other clinical factors are taken into account (Gifford et al., 2018).

In summary, while there seems to be some relationship between spectral resolution and speech understanding in CI users, the proportion of variance it explains becomes rather small when speech perception measures that more closely resemble real world environments are tested and other factors influencing variability are taken into account. In light of this failure for spectral resolution to account for the majority of variance seen in hearing outcomes of CI users, some researchers have decided to look beyond the periphery of the auditory system to find answers.

1.2 Review of Non-peripheral Factors Impacting Variability

1.2.1 Working memory and hearing loss

One non-peripheral factor that has been identified as showing potential for predicting variability in hearing outcomes for postlingual adult CI users is working memory. Working memory is defined by Pichora-Fuller et al. (2016) in a theoretical analysis of effortful listening as "the retention of information in conscious awareness when this information is not present in the environment, for its manipulation and use in guiding behavior." In the context of understanding degraded speech, words or fragments of words might be stored in working memory until the completion of a sentence or phrase, at which time they may be pieced together by the listener. It logically follows that individuals with hearing loss who have superior working memory abilities may be better able to leverage working memory to understand speech. While this theory has only been explored in the adult CI population by a handful of researchers (Lyxell et al., 1998; Moberly, Harris, Boyce, & Nittrouer, 2017; Moberly, Houston, et al., 2017), the connection between working memory and hearing loss more generally is much better understood.

A meta-analysis of 20 studies (Akeroyd, 2008) looking at the influence of various cognitive factors on speech understanding in adults with mild-to-moderate hearing loss, including general intelligence, scholastic achievement, processing speed and long- and short-term memory, found working memory tests to be the most informative predictors of hearing outcomes. Though the speech materials and cognitive measurements used across studies were highly variable, most authors that measured working memory in some way found it to be a significant secondary predictor of speech understanding in hearing-impaired adults, with the degree of hearing loss accounting for the most variance. The reading span task was identified as the most viable measurement for working memory, accounting for up to 25% of variability in speech performance in some studies (Foo, Rudner, Rönnberg, & Lunner, 2007; Lunner, 2003; Rudner, Foo, Rönnberg, & Lunner, 2007). The reading span task can vary slightly in exact parameters of presentation, but generally probes the ability of an individual to hold a string of letters in memory while completing a concurrent verbal decision-making task. The stronger the working memory capabilities of an individual, the more letters they will be able to store and recall at the end of each trial. This task is thought to translate rather well to the perception of degraded speech where phonemes or individual words might have to be stored in memory and then reassembled at a later time point.

A more recent meta-analysis, which reviewed 25 studies comparing speech perception in noise and cognitive factors in individuals with normal to moderate hearing loss, found working memory to be a significant predictor of hearing outcomes but that it accounted for less variance (8%) across studies than had been previously reported (Dryden, Allen, Henshaw, & Heinrich, 2017). The smaller overall proportion of variance accounted for by working memory found in this study could be due to the fact that many of the studies included in this meta-analysis used subjects with normal or very mild levels of hearing loss. It is possible that working memory contributes more to variability in speech outcomes in individuals with more severe hearing loss attempting to understand speech that is more degraded. Dryden et al. also found that working memory was a more consistent predictor of variability for speech signals that were longer and more complex, such as sentences in noise, which are more characteristic of speech encountered in real-world environments. Processing speed, which has been shown to decrease in older adults, was also a significant predictor of outcomes in speech understanding.

Since the trend towards implanting increasingly older patients with severe hearing loss has only taken hold relatively recently, the vast majority of research exploring the effects of cognitive factors like working memory and processing speed in CI users has been done in children. The impact of these cognitive factors on hearing outcomes and vice versa is a particularly hot topic in children due to the knowledge of, and pressure imposed by, critical periods in language development that are believed to occur throughout childhood and into adolescence. Several studies in pediatric CI users have reported deficits in working memory when compared to children with normal hearing, with these deficits often correlating with speech perception and verbal knowledge (Beer et al., 2014; Bharadwaj, Maricle, Green, & Allman, 2015; Nittrouer, Caldwell-Tarr, Low, & Lowenstein, 2017; Tao et al., 2014). Tao et al. found that measures of working memory correlated significantly with the perception of sentences in quiet, accounting for almost 50% of variance in performance after controlling for demographic factors. This connection between working memory and verbal fluency or speech understanding in children with CIs is thought to stem from the difficulty in encoding and remembering degraded speech signals that do not map onto correct phonological representations of words that may or may not be acquired and stored in memory (Nittrouer et al., 2017). If this one-to-one mapping of spoken language and stored phonological representations is particularly inconsistent, the acquisition of vocabulary slows and in turn impairs verbal working memory. Since postlingual adult CI users do not experience these same difficulties in initial language acquisition, with hearing loss occurring later in life, it is possible that working memory might not be as large of a mediating factor in hearing outcomes. However, a large amount of learning and re-mapping of electric sound through a CI onto acoustic representations of words stored in memory must occur for postlingual CI users to succeed with the device.

In line with this logic, along with past findings in hearing impaired adults and pediatric CI users, some recent studies have examined the extent to which working memory abilities predict hearing outcomes in postlingual adults with CIs. One of the first studies to look at this connection was done by Lyxell et al. (1998) in a small group of adult CI users with early-generation devices. When the 15 participants tested in this study were divided into two groups based on reading span scores, those with better working memory on average also showed better speech perception. Although two participants that had lower than average working memory had above average speech perception, Lyxell et al. reasoned that working memory might not mediate performance to as great an extent for CI users who have better signal, or spectral resolution, through their CI to begin with. More recently, Moberly et al. (2017a, 2017b) found significant correlations between an auditory version of the reading span task, called listening span, and three different speech perception measures, with auditory working memory accounting for roughly 40% of variance in speech understanding in adult CI users. A measure of inhibition-concentration, or the ability to inhibit automatic responses and create responses using attention and reasoning, was also significantly correlated with both auditory working memory and speech perception of CI users in the study. Moberly et al. concluded that inhibition-concentration mediates auditory working memory and speech perception in CI users, indicating the importance of higher level processing skills in the processing and understanding of degraded speech.

1.2.2 General intelligence, hearing loss and aging

With findings like those of Moberly et al. (2017b) that show interplay between different cognitive factors and their effect on speech perception, some researchers are inclined to assume that some measure of general intelligence, or g factor (Spearman, 1904), might best explain the holistic influence of higher level cognition on speech understanding for those with hearing loss. However, studies that have used more general or composite measures of cognition have often failed to show connections to speech perception for adults with hearing impairment (Collison, Munson, & Carney, 2004; Dryden et al., 2017; Nuesse, Steenken, Neher, & Holube, 2018). Measures used in these studies included, or were similar to, tasks such as Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998), which uses pattern completion to assess non-verbal intellectual efficiency, and the Wechsler Abbreviated Scale of Intelligence (Wechsler, 2011), which assesses general intelligence through verbal comprehension and perceptual reasoning tasks. The two meta-analyses mentioned in the previous section (Akeroyd, 2008; Dryden et al., 2017) both found measures of general intelligence or IQ to be the least informative predictors of speech perception outcomes in the hearing impaired population. Differences in exact speech materials, cognitive measures and hearing criteria for participants used across experiments leave many questions unanswered in regards to the specific higher level process or combination of processes that most impacts hearing outcomes in postlingual adult CI users.

One population in which the connection between general cognition and hearing loss is better understood is the elderly. In a meta-analysis of 36 studies, Loughrey et al. (2018) found significant correlations between age-related hearing loss and both cognitive decline and dementia. Another large-scale analysis of 7385 individuals participating in the English Longitudinal Study of Aging (Ray, Popli, & Fell, 2018) reported a link between cognitive decline and hearing loss, with untreated hearing loss identified as a factor that may be driving this negative relationship. In a longitudinal study of older adults, Lin et al. (2013) found that those with hearing loss had a 24% increased risk of cognitive impairment compared to those with normal hearing. Individuals with hearing loss also showed a steeper decline in cognition over the six-year time frame during which the study was conducted. Given these consistent findings linking hearing loss to cognitive decline in elderly populations, it is not surprising that clinicians have expanded their CI selection criteria to include much older individuals with hearing loss, in recent years. A recent meta-analysis by Taljaard et al. (2016) of 33 studies looking at hearing impairment and cognitive function found that, on the group level, hearing interventions, such as hearing aids or CIs, slowed cognitive decline in older adults. However, the degree of cognitive impairment was significantly associated with the degree of hearing impairment, regardless of treatment. Thus, since hearing outcomes in adult CI users are highly variable, especially in older adults (Hast et al., 2015; Mahmoud & Ruckenstein, 2014), it remains unclear how general cognition affects hearing outcomes with CIs and how these hearing outcomes over time might alter cognitive function later in life.

1.2.3 Social isolation and hearing loss

The complex interaction between cognition and hearing loss in adults is often accompanied by social and psychological factors that can negatively impact quality of life. Much research has been done, particularly in older adults, revealing a connection between hearing loss, social isolation, and depression (Chia et al., 2007; Kim et al., 2017; Mick, Kawachi, & Lin, 2014; Pronk et al., 2011). In a cross-sectional and longitudinal study of 767 older adults living in community dwellings, Mikkola et al. (2016) found that fully mobile adults with hearing impairment spent significantly less time outside the home and were three times more likely to withdraw from leisure activities than equally mobile adults with good hearing. Severe hearing impairment was also found to significantly increase the risk of depression in 30,000 patients followed by Kim et al. (2017) for a period of 11 years, regardless of age. Since depression and social withdrawal often precede declines in overall health and quality of life (Shankar, McMunn, Demakakos, Hamer, & Steptoe, 2017; Stubbs et al., 2017), researchers and clinicians have long hoped that hearing interventions and treatment, with devices like hearing aids and CIs, may be able to mitigate this downward social and physical health trend.

Many studies have shown increases in cognitive health (Dawes et al., 2015), listening ability and overall health outcomes with the adoption of hearing aids (Kitterick & Ferguson, 2018). A recent study looking at the effects of auditory rehabilitation in 125 hearing-impaired adults reported patients making gains in hearing outcomes, short- and long-term memory, as well as reduced depression after receiving various hearing interventions (Castiglione et al., 2016). While these findings are encouraging, the causal link between hearing aids and improvements in quality of life is not always straightforward. The vast majority of studies looking at this relationship have analyzed changes in adults with hearing loss who have willingly chosen to purchase hearing aids and regularly participate in auditory rehabilitation. Though these individuals may see improvements in quality of life and hearing outcomes, it may simply be because those who are proactive about managing health problems have other innate qualities, such as perseverance or a positive outlook, that mediate these gains. In CI users, the picture is slightly different as the impact of the intervention is often based on data before and after implantation in the same individuals. This approach to analyzing the effectiveness of the intervention on overall well-being is problematic for a number of reasons, first and foremost being that the hearing loss required to qualify to receive a CI is so severe, that general improvements in quality of life are virtually inevitable post-implantation. A 2015 study (Mäki-Torkko et al.) comparing experiences of 101 CI users and their significant others pre- and post-implantation describes the two experiences as living in two different worlds, with post-implantation associated with increases in normality, autonomy and social engagement. CI users interviewed by Hughes et al. (2018) echoed these sentiments but also highlight the ongoing cognitive effort and social difficulties post-implantation. One participant described feeling isolated in group settings due to the increased effort and time needed to listen and assimilate to the rapid, continuous nature of dialogue. Other participants also eluded to the cost of increased listening effort on working memory, reporting "all that mental juggling seems to affect my memory because I am trying to listen to [other people] and trying to make some notes, I am trying to think what I want to say, and also remember what is going on."

So while it is clear that CIs significantly improve hearing outcomes, social engagement, and overall mental health of most individuals who previously experienced severe-to-profound hearing loss, the nature of the relationship between hearing outcomes and social factors *post*-implantation remains unclear. Given the large variability in hearing outcomes of postlingual adult CI users, it is fair to assume these differences in outcomes contribute to varying levels of success in social situations. What continues to be a mystery is whether these relative successes or failures in social situations directly impact the motivation, cognition, and effort necessary to succeed in understanding speech with the device. CI users in the Hughes et al. study (2018) describe having a "finite amount of energy [] that can be used up very quickly" depending on other health ailments, work commitments, and social engagements on any given day. Thus, if a CI user feels the need to prioritize work, or struggles with other energy-draining health complications, listening rehabilitation and social engagement might not get as much attention and may suffer, as a result. On the other hand, if a CI user is driven by the need for social connectedness (Hughes et al., 2018), the consistent act of participating in dynamic and difficult social situations may bolster speech understanding with the device.

1.2.4 Listening strategies and use of semantic context

Relatively little research has been devoted to the amount of time CI users spend engaging socially and its impact on speech perception, but the ways in which CI users attempt to understand speech in these settings has received more attention in recent years. For individuals with normal hearing, little thought or effort is needed to effectively communicate in everyday environments. However, when normally salient acoustic cues in speech are degraded and ambiguous due to hearing impairment, meaningful effort must be employed to successfully follow continuous speech. One tool that has been used by auditory researchers to quantify listening effort is pupil dilation, with the enlargement of an individual's pupils while performing a given task corresponding to increased effort applied during task completion (Naylor, Koelewijn, Zekveld, & Kramer, 2018; Zekveld, Kramer, & Festen, 2010). In a study by Wagner et al. (2016) looking at the differential processing and effort associated with understanding degraded speech, the authors found greater listening effort was needed to disambiguate similar-sounding phonemes when individuals listened to sentences through a vocoder, meant to simulate CI processing. This inability to easily differentiate phonemes in words early in a sentence also led to a delay in integrating the semantic information needed to understand the speech in its entirety. Winn (2016) found similar effects in actual CI users listening to sentences, with the highest level of effort exerted *after* the last word of the sentence, indicating that CI users waited to accumulate as much phoneme information as possible before using cognitive effort to piece it all together into something meaningful. Interestingly, the pattern of errors found in CI users when understanding sentences in noise also pointed towards more use of semantic context than phoneme information when making educated guesses about the last word of the sentence. For example, if a participant misheard the word ‘right’, they were more likely to guess a semantically similar word like ‘left’, than a phonetically similar word like ‘fight’.

Amichetti et al. (2018) looked at the use of semantic context more directly by varying the entropy, or likelihood, the final word in a sentence would occur, given the words proceeding it. The final words of sentences were time-gated, so that only a fraction of the start of the final word was revealed on the first trial, and increasingly larger fractions revealed in each proceeding trial until the participant correctly identified the word. Results showed that the more likely a word was to occur at the end of the sentence, the smaller the gate size needed to be for participants to correctly guess the word. However, older CI users needed much bigger gate sizes to correctly identify words in sentences when there were a greater number of phonetically-similar competing words that could also fit the sentence, compared to younger CI users. So while it is clear that all CI users rely on the use of semantic context when understanding speech, older CI users may depend on this strategy more heavily than younger CI users. Winn (2016) called into question the pervasive use of speech understanding as the sole clinical measure of hearing outcomes for CI users when individuals may be using very high levels of cognitive effort to achieve those levels of performance. Thus, the connection between listening effort, use of semantic context, and speech understanding remains unclear in the adult CI population.

Shafiro et al. (2016) looked at the role of context in explaining variability in hearing outcomes of CI users indirectly by attempting to create a non-speech measure of auditory cognition. The authors developed the Familiar Environmental Sound Test, which consisted of 25 environmental, non-speech sounds that were arranged in sequences of five sounds that were either contextually coherent or incoherent. For example, birds chirping and an alarm ringing would be coherent within the context of morning sounds, but birds chirping and the phone ringing would not be associated with a similar environmental context. After each five-sound sequence, participants were asked to correctly identify the sounds they heard. Similar to normal-hearing and hearing-impaired listeners, CI users were better at identifying sounds in coherent sequences than incoherent sequences. However, CI users were much worse at remembering sounds occurring later in incoherent sequences than any other group, indicating interference when context could not be leveraged to understand the sounds. Performance for coherent sound sequences also significantly correlated with speech perception in noise for the CI group, with Shafiro et al. postulating that the ability to leverage context plays a role in both environmental sound identification and speech perception for CI users. While this study provides some insight into the influence of context in speech perception, the ability of individual CI users to effectively leverage semantic context and its connection to actual speech perception outcomes has yet to be quantified in terms of the variability it may explain in this population.

As can be seen from the review of the influence of peripheral and non-peripheral factors on CI performance, a number of questions remain regarding the relative influence of both. In addition, little is known about the variability within the normal-hearing population when stimuli are degraded to an extent similar to that produced by CIs.

1.3 Overview of Chapters

The goal of this dissertation was to better understand the peripheral and non-peripheral factors that contribute to the variability in speech perception outcomes of post-lingually deafened adult CI users. While exploring differences in spectral resolution and how they might influence speech understanding in CI users (Chapter 2), we noticed a great deal of variability in young NH participants listening to degraded speech processed through vocoders with varying amounts of simulated spectral smearing. This variability seemed to be inherent to the difficulty of the auditory task itself and perhaps even increase with increasing degradation of the speech signal. To investigate between-subject differences in the understanding of degraded speech that might not be explained by peripheral factors but instead be more cognitive in nature, we conducted a study (Chapter 3) relating multiple measures of speech perception, spectral resolution, and higher-level processing in CI users, as well as two control groups of age-matched and young NH listeners. To further explore the leveraging of semantic context as a compensatory listening strategy in the adult CI population, we developed an original corpus of sentences without semantic context (Chapter 4) to allow for direct comparisons of performance on sentences with semantic context, matched for talkers, length, grammatical structure and vocabulary (Chapter 5). Finally, the relationship between compensatory listening strategies, social engagement, and speech perception outcomes in CI users was explored in a study conducted outside the laboratory (Chapter 6), in which CI users and age-matched NH listeners completed ecological momentary assessments (EMAs) during the natural course of their day-to-day lives. Chapter 7 discusses what has been learned from this body of work as a whole and identifies both clinical implications of this research and avenues for further exploration.

CHAPTER 2: SPEECH PERCEPTION AND SPECTRAL RESOLUTION

Chapter 2 is reprinted from:

O'Neill, E. R., Kreft, H. A., & Oxenham, A. J. (2019). Speech perception with spectrally non-overlapping maskers as measure of spectral resolution in cochlear implant users. *Journal of the Association for Research in Otolaryngology, 20,* 151-167.

Abstract

Poor spectral resolution contributes to the difficulties experienced by cochlear-implant (CI) users when listening to speech in noise. However, correlations between measures of spectral resolution and speech perception in noise have not always been found to be robust. It may be that the relationship between spectral resolution and speech perception in noise becomes clearer in conditions where the speech and noise are not spectrally matched, so that improved spectral resolution can assist in separating the speech from the masker. To test this prediction, speech intelligibility was measured with noise or tone maskers that were presented either in the same spectral channels as the speech or in interleaved spectral channels. Spectral resolution was estimated via a spectral ripple discrimination task. Results from vocoder simulations in normal-hearing listeners showed increasing differences in speech intelligibility between spectrally overlapped and interleaved maskers as well as improved spectral ripple discrimination with increasing spectral resolution. However, no clear differences were observed in CI users between performance with spectrally interleaved and overlapped maskers, or between tone and noise maskers. The results suggest that spectral resolution in current CIs is too poor to take advantage of the spectral separation produced by spectrally interleaved speech and maskers. Overall, the spectrally interleaved and tonal maskers produce a much larger difference in performance between normal-hearing listeners and CI users than do traditional speech-in-noise measures, and thus provide a more sensitive test of speech perception abilities for current and future implantable devices.

2.1 Introduction

The cochlear implant (CI) has been an extremely effective auditory solution for many individuals with severe-to-profound hearing loss (e.g. Zeng et al. 2008). Despite the success of the device, a major challenge for CI users remains the difficulty of understanding speech in the presence of background noise. One factor believed to be critical in limiting CI performance is poor spectral resolution, produced by the limited number of electrodes in the CI array, and by the extensive overlap in the electrical fields produced by neighboring electrodes. The overlap or spread of current means that increasing the number of electrodes does not necessarily increase the effective number of independent frequency channels, and so performance does not generally improve with increasing number of electrodes beyond about eight or ten (Friesen, Shannon, Baskent, & Wang, 2001). Studies with normal-hearing (NH) listeners have successfully simulated the effects of current spread by implementing different forms of spectral smearing within noise- or tone-excited envelope vocoder schemes (Bingabr, Espinoza-Varas, & Loizou, 2008; Crew, Galvin, & Fu, 2012; Qian Jie Fu & Nogaki, 2005; Grange, Culling, Harris, & Bergfeld, 2017; Mesnildrey & Macherey, 2015; Oxenham & Kreft, 2014).

Despite the intuitively obvious connection between speech perception in noise and spectral resolution, the relationship between the two measures has not always been clear at the level of individual listeners. Although a number of studies have reported correlations between spectral ripple discrimination thresholds (i.e., the highest spectral ripple rate, in ripples per octave, at which a phase reversal can be detected) and speech perception in quiet (Anderson et al., 2011; Drennan et al., 2014; Henry et al., 2005; Won et al., 2011), results have been more mixed for speech perception in noise, with some studies finding a significant correlation (Holden et al., 2016; Jeon, Turner, Karsten, Henry, & Gantz, 2015; Won et al., 2011; Zhou, 2017) and others not (Anderson et al., 2011). Other measures, involving spectral ripple detection have often found correlations between the minimum detectable ripple depth and speech perception in noise (Anderson et al., 2012; Litvak et al., 2007; Saoji, Litvak, Spahr, & Eddins, 2009b). However, these correlations tend to be significant at low ripple rates (such as 0.25 or 0.5 ripples per octave, rpo) but not at higher ripple rates (such as 1 or 2 rpo), which is the opposite of what would be expected if spectral resolution, rather than intensity resolution, were limiting speech perception in noise (Anderson et al., 2012). A more recent measure of spectral and/or intensity resolution that is most similar to spectral ripple detection (Azadpour & Mckay, 2012) was also found not to be significantly correlated with speech perception in noise. Interestingly, a recent publication by Gifford et al. (2018) found correlations between spectral modulation detection thresholds at 0.5 and 1 rpo and speech perception in both quiet and noise for adult but not pediatric CI users.

One reason for the lack of robust correlations between measures of spectral resolution and speech perception in noise may be the use of noise that is spectrally matched to the speech. Spectral resolution is likely to be most important when the speech and noise are *not* spectrally matched, so that better spectral resolution can help segregate speech from noise. Recent studies have shown that spectrally separating speech and noise leads to increased speech intelligibility for NH listeners, compared to performance when speech and noise overlap in the frequency domain (Apoux & Healy, 2010; Kidd Jr, Mason, & Gallun, 2005). It is already known that CI users are not able to take as much advantage of spectral gaps in a masker as NH listeners. For instance, Oxenham and Kreft (2014) found that a spectrally sparse masker, consisting of 16 logarithmically spaced pure tones produced as much speech masking in CI users as a speech-shaped noise with the same overall level and spectral envelope, whereas NH listeners exhibited a large release from masking. These findings suggest that conditions in which the masker and speech do not completely spectrally overlap may provide a more sensitive test of the effects of spectral resolution, and so may provide measures of speech recognition that correlate more closely with measures of spectral resolution than more typical measures of speech in spectrally overlapping noise.

The aim of this study was to test the prediction that measures of speech intelligibility in spectrally unmatched noise should be more sensitive to differences in spectral resolution in CI users than traditional speech-in-noise tests. Conditions in which the speech and masker were presented to the same CI electrodes were compared with conditions in which the speech and masker were presented to different (interleaved) electrodes. This same paradigm was also implemented using the virtual channels that were used in the processing schemes of the CI users. Both noise and tones were used as maskers (Experiments 1 and 2), and the results were compared to more direct measures of spectral resolution using spectral ripple discrimination (Experiment 3). In all cases, the results from CI users were compared with results from NH listeners using tone-excited envelope vocoders to simulate various degrees of current spread (Crew et al., 2012; Oxenham & Kreft, 2014).

2.2 Experiment 1: Speech Perception

2.2.1 Methods

2.2.1.1 Listeners

A total of 13 post-lingually deafened CI users and 24 NH listeners (three groups of eight participants) were tested. All participants were native speakers of American English. Individual details for the CI users are provided in Table 2.1.

Table 2.1

Individual subject information for CI users.

Subject Code	Gender	Age (Yrs)	CI use (Yrs)	Etiology	HL prior to implant (Yrs)	Speech Processing Strategy
C16	F	63	16	Unknown	1	MPS
D02	F	67	15	Unknown	1	HiRes Optima-P; ClearVoice MED
D10	F	63	14	Unknown	8	HiRes-S w/Fidelity 120; ClearVoice HIGH
D25	F	53	10	Meniere's disease	33	HiRes Optima-S;ClearVoice LOW
D26	F	57	8	Unknown	11	HiRes Optima-S; ClearVoice OFF
D27	F	65	7	Otosclerosis	13	HiRes-S w/Fidelity 120; ClearVoice OFF
D28	F	68	14	Familial Progressive SNHL	7	HiRes Optima-S; ClearVoice MED
D35	F	57	6	High Fever	?	HiRes Optima-S; ClearVoice MED
D39	M	69	8	Unknown	7	HiRes Optima-S; ClearVoice MED
D41	F	68	4	Familial Progressive SNHL	41	HiRes-S w/Fidelity 120; ClearVoice MED
D42	M	61	3	Familial Progressive SNHL	2	HiRes Optima-S; ClearVoice MED

D44	F	70	9	Familial Progressive SNHL	18	HiRes Optima-P; ClearVoice HIGH
D46	F	60	4	Meniere's Disease	13	HiRes-S w/Fidelity 120; ClearVoice MED
D47	F	59	4	Unknown	< 1	HiRes Optima-P; ClearVoice MED
D52	F	60	14	High Fever	4	HiRes Optima-S; ClearVoice MED
D55	F	58	7	Familial Progressive SNHL	?	HiRes Optima-S; ClearVoice OFF

To take part in the study, CI users were required to obtain at least 40% of keywords correct in sentences from the IEEE corpus (IEEE, 1969) in quiet. The eight CI users who met this criterion are indicated in Table 2.2. Among the NH listeners tested, 15 were male and nine were female, with ages ranging from 18 to 32 years. Normal hearing was defined as having pure-tone audiometric thresholds less than 20 dB hearing level (HL) at all octave frequencies between 250 and 8000 Hz with no reported history of hearing disorders. All experimental protocols were approved by the Institutional Review Board of the University of Minnesota, and all listeners provided written informed consent prior to participating.

Table 2.2.

Experiments completed by each CI user included in the study.

Subject Code	**Experiment 1**	**Experiment 2**	**Experiment 3**
C16	*		X
D02	*		X
D10	X	X	X
D25		X	
D26	X	X	X
D27	*		X
D28	X		X
D35	*		X
D39	X	X	X
D41	X	X	X
D42	*		X
D44	X		X
D46	X		X
D47	X	X	X
D52		X	
D55		X	

* = attempted but did not pass screening
X = =successfully completed

2.2.1.2 Stimuli

The speech materials were comprised of sentences taken from the IEEE speech corpus (IEEE, 1969), recorded by a single female talker. The sentences were presented in either a noise or tonal masker, or in quiet. The Gaussian noise was spectrally shaped to match the long-term spectrum of the IEEE speech corpus. The tone frequencies were selected to match the center frequencies of the CI electrodes of each individual CI user tested, and the amplitudes were equated in terms of their rms to produce the same output level from each channel of the CIs as the noise masker. The 16 center frequencies from the standard clinical map for Advanced Bionics CIs were used to generate the stimuli for the NH listeners. The center frequencies for the individual CI users and for the standard clinical map are shown in Table 2.3.

Table 2.3

Center frequencies (CFs) of each CI user's clinical map, as well as those used for the NH listeners that were used for the implementation of the vocoding.

Subject Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
All NH	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
C16	434	644	954	1414	2096	3108	6206	OFF	NA	NA	NA	NA	NA	NA	NA	NA
D02	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D10	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D25	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D26	OFF	386	463	556	668	804	965	1160	1394	1674	2012	2417	2904	3490	4193	6638
D27	386	463	556	668	804	965	1160	1394	1674	2012	2417	2904	3490	4193	6638	OFF
D28	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D35	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D39	386	463	556	668	804	965	1160	1394	1674	2012	2417	2904	3490	4193	6638	OFF
D41	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D42	338	472	576	700	852	1038	1264	1538	1872	2280	2776	3380	4114	6609	OFF	OFF
D44	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D46	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D47	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D52	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665
D55	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665

The speech and masker were then passed through a tone-excited envelope vocoder (Dorman, Loizou, Fitzke, & Tu, 1998; Whitmal, Poissant, Freyman, & Helfer, 2007). The stimulus was divided into 16 frequency subbands (with the exception of those CI users who had fewer than 16 active channels) with cutoff frequencies and center frequencies of each subband made equal to those in the clinical maps of the individual CI users. For the NH listeners, the standard Advanced Bionics clinical map was used to set the subband frequencies (see Table 2.3). The bandpass filters used to generate the subbands were high-order (947) FIR filters, generated with Matlab's *fir1* function, producing very little overlap between the spectral content of adjacent subbands and a flat frequency response (±0.05 dB) within the entire passband. The impulse responses from the filters were time-aligned, reaching their peaks at a delay of approximately 20 ms, independent of filter center frequency. Two conditions were generated, one termed "interleaved" and one termed "overlapping". In the overlapping condition, both the speech and the masker were mixed at the appropriate signal-to-masker ratio (SMR) and then passed through the even-numbered vocoder channels (i.e. 2, 4, … 16), resulting in eight equally spaced frequency subbands (top panel Figure 2.1). In the interleaved condition, the speech was passed through the even-numbered channels (as before), but the masker was passed through the odd-numbered channels, resulting in spectral separation between the speech and the masker (bottom panel Figure 2.1).

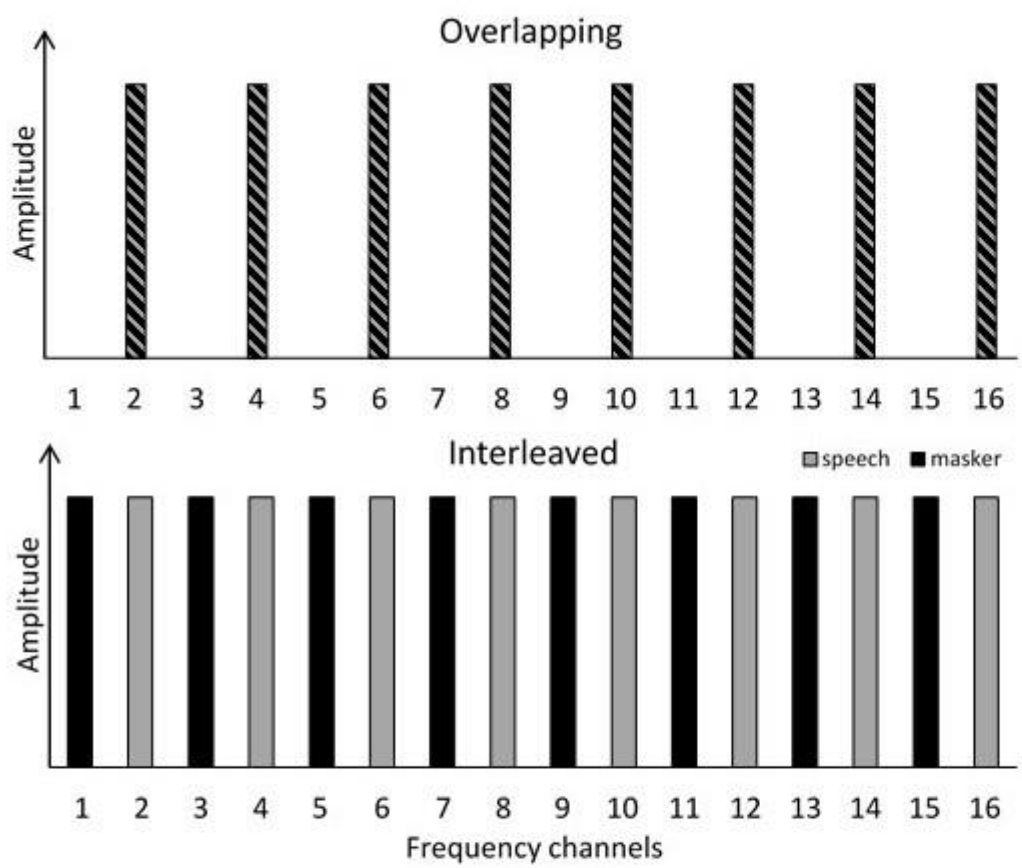

Figure 2.1

Schematic diagram of the two masker configurations used in Experiment 1. The top panel shows the condition in which speech and masker are overlapping in the even channels, and the bottom panel shows the condition in which speech and masker are interleaved with speech in the even channels and masker in the odd channels.

In both cases, the temporal envelope from each subband was extracted using a Hilbert transform, and the resulting envelope was lowpass filtered using a fourth-order Butterworth filter with a cutoff frequency of 50 Hz. This cutoff frequency was chosen to reduce possible voicing periodicity cues and to reduce the possibility (for NH listeners) that the vocoder produced spectrally resolved components via the amplitude modulation of the tonal carriers. The resulting temporal envelopes were used to modulate pure-tone carriers with frequencies corresponding to the center frequencies of each channel, which were then presented to the CI users and to the group of eight NH listeners who were assigned to the “No spread” conditions. Electrodograms (generated with software supplied by Advanced Bionics) showing the stimulation applied to each electrode for both the overlapping and interleaved conditions are shown in Figure 2.2 using the Optima stimulation strategy and the tone maskers. Since all but one of the CI participants used processing strategies that included some form of current steering (Fidelity 120 or Optima), there was some some degree of cross-talk between channels in both conditions. Most cross-talk is observed with the Optima strategy, which has a maximum assignment to one electrode of 75% of the current (with the other 25% to the other member of each electrode pair that constitute a virtual channel). The effects of cross-talk (as well as the limited filter resolution) can be seen most clearly in the overlapping condition, where some stimulus can be observed in the odd electrodes, despite no intended stimulation of these electrodes. However, despite some interaction between channels, the electrodograms show that the two configurations resulted in very different stimulation patterns (compare top and bottom panels of Figure 2.2). In particular, in the interleaved condition (top panel), the speech envelope is observed only in the even channels. In the odd channels, the current level representing the tone maskers is suppressed during the high-amplitude portions of the speech, presumably due to a combination of automatic gain control, compression, and current steering. No attempt was made to simulate this effect in the vocoder.

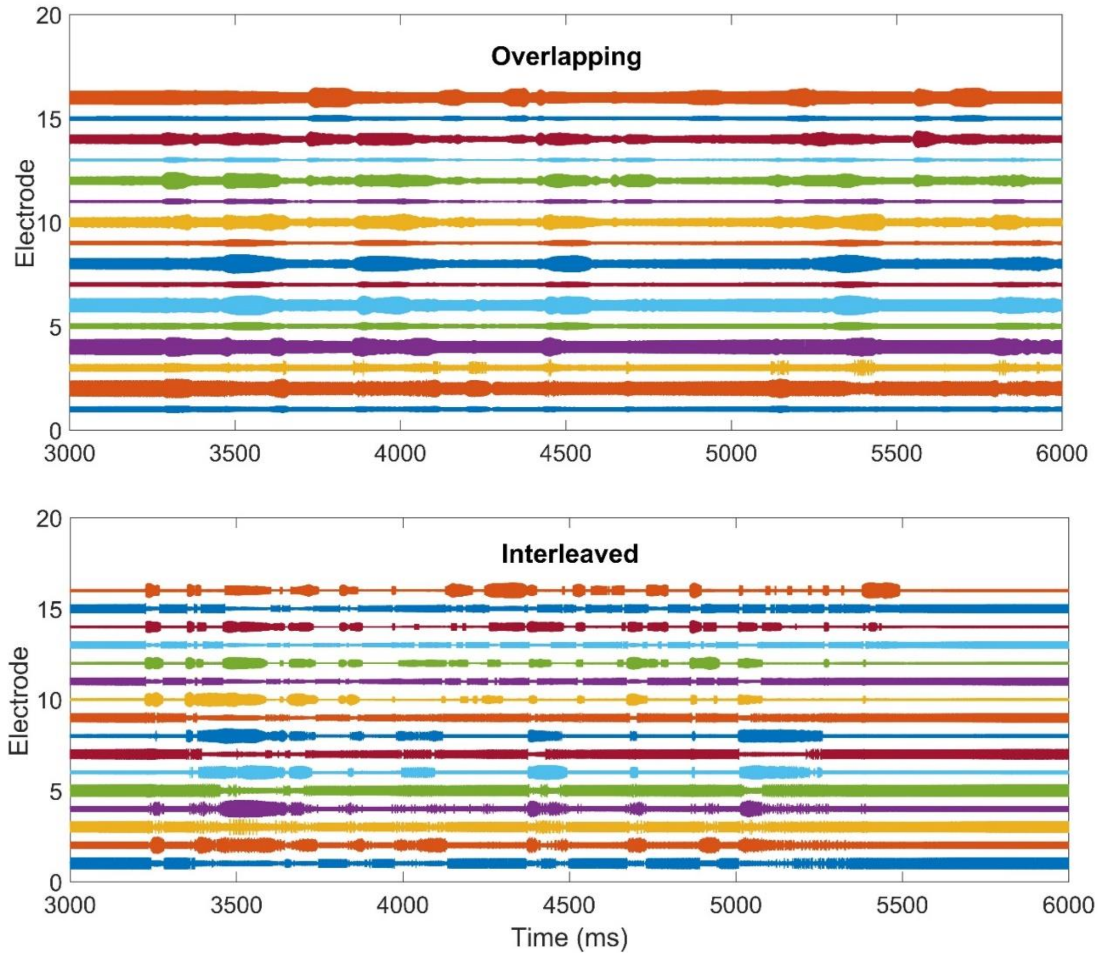

Figure 2.2

Electrodograms of the two masker configurations used in the present study. The top panel shows the condition in which speech and masker are overlapping in the even channels, and the bottom panel shows the condition in which speech and masker are interleaved with speech in the even channels and masker in the odd channels. The first sentence of the IEEE corpus and a tone masker with SMR of 5, processed by the Optima strategy, is shown in both panels.

For the other two groups of NH listeners, the effects of current spread were simulated via the vocoder in the same way as in Oxenham and Kreft (2014): each carrier was modulated by the weighted sum of the intensity envelopes from all 16 channels. The weights used in this sum were selected to produce slopes of either 24 dB/oct or 12 dB/oct to simulate different degrees of spectral smearing or current spread.

The level of the speech after filtering was 51 dB SPL for the CI users and was between 51 and 56 dB SPL for the NH listeners, depending on the degree of spectral smearing, as measured 1 m from the loudspeaker, corresponding to the position of the participant's head. The masker level was adjusted to produce the desired SMR, referring to the levels of the speech and masker before filtering. The masker was gated on 1 s before the beginning of each sentence and was gated off 1 s after the end of each sentence. The SMRs were selected in advance, based on pilot data, to span a range of performance between 0% and 100% word recognition for each condition. The resulting range was -15 to 10 dB SMR for the no spread and 24 dB/oct spread NH groups and 0 to 20 dB SMR for the 12 dB/oct spread NH group and for the CI group.

2.2.1.3 Procedure

The stimuli were generated and processed using MATLAB (The Mathworks, Natick, MA). The sounds were converted via a 24-bit digital-to-analog converter (L22, LynxStudio, Costa Mesa, CA) at a sampling rate of 22,050 Hz, and were presented via an amplifier and a single loudspeaker, placed approximately 1 m from the listener at 0° azimuth and level with the listener's head. The listeners were seated individually in a double-walled, sound-attenuating booth with approximate interior dimensions of 6'8" x 7'10" x 6'6". Bilateral CI users were instructed to use whichever processor they thought gave them better speech perception and to remove the other processor. One participant with a hearing aid in the ear contralateral to her CI was also instructed to remove it before beginning the experiment. The hearing provided by this contralateral ear without the hearing aid was deemed negligible, as a recent audiogram showed a flat severe sensorineural hearing loss with audiometric thresholds at octave frequencies between 250 and 8000 Hz of between 65 dB HL and 75 dB HL. Thus, the speech presented at an overall level of 51 dB SPL was inaudible without the hearing aid.

Listeners responded to sentences by typing what they heard on a computer keyboard. They were encouraged to guess individual words, even if they had not heard or understood the entire sentence. Sentences were scored for keywords correct as a proportion of the total number of keywords presented. Initial scoring was automatic, with each error then checked manually for potential spelling errors. Before the actual experiment took place, listeners were presented with two sentence lists (of 10 sentences each) of the HINT speech corpus (Nilsson, Soli, & Sullivan, 1994) to acclimate them to the stimuli before the scored sentences were presented. This procedure was repeated for each new masker type (noise or tone) and configuration (spectrally overlapping or interleaved).

In the actual experiment, two sentence lists of 10 sentences each were completed for each combination of masker type, SMR, and configuration (overlapping and interleaved). Each NH listener completed the experiment using one of three simulated spread conditions (no spread, 24 dB/oct spread, or 12 dB/oct spread). The stimuli for CI users were processed using a vocoder with no spread, with the center frequencies of each subband matching the center frequencies of each active channel in their map. The proportions of correct scores were converted to rationalized arcsine units (RAU) (Studebaker, 1985) to compensate for possible floor or ceiling effects before statistical analysis.

2.2.2 Results

The mean results from the three NH groups (no spread, 24 dB/oct spread, and 12 dB/oct spread) and the CI group are shown in the separate panels of Figure 2.3. The results from tone and noise maskers are denoted by circles and triangles, respectively. The filled symbols represent data from the conditions in which the speech and masker overlapped in the same frequency bands, and the open symbols represent data from conditions in which the speech and masker were presented to different interleaved frequency bands.

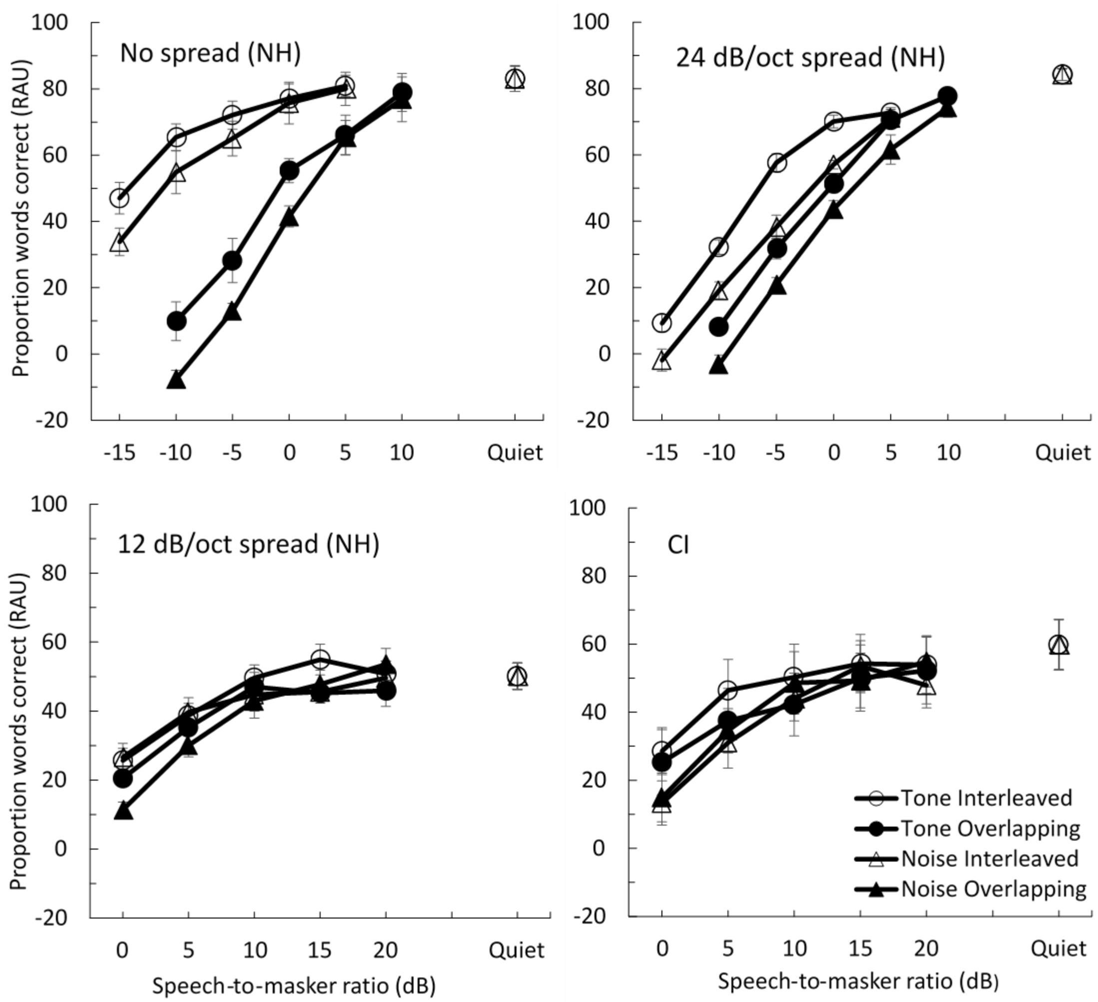

Figure 2.3

Speech perception for the CI listeners and the three groups of NH listeners are shown in the four panels. The RAU-transformed proportion of keywords from sentences reported correctly is plotted as a function of signal-to-masker ratio for four different combinations of two maskers (tone and noise) and two spectral configurations (interleaved and overlapped). Note the different ranges of signal-to-masker ratio in the upper and lower panels. Error bars represent ± 1 standard error of the mean between listeners.

As expected, increasing the amount of simulated spread in the NH listeners led to poorer speech perception. Also with increasing spread, the difference between the tone and noise maskers became less pronounced. This effect is expected, based on the results of Oxenham and Kreft (2014). They reasoned that the lack of difference at large spread values was due to the effective smoothing of the temporal envelope of the noise masker, due to the overlap between adjacent channels, making the noise masker more tone-like

(Oxenham & Kreft, 2014). The NH listeners also benefited less from the spectral separation of the speech and masker in the interleaved condition (i.e., difference between the overlapped and interleaved conditions) with increasing spread. For instance, with the no-spread NH group, there was a clear separation in performance in conditions with the tone and noise maskers (compare circles and triangles) for all but the highest SMRs, and the difference in performance between the spectrally interleaved and overlapped maskers was very large, reaching a mean difference of about 60 RAU at the lowest SMRs. In contrast, with the 12-dB/oct-spread group, performance was much poorer overall (note the different SMRs tested) and was very similar regardless of masker type or spectral overlap. Given the large differences between groups, in terms of pattern of results, the SMRs tested, and overall performance, separate repeated-measures ANOVAs were performed for each group.

For the no-spread NH group, a three-way repeated-measures ANOVA on the RAU-transformed proportion of words correctly reported confirmed significant main effects of masker type (tone or noise) [$F(1, 7)=7.8$, $P = 0.027$, partial $\eta^2=0.528$], condition (overlapping vs interleaved) [$F(1, 7)=186$, $P < 0.001$, partial $\eta^2=0.964$], and SMR [$F(3, 21)=290$, $P < 0.001$, partial $\eta^2=0.976$]. There were also interactions between masker type and SMR [$F(3, 21)=5.5$, $P = 0.006$, partial $\eta^2=0.439$], and between condition and SMR [$F(3, 21)=46$, $P < 0.001$, partial $\eta^2=0.867$].

The NH group listening through the vocoder with the 24 dB/oct spread showed a similar pattern of statistical outcomes, with a significant effect of masker type [$F(1, 7)=61.0$, $P < 0.001$, partial $\eta^2=0.897$], condition [$F(1, 7)=174$, $P < 0.001$, partial $\eta^2=0.961$], and SMR [$F(3, 21)=357$, $P < 0.001$, partial $\eta^2=.981$]. Results from this group also showed an interaction between condition and SMR [$F(3, 21)=10$, $P < 0.001$, partial $\eta^2=.597$]. Although the effect of spectral condition was significant, it was smaller than in the no-spread group, especially at the lower SMRs. For instance, at -5 dB SMR with the noise masker, the average increase in score from the overlapped to the interleaved condition was about 51 RAU in the no-spread group, whereas it was only 17 RAU in the 24 dB/oct spread group, despite similar levels of performance of both groups in the overlapped condition.

For the NH group listening through the vocoder with 12 dB/oct spread, there were only significant effects of condition [$F(1, 7)=8.8$, $P = 0.021$, partial $\eta^2=0.556$] and SMR [$F(4, 28)=66.6$, $P < 0.001$, partial $\eta^2=0.905$]; the effect of masker type was not significant [$F(1, 7)=0.62$, $P = 0.456$, partial $\eta^2=0.082$]. Similar to the other two NH groups, there was an interaction between condition and SMR [$F(4, 28)=3.7$, $P = 0.016$, partial $\eta^2=0.343$] but in contrast to the other two NH groups, there was also a three-way interaction between masker type, condition and SMR [$F(4, 28)=3.5$, $P = 0.019$, partial $\eta^2=0.334$]. Again, although the effect of condition reached significance, it was small, with the mean difference rarely exceeding 10 RAU for either the noise or tone maskers.

The results from the eight CI users resemble most closely those of the NH group listening through the vocoder with the 12 dB/oct spread. Despite the apparent similarity of the results from the two groups, the statistical analysis resulted in slightly different outcomes. The same three-way repeated-measures ANOVA performed on the data from the CI users confirmed a significant effect of SMR [$F(4, 28)=76.1$, $P < 0.001$, partial $\eta^2=0.916$], but no significant main effect of masker type [$F(1, 7)=3.9$, $P = 0.089$, partial $\eta^2=0.358$], and no main effect of condition (interleaved vs. overlapped) [$F(1, 7)=0.8$, $P = 0.401$, partial $\eta^2=0.102$]. There was a significant interaction between masker type and SMR [$F(4, 28)=5.1$, $P = 0.003$, partial $\eta^2=0.421$] and between masker type and condition [$F(1, 7)=6.0$, $P = 0.044$, partial $\eta^2=0.463$]. The interactions with masker type presumably reflect the fact that scores with the tone masker appear higher than scores with the noise masker for the interleaved conditions at SMRs of 0 and 5 dB, but only for the overlapped conditions at an SMR of 0 dB.

The effects of the interleaved and overlapped conditions in the different groups can be seen more clearly in Figure 2.4, which replots the data from Figure 2.3, but with the data from the different groups shown in the same panel. Results using the tone masker are presented in the top panel and results using the noise masker are shown in the bottom panel. Mean data for the no spread, 24 dB/oct spread, 12 dB/oct spread, and CI groups are represented by circles, triangles, squares, and diamonds, respectively. As before, the filled symbols represent data from the overlapped conditions and the open symbols represent data from the interleaved conditions. As noted above, the benefit gained by listeners in the spectrally interleaved conditions decreased with increasing spread in the NH group, and was essentially absent in the CI group.

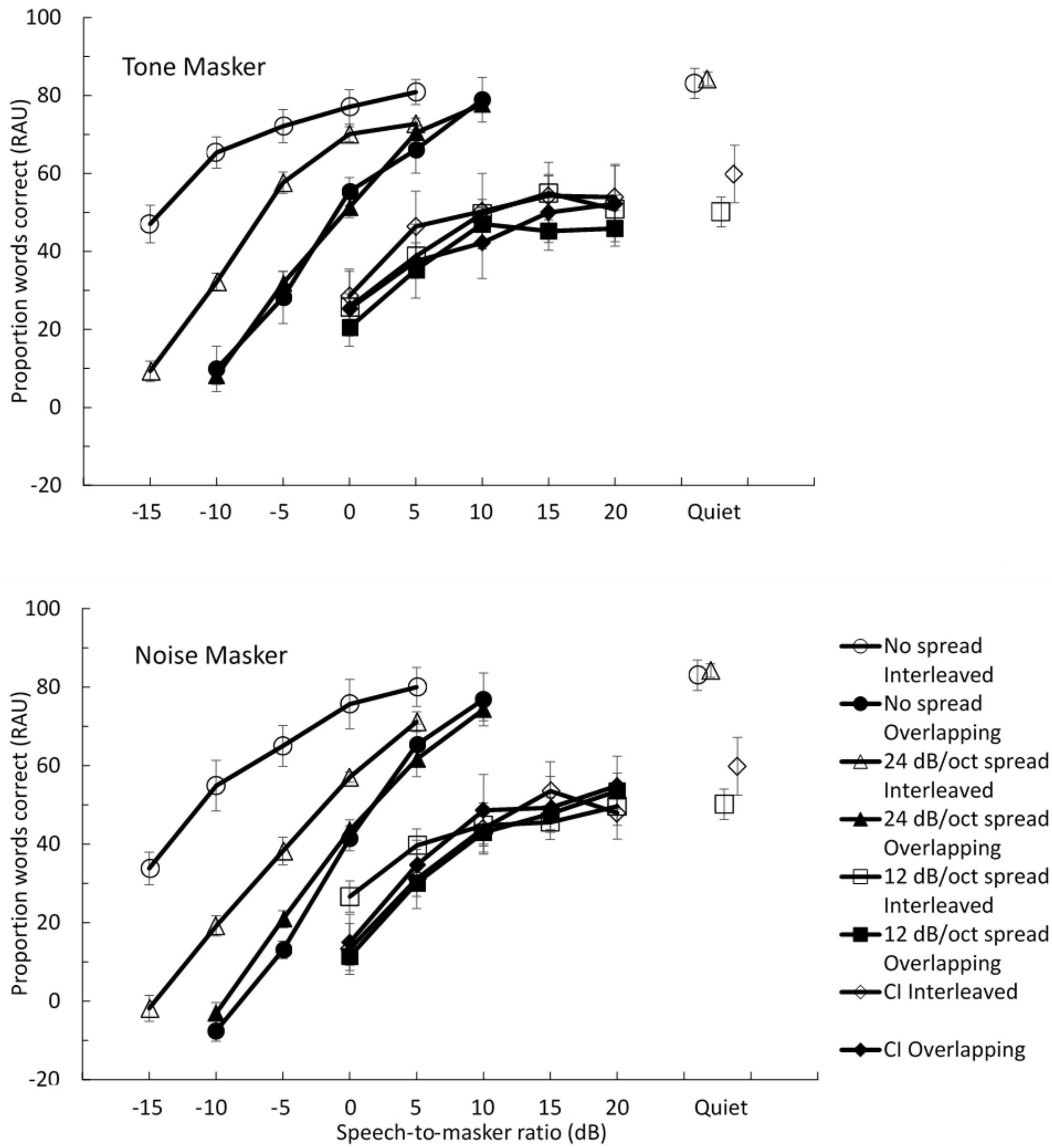

Figure 2.4

Results from Figure 2.3, replotted to facilitate comparisons between the four groups. Performance in conditions with tone maskers are shown in the top panel and performance in conditions with speech-shaped noise maskers are shown in the bottom panel. Open symbols represent interleaved conditions and filled symbols represent overlapping conditions. Error bars represent ± 1 standard error of the mean between listeners.

The results in quiet are shown at the right of each panel. Each participant completed two to four lists of 10 sentences in the quiet condition to establish a baseline average. The average across trials for each participant was used to calculate a mean for each group, shown at the right of each panel. A between-subjects one-way ANOVA on the performance in quiet showed a significant effect of group [$F(3,31) = 13.4$, $P < 0.001$]. Post-hoc contrasts (with Bonferroni correction and six possible contrasts yielding a criterion value, $\alpha = 0.05/6 = 0.0083$) revealed a significant difference between the 24 dB/oct spread group and the 12 dB/oct spread group ($P = 0.001$), but not between the no spread and 24 dB/oct spread groups ($P = 0.836$). The results from the CI group were significantly poorer than those of the no-spread and 24 dB/oct spread groups ($P = 0.002$ and $P = 0.001$, respectively) but were not significantly different from those of the 12 dB/oct group ($P = 0.154$).

2.2.3 Discussion

2.2.3.1 Benefit of spectrally interleaved maskers

The difference in speech understanding between conditions with overlapped and interleaved maskers can be viewed as spectral masking release. To investigate the relationship between spectral resolution (as manipulated through varying amounts of spread) and spectral masking release, we subtracted speech recognition RAU scores in the spectrally overlapped condition from the RAU scores in the interleaved condition. The upper panels of Figure 2.5 show this masking release for each group as a function of SMR. Considering the mean data from each NH group, the expected trend is clear: decreases in spectral resolution by increasing spread, from no spread to 24 dB/oct to 12 dB/oct (open circles, triangles, and squares, respectively), leads to less masking release. This is especially apparent at the lower SMRs, where performance is well below ceiling. As already observed in the raw data, the difference scores for the CI group are generally close to zero, indicating little or no spectral masking release.

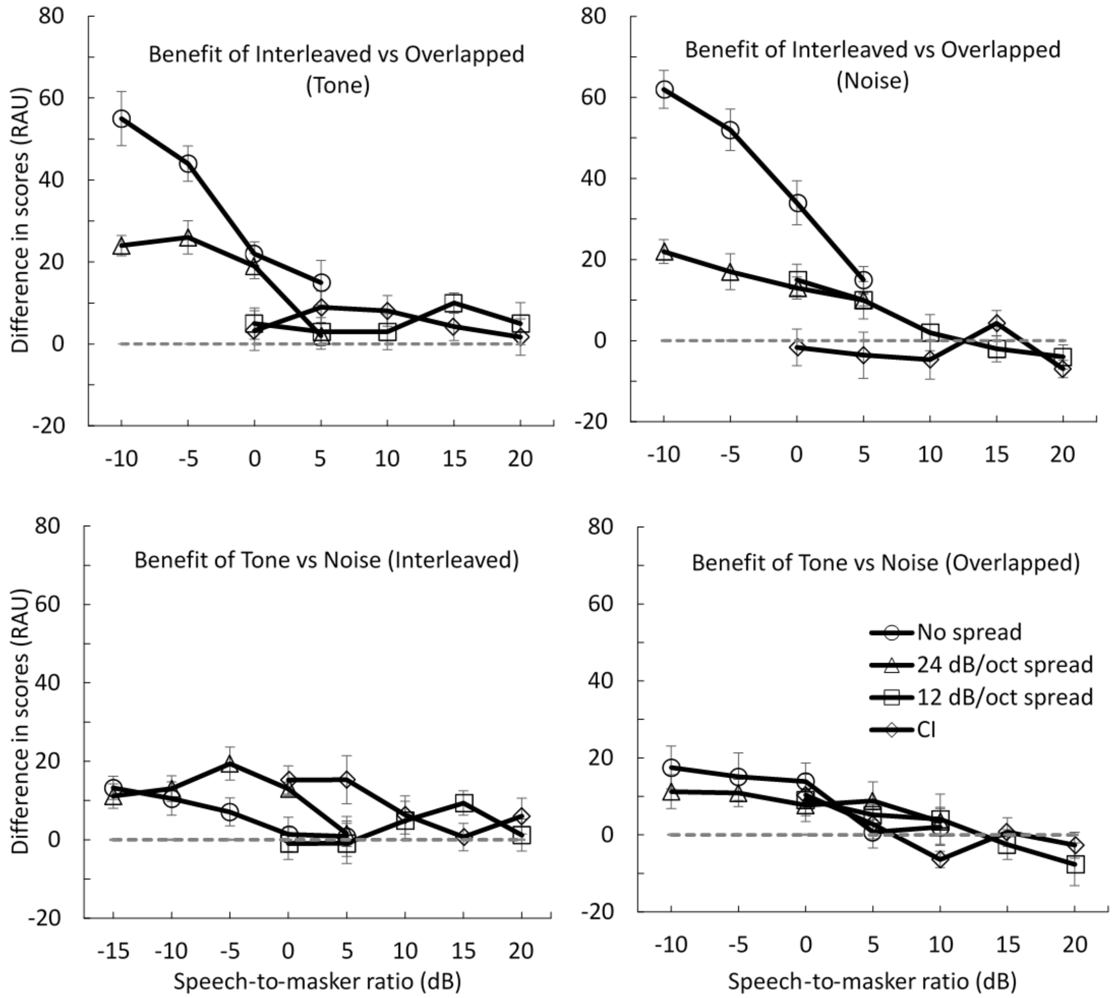

Figure 2.5

Speech perception for the CI listeners and the three groups of NH listeners are shown in the four panels. The benefit gained, or difference in RAU-transformed proportion of keywords from sentences reported correctly, is plotted as a function of speech-tomasker ratio. The top two plots show the benefit in speech perception for the interleaved vs overlapping condition, for each masker (tone and noise), respectively. The bottom two plots show the benefit in speech perception for the tone masker over the noise masker, for each condition (interleaved and overlapped), respectively. Results from the no spread, 24 dB/oct spread, 12 dB/oct spread, and CI groups are represented by open circles, triangles, squares, and diamonds. Note the different ranges of speech-to-masker ratio across groups. Error bars represent ±1 standard error of the mean between listeners.

2.2.3.2 Benefit of tone over noise maskers

Earlier studies have found that inherent temporal-envelope fluctuations in steady-state noise can account for a substantial proportion of the masking of speech in NH listeners (Stone et al. 2011, 2012; Stone and Moore 2014). A more recent study found that the same was not true for CI users, who exhibited as much masking with tone maskers that had no inherent fluctuations as they did with noise maskers (Oxenham & Kreft, 2014). The difference appeared to be due to loss of spectral resolution leading to interactions between the temporal envelopes from neighboring channels, which in turn resulted in an effective smoothing of the temporal-envelope fluctuations. If spectral resolution limits the effect of inherent masker fluctuations, then we may expect a relationship between spectral spread and the benefit of tone over noise maskers. This relationship is shown in the lower panels of Fig. 5, where speech scores with the noise masker are subtracted from speech scores with the tone masker.

The results from the no-spread and 24 dB/oct NH groups (circles and triangles) show modest benefits of the tone over noise maskers, particularly at the lower SMRs, where performance is well below ceiling. There is also increased benefit for the no-spread group, over the 24 dB/oct spread and 12 dB/oct spread groups, when speech and masker were overlapping. The CI group showed a modest benefit of the tone over the noise masker at lower SMRs in the interleaved condition, but when performance is considered across all conditions, no benefit of tone over noise maskers was observed, consistent with results from an earlier study that used all channels and only spectrally overlapped maskers (Oxenham & Kreft, 2014).

2.3 Experiment 2: Speech Perception with Virtual Channels

2.3.1 Rationale

All eight CI users from Experiment 1 used processing strategies that involved current steering (see Table 1). This implies that although each electrode is assigned to the frequencies listed in Table 1, the virtual channels produced by simultaneous stimulation of adjacent electrodes actually have center frequencies at the midpoints between the listed frequencies. In other words, by stimulating at the electrode frequencies, we were actually stimulating at the corner frequencies of the virtual channels. Theoretically and empirically, our approach seemed most appropriate, in terms of stimulating the actual electrodes present in the device. Nevertheless, it is possible that by bypassing the virtual-channel design of the CIs, we may have inadvertently lost some of the spectral resolution capabilities of the device. For this reason, we repeated critical conditions from Experiment 1 with a map that was centered on the virtual channels rather than the electrode center frequencies.

2.3.2 Methods

2.3.2.1 Participants

A total of eight CI users completed this experiment, with five of the eight having also participated in Experiment 1 (see Table 2.2). The other three CI users who completed Experiment 1 were not able to return and participate in this experiment, so three different CI users were recruited. As shown in Table 2.1, the CIs of all eight participants in this experiment used a current steering stimulation strategy. All participants were native speakers of American English. Individual details for the CI users are provided in Table 2.1.

2.3.2.2 Stimuli

The speech materials consisted of different lists of sentences from the same IEEE corpus (IEEE, 1969) used in Experiment 1. The sentences were presented in a noise masker or in quiet. The tone masker from Experiment 1 was not included, since the results showed no difference in performance between tone and noise maskers for CI users (see Figure 2.3). The Gaussian noise was again spectrally shaped to match the long-term spectrum of the IEEE speech corpus.

The speech and the noise masker were then passed through the same tone-excited envelope vocoder used in Experiment 1 (Dorman et al., 1998; Whitmal et al., 2007). In one case, the same channel frequencies were used as in Experiment 1; in the other case, the center frequencies of each channel corresponded to the midpoint between the center frequencies of each pair of active electrodes. In this way, the tone frequencies were placed at the center frequencies of virtual channels of each device. For CI users who had all 16 electrodes active in their clinical maps, 15 midpoint center frequencies were calculated and used to create the two different conditions, interleaved and overlapping, tested in Experiment 1. In the overlapping condition, both the speech and the masker were mixed at the appropriate signal-to-noise ratio (SNR) before filtering and were then passed through the even-numbered "virtual" channels (i.e. 2, 4, … 14), resulting in seven equally spaced frequency subbands. In the interleaved condition, the speech was passed through the even-numbered virtual channels (as before), but the noise masker was passed through the odd-numbered virtual channels (i.e. 1, 3, … 15), resulting in spectral separation between the speech and the masker. Electrodograms showing the stimulation applied to each electrode for both conditions (overlapping and interleaved) and channel manipulations (virtual and standard) are shown in Figure 2.6 using the Optima stimulation strategy and the noise masker. As expected, the difference between interleaved and overlapping stimulation is much less clear, because all stimuli are presented to all electrodes. However, the fine timing of the electrical pulses are such that, in the interleaved condition, the masker is distributed between half the electrode pairs (i.e., electrodes 1 and 2, 3 and 4, etc.), whereas the speech is distributed between the other half of the electrode pairs (i.e., 2 and 3, 4 and 5, etc.).

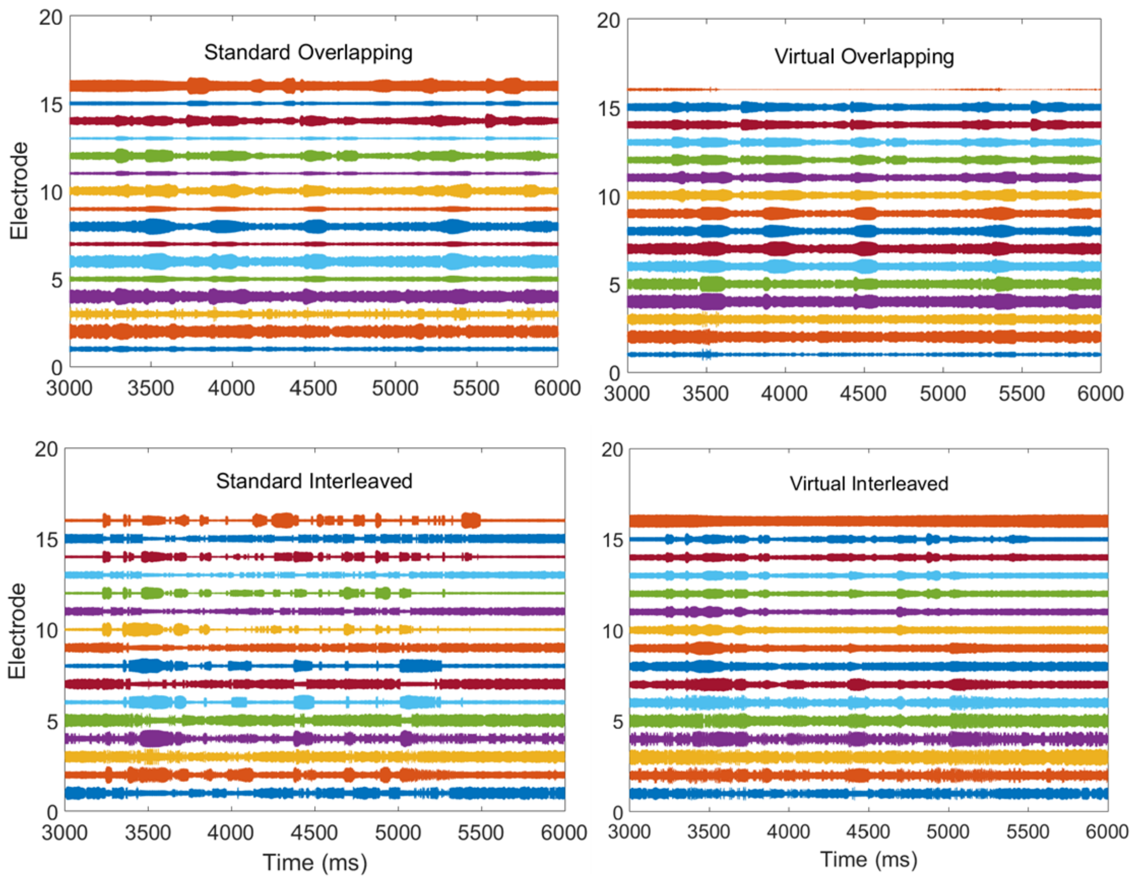

Figure 2.6

Electrodograms of the two masker and two channel configurations used in the experiment 2. The top two plots show the condition in which speech and masker are overlapping in the same channels, and the bottom two plots show the condition in which speech and masker are interleaved in alternating channels. The plots on the left show the standard channel configuration in which stimuli is presented to center frequencies of actual electrodes, and the plots on the right show the virtual channel configuration in which stimuli is presented to the midpoint frequency between electrodes. The first sentence of the IEEE corpus and a noise masker with SNR of 5, processed by the Optima strategy, is shown in all plots.

The level of the speech after filtering was 51 dB SPL, as measured 1 m from the loudspeaker, corresponding to the position of the participant's head. The noise level was adjusted to produce the desired signal-to-noise ratio (SNR), referring to the levels of the speech and noise before filtering. The noise was gated on 1 s before the beginning of each sentence and was gated off 1 s after the end of each sentence. We tested SMRs of 5 and 15 dB and quiet to replicate a representative sample of SMRs used in Experiment 1.

2.3.2.3 Procedure

The stimuli were generated, processed, and presented using the same procedure as in Experiment 1. The listeners were seated individually in a single-walled, sound-attenuating booth, with approximate interior dimensions of 6’ x 5’4” x 6’6”, located in an isolated quiet room. The CI users who had previously participated were instructed to use the same processor they used in Experiment 1 and to remove the other processor (if applicable). The new participants used the CI that they felt gave them better speech perception and were asked to remove the other processor. One participant with a hearing aid in the ear contralateral to her CI was also instructed to remove it before beginning the experiment and insert an ear plug in that ear, to avoid any possible influence of residual hearing.

As in Experiment 1, participants were asked to type what they heard and to guess if they were unsure. Scoring was conducted in the same way as in Experiment 1 and the percentage of correctly identified keywords was similarly transformed into RAU for statistical analysis.

2.3.3 Results and Discussion

Mean results for the CI users are shown Figure 2.7. The results from virtual and standard channel configurations are denoted by squares and triangles, respectively. The filled symbols represent data from the conditions in which the speech and masker were overlapping in the same frequency bands, and the open symbols represent data from conditions in which the speech and masker were presented to different interleaved frequency bands.

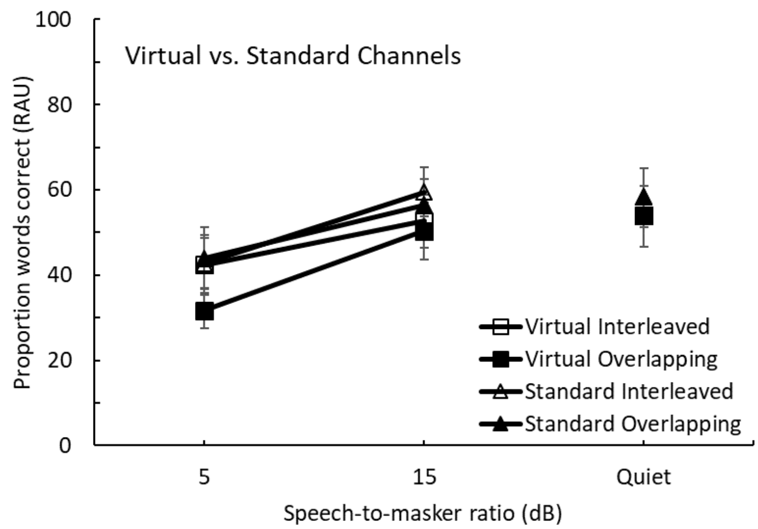

Figure 2.7

Speech perception for CI listeners is shown. The RAU-transformed proportion of keywords from sentences reported correctly is plotted as a function of signal-to-masker ratio for four different combinations of two channel configurations (virtual and standard) and two spectral configurations (interleaved and overlapped). Error bars represent ± 1 standard error of the mean between listeners.

As expected, based on data from Experiment 1, results from the interleaved and overlapping conditions were very similar in both the virtual and standard channel configurations. A repeated-measures ANOVA on data from the noise conditions (excluding measurements in quiet) showed a significant main effect of channel configuration (virtual vs. standard) [$F(1,7)=18.01$, $P=0.004$] and SMR [$F(1,7)=31.76$, $P=0.001$], but no main effect of condition (interleaved vs. overlapping) [$F(1,7)=4.62$, $P=0.069$]. Performance with virtual channels was overall poorer than performance using standard channels, which was expected, given that the virtual channel configuration contained one fewer active speech channel (7) than the standard configuration (8) in both interleaved and overlapping conditions. Overall performance also decreased with decreasing SMR, as anticipated.

The same ANOVA revealed no significant interactions between channel configuration, condition, or SMR ($P > 0.1$ in all cases). Since there was no significant benefit to speech perception of the spectrally interleaved masker in either the standard or the virtual channel configuration and no interactions between channel configuration and either condition or SMR (or both), it seems that the results and conclusions from

Experiment 1 were not substantively affected by the current steering strategies inherent to the CI processors of participants. Overall, the results from this experiment confirm that the methodology used for CI users in Experiment 1 was not confounded by the current steering strategies active in the processors of the participants. Based on this finding, all subsequent experiments in this study use only the standard channel configuration, as shown in Table 2.3.

2.4 Experiment 3: Spectral Ripple Discrimination

2.4.1 Rationale

Simulated decreases in spectral resolution led to decreased speech intelligibility, decreased difference between tone and noise maskers, and decreased release from masking when the masker was spectrally interleaved rather than overlapping with the target speech. In this experiment, we provided a direct measure of spectral resolution by measuring thresholds for spectral ripple discrimination in the CI users, as well as in the NH listeners through the same vocoders that were used in Experiment 1. Spectral ripples have been used to measure auditory spectral resolution for several decades (Wilson and Evans 1971; Houtgast 1977; Supin et al. 1994; Henry and Turner 2003; Anderson et al. 2012). Spectral ripple discrimination provides a rapid estimate of spectral resolution, which has been shown to correlate well with the more time-consuming psychophysical spatial tuning curves in CI users, when targeting the same spectral region (Anderson et al., 2011). Its estimates also match well with those from spectral ripple detection thresholds at comparable ripple rates, without being susceptible to the same temporal-envelope confounds at high ripple rates (Anderson et al., 2012). Thus, it remains a relatively well-validated measure of spectral resolution in both acoustic and electric hearing.

2.4.2 Methods

2.4.2.1 Participants

The same 13 CI users who completed the screening for Experiment 1 also took part in this experiment (see Table 2.2). The same 24 NH listeners were tested in conditions corresponding to their assigned groups from Experiment 1 (no spread, 24 dB/oct spread, and 12 dB/oct spread).

2.4.2.2 Stimuli

Spectrally rippled noise was generated using MATLAB (Mathworks, Natick, MA). Gaussian broadband (350-5600 Hz) noise was spectrally modulated, with sinusoidal variations in level (dB) on a log-frequency axis (as in Litvak et al. 2007) using the equation:

$$X(f) = 10^{(D/2)\sin\{2\pi[\log_2(f/L)]f_S+\theta\}/20}$$

Where $X(f)$ is the amplitude at frequency f (in Hz), D is the spectral depth or peak-to-valley ratio (in dB), L is the lower cut-off frequency of the noise pass band (350 Hz in this case), f_s is the spectral modulation frequency (in ripples per octave), and θ is the starting phase of the ripple function. Logarithmic frequency and intensity units were used as these are generally considered to be more perceptually relevant than linear units. Sinusoidal modulation was used, as this lends itself to linear systems analysis and has been used in a number of studies of spectral modulation perception in normal (acoustic) hearing (Eddins & Bero, 2007; Saoji & Eddins, 2007). The peak-to-valley ratio of the stimuli was held constant at 30 dB. This large peak-to-valley ratio provides a large spectral contrast in the stimulus, so that perceptual performance is most likely to be determined by the spectral resolution of the listener, just as the “notched-noise” method (Glasberg & Moore, 1990; Patterson, 1976) uses an “infinite” peak-to-valley ratio between the passband and stopband of the noise to probe frequency selectivity. The stimulus duration was 400 ms, including 20-ms raised-cosine onset and offset ramps. For NH listeners, the stimulus was either left unaltered or was passed through one of the same tone-excited envelope vocoders used in Experiment 1, with 16 frequency channels that had effectively no spectral spread, or that produced spectral spread equivalent to 24 or 12 dB/oct. The CI users were presented with the stimulus unaltered in one condition and passed through the same vocoder, without spread, in another condition.

The stimuli were presented using the same setup as in Experiment 1, via a loudspeaker positioned at approximately head height and about 1 m from the participant in a double-walled, sound attenuating booth. The average sound level of the noise was set to 60 dBA when measured at the location corresponding to the participant's head. In order to reduce any possible cues related to loudness, the noise level was roved across intervals within each trial by ±3 dB. The starting phase of the spectral modulation was selected at random with uniform distribution for each trial to reduce the potential for any consistent local intensity cues that fixed-phase stimuli might create.

2.4.2.3 Procedure

Bilateral CI users were instructed to use the same processor they chose to use in Experiment 1 (or in the screening for Experiment 1 if they failed to meet the speech perception requirements) and to remove the other processor. As in Experiment 1, one participant with a hearing aid in the ear contralateral to her CI was asked to remove it before beginning the experiment. All CI users were asked to use processor settings (volume, sensitivity, program) typical of everyday use. A three-interval, three-alternative forced-choice procedure was used. All three intervals contained spectrally rippled noise. In each trial, two intervals contained spectral ripples that had the same starting phase (selected at random from a uniform distribution on each trial), and in the remaining interval the phase of the ripple was reversed (180° phase shift). The order of the three intervals (two same, one reversed) was randomized on every trial. Listeners were instructed to choose the interval that sounded different from the other two. Correct-answer feedback was provided after each trial. The first trial of each run started at a ripple rate of 0.25 ripples per octave (rpo), corresponding to a single ripple across the 4-octave passband. In each successive trial, the ripple rate was varied adaptively using a 2-down, 1-up rule, with rpo initially increasing or decreasing by a factor of 1.41. After the first two reversals, the step size changed to a factor of 1.19 and decreased again to a factor of 1.09 after two more reversals. Each run was considered complete after a total of ten reversals, and the geometric mean ripple rate at the last six reversal points were used to determine the threshold. The adaptive procedure allowed for ripple rates below 0.25 rpo, but such low rates were never needed by the participants. Each CI user completed six runs under non-vocoded conditions and six runs using a vocoder without any additional spread (same as the no-spread NH group). Each NH listener completed six runs under non-vocoded conditions and six runs in their respective vocoded condition (depending on the group to which they were assigned). In each condition, the first two runs for each participant were considered practice, and the last four thresholds were used to compute a geometric mean threshold for each individual participant.

2.4.3 Results and Discussion

Individual thresholds for the 13 CI users, and mean thresholds for all groups from Experiment 1 (NH and CI) are shown in Fig. 8. Thresholds for individual CI users are represented by gray bars and group averages are shown in black. The pattern of thresholds using the unprocessed stimuli were similar, but somewhat higher overall [mean of 1.9 vs. 1.4 ripples/oct; paired t-test: $t(12) = 3.76$, $P = 0.003$]. Individual gray bars with a diagonal pattern represent CI users who did not have speech perception scores high enough to participate in Experiment 1. Error bars represent one standard deviation (not standard error) of the mean of the group averages to provide a sense of the spread of the individual thresholds.

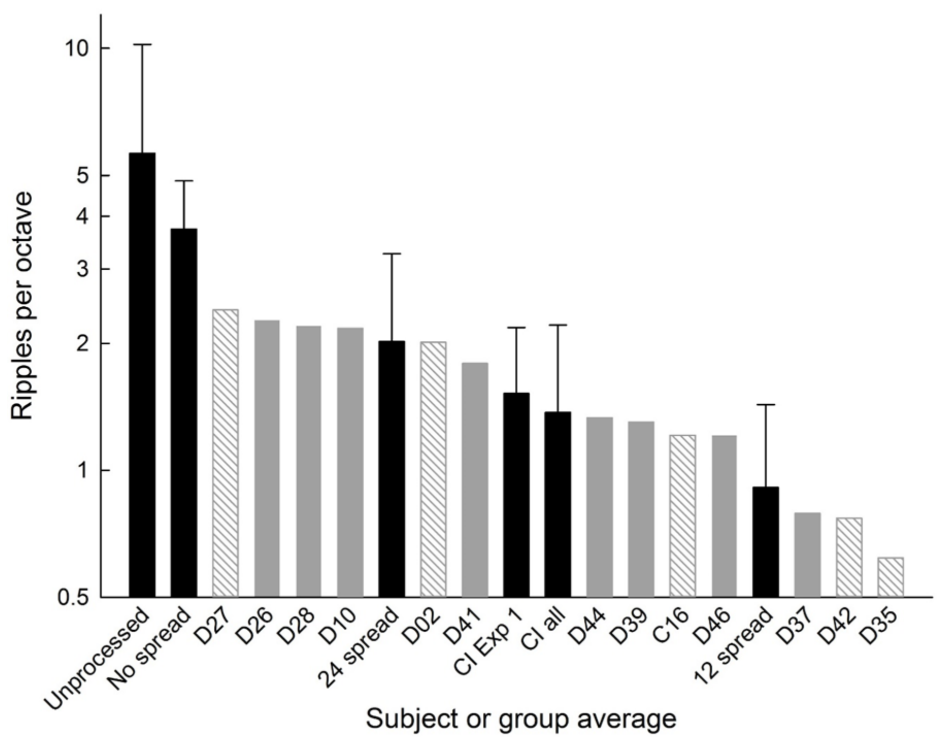

Figure 2.8

Individual spectral ripple discrimination thresholds for the CI users, along with mean thresholds for the CI group and the three NH groups. The gray bars represent individual ripple discrimination thresholds for CI users. Those with diagonal shading represent CI users who did not achieve high enough speech performance to participate in Experiment 1. The black bars represent mean ripple discrimination thresholds for each group tested in Experiment 1. Error bars represent 1 standard deviation above the mean between listeners for each group tested in Experiment 1.

Interestingly, CI users who had speech perception scores too poor to be included in Experiment 1 did not have noticeably poorer spectral ripple thresholds when compared to the CI users who were included. A two-sample t-test (equal variances assumed) found no significant difference in the mean ripple rate at threshold between the two groups [$t(11) = 0.97$, $P = 0.104$].

To confirm the apparent differences between the groups, a between-subjects one-way ANOVA was conducted on the log-transformed ripple thresholds with the CI group that completed the speech experiment and the three normal-hearing groups under vocoded conditions. There was a significant effect of group [$F(3,28) = 17.6$, $P < 0.001$]. Contrast analysis (with Bonferroni correction and six possible contrasts yielding a criterion value, $\alpha = 0.05/6 = 0.0083$) revealed that the mean threshold for the CI group was significantly different from that of the no-spread ($P < 0.001$) groups but not significantly different from that of the 24 dB/oct spread group ($P = 0.156$) or the 12 dB/oct spread group ($P = 0.016$). Considering only the NH groups, a separate ANOVA revealed a significant linear trend ($P < 0.001$), confirming the systematic improvement in thresholds as spectral resolution improved from 12 dB/oct to 24 dB/oct to the no-spread condition.

Finally, to determine the extent to which ripple discrimination thresholds reflect underlying spectral resolution, the performance of the NH group in the no-spread condition was compared with that expected from ideal analysis. With 16 channels, logarithmically spaced from 333 to 6665 Hz (4.3 octaves), each filter has a bandwidth of 0.27 octaves. Ripples should no longer be reliably discriminable when the ripple period (1/ripple rate) is less than or equal to the filter bandwidth, or when the ripple rate is 1/0.27 or 3.7 ripples per octave. As can be seen in our data, the no-spread condition has a mean threshold of 3.73 ripples per octave, which corresponds well to the theoretically predicted threshold of 3.7 ripples per octave. In summary, the ripple discrimination measure appears to provide an accurate assessment of underlying spectral resolution, at least in the NH groups.

2.5 Comparing Speech Perception and Spectral Resolution

As outlined in the introduction, our prediction was that speech perception should be related to spectral resolution, and that the relationship between speech perception in noise and spectral resolution should be stronger when the speech and noise are not spectrally matched (e.g., when they are spectrally interleaved). This section considers the various hypothesized relationships in light of the current data.

2.5.1 Speech perception in quiet and ripple discrimination thresholds

The relationship between spectral resolution and speech perception in quiet can be seen in Figures 2.4 and 2.5. The RAU-transformed proportion of keywords reported correctly in quiet are shown on the far right of the panels in Figure 2.4. As expected, speech perception improved for the NH listeners as the spectral spread was reduced from 12 to 24 dB/oct, but there was no further improvement between the 24 dB/oct condition and the no-spread condition. In contrast, spectral ripple discrimination thresholds improved steadily from the 12-dB/oct through the 24-dB/oct to the no-spread condition (Figure 2.8). Thus, speech perception in quiet appeared to reach a plateau in performance in the 24 dB/oct condition, whereas spectral ripple discrimination continued to improve. The fact that performance was still well below perfect (less than 85 RAU), suggests that was not solely a ceiling effect. This outcome is consistent with earlier studies showing that relatively few independent spectral channels are required for speech perception in quiet (e.g. Shannon et al. 1995).

2.5.2 Masked speech perception with spectrally overlapped and interleaved maskers

Summary speech recognition scores with spectrally matched (overlapped) maskers are represented by filled symbols in Figure 2.4. The data from the three NH groups showed improved speech perception as spectral resolution was increased by reducing spread from

12 to 24 dB/oct, but no further improvement was observed as the spread was reduced from 24 dB/oct to no spread for either the tone or noise maskers, in line with the results in quiet.

Speech recognition scores for spectrally interleaved maskers are represented by open symbols in Figure 2.4. Considering the mean data from the three NH groups, a lawful relationship can be observed, which appears stronger than that found for the spectrally overlapped maskers. In particular, speech perception continued to improve with increased spectral resolution from 12 dB/oct to 24 dB/oct to the no-spread condition. Considering just the three NH groups at a common SMR of 0 dB, a between-subjects, one-way ANOVA showed a significant interaction [$F(2,21) = 9.9$, $P = 0.001$] between condition (interleaved vs overlapping) and the amount of spread (no spread, 24 dB/oct spread and 12 dB/oct spread), confirming that speech perception was more strongly affected by the amount of spread used in the NH vocoder simulations in the interleaved than in the overlapped conditions.

For CI users, the results were very similar for the overlapped and interleaved conditions (Figure 2.4), mirroring speech perception results for the 12 dB/oct spread NH group. This lack of benefit of spectral separation is somewhat unexpected, given that ripple discrimination thresholds for CI users were comparable to those of the 24 dB/oct spread NH group, which saw a marked improvement in speech perception in the interleaved over the overlapped condition.

2.6 Discussion

2.6.1 Spectral resolution and spectral release from masking

The aim of this study was to test whether speech perception in spectrally unmatched maskers (in this case interleaved maskers) would provide a more sensitive test of spectral resolution than the more traditional speech perception in spectrally matched maskers. This hypothesis was supported by the results from the NH listeners under the different vocoder conditions. For instance, as shown most clearly in Figure 2.4, changing from the 24-dB/oct-spread to the no-spread condition produced essentially no change in performance in the overlapped conditions (compare filled circles and triangles), but resulted in a substantial improvement in performance in the interleaved conditions (compare open circles and triangles), suggesting that the interleaved conditions were more sensitive to changes in spectral resolution. Stated another way, changing from the overlapped to the interleaved masker configuration resulted in a much smaller spectral release from masking with 24-dB/oct slopes than in the no-spread condition. The trend continued, with the 12-dB/oct slopes showing even less spectral masking release. Because spectral ripple discrimination thresholds showed the same trend of improvement with increasing spectral resolution, there was a strong relationship between the amount of spectral masking release and spectral ripple discrimination, when observed at a group level.

The speech perception results from the CI users most closely resembled those of the NH listeners in the 12-dB/oct spread condition, with little or no spectral release from masking on average. Therefore, although spectrally interleaved maskers provide a more sensitive measure of spectral resolution overall, the spectral resolution observed with current CI users appears to be too poor to take advantage of the interleaving. Thus, a key hypothesis of our study, that spectrally unmatched maskers would provide a more sensitive test of spectral resolution in CI users, was not supported. It is possible that greater physical separation between speech and masker channels (by, for instance, stimulating only every third or fourth electrode with speech) would lead to larger effects of interleaving; however, this approach would also further decrease the number of active speech channels and thus further degrade overall performance.

2.6.2 Spectral ripple discrimination as a measure of spectral resolution

Spectral ripple discrimination has been used as a measure of spectral resolution for several decades, and the current results support its use, by showing that spectral ripple discrimination thresholds from the NH groups decrease with progressively decreasing spectral resolution from the no-spread to the 24 dB/oct and to 12 dB/oct spread conditions. There has been some consideration of the potential perceptual cues associated with detection a change in the spectral ripple phase. For instance, Anderson et al. (2011) tested whether the spectral edge of the stimulus could provide an additional cue, and found no significant difference in thresholds with spectrally smoothed edges. Introspection suggests that it involves a timbral cue similar to brightness, as a change in the spectral centroid may occur, which may also be perceived as pitch (Allen & Oxenham, 2014). Regardless of the particular cue, any change that is discriminable suggests that at least one place along the cochlear partition responds differently to the two stimuli, which in turn suggests that the stimuli can be spectrally resolved to some extent. Note, of course, that the detection of a change in ripple phase does not imply a full reconstruction of the ripple, which would require spatial sampling at a density that exceeds twice the maximum ripple rate. Consequently, ripple discrimination thresholds may not directly reflect the spectral resolution needed to perceive speech (as in Experiment 1), since speech covers a wide range of frequencies. Thus, as suggested by Anderson et al. (2011), broadband spectral ripple discrimination likely reflects the best spectral resolution found within the cochlea, whereas speech perception is more likely to reflect an aggregate of resolution along the length of the cochlea.

There was also considerable individual variability in thresholds among the NH listeners, as shown by the between-subjects standard deviations in Figure 2.8, despite the fact that the spectral resolution was held constant within each group of NH listeners. As with spectral ripple detection thresholds (Anderson et al., 2011), and spectro-temporal ripple detection or discrimination thresholds (Aronoff & Landsberger, 2013; Bernstein et al., 2013), spectral ripple discrimination thresholds depend not only on spectral resolution, but also on intensity resolution, or the ability to detect intensity differences between channels after filtering (Anderson et al., 2012). The potential conflating of spectral resolution with "detection efficiency" in such measures has also been pointed out for earlier measures of spectral resolution, such as the critical ratio (Patterson, Nimmo-Smith, Weber, & Milroy, 1982). Thus, some of the variability between NH listeners, as well as between CI users, may reflect differences in intensity processing rather than spectral resolution (Azadpour & Mckay, 2012). Because the individual differences between NH listeners are of a similar magnitude as the individual differences between CI users, our results suggest that differences in ripple discrimination thresholds between CI users are unlikely to be due solely to differences in spectral resolution.

2.6.3 Individual differences in CI users and NH listeners

Many studies have noted high levels of inter-individual variability in CI users when performing speech tasks (e.g. Henry and Turner 2003; Fu and Nogaki 2005; Fu et al. 2013). Indeed much effort has gone into attempts to account for individual differences in terms of perceptual and physiological measures of factors such as spectral resolution (affected by the electrode-neural interface) and neural survival. However, few studies have considered the variance within the NH population when performing equally challenging auditory tasks.

In the speech perception tasks, our results appear to support this emphasis on individual variability within the CI population. Inspection of Figure 2.3 reveals generally larger between-subjects variability in the CI group than in any of the NH groups. To quantify this difference, the mean estimated variance from each of the conditions was pooled across all conditions. The mean variance in the CI group was 543.8, compared with 196.8, 55.2, and 116.9 in the NH groups with the no-spread, 24-dB/oct spread, and 12 dB/oct spread, respectively. Thus, the variance of the CI was considerably larger. However, because our CI group was also considerably older on average than our NH groups, we cannot rule out the possibility that the larger variance in the CI group was due to age, rather than anything specific to CI processing (Landsberger, Padilla, Martinez, & Eisenberg, 2017).

In contrast, the between-subject variance in the spectral ripple discrimination task was similar across all four groups, as indicated by the similar size of the error bars in Figure 2.8. This outcome is particularly surprising, as differences in spectral resolution between individual CI users have been postulated to play an important role in predicting speech perception in quiet and in noise (Henry & Turner, 2003; Henry et al., 2005; Won et al., 2007). Yet in the three NH groups, the spectral resolution of listeners within each of the NH groups was held constant via the vocoder signal processing, but the amount of between-subject variability was similar to that found among the CI users.

Another interesting outcome is that the inter-individual variability is similar across the CI and NH groups in the *differences* between various speech perception measures. For instance, as shown by the error bars in Figure 2.5, variability in the benefit of interleaved over overlapped maskers (upper plots), or in the benefit of tone over noise maskers (lower plots), is again quite similar between the groups. This observation suggests it is unlikely that physiological factors, such as differences in effective current spread or neural survival, dominate individual differences in such measures of speech masking release in CI users. Further research with groups that are matched for age and cognitive function will be required to determine the extent to which inter-individual variability in CI users is quantitatively different from that observed in the NH population.

2.7 Conclusions

Speech perception in spectrally overlapped and interleaved tone or noise maskers was measured in CI users, and the results were compared with those from NH participants listening through tone-excited vocoders designed to simulate various degrees of current spread or channel interaction. The main findings can be summarized as follows:

- In the NH groups, decreasing spectral resolution led to decreasing speech perception in quiet and in the presence of maskers. The effect of decreased spectral resolution was greater for the spectrally interleaved maskers than for the spectrally overlapped maskers, resulting in increased spectral masking release with increased spectral resolution.
- The CI users on average showed little or no spectral masking release from the spectrally interleaved maskers, similar to the findings from the NH group with the simulated 12 dB/oct spectral spread. This outcome was not affected by whether the frequency channels were selected based on the frequencies allocated to each electrode or on the processors' virtual channel center frequencies, based on the CI's current steering algorithms.
- Spectrally interleaved and tonal maskers provide greater sensitivity to improvements in spectral resolution than do spectrally matched noise maskers. Although the resolution of current CIs is not sufficient to take advantage of this sensitivity, such measures may be useful in testing future generations of CIs, as well as testing listeners with hearing loss, whose spectral resolution may be intermediate to that of normal-hearing listeners and CI users.

CHAPTER 3: COGNITIVE FACTORS AND SPEECH PERCEPTION

Chapter 3 is reprinted from:

O'Neill, E. R., Kreft, H. A., & Oxenham, A. J. (2019) Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions. *Journal of the Acoustical Society of America, 146,* 195-210.

Abstract

This study examined the contribution of perceptual and cognitive factors to speech-perception abilities in cochlear-implant (CI) users. Thirty CI users were tested on word intelligibility in sentences with and without semantic context, presented in quiet and noise. Performance was compared with measures of spectral-ripple detection and discrimination, thought to reflect peripheral processing, as well as with cognitive measures of working memory and non-verbal intelligence. Thirty age-matched and 30 younger normal-hearing (NH) adults also participated, listening via tone-excited vocoders, adjusted to produce mean performance for speech in noise comparable to that of the CI group. Results suggest that CI users may rely more heavily on semantic context than younger or older NH listeners, and that non-auditory working memory explains significant variance in the CI and age-matched NH groups. Between-subject variability in spectral-ripple detection thresholds was similar across groups, despite the spectral resolution for all NH listeners being limited by the same vocoder, whereas speech perception scores were more variable between CI users than between NH listeners. The results highlight the potential importance of central factors in explaining individual differences in CI users, and question the extent to which standard measures of spectral resolution in CIs reflect purely peripheral processing.

3.1 Introduction

The cochlear implant (CI) has been a beneficial and often life-changing treatment for individuals with profound sensorineural hearing loss. However, despite advances in processing strategies, current-steering options, electrode arrays, and surgical techniques, clinicians continue to see a very wide range of outcomes for individual CI users. Many studies exploring the speech-perception abilities of CI users have shown word and sentence recognition ranging from 0 to 100 percent across individuals on any given task (Hast et al., 2015; Lenarz, Sönmez, Joseph, Büchner, & Lenarz, 2012; Mahmoud & Ruckenstein, 2014). Individual patient factors, such as CI experience, age at implantation, duration of deafness and etiology, have been shown to account for very little variance in performance (less than 10 percent) in large samples of CI users (Blamey et al., 1996; Blamey et al., 2013). A number of studies have explored the association between cross-modal plasticity and CI outcomes with some finding these changes to be adaptive (Anderson, Wiggins, Kitterick, & Hartley, 2017; Rouger et al., 2012; Strelnikov et al., 2013) and others finding them maladaptive (e.g., Lee et al., 2003; Sandmann et al., 2012; Zhou et al., 2018). Anatomical, physiological and surgical factors that may influence spectral resolution and signal quality, such as etiology of hearing loss, degree of neural survival, electrode-neural distance, and insertion depth of the electrode array, have been studied extensively. While some studies have shown that specific surgical factors account for some variance in hearing outcomes (Aschendorff, Kromeier, Klenzner, & Laszig, 2007; Finley et al., 2008; Holden et al., 2013; O'Connell, Hunter, & Wanna, 2016; Skinner et al., 2007; Wanna et al., 2014), other studies have not (Van Der Marel et al., 2015).

The limited number of electrodes and the effects of electrical field spread, due to the distance between the electrode and the spiral ganglion cells, lead to limited spectral resolution in CIs. It has often been shown that speech perception, particularly in noise, is strongly influenced by spectral resolution (Dorman et al., 1998; Friesen et al., 2001). Perhaps because differences in spectral resolution between CI users can be complex in nature and measures of spectral resolution are limited in their sensitivity, this relationship has not always been clear at the level of individual CI users. Although some studies in CI users have found a correlation between measures of spectral resolution (e.g., spectral-ripple discrimination or spatial tuning curves) and speech perception in quiet (Anderson et al., 2011; Henry & Turner, 2003; Henry et al., 2005) and in noise (Gifford et al., 2018; Won et al., 2007), others have not (Anderson et al., 2012). These apparent discrepancies may be in part because most measures of spectral resolution, such as spectral-ripple detection (Anderson et al., 2012; Gifford et al., 2018), spectral-ripple discrimination

(Anderson et al., 2011, 2012; Henry & Turner, 2003; Henry et al., 2005; Won et al., 2007), and spectrotemporal-modulation detection (Choi et al., 2018; Won et al., 2015), use broadband stimuli, but can be performed using a very limited portion of the entire spectrum (e.g., Anderson et al., 2011). In contrast, speech perception benefits from information across the entire spectrum, meaning that good spectral resolution at just one cochlear location will not be sufficient to provide good intelligibility. Also, although a recent large-scale study by Gifford et al. (2018) showed a positive correlation between spectral-modulation detection thresholds and sentence perception, the variability around the trendline was very large, making it difficult to predict speech performance on an individual basis.

The field of CI research has focused primarily on possible peripheral differences between users, with less attention being paid to differences at higher levels of processing. Studies with hearing-aid (HA) users have pointed to working memory and various cognitive abilities as possible factors contributing to hearing outcomes, with better working memory and cognitive abilities associated with better speech perception (Akeroyd, 2008; Arehart et al., 2013; Lunner, 2003; Rönnberg et al., 2013). A large-scale study of older adults with hearing loss also found that visual measures of cognitive-linguistic processing and environmental sound identification accounted for the most variance in aided speech understanding (Humes, Kidd, & Lentz, 2013). However, the correlations between cognitive factors and speech perception have been less clear in studies with adult CI users. A recent study with CI users by Moberly et al. (2017) found correlations between speech perception in noise and cognitive control, as well as with auditory but not visual working memory. Heydebrand et al. (2007) also found that better verbal working memory was associated with improvement in word recognition six months after CI activation, but general cognitive ability and processing speed were not. In the same vein, Hua et al. (2017) found correlations between some, but not all, measures of cognitive skills and working memory and the perception of words and sentences in quiet and in noise in bimodal listeners (those with a CI in one ear and a HA in the other).

In lieu of exploring correlations with cognitive measures, other researchers have looked at the use of semantic context as a way to tap into "top-down" processes that may affect speech perception in hearing-impaired listeners. It is known that semantic context is leveraged in commonly encountered acoustic environments (e.g. Baskent et al., 2016; Bhargava et al., 2014; Marslen-Wilson and Welsh, 1978; Signoret et al., 2018), and it may be that CI users rely more heavily on such context to compensate for the degraded input. One study with older hearing-impaired adults suggested that they rely more heavily on semantic context than older adults with normal hearing when performing speech-perception tasks, with the perception of low-context or semantically anomalous sentences requiring more cognitive effort (Moradi, Lidestam, Hällgren, & Rönnberg, 2014). Recent work by Winn (2016) used pupillometry measures to demonstrate decreased listening effort for high-predictability versus low-predictability sentences in CI users and NH listeners. Interestingly, the effect of sentence predictability on pupil diameter was smaller and occurred later after the stimulus offset in CI users and NH participants listening to vocoded speech than in NH participants listening to unprocessed speech. The pattern of errors shown in CI users also suggested more semantic than phonemic substitutions in cases where participants incorrectly guessed a missed word in a sentence, indicating a heavier reliance on context. Another recent study in CI users found a significant difference in gate size (i.e., duration or proportion of the word presented) required for recognition of the final words of sentences in cases of high versus low expectancy and entropy (Amichetti et al., 2018). Modeling of contextual information in words and sentences by Dingemanse and Goedegebure (2019) has also suggested that CI users make more use of context in speech recognition than NH listeners. Although these studies provide some insight into the use of semantic context in hearing-impaired listeners and CI users, the direct benefit in speech perception for full sentences with semantic context compared to those without has yet to be studied in CI users.

Finally, the individual variability in speech-perception abilities among younger and older NH listeners has rarely been studied under acoustic conditions that are sufficiently degraded to produce average performance similar to that observed in CI users. Thus, it remains unclear how much the use of context information differs, and how much more variable speech perception is between CI users than between NH listeners, under similarly degraded conditions.

In this study we attempted to assess the relative importance of perceptual and cognitive factors in predicting speech perception in CI users by using a diverse set of psychoacoustic and cognitive measures. The results were compared with those from two different NH control groups: one was age-matched to our CI participants and the other consisted of young NH listeners, mostly undergraduate students, similar to those most commonly used in the comparison groups of earlier studies. The NH participants were presented with materials via a tone-excited vocoder that was designed to simulate the effects of loss of spectral resolution and to produce performance for speech perception in noise that was comparable to that found for CI users. Speech materials consisted of syntactically correct sentences that were either semantically coherent (context) or incoherent (nonsense), presented both in quiet and in noise. Psychoacoustic measures included broadband spectral-ripple detection and discrimination. All participants also completed two different cognitive tests: a reading-span test, as a measure of verbal (visual) working memory (Conway et al., 2005), and Raven's Advanced Progressive Matrices, as a measure of non-verbal intelligence (Raven et al., 1998).

If variability in speech understanding among CI users is due primarily to individual differences in spectral resolution, then measures of spectral resolution should correlate more strongly with speech perception than the measures of cognitive performance. In contrast, for the NH listeners, where spectral resolution will be limited by the signal processing of the vocoder rather than the individual peripheral auditory systems, it may be that cognitive factors will explain any variance between listeners. In addition, we predicted that better working memory and cognitive function would increase the difference in performance between context and nonsense sentences because working memory is required to take advantage of semantic context (Signoret et al., 2018). Our overall hypothesis was that context sentences may depend more on cognitive function, whereas nonsense sentences may increase reliance on bottom-up processes (Mattys, White, & Melhorn, 2005), and may thus be mediated more by peripheral factors, as reflected by our measures of spectral resolution.

3.2 General Methods

3.2.1 Participants

A total of 30 adult CI users (23 females and 7 males) ranging in age from 20 to 80 years (mean = 61.5 years; standard deviation, SD = 12.8) were tested. All CI users had at least one year of CI use, with experience ranging between 1 and 28 years (mean = 10.4 years; SD = 7.3). The duration of hearing loss prior to implantation varied between CI users from less than a year to 41 years (mean = 10.2 years; SD = 10.8). Twenty-two of the CI users used Advanced Bionics devices, five used Cochlear devices, and three used Med-El devices. All CI users were post-lingually deafened, with the exception of one CI user who was deafened peri-lingually, followed by immediate implantation and strictly oral instruction. A group of 30 NH adults (23 females and 7 males), age-matched to the CI group with ages ranging from 20 to 78 years (mean = 61.5; SD = 12.7) were also tested. An additional control group was tested, consisting of 30 NH young adults (21 females and 9 males) ranging in age from 18 to 30 years (mean = 20.6; SD = 2.5). All participants were native speakers of American English. Normal hearing for the young listeners was defined as having pure-tone audiometric thresholds less than 20 dB hearing level (HL) at all octave frequencies between 250 and 8000 Hz with no reported history of hearing disorders. Normal hearing for the age-matched listeners was defined as having pure-tone audiometric thresholds less than 20 dB HL at all octave frequencies between 250 and 2000 Hz and no more than 30 dB HL at 4000 Hz and 6000 Hz, with no reported history of hearing disorders. Since close age-matching with the CI group was a priority, this audiometric criteria was a compromise that allowed us to successfully recruit older participants with relatively normal hearing. The average threshold for age-matched listeners at 8000 Hz was 22 dB HL (SD = 15), with individual thresholds ranging from 0 to 60 dB HL.

All the CI users listened with one CI. Bilateral CI users were instructed to use whichever processor they thought gave them better speech perception and to remove the other processor. Four unilateral CI users with some residual hearing in their contralateral ear were instructed to remove hearing aids and to insert an ear plug in the non-CI ear, which was worn for the entirety of the experiment. No CI user with residual hearing had unaided audiometric thresholds better than 70 dB HL at any frequency, which should have rendered any of our stimuli inaudible, particular after the 10-15 dB of attenuation expected by the ear plug. All four participants with residual hearing self-reported that their CI ear was better for speech perception than their HA ear. All CI users were asked to use processor settings (volume, sensitivity, program, noise reduction, directional microphones) typical of their everyday use.

All experimental protocols were approved by the Institutional Review Board of the University of Minnesota, and all listeners provided informed written consent prior to participating. The same 90 participants completed all experiments in this study.

3.2.2 Order of Experiments

The experiments in this study were completed in the order in which they appear: speech perception was completed first (context then nonsense sentences), followed by the cognitive measures (working memory then non-verbal intelligence), and finally measures of spectral resolution (spectral-ripple discrimination then detection). Generally, participants completed all testing in two 2-h sessions within a two-week period. All speech-perception testing was completed in the first session and measures of cognition and spectral resolution were completed in the second session. A small number of participants (5 CI users, 6 age-matched NH listeners, and 2 young NH listeners) completed the testing in three sessions due to lack of schedule flexibility and differences in testing pace.

3.3 Experiment 1: Speech Perception with Context and Nonsense Sentences

3.3.1 Stimuli

3.3.1.1 Context Sentences

The context speech materials were sentences taken from the Harvard-IEEE speech corpus (Rothauser et al., 1969), recorded by a single female talker. An example of a context sentences is “A *rod* is *used* to *catch pink salmon*,” with keywords in Italics. Each context sentence contained five keywords and there were 10 sentences per list.

3.3.1.2 Nonsense Sentences

The nonsense speech materials were sentences taken from the Helfer (1997) lists of nonsense sentences, recorded by a different single female talker, as used by Freyman et al. (2013) and Ruggles et al. (2014) among others. The nonsense sentences were created using common one- and two-syllable nouns and verbs taken from the Thorndike-Lorge lists of most common words (Thorndike & Lorge, 1945), with each sentence containing three key words and between five and seven total words. Each nonsense sentence was constructed either in the form, article *noun* (auxillary verb) *verb* (preposition) article *noun*, or *verb* article *noun* preposition article *noun*, where italicized words are key words and items in parentheses occur in some, but not all, sentences. An example of a nonsense sentence was "A *shop* can *frame* a *dog*," with keywords in italics. Each nonsense sentence contained three keywords and there were 10 sentences per list.

3.3.1.3 Signal Processing

Both context and nonsense sentences were presented in quiet and in Gaussian noise, spectrally shaped to match the long-term spectrum of each speech corpus. The speech and noise were mixed at the appropriate SNR before further processing and presentation to the participants. For the NH listeners, the mixture was passed through a 16-channel tone-excited vocoder with the center frequencies taken from the Advanced Bionics standard clinical map. For the CI users, the stimuli were divided into subbands based on each individual CI user's device and clinical map. The number of channels ranged from 7 to 22, depending on the type of CI processor and the number of deactivated electrodes in each CI.

The bandpass filters used to generate the subbands were high-order (947) FIR filters, generated with the *fir1* function in Matlab (Mathworks, Natick, MA), producing very little overlap between the spectral content of adjacent subbands and a flat frequency response (±0.05 dB) within the entire passband. The impulse responses from the linear-phase filters were time-aligned, reaching their peaks at a delay of approximately 20 ms, independent of filter center frequency. The temporal envelope from each subband was then extracted using a Hilbert transform, and the resulting envelope was lowpass filtered using a fourth-order Butterworth filter with a cutoff frequency of 50 Hz. This cutoff frequency was chosen to reduce possible voicing periodicity cues and to reduce the possibility (for NH listeners) that the vocoder produced spectrally resolved components via the amplitude modulation of the tonal carriers. The resulting temporal envelopes were used to modulate pure-tone carriers with frequencies corresponding to the center frequencies of each channel, which were then presented to the CI users. Stimuli were processed this way for the CI users to maintain consistency in vocoder processing among groups of listeners, so that the filtered envelopes were lowpass filtered at 50 Hz for all conditions and groups. A previous study showed no performance differences in CI users when comparing similarly vocoded to unprocessed stimuli (Oxenham & Kreft, 2014). For the NH listeners, the effects of current spread were simulated via the vocoder by modulating each carrier by the weighted sum of the intensity envelopes from all 16 channels (Oxenham & Kreft, 2014). The weights used in this sum were selected to produce attenuation slopes of 12 dB/octave on either side of the center frequency, to simulate sufficient spectral smearing for speech perception to approximate the average performance of CI users.

The speech was adjusted to a root mean square (rms) level of 65 dB SPL, as measured at the participant's head, and the noise level was adjusted to produce the desired SNR. The noise was gated on 1 s before the beginning of each sentence and gated off 1 s after the end of each sentence. The SNRs were selected in advance, based on previous studies (Oxenham & Kreft, 2014), to avoid ceiling and floor effects in performance, and were the same for both the context and the nonsense sentences, to facilitate direct comparisons between the two speech corpora.

3.3.2 Procedure

The stimuli were generated using Matlab and converted via a 24-bit digital-to-analog converter (L22, LynxStudio, Costa Mesa, CA) at a sampling rate of 22,050 Hz. The sounds were presented in a single-walled, sound-attenuating booth located in a quiet room via an amplifier and a single loudspeaker, placed approximately 1 m from the listener at 0° azimuth.

Listeners responded to sentences by typing what they heard on a computer keyboard. One of the oldest CI users did not have adequate typing proficiency to enter responses via a keyboard and so instead spoke the responses into a lapel microphone. The spoken answers were recorded and stored as Windows Media Audio (WMA) files and were later listened to and scored offline. Participants were encouraged to guess individual words, even if they had not heard or understood the entire sentence. Instructions were given orally and participants were asked if they had any questions about procedures before beginning the task. Sentences were scored for keywords correct as a proportion of the total number of keywords presented. Initial scoring was automatic, with each error then checked manually for potential spelling errors or homophones (e.g., wait and weight), which were marked as correct. Before the actual experiment took place, NH listeners were presented with two sentence lists of 20 sentences each from the AzBio speech corpus (Spahr et al., 2012) to acclimate them to the vocoded stimuli before the scored sentences were presented. During this training phase, each sentence was presented visually on the computer screen while the audio was played from the speaker. Participants were instructed to listen to each sentence and try to mentally map what they were hearing with the actual words of the sentence presented on the screen, similar to the procedure used by Litvak et al. (2007). The listeners did not type any responses during this phase. This training phase was included to acclimatize NH listeners to the vocoded speech and avoid data contamination due to its initial novelty (McGettigan, Rosen, & Scott, 2008; O'Neill, Kreft, & Oxenham, 2019b). Following this training phase, but before the actual testing began, both CI and NH participants completed four lists of each sentence corpus (context and nonsense) in quiet to become comfortable with the procedure.

In the actual experiment, two lists of 10 sentences each were completed for each SNR for the context sentences and three lists of 10 sentences each were completed for each SNR for the nonsense sentences. This design allowed for a comparable number of keywords to be tested from each corpus; because each context sentence contains five keywords and each nonsense sentence contains only three keywords, each participant was presented with 100 keywords per SNR for the context sentences and 90 keywords per SNR for the nonsense sentences. Every participant completed the context sentences first, followed by the nonsense sentences. After the training block of four lists of 10 sentences in quiet, the sentences for each scored condition were presented first at an SNR of 20 dB and then at decreasing SNRs in 5-dB steps to a final SNR of 0 dB. Scored sentences presented in quiet were completed as the final block of each condition. The proportions of correct words in each condition were converted to rationalized arcsine units (RAU) (Studebaker, 1985) before statistical analysis, to mitigate some of the potential effects of floor or ceiling performance.

3.3.3 Results

The mean RAU-transformed proportion of correct keywords from both context and nonsense sentences for the CI group, the age-matched NH group, and the young NH group are shown in Figure 3.1. Mean performance in any group always fell between 10 and 73, meaning that the RAU transform (which has the greatest effects at values close to 0 and 100 percent) did not substantively affect the statistical conclusions. The scores from the context sentences and nonsense sentences are denoted by filled and open symbols, respectively.

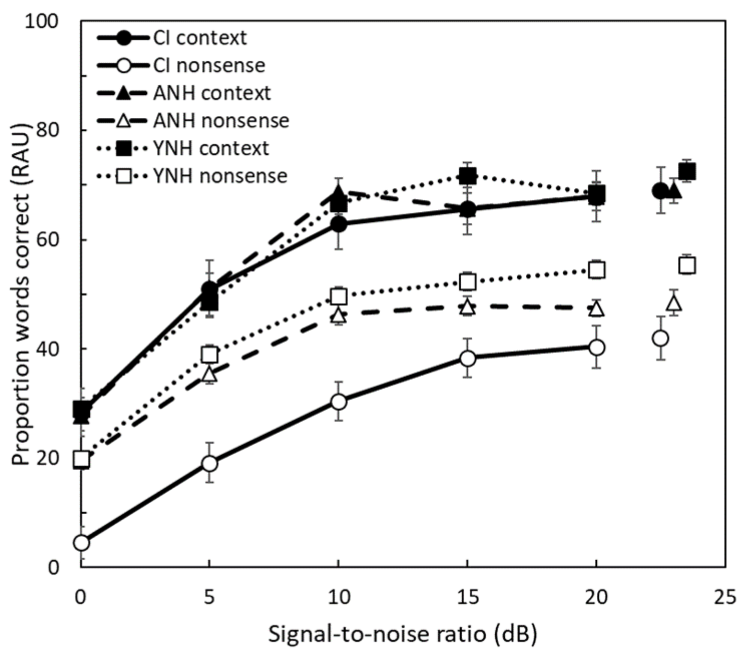

Figure 3.1

Speech perception for CI users and both groups of NH listeners. The RAU-transformed proportion of keywords from sentences reported correctly is plotted as a function of signal-to-noise ratio for both context (IEEE) and nonsense sentences. Error bars represent ±1 standard error of the mean between listeners.

As expected, performance was similar across all groups for the context sentences, confirming that the 16-channel vocoder with 12 dB/octave spread produces speech perception in NH listeners (in quiet and in noise) that is comparable to that of CI users. Performance for all groups was also poorer for the nonsense than for the context sentences and worsened with decreasing SNR, as expected. Age did not seem to affect performance, as data from the two NH groups were similar across all conditions. Most interestingly, the CI users' performance in the nonsense sentences appeared to be poorer than that of the other two groups.

A repeated-measures ANOVA on the RAU-transformed data, with sentence material (context vs nonsense) and SNR as within-subjects factors and group as a between-subjects factor, confirmed a significant main effect of sentence material [$F(1,87)=541.3$, $P<0.001$, $\eta_p^2=0.862$], SNR [$F(5,435)=744.4$, $P<0.001$, $\eta_p^2=0.895$], and group [$F(2,87)=3.2$, $P=0.046$, $\eta_p^2=0.068$]. There were also significant interactions between group and sentence material [$F(2,87)=23.97$, $P<0.001$, $\eta_p^2=0.355$], between group and SNR [$F(10,435)=3.07$, $P=0.002$, $\eta_p^2=0.066$], and between SNR and sentence material [$F(5,435)=11.64$, $P<0.001$, $\eta_p^2=0.118$]. The three-way interaction between sentence material, SNR, and group was also significant [$F(10,435)=2.94$, $P=0.001$, $\eta_p^2=0.063$].

To further examine these interactions, a series of pairwise comparisons with Bonferroni correction for multiple comparisons was performed. A pairwise comparison between sentence material and group (corrected $\alpha=0.025$) showed no significant effect of group for context sentences [$F(2,87)=0.12$, $P=0.89$], but did show a significant effect of group for nonsense sentences [$F(2,87)=12.76$, $P<0.001$]. Thus performance was similar between groups for the sentences with context but differed across groups when context was absent. Specifically, performance for the nonsense sentences was significantly poorer for the CI group when compared to both age-matched ($P=0.001$) and young ($P<0.001$) NH groups (corrected $\alpha=0.008$). However, performance on nonsense sentences did not differ significantly between young and age-matched NH listeners ($P=0.189$). Pairwise comparisons also confirmed that the effect of sentence material was significant for all three groups [CI group: $F(1,87)=359.1$, $P<0.001$; age-matched NH group:

$F(1,87)=137.77$, $P<0.001$; young NH group: $F(1,87)=92.34$, $P<0.001$; corrected $\alpha=0.017$], reflecting poorer performance for all groups on sentences without context when compared to those with context.

Other pairwise comparisons examining the interaction between sentence material, SNR, and group (corrected $\alpha=0.0014$) showed a significant difference in performance with nonsense sentences between the CI group and the age-matched NH group at poorer SNRs, but not at more favorable SNRs. For example, the CI group performed significantly more poorly than the age-matched NH group on nonsense sentences at SNRs of 0 ($P<0.001$) and 5 dB ($P<0.001$), but performed similarly at 20 dB ($P=0.058$) and in quiet ($P=0.109$). Performance with nonsense sentences for the CI group was significantly poorer than for the young NH group at both lower [0 ($P<0.001$) and 5 dB ($P<0.001$)] and higher [20 dB ($P<0.001$)] SNRs, as well as in quiet ($P=0.001$).

3.4 Experiment 2: Working Memory and Non-verbal Intelligence

3.4.1 Methods

3.4.1.1 Stimuli and procedure

To measure working memory, a reading-span task (Conway et al., 2005) was administered (current URL: https://ubiq-x.gitlab.io/rspan/). The task consisted of both individual letters and sentences, presented visually on a computer screen in an alternating fashion. The subset of letters used included F, H, J, K, L, N, P, Q, R, S, T, and Y. Each sentence varied in length from 10 to 15 words and could either be presented as congruent or incongruent contextually. In congruent sentences, all of the words in the sentence adhered to a meaningful context, whereas incongruent sentences contained one word that violated the contextual meaning of the sentence, resulting in a sentence that did not make logical sense, such as "The athlete broke his lunchbox and could not participate in the race." Each letter was presented for 1 s, followed by a sentence presented for the average amount of time it took each individual participant to judge the correctness of each sentence in an initial practice block. The number of pairs of sentences and letters presented in succession varied from two to seven and was randomized across trials. Each different number of pairs (two through seven) was presented three times for a total of 18 trials. No time limit was imposed while participants had to recall letters seen within a trial and subsequent trials did not initiate until the participant manually moved the procedure forward. A schematic diagram of the procedure is shown in Figure 3.2.

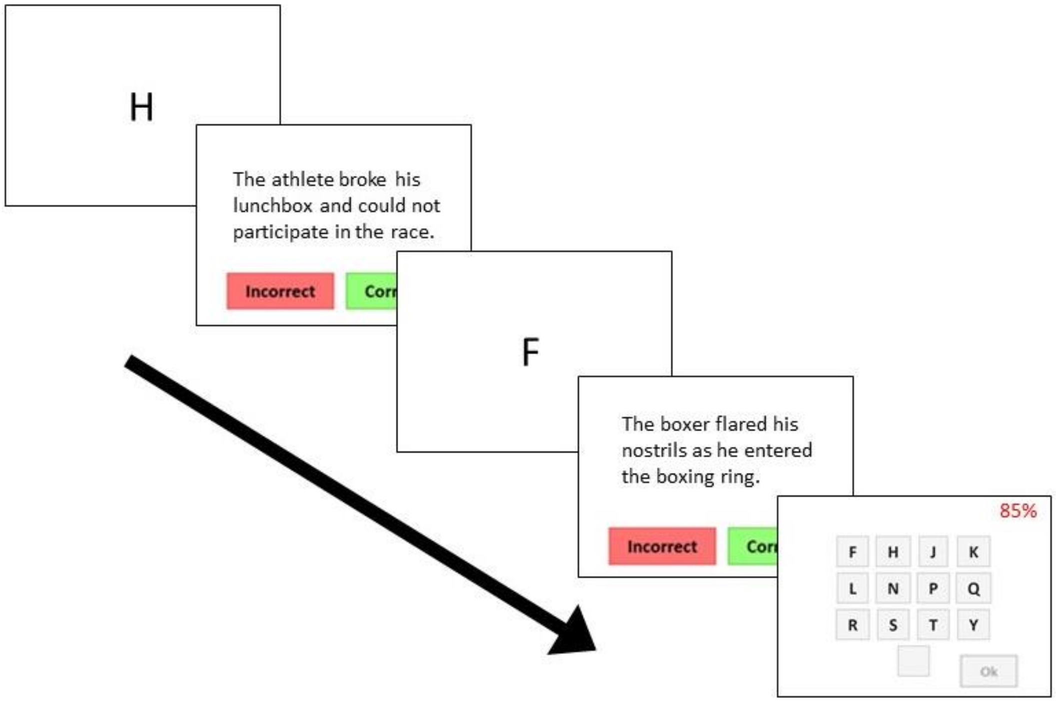

Figure 3.2

Schematic diagram of the reading span task used to measure visual working memory. The 85% in the upper right corner of the last (front-most) panel represents the minimum percent correct participants had to achieve on sentence decisions by the end of the task for results to be considered valid.

All participants completed the reading-span task by following instructions presented visually on a computer screen, while seated alone in a quiet room. Before starting the task, the participants were given an overview of the procedure by an experimenter to avoid any potential confusion and to give the participants an opportunity to ask questions. The participants were instructed to read all the directions presented on the computer screen and were asked to complete a series of training blocks designed to help familiarize them with all parts of the task. The training consisted of three blocks. In the first block, the participants were presented only with letters and were instructed to recall them in the order they were presented. The second block consisted only of sentences during which the participants had to decide whether each sentence was correct or incorrect contextually. The final block alternated the letters and sentences and participants practiced recalling the letters presented while intermittently reading sentences and determining if they were contextually congruent or incongruent. Once the training blocks were complete, the participants were informed that the actual experiment was about to begin.

Each of the 18 blocks of the actual reading-span task consisted of alternating presentations of letters and sentences followed by a screen where participants were asked to recall the letters they saw within that block in the same order they were presented. Participants used a mouse to click on the letters they saw within a given trial on a screen displaying all possible letters (i.e. F, H, J, K, L, N, P, Q, R, S, T, and Y). Each response screen also included a percentage in the upper right corner which represented their overall percent correct on sentence decisions and updated with each successive trial. Participants were told at the start of the test that they had to score at least 85% overall for their scores to be valid upon completion.

Once participants completed the task, letter recall scores were generated, along with the overall percent correct on sentence decisions. A handful of participants did not achieve 85% in their first attempt but, upon re-instruction, maintained this threshold of performance on their second attempt. Two CI users only achieved maximum scores of 84% and 83%, despite multiple repetitions of the task, and reported having difficulty switching between the tasks. Because their performance was close to the cut-off of 85% and because it was clear that the low performance was not because they were strategically ignoring the sentences, the scores from both participants were included in the overall analysis.

The letter recall score, termed the partial-credit unit score, is the percentage of letters correctly recalled in the correct serial position, averaged across trials. For example, if a series of letters presented was L, F, Q, T, K and a participant responded F, L, Q, R, K, they would receive a recall score of 40% on this trial, since only letters Q and K were recalled correctly in the correct serial position. The partial-unit score has been found to be the most reliable and psychometrically sound scoring method for this kind of reading-span task (Conway et al., 2005).

To measure non-verbal intelligence, the paper-and-pencil version of Raven's Advanced Progressive Matrices was used (Raven et al., 1998). The test consisted of 36 matrices, each with eight possible answers. Each matrix problem has eight different combinations of shapes and textures shown in a 3-by-3 grid, with the ninth configuration absent. One of the eight possible answers for each question correctly completes the pattern formed by the combination of shapes in the grid. The test started with easier matrices and became progressively more difficult. Each matrix appeared on a single page in a three-ring binder and participants recorded their answers on a separate sheet of paper. Before beginning the actual test, participants were given two practice matrices by the experimenter. If a participant understood the task and answered both problems correctly, they could begin the actual experiment. If a participant answered the first practice problem incorrectly, the experimenter would walk them through the second practice problem to make sure they understood the task before moving on to the experiment. All participants were given 30 minutes to complete the experiment and were instructed to answer as many problems correctly as possible in the given time frame. Participants were seated alone at a desk in a quiet room. After 30 minutes had elapsed, the experimenter returned and instructed participants to put down their pencil and stop taking the test. The test was scored for the total number of correct answers, regardless of the total number of questions answered.

3.4.2 Results

The mean results for the reading-span task from the CI group and both NH groups (age-matched and young) are shown in panel A of Figure 3.3. The white, gray, and black bars represent the mean partial unit scores for the CI group, age-matched NH group, and young NH group, respectively. Error bars represent one standard error of the mean between participants.

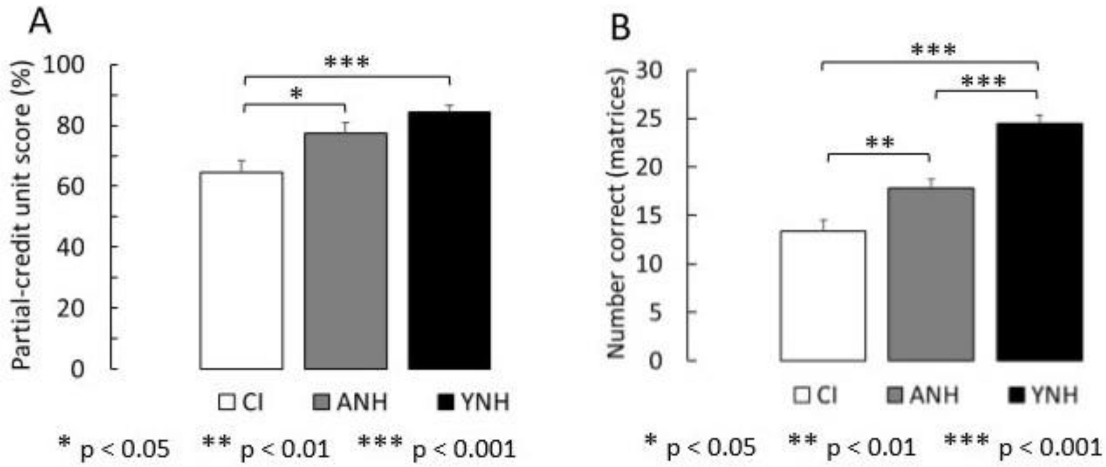

Figure 3.3

Group-mean scores for working memory and non-verbal intelligence are shown in panels A and B, respectively. Bars in panel A represent mean partial unit scores (proportion of letters within each trial recalled correctly, averaged across trials) on the reading span task for CI users and both NH groups. Bars in panel B represent the mean number of correctly answered matrix problems (in 30 minutes) for each group. Error bars represent one standard error of the mean between subjects.

A one-way, between-subjects ANOVA performed on data from all three groups revealed a significant effect of group [$F(2,87)=8.93$, $P<0.001$, $\eta_p^2=0.17$]. Post-hoc comparisons with Bonferroni correction for multiple comparisons (corrected $\alpha=0.017$) showed a significant difference between working memory for CI users and both the young ($P<0.001$) and age-matched ($P=0.008$) NH groups, but not between age-matched and young NH groups ($P=0.146$). It is noteworthy that verbal working memory scores on a non-auditory task were significantly poorer for CI users than for both groups of NH listeners. Anecdotally, a number of the CI users who scored lower than average seemed to experience more difficulty with task-switching (between retaining letters and determining sentence coherence) than just with retaining letters in memory. Although we have no direct measure of this difficulty in the current experiment, other studies have shown impairments in cognitive control among older hearing-impaired adults (Lash, Rogers, Zoller, & Wingfield, 2013b; Lash & Wingfield, 2014) that may compound deficits in working memory.

The mean results for Raven's Advanced Progressive Matrices from the CI group and both NH groups (age-matched and young) are shown in panel B of Figure 3.3. The white, gray, and black bars represent data from the CI group, age-matched NH group and young NH group, respectively. The bars represent the mean number of correctly answered matrix problems (in 30 minutes) for each group. Error bars represent one standard error of the mean between participants. According to smoothed detailed norms for the U.S. (Raven et al., 1998) on the same test but with no time limit, the young NH group-average score of 24 falls in the 64th percentile for participants aged 18 to 22, perhaps because all participants in this group were university students. The age-matched NH group-average score of 18 falls in the 56th percentile for individuals between the ages of 58 and 62, indicating the group as a whole scored slightly above average. The CI group-average score of 13 represents a score that corresponds to the 32nd percentile for individuals between the ages of 58 and 62. Although these comparisons are not exact, given that a 30-minute time limit was imposed on participants in this experiment, population data (Raven et al., 1998) indicate that scores only improve by one, on average, when a time limit is not imposed. Thus, even if group average scores were increased by one for each group, scores for the CI group would still be below the 50th percentile and those for the age-matched NH group would still be above.

A one-way, between-subjects ANOVA on scores from all three groups showed a significant effect of group [$F(2,87)=32.94$, $P<0.001$, $\eta_p^2=0.431$]. Post-hoc comparisons with Bonferroni correction (corrected $\alpha=0.017$) revealed a significant difference between CI users and both the age-matched NH ($P=0.002$) and young NH ($P<0.001$) groups, as well as between age-matched and young NH listeners ($P<0.001$). The difference between the CI users and the age-matched NH group on this measure of non-verbal intelligence was not expected, and points to factors other than age that might be influencing speech-perception performance differentially between NH listeners and CI users.

Some studies have found hearing loss to be associated with accelerated cognitive decline in older adults (Claes et al., 2018; Lin, Ph, et al., 2013; Livingston et al., 2017), which could be reflected in our results. However, since we did not control for socio-economic status, level of education, or sampling bias, and considering the relatively small number of participants in this study, it should not yet be concluded that CI users as a group tend to have lower scores on such measures of non-verbal intelligence.

3.5 Experiment 3: Spectral-ripple Discrimination and Detection

3.5.1 Methods

3.5.1.1 Stimuli

Spectrally rippled noise was generated using Matlab. Gaussian broadband (350-5600 Hz) noise was spectrally modulated, with sinusoidal variations in level (dB) on a log-frequency axis (as in Litvak et al. 2007) using the equation:

$$X(f) = 10^{(D/2)\sin\{2\pi[\log_2(f/L)]f_s+\theta\}/20}$$

where $X(f)$ is the amplitude at frequency f (in Hz), D is the spectral depth or peak-to-valley ratio (in dB), L is the lower cut-off frequency of the noise pass band (350 Hz in this case), f_s is the spectral modulation frequency (in ripples per octave), and θ is the starting phase of the ripple function.

The ripple-discrimination task involved spectrally-rippled stimuli with a fixed peak-to-valley ratio of 30 dB, while the ripple rate was varied adaptively to track the highest ripple rate or density, in ripples per octave (rpo), at which a phase reversal of the ripples is detectable. This threshold is thought to provide a measure of the limits of spectral resolution, and does appear to suffer from the potential confounds of temporal-envelope cues, which can affect spectral-ripple detection at high ripple rates (Anderson et al., 2012; Nechaev, Milekhina, & Supin, 2019). The ripple-detection task involved measuring the minimum detectable peak-to-valley ratio at a fixed ripple rate of 0.25, 0.5, 1, or 2 rpo. This task was included because ripple detection thresholds, particularly at low rates (0.25 and 0.5 rpo), have been shown to be correlated to the recognition of speech sounds (Anderson et al., 2012; McKay et al., 2018; Saoji et al., 2009a).

The duration of each stimulus was 400 ms, including 20-ms raised-cosine onset and offset ramps. For NH listeners, the stimulus was passed through the same tone-excited envelope vocoder used in Experiment 1, with 16 frequency channels that produced spectral spread equivalent to 12 dB/oct. The CI users were presented with the stimulus unaltered. Both ripple detection and discrimination have been used in previous studies (e.g. Drennan et al., 2014; Henry et al., 2005; Won et al., 2011) and both have been shown to correlate with each other and with more direct measures of spectral resolution, such as spatial tuning curves (Anderson et al., 2011).

The stimuli were presented using the same setup as in Experiment 1, via a loudspeaker positioned at approximately head height and about 1 m from the participant in a single-walled, sound attenuating booth. The average sound level of the noise was set to 60 dBA when measured at the location corresponding to the participant's head. To reduce any possible cues related to overall loudness, the noise level was roved across intervals within each trial by ±3 dB. The starting phase of the spectral modulation was selected at random with uniform distribution for each trial to reduce the potential for any consistent local intensity cues that fixed-phase stimuli might create.

3.5.1.2 Procedure

A three-interval, three-alternative forced-choice procedure was used for both tasks. Correct-answer feedback was provided after each trial. For the ripple-discrimination task, all three intervals contained spectrally rippled noise. In each trial, two intervals contained spectral ripples that had the same starting phase (selected at random from a uniform distribution on each trial), and in the remaining interval the phase of the ripple was reversed (180° phase shift). The order of the three intervals (two same, one reversed) was randomized on every trial with equal a priori probability. Listeners were instructed to choose the interval that sounded different from the other two. The first trial of each run started at a ripple rate of 0.25 rpo, corresponding to a single ripple across the 4-octave passband. In each successive trial, the ripple rate was varied adaptively using a 2-down, 1-up rule, with rpo initially increasing or decreasing by a factor of 1.41. After the first two reversals, the step size changed to a factor of 1.19 and decreased again to a factor of 1.09 after two more reversals. Each run was considered complete after a total of ten reversals, and the geometric mean ripple rate at the last six reversal points were used to determine the threshold.

For the ripple-detection task, one of the three intervals had a spectral ripple and the other two intervals contained spectrally flat noise. Listeners were instructed to select the interval that sounded different (i.e., the one with spectral modulation). The first trial of each run started at a peak-to-valley ratio of 20 dB. In each successive trial, the ripple depth was varied adaptively using a 2-down, 1-up rule, with the peak-to-valley ratio initially increasing or decreasing by 4 dB. After the first two reversals, the step size changed to 2 dB and decreased again to 0.5 dB after two more reversals. Each run was considered complete after a total of ten reversals, and the mean peak-to-valley ratio at the last six reversal points were used to determine the threshold. The maximum peak-to-valley ratio allowed was set to 50 dB. If participants failed to accurately respond at 50 dB on six trials, the run was terminated and 50 dB was recorded as the ripple detection threshold for that run. This occurred occasionally at the highest ripple rate (2.0 rpo).

All participants completed four runs of the ripple-discrimination task, followed by four runs of each ripple rate (0.25, 0.5, 1.0 and 2.0) in the ripple-detection task. The ripple rates in the ripple-detection task were randomized within blocks, with each block containing each ripple rate once. For each ripple task (discrimination and detection) and condition (different ripple rates in the detection task), the first run for each participant was considered practice, and the last three thresholds were used to compute a mean threshold for each individual participant. Instructions were given orally and participants were asked if they had any questions about procedures before beginning each task.

3.5.2 Results

The mean results for the ripple-discrimination task from the CI group and both NH groups (age-matched and young) are shown in panel A of Figure 3.4. The white, gray, and black bars represent the CI group, age-matched NH group and young NH group, respectively. The bars represent the mean ripple discrimination threshold (in rpo) for each group. Error bars represent one standard error of the mean between participants.

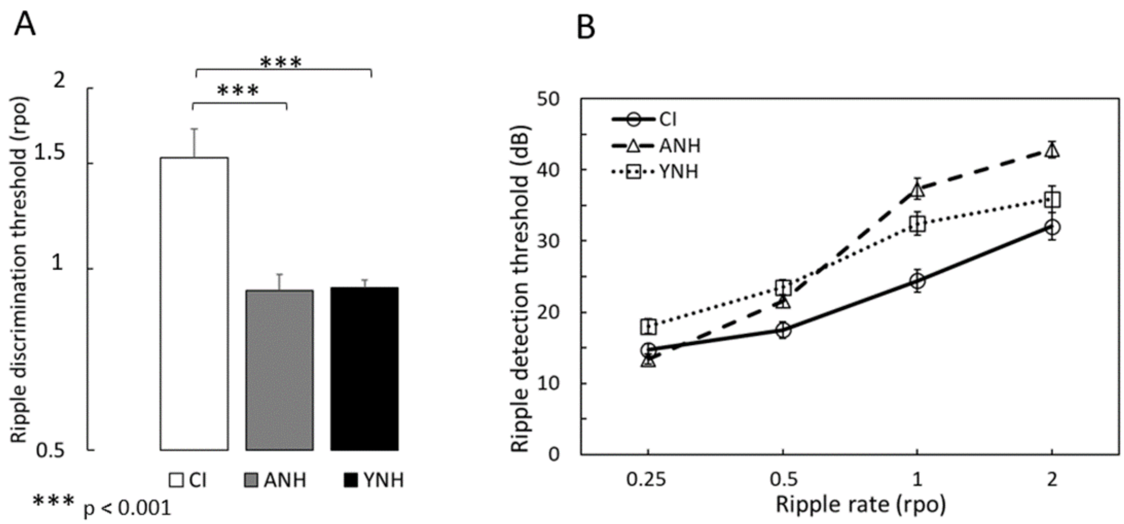

Figure 3.4

Group means for the ripple discrimination and detection tasks are shown in panels A and B, respectively. Bars in panel A represent mean ripple discrimination thresholds (in ripples per octave) for CI users and both groups of NH listeners. Panel B shows mean ripple detection thresholds (in dB) for four different ripple rates. Error bars represent one standard error of the mean between subjects.

Contrary to what might be expected based on results of Experiment 1, spectral discrimination thresholds were better for the CI group than for both NH groups listening through a vocoder with 12 dB per octave spread. Discrimination thresholds from the NH groups were very similar, with the age-matched group averaging 0.92 rpo and the young group averaging 0.93 rpo. The average ripple rate at threshold for the CI group was 1.53 rpo. A one-way, between-subjects ANOVA confirmed a significant effect of group [$F(2,87)=15.07$, $P<0.001$, $\eta_p^2=0.257$], with post-hoc comparisons with Bonferroni correction (corrected $\alpha=0.017$) showing no significant difference between NH groups ($P=0.894$), but a significant difference between thresholds for CI users and both age-matched ($P<0.001$) and young ($P<0.001$) NH groups. No differences were observed in thresholds of CI users with different processing strategies (e.g. n-of-m vs. CIS), although the number of participants using each strategy was too small to perform a formal statistical analysis.

The mean results from all three groups for the ripple-detection task are shown in panel B of Figure 2.4. Results from each group are shown with different symbols, with data from the CI group in circles, the age-matched NH group in triangles, and the young

NH group in squares. The symbols represent mean ripple-detection thresholds in dB for four different ripple rates, including 0.25, 0.5, 1.0 and 2.0 rpo. Consistent with results from the ripple-discrimination task, CI users had lower (better) ripple-detection thresholds than young NH listeners at all ripple rates. Results from young NH listeners follow a similar trend to that of CI users, with detection thresholds increasing with increasing ripple rate, resulting in a difference in thresholds of about 17 dB between the 0.25 rpo and the 2.0 rpo conditions. However, results from the age-matched NH group differ somewhat from both the CI group and young NH group, with thresholds increasing more rapidly with increasing ripple rate. The mean threshold for the age-matched NH group at 0.25 rpo is about 1 dB lower than the mean threshold for the CI group at the same rate but almost 11 dB higher than for the CI group at 2.0 rpo. The difference in average threshold between the lowest (0.25) and highest (2.0) ripple rates tested is about 30 dB, which is almost double the difference in thresholds found for the other two groups. In fact, this difference may be underestimated, since 24 age-matched NH participants were not able to reliably detect the ripples at 2 rpo on at least one of three averaged runs, even when the ripple depth was set at its maximum of 50 dB. Only 10 of the CI users and 16 of the younger NH listeners were not able to detect the ripple under the same conditions.

A repeated-measures ANOVA on the detection thresholds with ripple rate as a within-subjects factor and group as a between-subjects factor revealed a significant effect of ripple rate [$F(3,261)=329.18$, $P<0.001$, $\eta_p^2=0.791$], a significant effect of group [$F(2,87)=10.54$, $P<0.001$, $\eta_p^2=0.195$] and a significant interaction between ripple rate and group [$F(6,261)=14.70$, $P<0.001$, $\eta_p^2=0.253$]. Post hoc comparisons with Bonferroni correction for multiple comparisons (corrected $\alpha=0.0042$) showed that detection thresholds for the CI group and age-matched NH group were not significantly different at lower ripple rates [0.25 rpo ($P=0.29$) and 0.5 rpo ($P=0.008$)], but were significantly different at higher ripple rates [1.0 rpo ($P<0.001$) and 2.0 rpo ($P<0.001$)]. Thresholds for the young NH group were not significantly different from the CI group at the highest [2.0 rpo ($P=0.112$)] and lowest ripple rates [0.25 rpo ($P=0.009$)], but were significantly different at ripple rates of 0.5 ($P<0.001$) and 1.0 ($P=0.001$) rpo.

3.6 Individual Differences

3.6.1 Comparing within-group variances

To compare the amount of variance in each experimental measure for the CI group and both NH groups, the factor by which the variance was greater between each pair of groups was calculated and Levene's Test for Equality of Variances was performed. Results from these comparisons are shown in Table 3.1.

Table 3.1

Factor by which the variance was greater for the first group (listed) than the second group for different measures. Levene's Test for Equality of Variances was also performed for each comparison to calculate which differences were significant.

Task	**CI vs. YNH**	**CI vs. ANH**	**ANH vs. YNH**
Context Quiet	4.34**	3.41*	1.27
Context Noise Average	6.14***	3.69**	1.66
Nonsense Quiet	5.17***	2.89**	1.79
Nonsense Noise Average	6.55***	4.59***	1.43
Working Memory	2.95***	1.10	2.67
Non-verbal Intelligence	1.91	1.56	1.22
Ripple Discrimination	5.07***	4.58**	1.11
Ripple Detection average	1.30	1.82	0.72

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

The CI users had significantly greater between-subject variance on all four speech measures (context and nonsense sentences in quiet and in noise) compared to the age-matched NH group and young NH group. It should be noted that the magnitude by which the variance was greater for the CI group when compared to the young NH group was larger than when compared to the age-matched NH group, suggesting more variance among older participants overall. However, the variance for the age-matched NH group was not significantly greater than the young NH group on any of the speech perception measures. Interestingly, CI users did not show significantly more between-subject variance on the ripple-detection task than the age-matched or young NH listeners. The CI group did show significantly more variance on the ripple-discrimination task, although average thresholds for this measure were also quantitatively different from those of both NH groups, with the average for CI users being significantly better. On the ripple-detection task, however, thresholds for CI users were much more similar to those of the NH groups and did not show increased variance.

Another noteworthy difference in variance between groups was that the CI group had significantly more variance on the reading-span task, measuring working memory, than the young NH group but not the age-matched NH group. If working memory does indeed play a significant role in performance for CI users, this increase in variability when compared to young NH listeners further supports the use of age-matched controls in studies of CI users.

3.6.2 Correlations between measures of speech perception, spectral resolution, and cognitive function

The possible influence of both peripheral and central factors on speech-perception performance was explored by correlating speech perception scores with the measures of spectral resolution and cognitive function. Because of the large number of potential correlations, the correlations were restricted to those based on *a priori* hypotheses for CI users (e.g., the relationship between speech perception and spectral resolution) and on the findings from an exploratory principal component analysis (PCA) of the data. The factors included in the PCA analysis were context and nonsense speech perception, working memory, non-verbal intelligence, spectral-ripple discrimination and detection, age, and years of CI use (for the CI group only). Pearson's r values and corresponding P values for various comparisons, selected based on our initial hypotheses and on the PCA loadings, are shown in Table 3.2. To summarize speech performance, scores were averaged for each participant across all SNRs (including quiet) to produce an overall speech score for each type of speech material (context and nonsense). Ripple-detection thresholds were also averaged across ripple rates where thresholds for all groups were above floor performance (i.e. 0.25, 0.5, and 1.0 ripples/octave).

Table 3.2

Correlations between experimental measures for CI users and both NH groups. Scores on context and nonsense sentences are averaged for a general speech perception measure in the bottom half of the table. Pearson's r values that have p-values less than 0.05 are bolded to highlight significant correlations.

	CI		ANH		YNH	
Correlations between experimental measures	**R**	**p-value**	**R**	**p-value**	**R**	**p-value**
Context Sentences vs Nonsense Sentences	**0.887**	**<0.001**	**0.847**	**<0.001**	**0.865**	**<0.001**
Ripple Discrimination vs Ripple Detection	**-0.826**	**<0.001**	**-0.648**	**<0.001**	**-0.563**	**0.001**
Working Memory vs Non-verbal Intelligence	**0.417**	**0.022**	**0.712**	**<0.001**	**0.664**	**<0.001**
Speech Perception vs Ripple Discrimination	**0.556**	**0.001**	0.340	0.066	0.315	0.090
Speech Perception vs Ripple Detection	**-0.529**	**0.003**	**-0.514**	**0.004**	-0.273	0.144
Speech Perception vs Working Memory	**0.430**	**0.018**	**0.447**	**0.013**	0.104	0.585
Speech Perception vs Non-verbal Intelligence	0.319	0.086	**0.585**	**0.001**	0.198	0.293
Ripple Discrimination vs Working Memory	0.168	0.374	**0.525**	**0.003**	**0.552**	**0.002**
Ripple Detection vs Working Memory	-0.217	0.250	**-0.501**	**0.005**	-0.311	0.095
Ripple Discrimination vs Non-verbal Intelligence	0.170	0.369	0.315	0.090	**0.628**	**<0.001**
Ripple Detection vs Non-verbal Intelligence	-0.341	0.065	-0.241	0.199	**-0.435**	**0.016**

Considering first the measures of speech perception, strong correlations were found in all three groups between performance on context sentences and nonsense sentences (CI group: r=0.887, P<0.001; age-matched NH group: r=0.847, P<0.001; young NH group: r=0.865, P<0.001). Perhaps not surprisingly, given these very high correlations, there were no correlations observed between the *difference* in performance between context and nonsense sentences and either of the cognitive measures for any group ($p > 0.58$ in all cases). Because of the high correlations between the two speech measures, we averaged performance on context and nonsense sentences to create a single global measure of speech perception for each participant that was used in the remainder of the correlations.

Significant correlations were observed between the measures of speech perception and spectral resolution. The CI group showed significant correlations when comparing speech perception and ripple-discrimination thresholds (r=0.556, P=0.001) and ripple-detection thresholds (r=-0.529, P=0.003). However, the age-matched NH group also showed significant correlations, of roughly the same magnitude, between speech perception and ripple-detection thresholds (r=-0.514, P=0.004). The young NH group showed a different trend, with speech perception not significantly correlating with ripple-detection thresholds (r=-0.273, P=0.144).

In line with previous research indicating a link between working memory and speech perception for older hearing-impaired listeners (Akeroyd, 2008; Lunner, 2003; Zekveld, Rudner, Johnsrude, Heslenfeld, & Rönnberg, 2012), speech perception was significantly correlated with scores of verbal working memory for CI users (r=0.430, P=0.018) and age-matched NH listeners (r=0.447, P=0.013), but not for young NH listeners (r=0.104, P=0.585). The significant correlation observed for the CI users and age-matched NH listeners, but not younger NH listeners, suggests that both age and hearing loss affect the impact of working memory on speech perception in noise (Füllgrabe & Rosen, 2016).

Scores on Raven's Advanced Progressive Matrices did not show a significant correlation with speech perception for CI users (r=0.319, P=0.086), indicating non-verbal intelligence was not a strong predictor of speech performance. Interestingly, age-matched NH listeners did show a significant correlation between non-verbal intelligence and speech perception (r=0.585, P=0.001), whereas young NH listeners did not (r=0.198, P=0.283). Overall, no clear picture emerged relating non-verbal intelligence with speech performance across all three groups.

Correlations between speech-perception scores and the cognitive measures for both the CI users and the age-matched NH listeners support suggestions that cognitive factors can affect speech perception (Conway, Deocampo, Walk, Anaya, & Pisoni, 2014; Heydebrand et al., 2007; Moberly, Houston, et al., 2017). Interestingly, the proportion of variance accounted for was quite similar to that accounted for by the measures of spectral resolution. The lack of correlation between the spectral resolution and cognitive measures in the CI users suggests that these factors are accounting independently for variance. To pursue this question further, we devised a composite measure of spectral resolution. The composite score was derived by combining the within-group z-scores from the spectral-ripple discrimination thresholds and the average ripple-detection thresholds for ripple rates of 0.25, 0.5 and 1.0 rpo. In addition, a composite measure of cognitive performance was derived, which was the combined within-group z-scores from the two cognitive tests. A multiple linear regression analysis using the CI users' combined speech score as the dependent variable showed that the two composite measures accounted for 41.2% of the total variance. The cognitive and spectral resolution composite measures independently accounted for 23% and 56% of the 41.2% explained variance, respectively.

3.7 Discussion

The aim of this study was to analyze peripheral and cognitive factors that may influence speech perception of CI users, and to compare the results, in terms of overall performance and variability, with those from young and age-matched NH participants, listening through a vocoder designed to similarly limit speech perception. The measures included speech perception with context and nonsense sentences, spectral-ripple detection and discrimination, visual working memory, and non-verbal intelligence. The main findings and their implications are discussed below.

3.7.1 Greater reliance on semantic context for CI users than NH listeners

One striking finding from this study was the large decrement in speech intelligibility found for CI users when semantic context was not available. Whereas the difference in performance of the age-matched and young NH listeners between context and nonsense sentences was about 15 percentage points on average, the difference for the CI users was about 30 percentage points (see Figure 3.1). This difference was observed despite the fact that performance in the context sentences was very similar across all three listener groups, due to the use of a vocoder with the NH groups. One interpretation of this finding is that CI users have learned through experience to make more use of semantic context information than NH listeners, due perhaps to the fact that the CI users are continually presented with degraded auditory input. This interpretation is supported by findings from a recent study (Dingemanse & Goedegebure, 2019), showing that CI users made more use of contextual information in recognition of words and sentences than NH listeners. Because the same effects are not observed in NH listeners even when the stimuli are degraded by a vocoder, they may reflect longer-term adaptation of CI users to chronically degraded auditory input by learning to rely more heavily on context (Glick & Sharma, 2017). However, because our two sentence corpora also differed on other dimensions (e.g., different number of keywords, different talkers, etc.), other interpretations remain possible, as discussed in the section on limitations below.

3.7.2 Correlations between measures and variance within measures suggest peripheral and cognitive contributions to speech perception

The large variability in speech perception and spectral resolution within the population of CI users is well documented (Hast et al., 2015; Lenarz et al., 2012; Mahmoud & Ruckenstein, 2014); much less attention has been paid to variability within the NH population when the auditory input is degraded to simulate the average performance of CI users. Our results show much greater variability between CI users in measures of speech perception than between young or age-matched NH listeners, supporting the hypothesis that CI-specific factors (such as the electrode placement or neural survival) underlie a larger proportion of the variance observed. Comparisons of the within-group variances in the measures of spectral resolution were more mixed. The CI group had significantly more variance than the age-matched NH group for the ripple-discrimination thresholds. However, the amount of variance for thresholds averaged across ripple-detection rates was actually similar for the CI and age-matched NH groups.

Spectral ripple-detection thresholds are thought to reflect not only spectral resolution but also intensity resolution, as detection requires the ability to detect the level differences between the spectral peaks and valleys after auditory filtering or CI presentation (Anderson et al., 2012). From past studies, it appears as if ripple detection at low rates (< 1 rpo), i.e., those most likely to reflect intensity resolution, are best correlated with speech perception (Anderson et al., 2012; Litvak et al., 2007; McKay et al., 2018; Saoji et al., 2009a). A better measure of spectral resolution may therefore be the difference in thresholds between a low rate (i.e., 0.25 rpo) and a high rate (i.e., 1.0 rpo). However, even with this difference measure, the CI group still did not have significantly more variance than either NH group. Finally, the within-group variance in the two cognitive measures was similar between the two age-matched groups, and was less among the younger group. Thus, the variance of these cognitive measures did not differ with hearing status, once age was accounted for. Nevertheless, as mentioned in

Sec. 3.4.2, mean absolute performance in both tasks was lower among the CI users in this sample than in either NH group.

The strong correlation between the two measures of speech perception (context and nonsense sentences) in all three listener groups was perhaps not surprising, given the similarity in task and materials. Nevertheless, the results do not support our initial hypothesis that working memory and/or non-verbal intelligence may be more related to performance in context sentences, whereas measures of spectral resolution may be more related to performance in nonsense sentences. It is possible that understanding nonsense sentences required more cognitive resources than we initially predicted and that working memory and/or non-verbal intelligence may have mediated performance on both context and nonsense sentences. This interpretation seems plausible given the strong correlation between the two measures, even for the NH groups.

Correlations between the speech measures and the measures of spectral resolution were relatively high among the CI users ($r \approx 0.5$), confirming the relationship between measures of spectral resolution and speech perception that has been found in many other studies (Henry et al., 2005; Holden et al., 2016; Jeon et al., 2015; Won et al., 2011; Zhou, 2017). However, the similarly high correlations ($r = -0.514$) between spectral-ripple detection and speech perception in the age-matched NH higher listeners was puzzling, as spectral resolution in that group was limited by the vocoder, and thus should have been similar for all NH listeners. Again, this correlation remained when using a difference measure (difference of thresholds at 0.25 and 1.0 rpo, described above). In addition, the fact that some measures of spectral resolution were correlated with cognitive measures in the NH groups is a further indication that these behavioral measures of spectral resolution cannot be assumed to reflect solely peripheral processes (Neher, Lunner, Hopkins, & Moore, 2012).

Overall, the relatively high proportion of variance accounted for in CI users by the non-auditory cognitive measures suggests that non-peripheral factors account for a significant proportion of the variance observed in speech perception across the population of CI users. This result mirrors recent findings in a study of NH young adults, which showed stable individual differences that generalized across three types of degraded speech: noise-vocoded speech, time-compressed speech, and speech in babble noise

(Carbonell, 2017). It is possible that additional cognitive processes, such as inhibition-concentration (Moberly, Houston, et al., 2017) or cognitive control (Araneda et al., 2015), in conjunction with working memory and non-verbal intelligence, may be influencing these strong correlations. Mean performance for CI users was also poorer than that of our NH groups on measures of both working memory and non-verbal intelligence, which raises questions about cognitive load and perhaps cognitive decline in CI users over time. Although the CI population is different from age-matched NH listeners in a number of ways, the results from this study cast some doubt on the notion that predominantly peripheral factors account for variability in the hearing outcomes of CI users.

3.7.3 Vocoder fails to capture CI performance in NH listeners across auditory tasks

As a population, CI users come with a wide range of individual differences that are very difficult to control for in an experimental setting, including etiology of hearing loss, neural survival, duration of deafness, experience with hearing aids, exposure to American Sign Language, CI use in daily life, electrode placement, mapping, and more. The ability to simulate aspects of CI processing in NH listeners, therefore, provides an opportunity to study the contributions of implant processing, independently from these other factors (e.g., Bingabr et al., 2008; Crew et al., 2012; Dorman et al., 1998; Fu and Nogaki, 2005; Grange et al., 2017; Mesnildrey and Macherey, 2015; Shannon et al., 1995). Consistent with several earlier studies (e.g., Henry and Turner, 2003; Litvak et al., 2007; Oxenham and Kreft, 2014), our tone-excited vocoder with 12 dB per octave filter slopes to simulate current spread was successful in replicating average CI performance on the context sentences for both age-matched and young NH listeners. However, when the corpus of nonsense sentences was tested, performance between CI users and NH listeners was significantly different, with CI users’ performance being much poorer.

It could be that our vocoder actually underestimated the degree to which spectral degradation influences speech perception in CI users, as indicated by performance in nonsense sentences. Arguing against this hypothesis is the fact that the CI users generally outperformed the NH listeners on the measures of spectral resolution, suggesting that (if anything) the CI users experienced less spectral degradation than the NH listeners. This apparent mismatch between poorer speech perception and better spectral resolution of CI users relative to age-matched NH listeners may be explained by the potential for non-uniform spectral resolution along the length of the cochlea in the CI users, caused perhaps by uneven neural survival and/or unequal electrode-neural interface quality across the electrode array. Any such unevenness would impact speech perception but may not affect our measures of spectral resolution, where performance could be based on the single cochlear location that provides the most information (Anderson et al., 2011; O'Neill et al., 2019b).

The fact that the same vocoder was not able to match performance between task types, or even between different speech corpora, shows that the vocoder fails to capture some important aspects of CI perception that are currently not fully understood. It may be that simulating uneven spectral resolution across the frequency spectrum within a vocoder could potentially bring the measures of speech and spectral resolution in line.

3.7.4 Limitations

One important conclusion that could be drawn from this study is that CI users rely more heavily on semantic context for speech understanding than NH listeners. However, before we accept this interpretation, other differences between the corpora should be considered. First, the sentences within each corpus had different structures, with the IEEE context sentences having five keywords per sentence, compared with the three keywords per sentence of the nonsense sentences. Second, the vocabulary was not matched between the corpora, with the words in the nonsense sentences being generally simpler than those used in the IEEE corpus. Third, each corpus was recorded using a different single female talker, meaning that idiosyncratic differences in intelligibility between the two talkers cannot be ruled out. Finally, our lack of counterbalancing of speech materials (having always presented context sentences first, followed by nonsense sentences) could have disproportionately affected performance on the nonsense sentences, if the degree of fatigue differed between groups. The future use of better-matched corpora, with similar vocabulary and sentence structure, spoken by the same talkers, and with counterbalanced presentation order could address these concerns. Finally, it may be that the absence of context changes the relative perceptual weights assigned to certain aspects of the speech (e.g., between consonants and vowels), which may in turn be represented differently by the vocoder and the actual CI.

When interpreting the apparently lower cognitive scores of the CI users, relative to the age-matched NH group, it should be noted that NH participants were recruited primarily by way of existing connections to the University of Minnesota, either as students, alumni, former faculty, or with a general interest or involvement in higher education. In contrast, CI users were recruited based solely on the basis of hearing history and current use of at least one CI. These different sampling strategies could have resulted in groups of NH listeners that were skewed towards higher cognitive function in relation to the general population. In addition, since the reading span test was verbal in nature, factors such as verbal fluency could have influenced working memory scores reported.

Finally, since the audiometric criteria used for our age-matched NH group was modified to facilitate close age-matching with older CI users, we cannot rule out distal effects of mild high-frequency hearing loss in our results (Yeend, Beach, & Sharma, 2019). However, this appears unlikely as the average threshold for age-matched NH listeners at 8000 Hz was 22 dB HL and individual high-frequency thresholds did not correlate significantly with the vocoded speech perception scores.

3.8 Conclusions

Both cognitive and peripheral factors that may influence speech perception in CI users were explored in 30 CI users, as well as in 30 age-matched and 30 young NH adults listening through a 16-channel vocoder that simulated substantial current spread, using filters with 12 dB per octave slopes. The main findings can be summarized as follows:

- CI users performed more poorly on sentences lacking semantic context than either NH group listening to degraded speech stimuli processed through a vocoder. This may indicate the importance of effective listening strategies gained over time, and a greater reliance of CI users on context to aid in speech understanding in everyday environments.
- Between-subject variance was greater for CI users than for either group of NH listeners in speech perception for speech stimuli both with and without context, as well as for some, but not all measures, of spectral resolution.

- Correlations involving measures of spectral resolution in NH listeners, even though resolution was limited by the vocoder, and not by individual differences, suggest that these measures of spectral resolution capture more than just peripheral contributions to perception.
- Significant correlations in CI users and age-matched NH listeners between measures of speech perception and cognitive factors highlight the influence of non-peripheral factors in understanding degraded speech.
- The vocoder used to process auditory stimuli for both groups of NH listeners accurately reflected performance in CI users for context sentences, but not for nonsense sentences or for measures of spectral resolution. This result suggests that current vocoders fail to capture important aspects of CI performance.

CHAPTER 4: DEVELOPMENT OF SPEECH CORPUS WITHOUT SEMANTIC CONTEXT

Chapter 4 is reprinted from:

O'Neill, E. R., Parke, M. N., Kreft, H. A., & Oxenham, A. J. (2020). Development and validation of sentences without semantic context to complement the Basic English Lexicon sentences. *Journal of Speech, Language, and Hearing Research* (in publication).

Abstract

Purpose: The goal of this study was to develop and validate a new corpus of sentences without semantic context to facilitate research aimed at isolating the effects of semantic context in speech perception.

Method: The newly developed corpus contains nonsensical sentences but is matched in vocabulary and syntactic structure to the existing Basic English Lexicon (BEL) corpus. It consists of 20 lists, with each list containing 25 sentences and each sentence having four keywords. Each new list contains the same keywords as the respective list in the original BEL corpus, but the keywords within each list are scrambled across sentences to eliminate semantic context within each sentence, while maintaining the original syntactic structure. All sentences in the original and nonsense BEL corpus were recorded by the same two male and two female talkers.

Results: Mean intelligibility scores for each list were estimated by calculating the mean percent of correct keywords achieved by 40 normal-hearing listeners for one male and one female talker. Although small but significant differences were found between some pairs of lists, mean performance for all 20 lists fell within the 95% confidence intervals.

Conclusions: Lists in the newly developed nonsense corpus are reasonably well equated for difficulty and can be used interchangeably in a randomized experimental design. Both the original and nonsense BEL sentences, recorded using the same four talkers, are available for general, non-commercial use.

4.1 Introduction

A number of different materials have been used in clinical and research settings to assess the ability of hearing-impaired listeners to understand speech. These include sentence lists, such as the Hearing in Noise Test (HINT) (Nilsson, Soli, & Sullivan, 1994), AzBio sentences (Spahr et al., 2012), the Bamford-Kowal-Bench Speech-in-Noise Test (BKB-SIN) (Bench, Kowal, & Bamford, 1979), and the Quick Speech-in-Noise Test (QuickSIN) (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004), as well as isolated words, such as the CNC word list (Nilsson, McCaw, & Soli, 1996) and spondees (e.g. Harris, 1991; Turner et al., 2004). Sentences have some advantages over isolated words, in that they incorporate the coarticulation between words that occurs in natural communication. They also include varying degrees of semantic context that makes many words in sentences predictable to some extent, based on the preceding or following words. However, although the inclusion of semantic context has ecological validity, it also imposes a degree of uncertainty regarding exactly what was actually heard, as opposed to simply inferred by the participants. This process of inferring, or "filling in," may inflate the estimated audibility of speech, and may potentially discount any additional listening effort exerted to achieve a given level of speech understanding (e.g. Sarampalis et al., 2009; Winn, 2016).

Interest in listening effort has grown in recent years in response to individuals with hearing loss and cochlear implants (CIs) reporting high levels of mental fatigue associated with daily listening (Hughes et al., 2018). Increased listening effort has also been inferred from laboratory studies using measures such as pupil dilation (e.g. Beatty, 1982; Kahneman and Beatty, 1966), which may result at least in part from increased semantic inference or "filling in" on the part of CI users (Winn, 2016). Despite the growing interest in the connection between the use of semantic context and listening effort, the speech materials currently available to explore this issue remain limited. Most commonly, a set of sentences is used for which the final word in each sentence is either predictable or not, based on the preceding words in the sentence (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Lash et al., 2013b). This type of sentence provides a controlled method for analyzing the effect of semantic context, but places artificial importance on the final word of each sentence that is not directly comparable to the ways in which semantic context operates in conversational speech typical of everyday environments. For example, hearing-impaired listeners may use context occurring later in a sentence to resolve ambiguities occurring earlier in a sentence. In addition, the resulting measure of speech perception is limited to just one word per sentence, making it a relatively inefficient method to estimate the overall intelligibility.

Some research has tackled this problem from a different angle, using stimuli that eliminate semantic context by scrambling all the words in a sentence, but violate typical grammatical structure and prosody in the process (Boothroyd & Nittrouer, 1988). Other studies have maintained correct grammatical structure but have removed semantic context by replacing nouns and verbs with novel tokens, such as pseudo-words (Carroll, 1883; Yamada & Neville, 2007). Many studies also utilize matrix sentences which maintain a consistent sentence structure (e.g. proper name + verb + number + color + noun) (e.g. Bolia et al., 2000; Hagerman, 1982; Kollmeier et al., 2015), but the closed-set nature of the measure and limited number of choices for each word category make results hard to interpret in terms of everyday speech perception outcomes. A perhaps more ecologically valid nonsense sentence corpus has been used to examine audio-visual cues on speech perception (Helfer, 1997) and to study the intelligibility of whispered speech (Freyman, Griffin, & Oxenham, 2012; Ruggles et al., 2014). It contains sentences with English words, typical grammar, syntax, and prosody, but no semantic context. These sentences thus sound correct, but do not make any logical sense. These "nonsense sentences" provide more information than sentences with only one target word and arguably provide more ecological validity than sentences with invalid syntax or sentences selected from a closed and known set of words in each position.

Although a nonsense sentence corpus offers the opportunity to study sentence perception in the absence of semantic context, it is difficult to quantify the effect of such semantic context because there is currently no equated sentence corpus containing semantic context for comparison. Specifically, if performance on these nonsense sentences is compared to performance on another existing corpus with semantic context, such as the AzBio (Spahr et al., 2012) or IEEE (Rothauser et al., 1969) sentences, the influence of different talkers, vocabulary, number of keywords, and sentence length inherent to the different corpora cannot be eliminated as confounds (O'Neill, Kreft, & Oxenham, 2019a).

The purpose of this study was to develop, record, and validate a new nonsense sentence corpus, matched for talkers, vocabulary, number of keywords, and sentence length to that of an existing sentence corpus. The Basic English Lexicon (BEL) sentence corpus (Calandruccio & Smiljanic, 2012) was employed due to its simple vocabulary and sentence structure, as well as its high degree of semantic predictability. The new nonsense sentence corpus was developed in such a way that the keywords within each list were maintained from the original corpus, as was the grammatical structure of each sentence. Keywords within each list were scrambled across sentences, so the final lists consisted of sentences with typical grammatical structure but without semantic context. The resulting sentences were therefore correct and fully formed English sentences, but were also extremely unlikely and unpredictable contextually. Both the original and new nonsense BEL sentences were then recorded by the same two male and two female talkers. Finally, normal-hearing participants were tested on the nonsense sentences to ensure that the keywords were not predictable and that the lists were balanced for intelligibility in the presence of background noise.

4.2 Method

4.2.1 Sentence development

To maintain lexical and grammatical consistency with the original BEL sentence corpus, all lists in the nonsense BEL corpus contained the same vocabulary, keywords, and sentence structures as the original 20 lists of 25 sentences each. The original BEL sentence corpus utilizes different variations of a basic syntactic framework that consists of combinations of the following word categories: determiner (D), adjective (A), noun (N), pronoun (Pro), adverb (Adv), verb (V), and preposition (P). Although a number of different combinations of these basic word categories were used to create various syntactic structures, 12 variants account for 70% of all syntactic frameworks, with the remaining 30% of variants being slightly less common but still basic in the sense that they lacked syntactic complexity, such as embedded or proposed dependent clauses. An example of one of the common syntactic structures used is DANVA, with a corresponding sentence being, "*The boiled fish smells bad*." To ensure that all sentence structures in the nonsense corpus matched those used in the original BEL corpus, the original syntactic structure for every sentence was identified, and each word was sorted into its appropriate word category. Once all of the words from sentences within a list were sorted into word categories, these words were randomized within each word category and added back into the prescribed sentence structures to create novel nonsense sentences. This process was done manually to ensure the agreement of grammatical features, such as verb tense and noun plurality. The resulting 20 novel lists contained the same 25 sentence structures and 100 keywords as the original BEL sentence list from which it was derived, but was devoid of any semantic context. An example of one such nonsense list is shown in Table 4.1. The complete list of all 500 nonsense sentences can be found online together with the recorded sentences (https://drive.google.com/drive/folders/1H3V_4oD-7sVttKDycqcQPsn7_OMVlL3P).

Table 4.1

List 11 of the original BEL corpus and BEL nonsense corpus. The syntactic structure for each sentence is shown in the far left column with word categories as follows: D = determiner, A = adjective, N = noun, V = verb, P = preposition, Adv = adverb, Pro = pronoun. Keywords are in uppercase and the four bold words indicate an example of how words were redistributed from one original sentence across the nonsense sentences.

Syntactic structure	Original sentence using described structure	Nonsense sentence using same structure
DNVPAN	The **MEETING STARTS** in **TWENTY MINUTES**.	The WINDOWS LEARNED in BROWN SECRETARY.
DNVAN	The CUSTOMERS HATE BLACK TEA.	The MEAL PLANNED **TWENTY** KIDS.
DANVA	The SICK PERSON FEELS BETTER.	That COOL ROOM DRINKS HERE.
DANVAdvA	That BROWN BIRD is ALWAYS HERE.	A DANGEROUS BIRD was REALLY ORANGE.
DANVProAN	The THREE COUSINS did their MATH HOMEWORK.	The TWO GLOVES had their ENGLISH HORSE.
DANVDN	The DARK CLOUD COVERED the SKY.	The CHICKEN MOVIE CLIMBED the SON.
DANVN	The GROCERY STORE SELLS FOOD.	The GROCERY PERSON NEEDS FARM.
DNVPDAN	The MOVIE STARTED in the SMALL ROOM.	The CAKE BIT in the BETTER SOUP.
DANVDAN	The CHICKEN SOUP was a TASTY MEAL.	The DIFFICULT JUICE was the BIRTHDAY TEST.
DANVAandA	The COOL NIGHT was COMFORTABLE and RELAXING.	The TASTY NIGHT was THREE and DARK.
DANVVAdv	The BIRTHDAY CARD was SENT LATE.	The TROUBLED GRADE is SENT EASILY.
DNVNAdv	The SECRETARY LEARNED SPANISH EASILY.	A RADIO FEELS PROFESSOR LOUDLY.
DANVPDN	The WHITE HORSE LIVES on a FARM.	The SMALL TEA SELLS over an APARTMENT.
ProNVAN	Our APARTMENT NEEDS MORE WINDOWS.	Our HOMEWORK BUYS MORE SKY.
ProNVAN	Our MOTHER DRINKS ORANGE JUICE.	Our RABBIT STARTED SICK CUSTOMERS.
DANVDN	The DANGEROUS SNAKE BIT the RABBIT.	A COMFORTABLE BOYFRIEND COVERED the COUSINS.
DNVDAN	The PROFESSOR GAVE an UNFAIR GRADE.	The SNAKE PLAYED the UNFAIR WEDDING.
ProVANPDN	They PLAYED FAST MUSIC on the RADIO.	They HATE GREEN **MINUTES** on the CARD.
DANVAdvA	That ENGLISH TEST was REALLY DIFFICULT.	The MATH STORE was ALWAYS RELAXING.
DNandNVProN	The BOYFRIEND and GIRLFRIEND PLANNED their WEDDING.	The SISTER and CLOUD SCREAMED their MOTHER.
DNVAdvPDN	The KIDS SCREAMED LOUDLY in the PARK.	The **MEETING** LIVES LATE on the FENCE.
DANVN	The TROUBLED SON STOLE MONEY.	The WHITE MONEY GAVE KITTEN.
ProNVNAdv	His SISTER BUYS CAKE DAILY.	His SNOWMAN STOLE SPANISH DAILY.
DANVPDN	A LITTLE KITTEN CRIMBED over the FENCE.	The LITTLE GIRLFRIEND **STARTS** in that MUSIC.
DNVAAN	The SNOWMAN had TWO GREEN GLOVES.	The PARK did FAST BLACK FOOD.

4.2.2 Sentence recordings

Four native speakers of American English, two female (aged 20 and 62 yrs) and two male (aged 26 and 63 yrs), recorded all 500 original BEL sentences, as well as all 500 newly created nonsense BEL sentences. All sentences were digitally recorded in a single-walled, sound-attenuating booth located in a quiet room, at a 22050 Hz sampling rate with 16-bit resolution, using a PMD670 solid state recorder (Marantz, Mahwah, NJ). Talkers were seated approximately 12 inches from a stationary microphone (ME64, Sennheiser, Old Lyme, CT) and instructed to keep their backs against the back of the chair, to maintain a constant distance from the microphone. Talkers were also instructed to maintain a stable level of speech and to read each sentence in a natural conversational manner. Each list of 25 sentences was printed on a separate sheet of paper, and talkers were instructed to pause in between lists, to avoid any sound contamination of the sentences due to page turns. Talkers were also instructed to take slightly longer than natural pauses in between each sentence, to aid with the subsequent splicing process. Finally, if the talkers hesitated or made a noticeable error, they were instructed to pause and repeat the sentence.

After the initial recording session by each talker, the sound files were digitally edited using Audacity software (free audio editor, Version 2.3.3, 2019) and each sentence was spliced and saved as an individual sound file. Each audio file was then cross-checked with the text of each sentence to ensure word accuracy, and was also checked for any audible distortions, microphone pops, clipping, or ambient sound. After the initial editing, each talker re-recorded any flagged sentences and the editing process was repeated until all sentence recordings were deemed adequate. The average sentence duration for recordings of the original and nonsense BEL corpora was 2.27 s and 2.47 s, respectively. A two-way repeated-measures ANOVA (with a Huyhn-Feldt correction for lack of sphericity) with average sentence duration per list as a dependent variable and talker and corpus as factors, showed a significant effect of corpus (original vs nonsense) [$F(1, 19) = 196.8$, $P < 0.001$, $\eta_p^2=0.912$], talker [$F(2.5, 47.9) = 234.9$, $P < 0.001$, $\eta_p^2=0.925$], and a significant interaction between corpus and talker [$F(2, 38.9) = 7.8$, $P = 0.001$, $\eta_p^2=0.291$]. On average, sentence durations were longer for the nonsense corpus than the original corpus, but each talker also spoke at different rates, which impacted overall durations for each corpus differently. Recordings from all four talkers for both the original BEL corpus and the nonsense BEL corpus are available for download from https://drive.google.com/drive/folders/1H3V_4oD-7sVttKDycqcQPsn7_OMVlL3P.

4.2.3 Evaluation of predictability

To ensure the original BEL corpus and nonsense BEL corpus did indeed differ in predictability due to semantic context, a "fill-in-the-blank" test was administered to 20 native speakers of American English. The participants (14 females, 6 males) were mostly undergraduate students and ranged in age from 19 to 30 years (mean = 20.9 yrs; standard deviation (SD) = 2.7). All experimental protocols were approved by the Institutional Review Board of the University of Minnesota, and all listeners provided informed written consent prior to participating. For each sentence in the original and nonsense BEL corpora, one keyword was randomly omitted and replaced by a blank. Sentences from two original lists and two nonsense lists were then combined in random order to create a 100-sentence fill-in-the-blank test for each participant. This was done for all lists of sentences so that each participant completed four lists and each of the 40 lists was completed by two participants. An excerpt from one of the fill-in-the-blank tests is shown in Table 4.2.

Table 4.2

Example sentences from the fill in the blank test. The left column shows how the sentences appeared on the fill in the blank test. The middle column shows the missing word from each sentence and the far right column denoted the corpus to which each sentence belongs.

Sentences shown to participants	Missing keyword	Sentence type
1. The hungry teenagers eat _______.	snacks	Original
2. A shy _______ traveled the people.	cousin	Nonsense
3. The woman was weird in _______ problems.	many	Nonsense
4. The _______ band played in a concert.	popular	Original
5. The milk and cheese smelled _______.	horrible	Original
6. The thirsty _______ was excited and black.	dish	Nonsense
7. The people write after her _______.	salad	Nonsense
8. Her neighbors are _______ and not silver.	bright	Nonsense
9. The bears eat brown _______.	performer	Nonsense
10. The class broke _______ twins.	scary	Nonsense

The tests were administered on a computer in a quiet room, with participants typing their answers into a PDF document with an empty cell corresponding to each blank. Participants were instructed to fill in the blank in each sentence with the word they thought would fit best. Participants were told to type only one word per blank and to guess when they were unsure of the correct answer. Responses were scored in three ways. For the first scoring method, a response was only considered to be correct if it exactly matched the actual missing keyword. Since the correct verb tense was often ambiguous, the second scoring method also considered responses to be correct if the verb was correctly identified but the tense was incorrect. For example, if the missing keyword was “drinks”, but a participant answered “drank,” that response would be considered correct. In the final scoring method, any response which was a synonym, an antonym or in the same semantic category as the actual missing keyword, was considered correct. For example, if the missing keyword was “store,” but a participant responded with the word “shop,” the response would be marked as correct. A list of the original words along with the words that were accepted as having a similar semantic meaning is provided in Appendix A. Taken together, these three scoring methods provided a more nuanced interpretation of the data, especially when considering how ambiguities may have been resolved if the sentences had been spoken, rather than read.

4.2.4 List validation

To confirm that the 20 newly created lists of nonsense BEL sentences were equated for difficulty and could be used interchangeably in experimental design, 40 young, normal-hearing (NH) adults listened to all 20 lists of nonsense sentences, in background noise. Only three had previously participated in the written validation test described above. Twenty participants (19 females, 1 male) ranging in age from 18 to 22 years (mean = 19.5 yrs; SD = 2.2) listened to sentences recorded by the older female talker and 20 participants (15 females, 5 males) ranging in age from 18 to 24 years (mean = 20.9 yrs; SD = 2.7) listened to sentences recorded by the younger male talker. Though traditionally validation studies have been conducted using only one talker (e.g. Calandruccio and Smiljanic, 2012; Nilsson et al., 1994), we chose to include two of the four recorded talkers for a more robust, gender-balanced analysis of possible list-level differences in performance, while staying within reasonable time constraints for data collection. Normal hearing was defined as having pure-tone audiometric thresholds less than 20 dB hearing level (HL) at all octave frequencies between 250 and 8000 Hz with no reported history of hearing disorders. All experimental protocols were approved by the Institutional Review Board of the University of Minnesota, and all listeners provided informed written consent prior to participating.

All sentences were presented in Gaussian noise, spectrally shaped to match the long-term spectrum of the nonsense BEL corpus as recorded by either the older female talker or the younger male talker. Thus, the speech-shaped noise differed for the two groups of participants, but was mixed with the speech at the same signal-to-noise ratio (SNR) of -4 dB. This SNR was selected based on pilot testing to avoid floor and ceiling effects and to facilitate direct comparisons between lists and talkers. The speech was presented at a root-mean-square (rms) level of 65 dB sound pressure level (SPL), as measured at the position corresponding to the participant’s head, and the noise level was adjusted to produce the desired SNR. The noise was gated on 1 s before the beginning of each sentence and gated off 1 s after the end of each sentence. The stimuli were generated using MATLAB and converted via a 24-bit digital-to-analog converter (E22, Lynx Studio Technology, Costa Mesa, CA) at a sampling rate of 22050 Hz. The sounds were presented in a single-walled, sound-attenuating booth located in a quiet room via an amplifier and a single loudspeaker, placed approximately 1 m from the listener at 0° azimuth.

Participants responded to sentences by typing what they heard on a computer keyboard. Listeners were encouraged to guess individual words, even if they had not heard or understood the entire sentence. Instructions were given orally and participants were asked if they had any questions about procedures before beginning the task. Sentences were scored for keywords correct as a proportion of the total number of keywords presented. Initial scoring was automatic, with each error then checked manually for potential spelling errors or homophones (e.g. wait and weight), which were marked as correct. The 20 lists of nonsense BEL sentences were ordered randomly and split into four blocks (each containing five lists) for each participant. All testing was completed in one 2-h session per participant with a short break after completion of the first two blocks.

4.3 Results and Discussion

4.3.1 Predictability of nonsense and original BEL sentences

The mean proportion of correctly "filled in" words for each list of original and nonsense BEL sentences, scored using the three different scoring methods described above, is shown in Figure 4.1. Purple bars represent the mean proportion of reported keywords that were exactly correct, blue bars show the proportion of keywords that were either exactly correct or contained the right word but wrong tense or plurality, and orange bars show the proportion of reported keywords that were either exactly correct, correct except for an error in tense or plurality, or had the same semantic meaning as the missing keyword. Solid and hatched bars represent results for the original and nonsense BEL sentences, respectively.

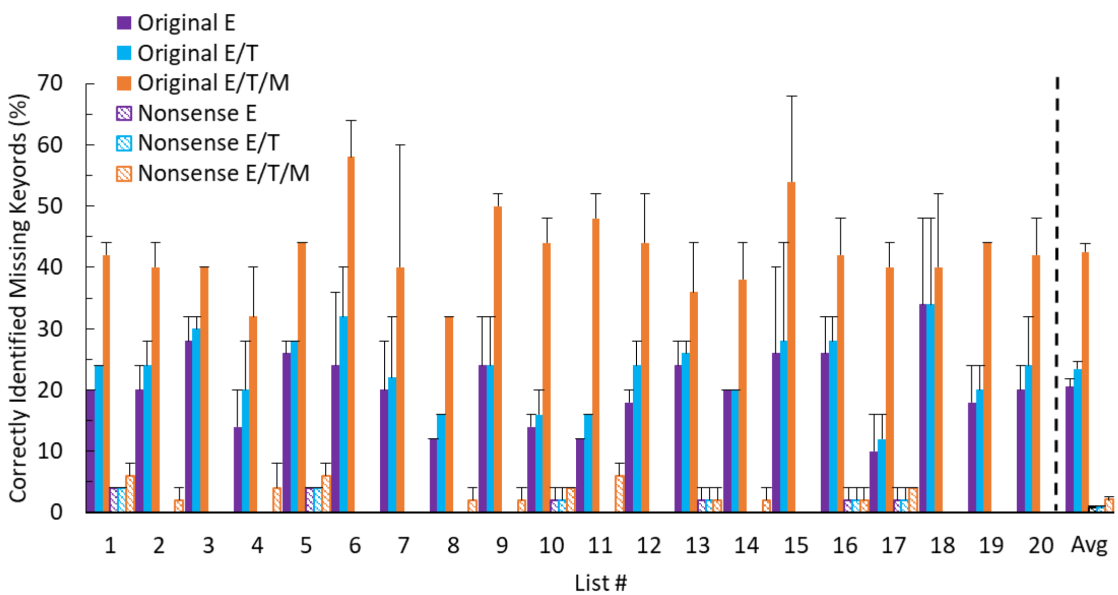

Figure 4.1

Mean proportion of correctly "filled in" words for each list of original BEL sentences and nonsense BEL sentences, scored using three methods. Purple bars represent the mean proportion of reported keywords that were exactly correct (E), blue bars show the proportion of keywords that were either exactly correctly or contained the right word but wrong tense or plurality (E/T), and orange bars show the proportion of reported keywords that were either exactly correct, correct except for an error in tense or plurality, or had the same semantic meaning as the missing keyword (E/T/M). Results for the original BEL sentences are shown by solid bars and those for the nonsense BEL sentences are shown by hatched bars. Error bars represent 1 standard error of the mean between listeners.

As expected, participants were able to correctly guess the missing keywords (allowing tense or plurality errors) at much higher rates for the original BEL sentences (avg = 22%) than for the nonsense BEL sentences (avg < 1%). Because of the highly non-normal distribution of scores (with many scores at zero), a non-parametric Wilcoxon signed rank test was used, which confirmed that these rates were significantly different from each other ($z = -5.53$; $p < 0.001$). A significant difference was also found for the rates that counted semantically similar words as correct (43% and 2% for original and nonsense BEL sentences, respectively; $z = -5.52$, $p < 0.001$). These results, showing correct response rates of 2% or less in all cases, confirm that the new nonsense BEL sentences provide minimal semantic context.

4.3.2 Speech intelligibility in noise of nonsense BEL sentences

Speech perception results for all 20 lists of nonsense BEL sentences, as recorded by an older female talker and a younger male talker, are shown in Figure 4.2. The blue squares and red circles represent individual participant scores for the female and male talker, respectively. The solid colored bars show list averages for the female and male talker; the solid black bars show the means across both talkers; the three dotted lines represent performance averaged across all 20 lists.

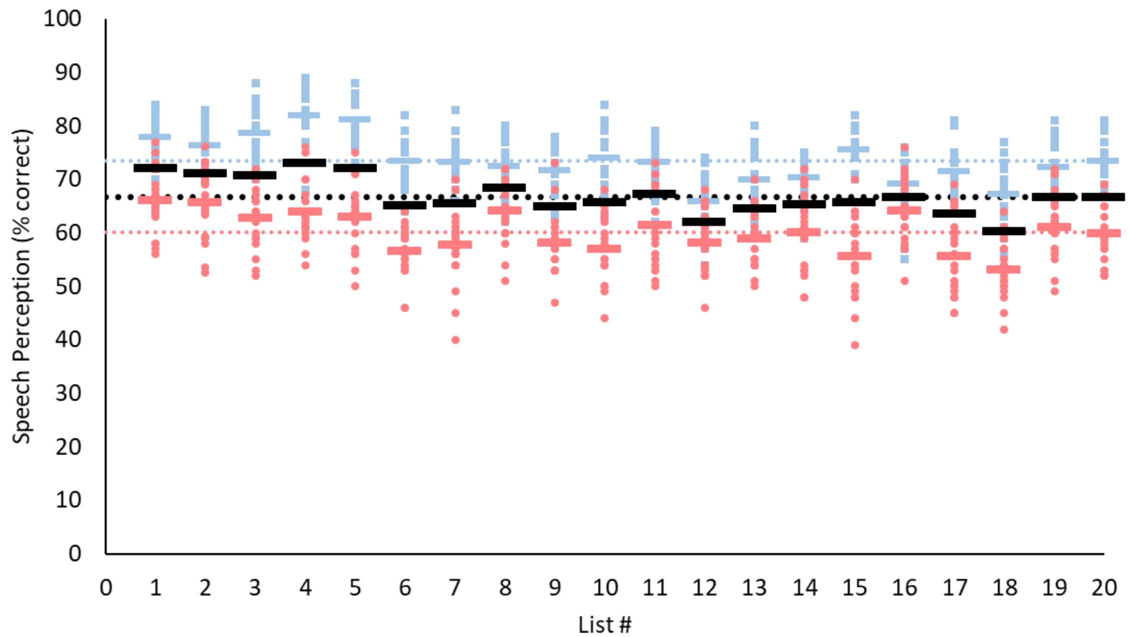

Figure 4.2

Speech perception for nonsense BEL sentences recorded by an older female talker and a younger male talker. Red circles and blue squares show individual data for the male and female talker, respectively. The red and blue bars indicate mean performance for each list for the male and female talkers, and black bars show list averages across talker.

Because the scores were all far from floor or ceiling and were normally distributed, a repeated-measures analysis of variance (ANOVA) was carried out on the proportion correct scores, with a between-subjects factor of talker (younger male or older female) and a within-subjects factor of list. There was a significant main effect of talker [$F(1,38) = 126.4$, $P < 0.001$, $\eta_p^2=0.769$], a significant main effect of list [$F(1,19) = 24.5$, $P < 0.001$, $\eta_p^2=0.392$] and a significant interaction between talker and list [$F(1,19) = 7.8$, $P < 0.001$, $\eta_p^2=0.171$]. The main effect of talker confirmed the higher scores achieved with the older female talker (avg = 73%, SD = 5.0, range 54 to 89%) than the younger male talker (avg = 60%, SD = 6.2, range 39 to 77%).

To illustrate performance differences between lists, independent of talker differences, scores are replotted in Figure 4.3, relative to each participant's mean score across sentence lists, with the horizontal bars representing the percentage point deviation from the mean for each talker (blue or red for female or male, respectively) or the mean across talkers (black).

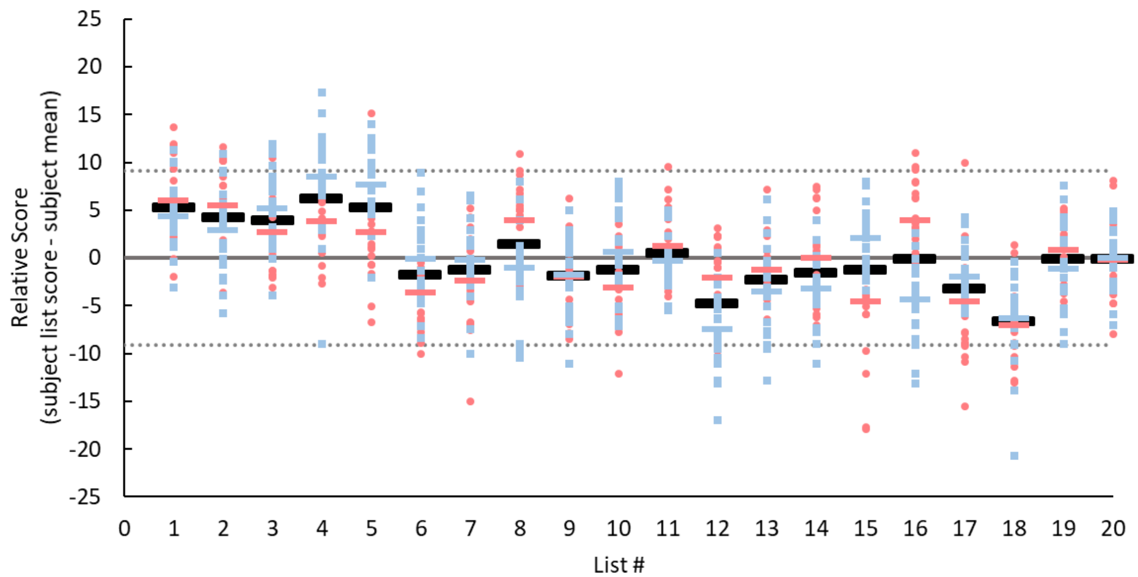

Figure 4.3

Scores for nonsense BEL sentences recorded by an older female talker (blue) and a young male talker (red), plotted relative to each participant's mean across lists. The red and blue bars indicate mean relative scores for each list for the male and female talkers, respectively, and black bars show averages across the two talkers. The horizontal dotted lines show the 95% confidence interval for performance on all lists across talkers.

As shown in Figure 4.3, some lists produced consistently better-than-average (e.g., List 1) or worse-than-average (e.g., List 18) performance for both talkers, whereas others (e.g., List 16) produced different results depending on the talker. When averaged across talkers, no sentence list produced performance that was more than 7 percentage points away from the mean, and no sentence list was outside the 95% confidence intervals. The only lists for which mean performance deviated from overall average performance by more than 5% for both talkers were lists 1 and 18. Interestingly, in a validation study of the original BEL sentences (Rimikis, Smiljanic, & Calandruccio, 2013), performance for non-native English speakers was also better than average for List 1 and poorest for List 18. Therefore, the differences observed in nonsense lists 1 and 18 may be due to differences in vocabulary specific to these lists, rather than any effects of word scrambling.

4.4 Conclusions

A corpus of 500 syntactically correct but semantically incongruent nonsense sentences, matched for vocabulary, number of keywords, sentence length and talker to the existing BEL corpus was developed and validated for list equivalency in young normal-hearing listeners. In conjunction with the original BEL corpus, this new nonsense BEL corpus can be used to research questions related to the role of semantic context for speech perception and listening effort. With further validation in listeners with hearing loss or cochlear implants, the lists could also be used for clinical testing. The recordings for both the original and nonsense BEL sentences, and text for the nonsense BEL sentences, are available for download and non-commercial use at https://drive.google.com/drive/folders/1H3V_4oD-7sVttKDycqcQPsn7_OMVlL3P.

CHAPTER 5: SEMANTIC CONTEXT AND TALKER VARIABILITY IN SPEECH PERCEPTION

Abstract

This study assessed the impact of semantic context and talker variability on speech perception by cochlear-implant (CI) users and compared between-subjects variance in performance with that found in normal-hearing (NH) listeners under vocoded conditions. Thirty post-lingually deafened adult CI users were tested, along with 30 age-matched and 30 younger NH listeners, on sentences with and without semantic context, presented in quiet and noise, spoken by four different talkers. Additional measures included working memory, non-verbal intelligence, and spectral-ripple detection and discrimination. Semantic context and between-talker differences influenced speech perception to similar degrees for both CI users and NH listeners. The between-subjects variance for speech perception in noise was similarly large within both NH groups, and was smaller than that within the CI group. Spectral-ripple detection and discrimination in CI users was significantly correlated with speech perception, but a single set of vocoder parameters for NH listeners was not able to capture average CI performance in both speech and spectral-ripple tasks. Results suggest that semantic context is used similarly by CI users and NH users when average performance is equated, and confirm that only some of the variability found within the CI population is likely accounted for by peripheral CI-related factors.

5.1 Introduction

Cochlear implants (CIs) have provided many profoundly hearing-impaired individuals with the ability to converse orally with their peers, often without the aid of visual cues. Nevertheless, understanding speech processed through a CI remains much more difficult than with normal hearing. A large-scale study of 175 unilateral and bilateral CI users (Gifford, Shallop, & Peterson, 2008) found average performance for CNC monosyllabic word recognition in quiet to be 60%. However, the average score is always representative, as speech understanding abilities vary widely within the post-lingual adult CI population, with many studies reporting word- and sentence-recognition scores ranging from 0 to 100 percent across individuals on any given task (Hast et al., 2015; Lenarz et al., 2012; Mahmoud & Ruckenstein, 2014).

Studies seeking to explain these individual differences in large samples of CI users (e.g., Blamey et al., 1996, 2013; Zhao et al., 2020) have found that typically identified factors, such as CI experience, age at implantation, duration of deafness, and etiology, together account for very little of the variance in performance (often 10 percent or less). A recent large-scale study by Gifford et al. (2018) found that a behavioral measure of spectral resolution accounted for a significant portion of variance in speech perception abilities of CI users (25%), but that the association was not sufficiently strong to accurately predict the performance of individual listeners. In addition to this largely unexplained variance in speech understanding among CI users, speech perception has also been found to improve slowly over time, with individuals showing different rates of improvement, reaching peak performance anywhere from 1 to 3 years after implantation (Cusumano et al., 2017; Lenarz et al., 2012; Ruffin et al., 2007).

Since many peripheral and clinical factors, such as the mechanics of the device, surgical factors, underlying spectral resolution, etiology, duration of deafness, and age at implantation, remain constant after implantation or are unlikely to improve, higher-level factors are thought to underlie enhancements in speech perception over time. A number of studies have explored the association between cross-modal plasticity and CI outcomes, with some finding these changes to be adaptive (Anderson et al., 2017; Rouger et al., 2012; Strelnikov et al., 2013) and others finding them maladaptive (e.g., Lee et al., 2003; Sandmann et al., 2012; Zhou et al., 2018). Cognitive factors, including working memory and cognitive control, have been found to be associated with speech perception outcomes in hearing-impaired individuals with hearing aids (Akeroyd, 2008; Arehart et al., 2013; Lunner, 2003; Rönnberg et al., 2013), but their role in the CI population is less well established. A recent study of CI users by Moberly et al. (2017) found correlations between speech perception in noise and cognitive control, as well as with auditory but not visual working memory. Heydebrand et al. (2007) also found that better verbal working memory was associated with improvement in word recognition six months after CI activation, but general cognitive ability and processing speed were not. In the same vein, Hua et al. (2017) found correlations between some, but not all, measures of cognitive skills and working memory and the perception of words and sentences in quiet and in noise in bimodal listeners (those with a CI in one ear and a HA in the other). Correlations were also found between speech perception and visual working memory, but not non-verbal intelligence, in CI users in a recent study by O'Neill et al. (2019a).

Another way to study the influence of more central processes is to explore the influence of semantic context on speech intelligibility scores. It is known that semantic context is leveraged in commonly encountered acoustic environments (e.g. Baskent et al., 2016; Bhargava et al., 2014; Marslen-Wilson and Welsh, 1978; Signoret et al., 2018), and it may be that listeners rely more heavily on such context in more difficult listening situations to compensate for reduced clarity in the speech signal. One study with older hearing-impaired adults suggested that they rely more heavily on semantic context than older adults with normal hearing when performing speech-perception tasks, with the perception of low-context or semantically anomalous sentences requiring more cognitive effort (Moradi et al., 2014). A similar conclusion was reached by Dingemanse and Goedegebure (2019) when comparing their data from CI users with earlier data from NH listeners. However, in both cases the performance of NH listeners was compared under the same acoustic conditions as the CI users or hearing-impaired listeners, so it remained unclear whether the increased use of context by the patient groups was due to the acute degraded nature of the perceived stimuli, or to longer-term changes in listening strategy produced by hearing loss. O'Neill et al. (2019a) provided preliminary data to address this question by comparing the performance of CI users with that of age-matched NH listeners whose performance was degraded by vocoding the speech materials to approximate the average performance of the CI group. They found that the difference in performance between sentences with and without semantic context was significantly greater for the CI users than for the age-matched and young NH participants listening to speech through a vocoder. However, other differences between the two sentence corpora, such as vocabulary, grammatical structure, sentence length, talker, and the order of presentation, precluded any strong conclusions regarding the differential effect of semantic context.

To address the limitations of the earlier study, and to further explore the influence of non-peripheral factors on speech perception in CI users, we conducted a study with new sentence materials, specifically developed for that purpose. In addition to speech perception, participants also completed multiple psychoacoustic and cognitive measures to assess possible associations between these factors and speech understanding within the CI population. The results were compared with those from two different NH control groups: one was age-matched to our CI participants and the other consisted of young NH listeners, mostly undergraduate students, similar to those most commonly used in the comparison groups of earlier studies. The NH participants were presented with materials via a tone-excited vocoder that was designed to simulate the effects of loss of spectral resolution and to produce performance for speech perception in noise that was comparable to that found for CI users using sentences without semantic context. Speech materials consisted of syntactically correct sentences that were either semantically coherent (context) or incoherent (nonsense), presented both in quiet and in speech-shaped noise, by four different talkers. The context and nonsense sentences were matched for vocabulary, length, grammatical structure, and talkers. Psychoacoustic measures of spectral resolution were completed by CI users and age-matched NH listeners and included broadband spectral-ripple detection and discrimination. These two groups of participants also completed two different cognitive tests: a reading-span test as a measure of verbal (visual) working memory (Conway et al., 2005), and Raven’s Advanced Progressive Matrices as a measure of non-verbal intelligence (Raven et al., 1998).

If the reliance on or benefit gained from semantic context is solely dependent on the overall level of degradation of the audio signal, CI users and NH participants (listening to speech processed through a vocoder to match CI performance without semantic context), should show a similar benefit from the presence of context. However, if successful leveraging of semantic context improves over time, CI users should exhibit a larger benefit from the presence of context. Differences in the clarity of the speech signal due to talker variation or changes in SNR may also mediate the benefit gained from semantic context and shed light on cognitive effort more characteristic of real-world environments. Our final research question involved the study of within-group variance; although many researchers have highlighted the large variability in outcomes within the CI population, there has been relatively little study of variability within the NH population under similarly degraded conditions. We hypothesized that increasing the degradation of the vocoder to match a more "baseline" measure of performance for CI users (on sentences without semantic context), may increase the variability for speech understanding in the NH groups to levels more similar to those typically seen in the CI population.

5.2 General Methods

5.2.1 Participants

A total of 30 post-lingually deafened adult CI users (23 females and 7 males) ranging in age from 34 to 79 years (mean = 63.5 years; standard deviation, SD = 9.7) were tested. All CI users had at over 6 months of CI use, with experience ranging from 7 months to 22.5 years (mean = 8.9 years; SD = 5.8). The duration of hearing loss prior to implantation varied between CI users from less than a year to 42 years (mean = 14.4 years; SD = 14.3). Twenty of the CI users used Advanced Bionics devices, 7 used Cochlear devices, and 3 used Med-El devices. Twenty-three of the CI users had previously participated in our earlier study (O'Neill et al., 2019a). A group of 30 NH adults (21 females and 9 males), age-matched to the CI group with ages ranging from 32 to 78 years (mean = 63.4; SD = 9.5) were also tested, 16 of whom had participated in our earlier study (O'Neill et al., 2019a). An additional control group was tested, consisting of 30 NH young adults (24 females and 6 males) ranging in age from 18 to 27 years (mean = 20.9; SD = 2.2). All but one of these young participants were newly recruited. All participants were native speakers of American English. Normal hearing for the young listeners was defined as having pure-tone audiometric thresholds less than 20 dB hearing level (HL) at all octave frequencies between 250 and 8000 Hz with no reported history of hearing disorders. Normal hearing for the age-matched listeners was defined as having pure-tone audiometric thresholds less than 20 dB HL at all octave frequencies between 250 and 2000 Hz and no more than 30 dB HL at 4000 Hz and 6000 Hz, with no reported history of hearing disorders. Since close age-matching with the CI group was a priority, this audiometric criteria was a compromise that allowed us to successfully recruit older participants with relatively normal hearing.

All the CI users listened with one CI. Bilateral CI users were instructed to use whichever processor they thought gave them better speech perception and to remove the other processor. Four unilateral CI users with some residual hearing in their contralateral ear were instructed to remove hearing aids and to insert an ear plug in the non-CI ear, which was worn for the entirety of the experiment. All four participants with residual hearing self-reported that their CI ear was better for speech perception than their HA ear. All CI users were asked to use processor settings (volume, sensitivity, program, noise reduction, directional microphones) typical of their everyday use.

All experimental protocols were approved by the Institutional Review Board of the University of Minnesota, and all listeners provided written informed consent prior to participating. The same 60 CI users and age-matched NH participants completed all experiments in this study. The 30 young NH listeners only completed the speech perception portion of this study.

5.2.2 Order of Experiments

For all newly recruited participants, the experiments in this study were completed in the order in which they appear: speech perception was completed first (the order of completion for context and nonsense sentences was counterbalanced), followed by the cognitive measures (working memory then non-verbal intelligence), and finally measures of spectral resolution (spectral-ripple discrimination then detection). Generally, participants completed all speech perception testing in two 2-h sessions and all non-speech testing (working memory, non-verbal intelligence, spectral-ripple discrimination and detection) in a third 2-h session. Those participants who had taken part in our previous study (23 CI users, 16 age-matched NH listeners, and 1 young NH listener) had completed testing for the same measures of cognition and spectral resolution as part of an earlier study (O'Neill et al., 2019a) and so did not repeat this testing in the current study. The average elapsed time between data collection for the previous and current studies was 24.8 months. A small number of participants completed the testing in more than three sessions due to lack of schedule flexibility and differences in testing pace.

5.3 Experiment 1: Speech Perception with Context and Nonsense Sentences

5.3.1 Stimuli

5.3.1.1 Context sentences

The context speech materials were sentences taken from the Basic English Lexicon (BEL) speech corpus (Calandruccio & Smiljanic, 2012), recorded by four different talkers (2 female, aged 20 and 62, and 2 male, aged 26 and 63). An example of a context sentences is “The *park opens* in *eleven months*,” with keywords in italics. All 20 lists were recorded by each talker, with each list containing 25 sentences of four keywords each.

5.3.1.2 Nonsense sentences

The nonsense speech materials were taken from the newly developed BEL nonsense corpus (O’Neill et al., submitted), recorded by the same four talkers as the context sentences. This corpus contains the same vocabulary, keywords and sentence structures as the original BEL corpus, but with the keywords scrambled in such a way as to render the resulting sentences semantically unpredictable. Thus each list from the nonsense BEL corpus contains sentences that are grammatically and syntactically correct and all of the same keywords as the corresponding list in the original BEL corpus, but the semantic context has been removed. An example of a nonsense sentence is “The *girl ate busy dresses*,” with keywords in italics. As with the original BEL sentences, there were 20 lists of 25 sentences, with four keywords per sentence. The speech materials are available for download here: https://drive.google.com/drive/folders/1H3V_4oD-7sVttKDycqcQPsn7_OMVlL3P.

5.3.1.3 Signal processing

Both context and nonsense sentences were presented in quiet and in Gaussian noise, spectrally shaped to match the long-term spectrum of each speech corpus (concatenated across all four talkers). The speech and noise were mixed at the appropriate SNR before further processing and presentation to the participants. For the NH listeners, the mixture was passed through a 16-channel tone-excited vocoder with the center frequencies taken from the Advanced Bionics standard clinical map. For the CI users, the stimuli underwent no further processing.

The bandpass filters used to generate the subbands of the vocoder were high-order (947) FIR filters, generated with the *fir1* function in Matlab (Mathworks, Natick, MA), producing very little overlap between the spectral content of adjacent subbands and a flat frequency response (±0.05 dB) within the entire passband. The impulse responses from the linear-phase filters were time-aligned, reaching their peaks at a delay of approximately 20 ms, independent of filter center frequency. The temporal envelope from each subband was then extracted using a Hilbert transform, and the resulting envelope was lowpass filtered using a fourth-order Butterworth filter with a cutoff frequency of 50 Hz. This cutoff frequency was chosen to reduce potential voicing periodicity cues and to reduce the possibility that the vocoder produced spectrally resolved components via the amplitude modulation of the tonal carriers. The resulting temporal envelopes were used to modulate pure-tone carriers with frequencies corresponding to the center frequencies of each channel. The effects of current spread were also simulated via the vocoder by modulating each carrier by the weighted sum of the intensity envelopes from all 16 channels (Oxenham & Kreft, 2014). The weights used in this sum were selected to produce attenuation slopes of 8 dB/octave on either side of the center frequency, to simulate sufficient spectral smearing for speech perception to approximate the average performance of CI users on the nonsense sentences in noise.

The speech was adjusted to a root mean square (rms) level of 65 dB SPL, as measured at the location corresponding to the participant's head, and the noise level was adjusted to produce the desired SNR. The noise was gated on 1 s before the beginning of each sentence and gated off 1 s after the end of each sentence. The SNRs were selected in advance, based on previous studies (O'Neill et al., 2019a; Oxenham & Kreft, 2014), to avoid ceiling and floor effects in performance, and were the same for both the context and the nonsense sentences, to facilitate direct comparisons between the two speech corpora.

5.3.2 Procedure

The stimuli were generated using Matlab and converted via a 24-bit digital-to-analog converter (L22, LynxStudio, Costa Mesa, CA) at a sampling rate of 22,050 Hz. The sounds were presented in a single-walled, sound-attenuating booth located in a quiet room via an amplifier and a single loudspeaker, placed approximately 1 m from the listener at 0° azimuth.

Listeners responded to sentences by typing what they heard on a computer keyboard. Participants were encouraged to guess individual words, even if they had not heard or understood the entire sentence. Instructions were given orally and participants were asked if they had any questions about procedures before beginning the task. Sentences were scored for keywords correct as a proportion of the total number of keywords presented. Initial scoring was automatic, with each error then checked manually for potential spelling errors or homophones (e.g., wait and weight), which were marked as correct. Before the actual experiment took place, NH listeners were presented with two sentence lists of 20 sentences each from the AzBio speech corpus (Spahr et al., 2012) in quiet to acclimate them to the vocoded stimuli before the scored sentences were presented. During this training phase, each sentence was presented visually on the computer screen while the audio was played from the speaker. Participants were instructed to listen to each sentence and try to mentally map what they were hearing with the actual words of the sentence presented on the screen, similar to the procedure used by Litvak et al. (2007). The listeners did not type any responses during this phase. This training phase was included to acclimatize NH listeners to the vocoded speech and avoid data contamination due to its initial novelty (McGettigan et al., 2008; O'Neill et al., 2019b).

Following the training phase, two different general testing procedures were implemented in all participant groups. Half of the participants in each group completed training and testing for the context sentences first, followed by the nonsense sentences, and the other half of the participants completed training and testing for the nonsense sentences first, followed by the context sentences. Directly prior to the main test blocks, all participants completed four lists of context or nonsense sentences in quiet (corresponding to the subsequent sentence testing material) to become comfortable with the procedure. This additional practice phase was completed before each set of testing phase blocks (context and nonsense). The context sentences used for the context practice block were selected from the Harvard-IEEE speech corpus (Rothauser et al., 1969) and the nonsense sentences used for the nonsense practice block were selected from the Helfer (1997) lists of nonsense sentences.

After each practice block, four testing blocks of 100 sentences each were completed for each set of testing materials (context and nonsense). Each testing block contained four sentence lists of 25 sentences each recorded by one of the four talkers, and each of the four lists were presented at one of four SNRs (0 dB, 5 dB, 10 dB and quiet). The SNRs were always presented in the same order from easiest to hardest within each block, starting with quiet, followed by 10 dB, 5 dB, and 0 dB. The presentation order of the talkers was randomized across participants but held constant across testing materials (context and nonsense) for each participant. The proportions of correct words in each condition were converted to rationalized arcsine units (RAU) (Studebaker, 1985) before statistical analysis, to mitigate some of the potential effects of floor or ceiling performance.

5.3.3 Results

The mean RAU-transformed proportion of correct keywords, averaged across talkers and presentation order of materials, for the CI group, the age-matched NH group, and the young NH group are shown in Figure 5.1. The scores from the context sentences and nonsense sentences are denoted by filled and open symbols, respectively.

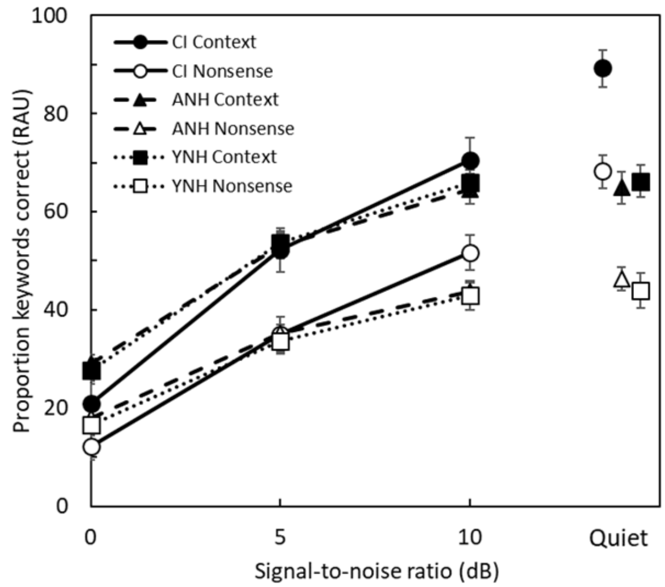

Figure 5.1

Speech perception for CI users and both groups of NH listeners averaged across talkers and presentation order of materials. The RAU-transformed proportion of keywords from sentences reported correctly is plotted as a function of signal-to-noise ratio for both context and nonsense sentences. Error bars represent ±1 standard error of the mean between listeners.

As expected, performance for all groups was poorer for the nonsense sentences than for the context sentences and worsened with decreasing SNR. Performance for the young and age-matched NH groups was also very similar for both context and nonsense sentences, which is consistent with results from our previous study (O'Neill et al., 2019a). In contrast to previous findings, the overall slope of the function across the SNRs tested differed between the CI group and both NH groups. More specifically, performance for the CI group was poorer at the most difficult SNR (0 dB) and better at the more favorable SNRs (10 dB and especially in quiet) when compared to performance for both NH groups. However, the difference in performance between context and nonsense sentences was very similar across all groups, with performance for all three groups improving 15-20 RAU, on average, with the addition of semantic context.

A repeated-measures analysis of variance (ANOVA) on the RAU-transformed data, with sentence material (context and nonsense), talker, and SNR (including quiet) as within-subjects factors and group and material presentation order as between-subjects factors, confirmed significant main effects of sentence material [$F(1,84)=846.1$, $P<0.001$, $\eta_p^2=0.910$], talker [$F(3,252)=479.6$, $P<0.001$, $\eta_p^2=0.851$], and SNR [$F(2.09,175.58)=1271.7$, $P<0.001$, $\eta_p^2=0.938$]. The effect of talker can be seen in Figure 5.2, which shows performance for all four talkers tested, averaged across the presentation order of materials. Results for the three groups are shown in the three panels of Figure 5.2, where each talker is represented by a different color and line type and performance on context and nonsense sentences is shown by filled and open symbols, respectively. In general, the most intelligible talker was the older female talker (F2) and the least intelligible was the older male talker (M2).

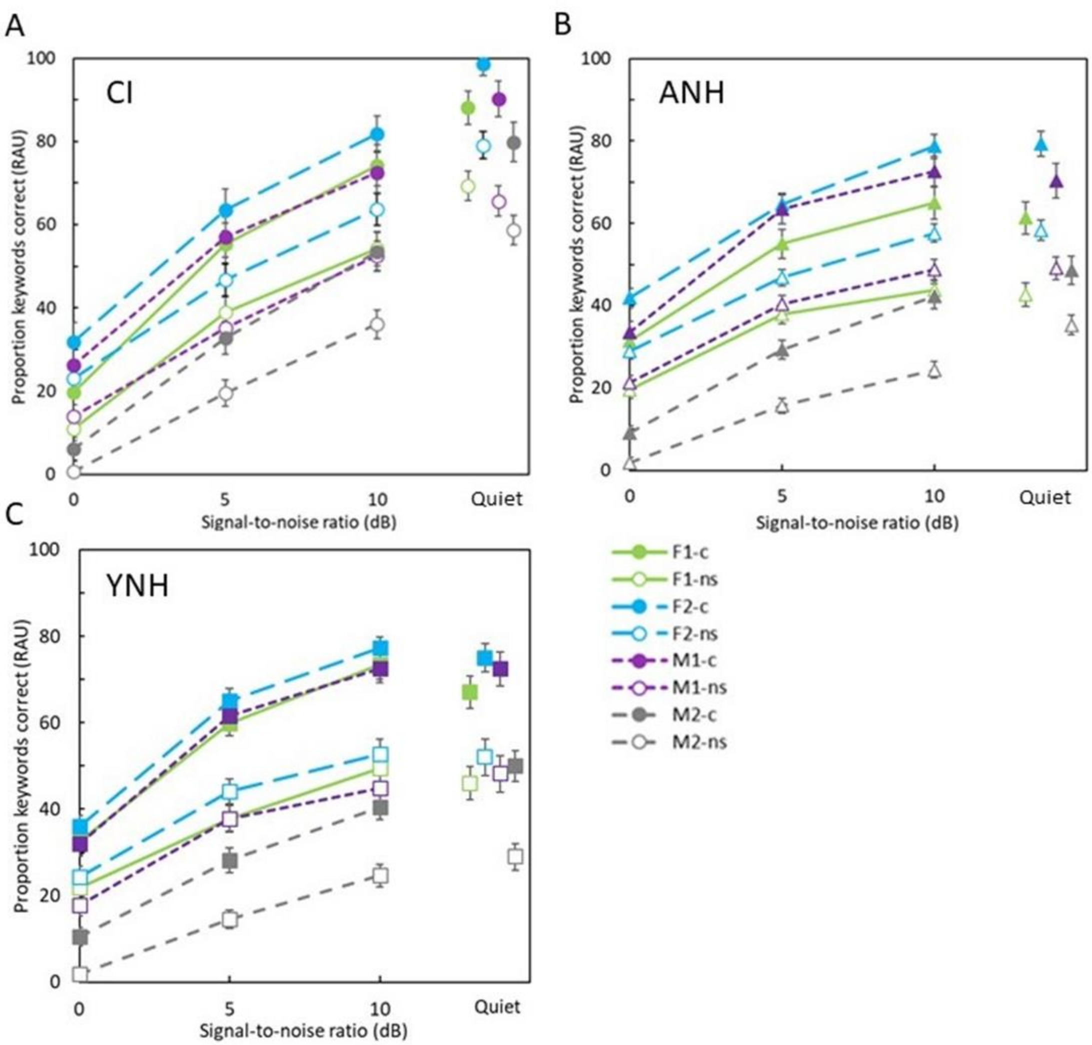

Figure 5.2

Speech perception for all four talkers tested, averaged across presentation order of materials. The RAU-transformed proportion of keywords from sentences reported correctly is plotted as a function of signal-to-noise ratio for both context and nonsense sentences. Panels A-C show data for the CI, age-matched NH, and young NH groups, respectively. Error bars represent ±1 standard error of the mean between listeners.

The repeated-measures ANOVA revealed a significant interaction between material and talker [$F(3,252)=15.5$, $P<0.001$, $\eta_p^2=0.156$] which was further explored in a series of pairwise comparisons with Bonferroni correction for multiple comparisons. The pairwise comparisons between sentence materials and talkers showed a significant difference between each talker and every other talker tested for performance on context sentences ($P<0.01$ for all comparisons), and a significant difference between every combination of talkers tested ($P<0.001$ for all comparisons), with the exception of the two young talkers ($P=1.000$), for performance on the nonsense sentences. Notably, the average difference in performance on context sentences when delivered by the least intelligible talker as compared to the most intelligible talker was 26, 34 and 31 RAU for the CI, age-matched NH, and young NH groups, respectively. Average differences in performance on nonsense sentences was 24, 29, and 26 RAU for the CI, age-matched NH, and young NH groups, respectively.

The repeated-measures ANOVA also revealed significant two-way interactions between sentence material and order [$F(1,84)=105.2$, $P<0.001$, $\eta_p^2=0.556$], sentence material and SNR [$F(3,252)=84.8$, $P<0.001$, $\eta_p^2=0.502$], talker and group [$F(6,252)=6.9$, $P<0.001$, $\eta_p^2=0.142$], talker and SNR [$F(8.81,739.84)=24.8$, $P<0.001$, $\eta_p^2=0.228$], SNR and order [$F(3,252)=4.0$, $P=0.008$, $\eta_p^2=0.046$], and SNR and group [$F(6,252)=66.2$, $P<0.001$, $\eta_p^2=0.612$]. Though multiple three- and four-way interactions reached significance, the only one that had an effect size (partial eta squared value) greater than 0.1 was a significant three-way interaction between sentence material, order, and group [$F(2,84)=13.2$, $P<0.001$, $\eta_p^2=0.239$].

The two-way interaction between sentence material and order, and the three-way interaction between sentence material, order and group, can most clearly be seen in Figure 5.3, where performance for all three groups on both context and nonsense sentences (averaged across talkers) is shown in two panels. Panels A and B show performance for all participants who completed context sentences first and nonsense sentences first, respectively. To explore the two- and three-way interactions more directly, separate repeated-measures ANOVAs were conducted for each of the three groups, with sentence material, talker and SNR as within-subjects variables and order as a between-subjects variable.

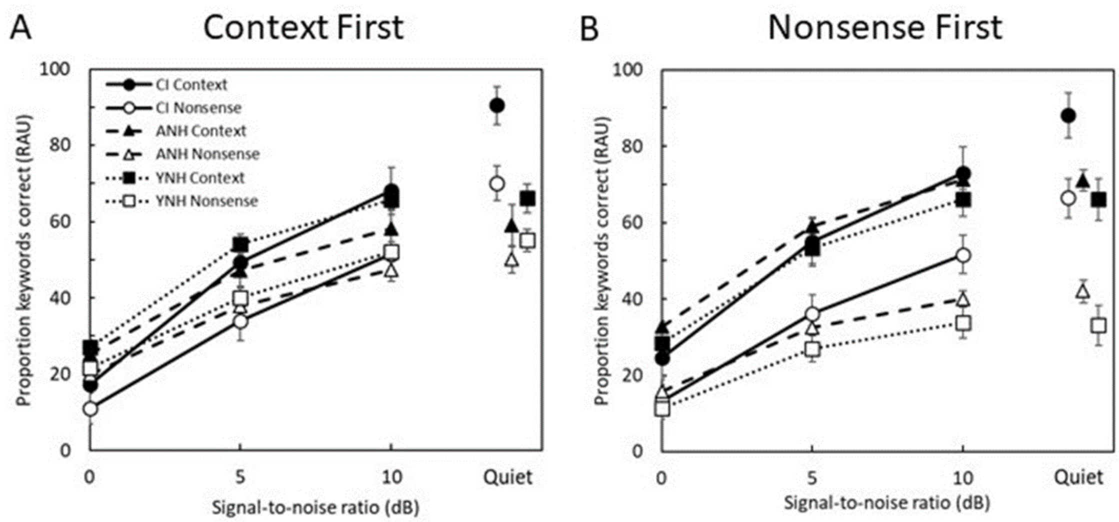

Figure 5.3

Speech perception for CI users and both groups of NH listeners averaged across talkers. Panel A shows data for participants who completed context sentences first, followed by nonsense sentences. Panel B shows data for participants who completed nonsense sentences first, followed by context sentences. The RAU-transformed proportion of keywords from sentences reported correctly is plotted as a function of signal-to-noise ratio for both context and nonsense sentences. Error bars represent ±1 standard error of the mean between listeners.

The ANOVA for the CI group, with Huynh-Feldt correction for lack of sphericity, also found significant main effects of sentence material [$F(1,28)=238.9$, $P<0.001$, $\eta_p^2=0.895$], talker [$F(2.82,78.93)=155.8$, $P<0.001$, $\eta_p^2=0.848$] and SNR [$F(2.14,59.96)=577.8$, $P<0.001$, $\eta_p^2=0.954$], as seen in the main ANOVA. However, there was no significant interaction between sentence material and presentation order [$F(1,28)=2.9$, $P=0.098$, $\eta_p^2=0.095$], indicating that CI users performed comparably on context and nonsense sentences, regardless of which material was tested first. There were also significant interactions between sentence material and talker [$F(3,84)=7.1$, $P<0.001$,

η_p^2=0.202], sentence material and SNR [F(3,84)=38.3, P<0.001, η_p^2=0.578], and talker and SNR [F(6.51,182.23)=9.0, P<0.001, η_p^2=0.244].

Similar to the CI group, the ANOVA performed on data from the age-matched NH group showed significant main effects of sentence material [F(1,28)=205.7, P<0.001, η_p^2=0.880], talker [F(2.80,78.39)=169.9, P<0.001, η_p^2=0.859], and SNR [F(2.34,65.59)=358.5, P<0.001, η_p^2=0.928], as well as interactions between sentence material and talker [F(3,84)=5.1, P=0.003, η_p^2=0.154], sentence material and SNR [F(3,84)=19.8, P<0.001, η_p^2=0.415], and talker and SNR [F(8.33,233.13)=12.7, P<0.001, η_p^2=0.311]. In contrast to the CI group, there was a significant interaction between sentence material and order [F(1,28)=52.1, P<0.001, η_p^2=0.651], and a significant three-way interaction between sentence material, SNR, and order [F(3,252)=4.1, P=0.009, η_p^2=0.129] for the age-matched NH group. Pairwise comparisons with Bonferroni correction for multiple comparisons revealed that performance on context sentences was significantly better (P=0.03) if completed second (after completion of the nonsense sentences), but that performance on nonsense sentences did not significantly differ between presentation orders (P=0.087).

In general, the ANOVA performed on data from the young NH group mirrored that of the age-matched NH group and showed significant main effects of sentence material [F(1,28)=503.2, P<0.001, η_p^2=0.947], talker [F(3,84)=164.9, P<0.001, η_p^2=0.855], and SNR [F(1.87,52.32)=366.1, P<0.001, η_p^2=0.929], as well as interactions between sentence material and talker [F(3,84)=5.5, P=0.002, η_p^2=0.164], sentence material and SNR [F(2.79,78.04)=31.4, P<0.001, η_p^2=0.528], and talker and SNR [F(9,252)=10.0, P<0.001, η_p^2=0.262]. There were also significant interactions between sentence material and order [F(1,28)=91.0, P<0.001, η_p^2=0.765] *and* talker and order [F(3,84)=3.3, P<0.024, η_p^2=0.106], as well as a significant three-way interaction between sentence material, SNR, and order [F(3,252)=7.5, P<0.001, η_p^2=0.211]. Interestingly, pairwise comparisons with Bonferroni correction for multiple comparisons revealed different behavior in the young and age-matched NH groups in relation to the interaction between sentence material and order. For the young NH group, performance on context sentences was not significantly different between the two presentation orders (P=0.951), but performance on nonsense sentences was significantly poorer if completed first (P=0.001).

Finally, to investigate whether the benefit from semantic context is modulated by differences in talker or SNR, a repeated-measures ANOVA on the *difference* in performance between context and nonsense sentences was conducted. Talker and SNR were within-subject variables and presentation order and group were between-subjects factors. The ANOVA on data from all three groups revealed significant main effects of talker [$F(3,252)=15.4$, $P<0.001$, $\eta_p^2=0.155$], SNR [$F(3,252)=84.9$, $P<0.001$, $\eta_p^2=0.503$], and presentation order [$F(1,84)=105.2$, $P<0.001$, $\eta_p^2=0.556$]. The main effect of talker indicates that the benefit gained from semantic context varied depending on the talker. Contrast analysis revealed that listeners gained significantly more benefit from semantic context when sentences were delivered by the young male talker when compared to every other talker (F1: P=0.002, F2: P=0.014, M2: P<0.001) and significantly less benefit when sentences were delivered by the older male talker when compared to every other talker (F1: P<0.001, F2: P<0.001, M1: P<0.001). The main effect of SNR also indicates that the benefit gained from the presence of semantic context varied depending on the overall SNR. Specifically, contrast analysis revealed that participants were better able to benefit from semantic context at more favorable SNRs (+10 dB and quiet) and were less able to use context at less favorable SNRs (0 dB and +5 dB), with the benefit from context being significantly greater at +5 than at 0 dB SNR (P<0.001) and also significantly greater at +10 than +5 dB SNR (P<0.001). Though a few interactions reached significance, only the interaction between group and presentation order had an effect size larger than 0.1 [$F(2,84)=13.1$, $P<0.001$, $\eta_p^2=0.237$], again reflecting the earlier observation that presentation order had a significant effect for both NH groups, but not the CI group.

5.4 Experiment 2: Working Memory and Non-verbal Intelligence

5.4.1 Methods

5.4.1.1 Stimuli and procedure

To measure working memory, a reading-span task (Conway et al., 2005) was administered (current URL: https://ubiq-x.gitlab.io/rspan/). Participants were instructed to remember a string of letters (varying in length from two to seven) while completing a distractor task in which they needed to read a sentence between each letter presentation and decide whether or not the sentence made logical sense. The presentation of letters and sentences was interleaved during each trail and the test contained a total of 18 trials. The final working memory score was the percentage of letters correctly recalled in the correct serial position, averaged across trials, for each participant. For more detailed information about this reading-span task, see O'Neill et al. (2019a).

To measure non-verbal intelligence, the paper-and-pencil version of Raven's Advanced Progressive Matrices was used (Raven et al., 1998). The test consisted of 36 matrices, each with eight possible answers. Each matrix problem has eight different combinations of shapes and textures shown in a 3-by-3 grid, with the ninth configuration absent. One of the eight possible answers for each question correctly completes the pattern formed by the combination of shapes in the grid. Participants were given 30 minutes to complete the test, which was scored for the total number of correct answers, regardless of the total number of questions answered. For more detailed information about this test, see O'Neill et al. (2019a).

5.4.2 Results

The mean results for the reading-span task from the CI group and age-matched NH group are shown in panel A of Figure 5.4 by gray and black bars, respectively. The bars represent the mean partial unit scores for each group. Error bars represent one standard error of the mean between participants.

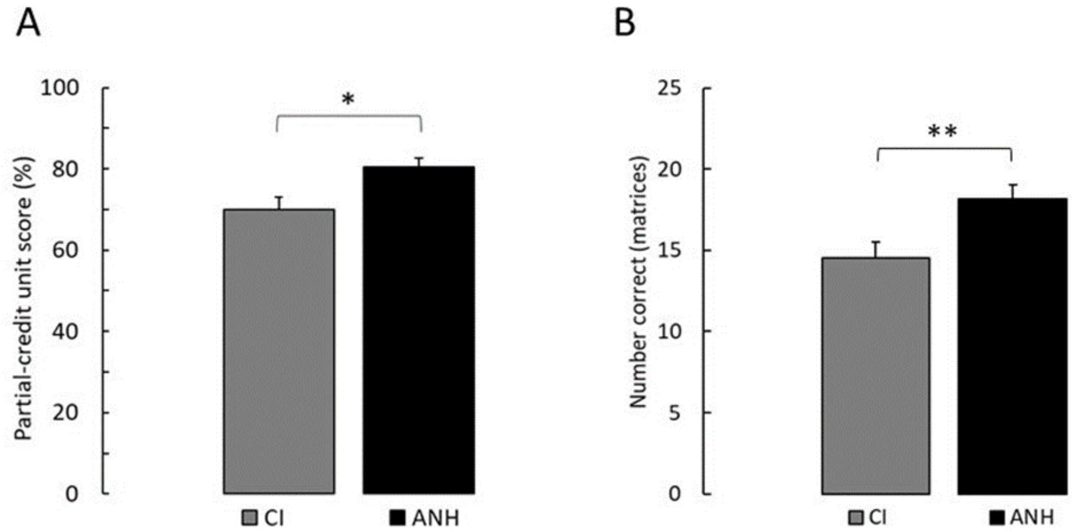

Figure 5.4

Group means for working memory and non-verbal intelligence are shown in panels A and B, respectively. Bars in panel A represent mean partial unit scores (proportion of letters within each trial recalled correctly, averaged across trials) for CI users and the age-matched NH group. Bars in panel B represent the mean number of correctly answered matrix problems (in 30 minutes) for both groups. Error bars represent one standard error of the mean between subjects and significant differences in scores between groups are shown by: * $p < 0.05$, ** $p < 0.01$.

An independent samples t-test, with equal variances not assumed, revealed a significant difference in scores for the CI group (mean=70.0, SD=16.7) and the age-matched NH group (mean=80.4, SD=13.0); t(54.7)=-2.67, *P*=0.010, Cohen's *d*=0.69, indicating that working memory was significantly poorer for the CI group than the age-matched NH group. This deficit in the CI group with respect to working memory is consistent with results from our previous study (O'Neill et al., 2019a) and other studies that have shown impairments in cognitive control among older hearing-impaired adults (Lash, Rogers, Zoller, & Wingfield, 2013a; Lash & Wingfield, 2014).

The mean results for Raven's Advanced Progressive Matrices from the CI group and age-matched NH group are shown in panel B of Figure 5.4 by gray and black bars, respectively. The bars represent the mean number of correctly answered matrix problems (in 30 minutes) for each group. Error bars represent one standard error of the mean between participants. An independent samples t-test revealed a significant difference in scores between the CI group (mean=14.5, SD=5.6) and the age-matched NH group (mean=18.2, SD=4.6); t(58)=-2.77, *P*=0.008, indicating that non-verbal intelligence was significantly poorer for the CI group than the age-matched NH group. This difference is again consistent with results from our previous study (O'Neill et al., 2019a) and other studies that have found hearing loss to be associated with accelerated cognitive decline in older adults (Claes et al., 2018; Lin, Yaffe, Xia, Xue, Harris, Purchase-Helzner, Satterfield, Ayonayon, Ferrucci, Simonsick, et al., 2013; Livingston et al., 2017).

5.5 Experiment 3: Spectral-Ripple Discrimination and Detection

5.5.1 Methods

5.5.1.1 Stimuli

Spectrally rippled noise was generated using Matlab. Gaussian broadband (350-5600 Hz) noise was spectrally modulated, with sinusoidal variations in level (dB) on a log-frequency axis (as in Litvak et al. 2007) using the equation:

$$X(f) = 10^{(D/2)\sin\{2\pi[\log_2(f/L)]f_S+\theta\}/20}$$

where *X*(*f*) is the amplitude at frequency *f* (in Hz), *D* is the spectral depth or peak-to-valley ratio (in dB), *L* is the lower cut-off frequency of the noise pass band (350 Hz in this case), f_s is the spectral modulation frequency (in ripples per octave, rpo), and θ is the starting phase of the ripple function.

The ripple-discrimination task involved spectrally rippled stimuli with a fixed peak-to-valley ratio of 30 dB, while the ripple rate was varied adaptively to track the highest ripple rate or density, in rpo, at which a phase reversal of the ripples is detectable. The ripple-detection task involved measuring the minimum detectable peak-to-valley ratio at a fixed ripple rate of 0.25, 0.5, 1, or 2 rpo. The duration of each stimulus was 400 ms, including 20-ms raised-cosine onset and offset ramps. For the age-matched NH listeners, the stimuli were passed through the same tone-excited envelope vocoder used in Experiment 1, with 16 frequency channels that produced spectral spread equivalent to 12 dB/oct. Since data for just over half of the age-matched NH listeners was collected in a previous study in which 12 dB/oct slopes were used, this same vocoder was used to maintain consistency with data collected in this experiment. The CI users were presented with the unaltered stimuli.

The stimuli were presented using the same setup as in Experiment 1, via a loudspeaker positioned at approximately head height and about 1 m from the participant in a single-walled, sound attenuating booth. The average sound level of the noise was set to 60 dBA when measured at the location corresponding to the participant's head. To reduce any possible cues related to overall loudness, the noise level was roved across intervals within each trial by ±3 dB. The starting phase of the spectral modulation was selected at random with uniform distribution for each trial to reduce the potential for any consistent local intensity cues that fixed-phase stimuli might create.

5.5.1.2 Procedure

A three-interval, three-alternative forced-choice procedure was used for both tasks and correct-answer feedback was provided after each trial. For the ripple-discrimination task, in each trial, two reference intervals contained spectrally rippled noise that had the same starting phase (selected at random from a uniform distribution on each trial), and in the remaining target interval the phase of the spectrally rippled noise was reversed (180° phase shift). The interval containing the target was randomized on every trial with equal a priori probability. Listeners were instructed to choose the interval that sounded different from the other two. The first trial of each run started at a ripple rate of 0.25 rpo, corresponding to a single ripple across the 4-octave passband. In each successive trial, the ripple rate was varied adaptively using a 2-down, 1-up rule, with rpo initially increasing or decreasing by a factor of 1.41. After the first two reversals, the step size changed to a factor of 1.19 and decreased again to a factor of 1.09 after two more reversals. Each run was considered complete after a total of ten reversals, and the geometric mean ripple rate at the last six reversal points was defined as threshold.

For the ripple-detection task, the target intervals contained a spectral ripple and the two reference intervals contained spectrally flat noise. Listeners were instructed to select the interval that sounded different (i.e., the one with spectral modulation). The first trial of each run started at a peak-to-valley ratio of 20 dB. In each successive trial, the ripple depth was varied adaptively using a 2-down, 1-up rule, with the peak-to-valley ratio initially increasing or decreasing by 4 dB. After the first two reversals, the step size changed to 2 dB and decreased again to 0.5 dB after two more reversals. Each run was considered complete after a total of ten reversals, and the mean peak-to-valley ratio at the last six reversal points was defined as threshold. The maximum peak-to-valley ratio allowed was set to 50 dB. If participants failed to accurately respond at 50 dB on six trials, the run was terminated and 50 dB was recorded as the ripple detection threshold for that run. This occurred occasionally at the highest ripple rate (2.0 rpo).

All participants completed four runs of the ripple-discrimination task, followed by four runs of each ripple rate (0.25, 0.5, 1.0 and 2.0) in the ripple-detection task. The ripple rates in the ripple-detection task were randomized within blocks, with each block containing each ripple rate once. For each ripple task (discrimination and detection) and condition (different ripple rates in the detection task), the first run for each participant was considered practice, and the last three thresholds were used to compute a mean threshold for each individual participant. Instructions were given orally and participants were asked if they had any questions about procedures before beginning each task.

5.5.2 Results

The mean results for the ripple-discrimination task from the CI group and age-matched NH group are shown in panel A of Figure 5.5, by gray and black bars, respectively. The bars represent the mean ripple discrimination threshold (in rpo) for each group. Error bars represent one standard error of the mean between participants.

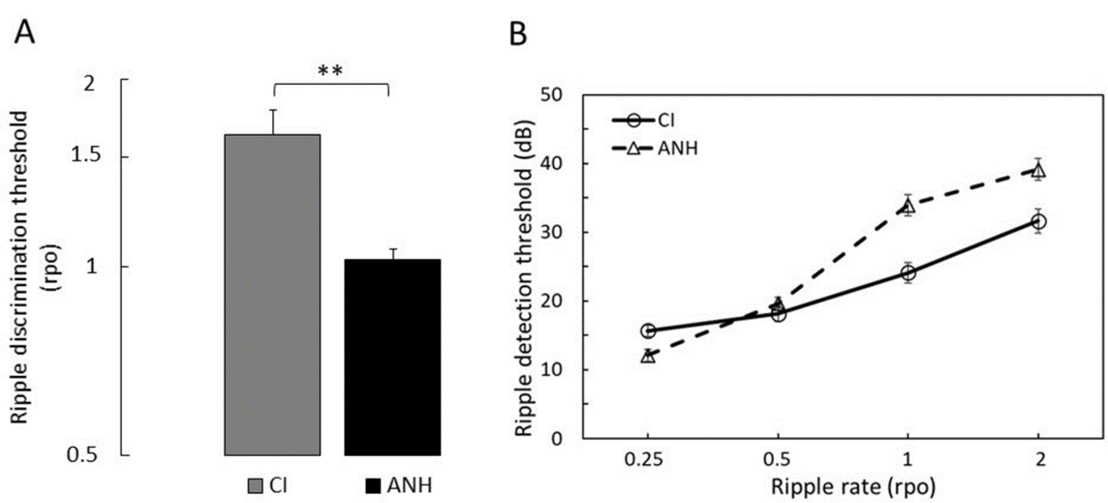

Figure 5.5

Group means for the ripple discrimination and detection tasks are shown in panels A and B, respectively. Bars in panel A represent mean ripple discrimination thresholds (in ripples per octave) for CI users and the age-matched NH listeners. Panel B shows mean ripple detection thresholds (in dB) for four different ripple rates. Error bars represent one standard error of the mean between subjects and significant differences in ripple discrimination thresholds are indicated by: ** $p < 0.01$.

Given that speech perception scores from Experiment 1 were similar for CI users and age-matched NH listeners, with speech processed through a vocoder with 8 dB/oct spread, it is interesting that ripple discrimination thresholds were poorer for the age-matched NH listeners (mean=1.03 rpo, SD=0.22) than the CI users (mean=1.63 rpo, SD=0.88) [$t(40.9)=2.94$, $P=0.005$, Cohen's $d=0.76$], even though a vocoder with better spectral resolution (12 dB/oct spread) was used. It seems likely that the difference between the two groups would have been even more pronounced had the vocoder with 8 dB/oct filter slopes been used. Ripple-discrimination thresholds in the CI group ranged from 0.36 to 3.96 rpo. The highest thresholds, implying the best spectral resolution, were still below the theoretical resolution limit of 1 ripple per filter bandwidth of the CI, as discussed in O'Neill et al. (2019a). Note that this is twice the limit imposed by the Nyquist-Shannon theorem for perfect reconstruction of the stimulus spectrum (see Resnick et al., 2020); however, the purpose of measures of spectral resolution, which is to test frequency selectivity in individual participants and not the sampling resolution of the system, does not require perfect reconstruction of the spectral envelope in order to detect a difference in two spectra. Therefore, the measure appears to provide a reasonable window into the limits of spectral resolution.

The mean thresholds from the CI group and age-matched NH group for the ripple-detection task are shown in panel B of Figure 5.5, by circles and triangles, respectively. When compared to the CI group, not only are thresholds for the age-matched NH group poorer at the three highest ripple rates (0.5, 1, and 2 rpo) but thresholds also increase more rapidly with increasing ripple rate. The mean threshold for the age-matched NH group at 0.25 rpo is 3.5 dB lower than the mean threshold for the CI group at the same rate, but almost 10 dB and 7.5 dB higher than thresholds for the CI group at higher rates of 1 and 2 rpo, respectively.

A repeated-measures ANOVA on the detection thresholds with ripple rate as a within-subjects factor and group as a between-subjects factor (with a Huynh-Feldt correction for lack of sphericity) revealed a significant effect of ripple rate [$F(2.30,133.14)=195.1$, $P<0.001$, $\eta_p^2=0.771$], a significant effect of group [$F(1,58)=7.2$, $P=0.009$, $\eta_p^2=0.110$] and a significant interaction between ripple rate and group [$F(2.30,133.14)=18.9$, $P<0.001$, $\eta_p^2=0.246$]. Post hoc comparisons with Bonferroni correction for multiple comparisons showed that detection thresholds for the CI group were significantly higher (poorer) than the age-matched NH group at the lowest ripple rate [0.25 rpo ($P=0.004$)], not significantly different at the 0.5 ripple rate ($P=0.321$), and significantly lower (better) at the highest ripple rates [1 rpo ($P<0.001$) and 2 rpo ($P=0.003$)].

5.6 Individual Differences

5.6.1 Comparing within-group variances

One important question we set out to answer was whether variability in performance between CI users is greater than that between NH listeners, when mean performance is approximately equated between the groups. To compare the amount of variance in each experimental measure for the CI group and both NH groups, the factor by which the variance was greater between each pair of groups was calculated and Levene's Test for Equality of Variances was performed. Results from these comparisons are shown in Table 5.1. To account for the difference in mean performance based on order of presentation, the individual speech scores were centered by subtracting the mean for the specific group and material presentation order to which the listener belonged. These centered values were then used in the variance analysis. The CI users had significantly greater between-subject variance compared to both NH groups on both speech-in-noise measures (context and nonsense), but not for either measure of speech in quiet. No significant differences were observed between the young and age-matched NH groups.

Table 5.1

Factor by which the variance was greater for the first group (listed) than the second group for different measures. Levene's Test for Equality of Variances was also performed for each comparison to calculate which differences were significant (* p < 0.05, ** p < 0.01, *** p < 0.001).

Significant difference in variance	**CI vs. ANH**	**CI vs. YNH**	**ANH vs. YNH**
Context quiet	1.57	1.35	0.86
Context noise average	3.59**	3.21**	0.9
Nonsense quiet	2.13	1.31	0.61
Nonsense noise average	4.15***	2.45**	0.58
Working Memory	1.66*	X	X
Non-verbal Intelligence	1.49	X	X
Ripple Discrimination	4.57**	X	X
Ripple Detection average	1.99	X	X

The CI users showed increased variance when compared to the age-matched NH group on the ripple-discrimination task, but not on the ripple-detection task. Ripple-discrimination thresholds were also significantly better for the CI group than the age-matched NH group, whereas ripple-detection thresholds were more similar between the two groups, especially at lower ripple rates. The CI users also showed significantly more variance and poorer overall scores on the reading span task (measuring working memory) when compared to age-matched NH listeners. Since working memory has been shown to correlate with speech perception in hearing-impaired listeners in previous studies (Akeroyd, 2008; Lunner, 2003; Moberly, Houston, et al., 2017; O'Neill et al., 2019a; Zekveld et al., 2012) and nearly reached significance in this study, this increased variance on a relevant cognitive measure may be contributing to overall levels of variability seen in the CI population, in general.

5.6.2 Correlations between measures of speech perception, spectral resolution, and cognitive function

The possible influence of both peripheral and central factors on speech perception was explored by correlating speech scores with the measures of spectral resolution and cognitive function. Due to the large number of potential correlations, and to allow for ease of comparison with our previous study, the correlations were restricted to those explored directly in our previous study. In addition, since the young NH listeners only completed Experiment 1, this group was not included in the correlational analyses. Pearson's *r* values and corresponding *P* values for various comparisons, for the CI and age-matched NH groups, are shown in Table 5.2. To summarize speech performance, scores were averaged for each participant across all SNRs (including quiet) and all talkers to produce an overall speech score for each type of speech material (context and nonsense). To account for differences in variance between groups, as well as for the effects of presentation order, the resulting individual speech scores were converted to standardized (z) scores. Ripple-detection thresholds were also averaged across ripple rates where thresholds for both groups were above floor performance (i.e. 0.25, 0.5, and 1.0 ripples/octave) to produce a single measure of ripple detection sensitivity for each participant.

Table 5.2

Correlations between experimental measures for CI users and age-matched NH listeners. Pearson's r values that have p-values less than 0.05 are bolded to represent significant correlations.

Correlations between experimental measures	**CI**		**ANH**	
	r	**p-value**	**r**	**p-value**
Context Sentences vs Nonsense Sentences	**0.983**	**<0.001**	**0.902**	**<0.001**
Ripple Discrimination vs Ripple Detection	**-0.788**	**<0.001**	**-0.567**	**0.001**
Working Memory vs Non-verbal Intelligence	**0.455**	**0.011**	0.183	0.333
Speech Perception vs Ripple Discrimination	**0.637**	**<0.001**	0.168	0.376
Speech Perception vs Ripple Detection	**-0.545**	**0.002**	-0.207	0.272
Speech Perception vs Working Memory	0.349	0.059	0.344	0.063
Speech Perception vs Non-verbal Intelligence	0.192	0.31	0.134	0.481
Ripple Discrimination vs Working Memory	0.162	0.391	**0.516**	**0.004**
Ripple Detection vs Working Memory	-0.128	0.5	**-0.418**	**0.022**
Ripple Discrimination vs Non-verbal Intelligence	-0.027	0.888	0.12	0.527
Ripple Detection vs Non-verbal Intelligence	-0.038	0.842	0.057	0.766

Very strong correlations were found for both groups between performance on context sentences and nonsense sentences (CI group: r=0.983; P<0.001; age-matched NH group: r=0.902, P<0.001). Given that the two speech materials were matched in every way except for the presence of semantic context, these high correlational values are perhaps not surprising, but they also imply that very little variance remained to be accounted for based on semantic context. Because of these high correlations, we averaged performance on context and nonsense sentences to create a single global measure of speech perception for each participant that was used in the remainder of the correlations.

Significant correlations were observed between the measures of speech perception and spectral resolution for the CI group, but not for the age-matched NH group. The CI group showed significant correlations when comparing speech perception and ripple-discrimination thresholds (r=0.637, P<0.001) and ripple-detection thresholds (r=-0.545, P=0.002). Interestingly, though measures of spectral resolution did not correlate with speech perception for the age-matched NH group, they did correlate with working memory. The age-matched NH group showed significant correlations between working memory and both ripple-discrimination thresholds (r=0.516, P=0.004) and ripple-detection thresholds (r=-0.418, P=0.022). While measures of spectral resolution most likely reflect differences in peripheral coding between CI users, the correlation between these measures and working memory in the age-matched NH group also indicate that cognitive factors may also be contributing to these differences in thresholds.

Consistent with the previously reported moderate relationship between working memory and speech perception in hearing-impaired listeners (Moberly, Harris, et al., 2017; Moberly, Houston, et al., 2017; O’Neill et al., 2019a), there was a trend towards a positive relationship between working memory and speech perception for both CI and age-matched NH groups that did not reach significance (CI group: r=0.349; P=0.059; age-matched NH group: r=0.344, P=0.063). This relationship did reach significance for both groups in our previous study (O’Neill et al., 2019a). The difference in outcomes may reflect the fact that the current study used shorter sentences that contained simpler vocabulary than those used in the previous study, the reduced role of working memory for these sentence materials is not altogether surprising. Scores on Raven’s Advanced Progressive Matrices did not show a significant correlation with speech perception for CI users (r=0.192; P=0.31) or age-matched NH listeners (r=0.134; P=0.481), indicating that measures of more general intelligence may not play a significant role in the perception of degraded speech, consistent with the results from earlier studies (Collison et al., 2004; Dryden et al., 2017; Nuesse et al., 2018).

5.7 Discussion

The aim of this study was to explore non-peripheral factors, specifically the role of semantic context and multiple talkers, which may influence speech perception of CI users. We compared the results from CI users, in terms of overall performance and variability, with those from young and age-matched NH participants, listening through a vocoder designed to similarly limit speech perception. The measures included speech perception with context and nonsense sentences by four different talkers, visual working memory, non-verbal intelligence, as well as spectral-ripple detection and discrimination. The main findings and their implications are discussed below.

5.7.1 Similar benefit of semantic context in NH and CI participants

A major question addressed by this study was whether CI users take more advantage of semantic context than NH listeners perceiving degraded speech acutely, as a result of more time spent dealing with challenging acoustic environments. Earlier studies had suggested that CI users rely more on context than NH listeners (Dingemanse & Goedegebure, 2019), but overall task difficulty had not been equated. Our previous study (O'Neill et al., 2019a) suggested a larger difference between context and nonsense sentences in CI users, but a number of other differences between the two sentence corpora (including different talkers, sentence complexity, and the fact that the context sentences were always presented first) precluded any definitive conclusion. In the present study, with the two sentence corpora equated in terms of complexity, vocabulary, talkers and presentation order, no overall difference was found between the CI users' and the NH listeners' ability to use semantic context. This outcome suggests that the degraded nature of the speech stimuli, rather than compensatory strategies gained through long-term exposure, may determine both NH and CI listeners' reliance on semantic context for speech understanding. This conclusion differs from the preliminary conclusions of

O'Neill et al. (2019a), which used a vocoder that simulated less spectral spread (12 dB/oct filter slopes) than the vocoder used with NH listeners in this study (8 dB/oct filter slopes). The previous study also did not use corpora matched for sentence complexity, vocabulary, or talkers, and did not counterbalance order of presentation (context sentences were always presented first in that study). It is unclear which of these factors account for the different outcomes.

One interesting finding from this study was that the benefit gained from semantic context seemed greater at higher overall levels of performance, whether due to greater individual talker intelligibility or higher SNR. The relatively smaller benefit from context seen when overall performance was poorer may be due, in part, to the compressive nature of the percent-correct speech measure, even after RAU transform. However, the nature of the measure is unlikely to account for the full effect, as the benefit from context remained high for speech perception in quiet, where performance was near ceiling in some cases. Instead, it suggests that a minimum degree of intelligibility is required to make use of context. This finding also has real world implications, suggesting that CI users may not be able to rely on semantic context to aid speech understanding in the most difficult listening environments.

5.7.2 Presentation order influences the apparent effect of semantic context

A surprising finding was the significant interaction between presentation order and sentence materials (context and nonsense sentences) for the two NH groups. For both groups, the difference in performance between the context and nonsense sentences was greater for those who were tested first on the nonsense sentences. The general pattern of results seems to be due to longer-term learning effects, with performance improving over time. This improvement occurred despite the fact that all NH listeners underwent considerable acclimatization prior to testing, including listening to 40 context sentences (AzBio) in quiet with the written versions presented simultaneously, as well as another 40 sentences of the actual task (with nonsense or context sentences). Given that the effect was not observed in the CI users, it seems reasonable to conclude that it was due to longer-term perceptual learning of the vocoded stimuli by the NH listeners.

Interestingly, for age-matched NH listeners, performance on nonsense sentences seemed independent of presentation order, whereas performance on context sentences was poorer if completed first. The opposite pattern was observed with the young NH group, where performance on context sentences was independent of presentation order but performance on nonsense sentences was significantly poorer if completed first. It remains unclear what accounts for this interaction between NH group and presentation order.

5.7.3 Differences in between-listener variability between CI users and NH listeners

The large variability in speech perception within the population of CI users is well documented (Hast et al., 2015; Lenarz et al., 2012; Mahmoud & Ruckenstein, 2014), but much less attention has been paid to variability within the NH population when the auditory input is degraded to simulate the average performance of CI users. Results from our previous study (O'Neill et al., 2019a) suggested higher levels of variability for speech perception in CI users than NH participants listening to vocoded speech. However, the vocoder used in that study was matched for performance of CI users on context sentences, and so may have underestimated actual listening difficulty. In this study, the vocoder used was instead matched for performance of CI users on nonsense sentences, and was thus more degraded and perhaps better simulated actual listening difficulty. Nevertheless, for speech in noise, where average performance was similar across listener groups, between-listener variance remained higher for the CI users than for the age-matched NH listeners by a factor of about 4 (implying larger standard deviations by a factor of 2), which is similar to what was observed in our previous study between these two groups (O'Neill et al., 2019a). Interestingly, although between-listener variance remained somewhat higher in the CI group than in either NH group for the speech scores in quiet, the differences did not reach statistical significance, in contrast to our previous study (O'Neill et al., 2019a), where variability within the CI group was higher for all measures of speech perception, including those in quiet. It is worth noting that the magnitude by which variance was greater for the CI group than the young NH group for measures of speech perception decreased on average by a factor of 2 in this study, when compared to the previous study. Since all speech stimuli was more degraded for NH listeners in this study than in the previous study, this finding may suggest that variance in speech understanding increases with increasing degradation and that some variability in performance is related to the difficulty of the auditory task, independent of variance caused by purely CI-related factors. In addition, the CI group showed more variability in working memory scores than the age-matched NH group, which could also contribute to overall variability in speech scores.

Between-listener variability in ripple-detection thresholds in the CI group was greater, but not significantly so, than that in the age-matched NH group. The statistically similar variance in the two groups is consistent with the results from our previous study (O'Neill et al., 2019a) and highlights the fact that performance in this task is not just dependent on peripheral spectral resolution (which was determined by the vocoder in the NH group), but instead also relies on intensity resolution (Anderson et al., 2012), as well as presumably other higher-level cognitive functions that play some role in performance on all psychophysical tasks.

5.7.4 Peripheral vs. central contributions to speech perception

Correlations between measures of speech perception and spectral resolution (ripple discrimination and detection) for CI users were similar in magnitude to those found in previous studies examining this relationship, including a recent large-scale study by Gifford et al. (2018), accounting for 25-40% of variance in speech understanding across users. This finding suggests that peripheral differences in spectral resolution do influence speech perception outcomes in CI users, to some extent. However, the significant correlations found between working memory and speech understanding for CI users and age-matched NH listeners in our previous study (O'Neill et al., 2019a) and correlations approaching significance between these measures in this study highlight the additional influence of higher-level factors in the perception of degraded speech. It is possible that the difference in speech materials between studies could have reduced the influence of working memory in our current study (and strength of its association with speech understanding), since the BEL sentences are shorter in length, are less syntactically complex, and contain simpler vocabulary than those used in our previous study. Our CI and age-matched NH participants were also slightly younger in this study, as compared to our previous study, since the four oldest CI users tested in the previous study did not complete this study. Eliminating these participants in their mid-to-late 70s who had poorer working memory also may have affected the strength of the correlation between working memory and speech perception seen in this study. Larger samples spanning a wider age range would be needed to assess more accurately the true contribution of working memory in the perception of degraded speech, especially for elderly CI users.

5.8 Conclusions

Speech perception using context and nonsense sentences was studied in 30 CI users, as well as in 30 age-matched and 30 young NH adults listening through a 16-channel vocoder that simulated substantial current spread, using filters with 8 dB/oct slopes. Speech scores were compared with performance in spectral-ripple detection and discrimination tasks, as well as in working-memory and nonverbal-intelligence tasks. The main findings can be summarized as follows:

- With overall speech performance in noise equated between groups, the benefit of semantic context was not greater for the CI users than for the NH listeners, suggesting the degradation of the auditory signal, rather than compensatory strategies gained through long-term exposure, may dictate the benefit gained from semantic context.
- There was evidence for longer-term learning in the NH groups, leading to an interaction between sentence material (context and nonsense sentences) and order of testing (nonsense sentences presented first or second). No such effect was found for the CI users, suggesting that substantial practice (over multiple hours) with vocoded stimuli would be needed to ensure stable performance by NH listeners with vocoded stimuli.
- The benefit gained from leveraging semantic context varied to some extent between talkers and with SNR. Specifically, the benefit from semantic context was limited when overall performance poor.
- Between-listener variability was large in the NH groups but was significantly greater in the CI group, at least for speech in noise. The large variance in the NH

groups suggests that factors not unique to the CI population contribute to variability in speech perception performance.

- The fact that the same vocoder cannot accurately capture CI performance in both speech and spectral-ripple detection and discrimination tasks suggests that current vocoders fail to capture some important aspects of CI performance.

CHAPTER 6: SOCIAL ENGAGEMENT AND LISTENING BEHAVIOR IN EVERYDAY LIFE

Chapter 6 reprinted from:

O'Neill, E. R., Basile, J. D., & Nelson, P. (2020) Individual hearing outcomes in cochlear-implant users influence social engagement and listening behavior in everyday life. *International Journal of Audiology* (in preparation).

Abstract

Cochlear implants (CIs) can improve the quality of life for people with profound sensorineural hearing loss, but individual users struggle to varying degrees with different listening environments. Daily communication challenges can negatively impact social well-being. In this study, our aim was to assess the social engagement of CI users in daily life and relate it to objective hearing outcomes. Ecological Momentary Assessments (EMAs) collected using a smartphone app were used to probe patterns of listening behavior in CI users and age-matched normal-hearing (NH) adults to detect differences in social engagement and overall quality of life. Speech perception measures, with accompanying difficulty ratings, were also performed to uncover possible correlations between objective and subjective listening behavior. Results suggest that poorer-performing CI users spend more time at home and less time conversing with others than higher-performing CI users and NH peers. Perception of listening difficulty is also very different for CI users and NH listeners, with CI users reporting little difficulty despite poor speech perception performance. Overall, the data suggest that systematic differences exist between how CI users and NH adults navigate and manipulate listening and social environments in everyday life.

6.1 Introduction

Cochlear implants (CIs) have been successful in restoring a sense of hearing for many individuals with profound hearing loss. In addition to an improvement in hearing outcomes, several studies have also shown an increase in overall quality of life post-implantation (e.g. Chung et al., 2012; Crowson et al., 2017; McRackan et al., 2018). A 2015 study (Mäki-Torkko et al.) comparing experiences of 101 CI users and their significant others pre- and post-implantation describes the two experiences as living in two different worlds, with post-implantation associated with increases in normality, autonomy and social engagement. This increase in quality of life is not altogether surprising, given the severe level of hearing loss required to qualify to receive a CI and the ability of the device to provide some auditory input for most users. However, the relative social engagement of CI users compared to their normal-hearing (NH) peers is much less understood and clear-cut.

Focus groups of CI users interviewed by Hughes et al. (2018) on topics related to listening effort, social engagement and overall quality of life highlighted ongoing social difficulties post-implantation. One participant described feeling isolated in group settings due to the increased effort and time needed to listen and assimilate to the rapid, continuous nature of dialogue. The extra time needed to process and understand what is being heard through CIs also often limited the ability of users to actively contribute to conversations, with many finding themselves more in an "observer" role than truly socially engaged (Hughes et al., 2018). The degraded speech signal delivered by CIs leads to a complex social landscape for CI users where assessing their own ability to hear successfully in a given environment, the cognitive demands of listening, the level of fatigue and anxiety associated with social engagement and level of social support contribute to the eventual decision of how to proceed in a given social situation (Pichora-Fuller, 2016).

Though relatively little is known about the association between hearing outcomes of CI users and social engagement post-implantation, many studies have shown hearing impairment in general to be associated with social isolation, depression, and cognitive decline in adults (e.g., Chia et al., 2007; Kim et al., 2017; Lin et al., 2013; Mick et al., 2014; Pronk et al., 2011). In a cross-sectional and longitudinal study of 767 older adults living in community dwellings, Mikkola et al. (2016) found that fully mobile adults with hearing impairment spent significantly less time outside the home and were three times more likely to withdraw from leisure activities than equally mobile adults with good hearing. Severe hearing impairment was also found to significantly increase the risk of depression in 30,000 patients followed by Kim et al. (2017) for a period of 11 years, regardless of age. Since depression and social withdrawal often precede declines in overall health and quality of life (Shankar et al., 2017; Stubbs et al., 2017), researchers and clinicians have long hoped that hearing interventions and treatment, with devices like hearing aids and CIs, may be able to mitigate this downward social and physical health trend.

Many studies have shown increases in cognitive health (Dawes et al., 2015), listening ability and overall health outcomes with the adoption of hearing aids (Kitterick & Ferguson, 2018). A recent study looking at the effects of auditory rehabilitation in 125 hearing-impaired adults reported patients making gains in hearing outcomes, short- and long-term memory, as well as reduced depression after receiving various hearing interventions (Castiglione et al., 2016). While these findings are encouraging, the causal link between hearing aids and improvements in quality of life is not always straightforward. The vast majority of studies looking at this relationship have analyzed changes in adults with hearing loss who have willingly chosen to purchase hearing aids and regularly participate in auditory rehabilitation. Though these individuals may see improvements in quality of life and hearing outcomes, it may simply be because those who are proactive about managing health problems have other innate qualities, such as perseverance or a positive outlook, that mediate these gains.

Given the wide range of hearing outcomes for individual CI users (Hast et al., 2015; Lenarz et al., 2012; Mahmoud & Ruckenstein, 2014), understanding the connection between speech understanding and social engagement is crucial. Since many adult CI users receive little or no formal training or rehabilitation with the device, it could be that individuals who seek out social activities more frequently achieve better speech understanding, facilitated by practice in difficult environments. On the other hand, differences in signal quality and speech perception abilities across users may cause those who struggle with the device initially to withdraw socially. CI users in the Hughes et al. study (2018) describe having a “finite amount of energy [] that can be used up very quickly” depending on other health ailments, work commitments, and social engagements on any given day. Thus, if a CI user feels the need to prioritize work, or struggles with other energy-draining health complications, listening rehabilitation and social engagement might not get as much attention and may suffer, as a result. On the other hand, if a CI user is driven by the need for social connectedness (Hughes et al., 2018), the consistent act of participating in dynamic and difficult social situations may bolster speech understanding with the device.

A logical first step to better understand the complex relationship between hearing outcomes and social engagement in CI users would seem to be to first establish that there is evidence for such an association. Therefore, the purpose of this study was to answer two basic questions: 1) Are individuals who achieve better hearing outcomes with CIs more socially engaged than those who achieve poorer outcomes? 2) Are CI users as a group less socially engaged than their NH peers? We assessed hearing outcomes by measuring speech understanding of complex, multi-talker sentences, presented in varying levels of background noise, in postlingual, adult CI users and age-matched NH listeners. We also asked participants to rate the difficulty of understanding these sentences in the lab and used these ratings as a benchmark when assessing ratings of speech understanding difficulty in actual social environments. Social engagement of CI users and NH adults was measured using ecological momentary assessments (EMAs). Two different surveys, assessing social engagement and listening behavior, were completed by participants via a smartphone app during the course of their everyday lives. We predicted that NH participants would be more socially engaged than CI users and that poorer performing CI users would be even less socially active than better performing CI users.

6.2 Experiment 1: Speech Perception and Difficulty Ratings

6.2.1 Methods

6.2.1.1 Participants

A total of 18 post-lingual adult CI users (14 females and 4 males) ranging in age from 37 to 79 years (mean = 62.3 years; standard deviation, SD = 9.5) participated. All CI users had at least two years of CI use, with experience ranging between 2 and 23 years (mean = 11.9 years, SD = 6.2). The duration of hearing loss prior to implantation varied between CI users from less than a year to 33 years (mean = 11 years; SD = 10.5). Fifteen of the CI users used Advanced Bionics devices, two used Med-El devices, and one used a Cochlear device. Eleven of the CI users were bilateral users, four were unilateral users, and three were bimodal users (one CI and one hearing aid). A group of 18 NH adults (12 females and 6 males), age-matched to the CI group with ages ranging from 39 to 77 years (mean

= 61.6 years; SD = 8.5) also participated. All participants were native English speakers, with the exception of one CI user and one NH participant, both of whom learned and spoke English regularly at a very young age. Normal hearing for the age-matched listeners was defined as having pure-tone audiometric thresholds less than 20 dB HL at all octave frequencies between 250 and 2000 Hz and no more than 30 dB HL at 4000 Hz and 6000 Hz, with no reported history of hearing disorders. Since close age-matching with the CI group was a priority, this audiometric criteria was a compromise that allowed us to successfully recruit older participants with relatively normal hearing. The average threshold for age-matched listeners at 8000 Hz was 18 dB HL (SD = 13), with individual thresholds ranging from 0 to 45 dB HL.

All experimental protocols were approved by the Institutional Review Board of the University of Minnesota, and all listeners provided informed written consent prior to participating. The same 36 participants completed all experiments in this study.

6.2.1.2 Stimuli and materials

The speech materials were sentences taken from the PRESTO speech corpus (Gilbert, Tamati, & Pisoni, 2013), recorded by male talkers with eight different American regional dialects. Each list contained nine sentences, with between three and six keywords per sentence. These sentences were used because each list contains a variety of sentence structures, sentence lengths, vocabulary, talkers and dialects, and thus closely resemble speech encountered in the everyday lives of CI users. The sentences were presented in quiet and in Gaussian noise, spectrally shaped to match the long-term spectrum of the speech corpus. The noise was gated on 1 s before the beginning of each sentence and gated off 1 s after the end of each sentence. Five signal-to-noise ratios (SNRs) were selected to reflect a range of noise levels typical of daily listening situations. The same SNRs (0, 5, 10, 15 and quiet) were used for both CI users and NH listeners to facilitate direct comparisons of both speech perception and difficulty ratings between groups.

Subjective difficulty ratings were completed after each block of sentences and recorded by having participants circle their chosen rating on a paper form. The rating scale had four options, which were “not difficult,” “somewhat difficult,” “difficult” and “very difficult.”

6.2.1.3 Procedure

The stimuli were generated using MATLAB and converted via a 24-bit digital-to-analog converter (L22, LynxStudio, Costa Mesa, CA) at a sampling rate of 22050 Hz. The sentences were presented in a single-walled, sound-attenuating booth located in a quiet room via an amplifier and a single loud speaker, placed approximately 1 m from the listener at 0° azimuth.

Listeners responded to sentences by typing what they heard on a computer keyboard. Participants were encouraged to guess individual words, even if they had not heard or understood the entire sentence. Instructions were given orally and participants were asked if they had any questions about procedures before beginning the task. Sentences were scored for keywords correct as a proportion of the total number of keywords presented. Initial scoring was automatic, with each error then checked manually for potential spelling errors or homophones (e.g., wait and weight), which were marked as correct. Two lists of nine sentences each were completed for each SNR tested. Lists for each SNR were blocked and blocks were randomized across participants. After the completion of each block, participants were instructed to give a difficulty rating based on how difficult they felt it was to understand speech within a given block.

6.2.2 Results

The proportion of correct keywords for PRESTO sentences at all SNRs tested and corresponding difficulty ratings is shown for individual CI users and age-matched NH listeners in Figure 6.1. Filled circles and open diamonds represent data for CI users and age-matched NH listeners, respectively.

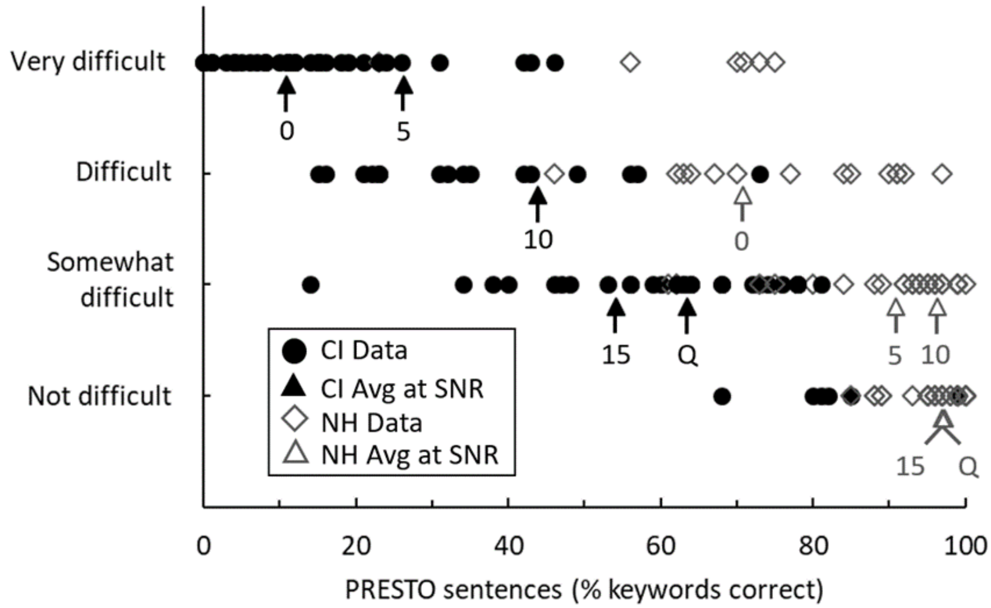

Figure 6.1

Speech perception for CI users and age-matched NH listeners. Percent correct keywords in PRESTO sentences is plotted as a function of difficulty rating by individual participants. Group average data for each SNR tested (0,5,10,15, and quiet) are marked with arrows.

One interesting finding is the extent to which the perception of difficulty varies across individuals, as a function of speech understanding. For example, one CI user rated a block in which she understood 14% of keywords in sentences as “somewhat difficult,” while an age-matched NH listener gave the same difficulty rating for a block in which she understood 100% of keywords. Among CI users, the rating “somewhat difficult” was assigned to intelligibility scores ranging from 14 to 81% and scores corresponding to the “difficult” rating ranged from 15 to 73%. In general, CI users were more likely to rate a given level of speech understanding as less difficult than NH listeners at a similar level of speech understanding. Put another way, the average speech perception score corresponding to a given rating was much poorer for CI users as a group, than NH listeners. This is more clearly demonstrated in Panel B of Figure 6.2 where average speech perception performance is plotted as a function of difficulty rating for each group. Overall, NH listeners were much less tolerant of any dip in intelligibility, often giving a rating of “somewhat difficult” or “difficult” as soon as just a handful or words were not perceived clearly. In contrast, performance for CI users had to dip much more significantly to render a rating of “somewhat difficult” or “difficult”.

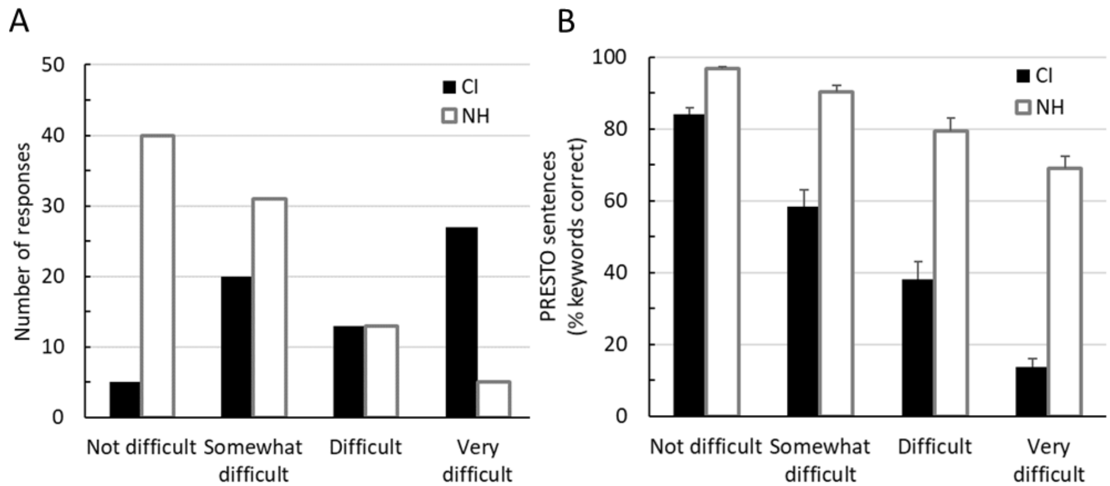

Figure 6.2

Number of responses and average speech perception performance for ratings of speech understanding (not difficult, somewhat difficult, difficult and very difficult) for PRESTO sentences. Panel A shows the total number of times each rating was selected and Panel B shows the average speech perception when each rating was selected. Results for the CI and age-matched NH groups are shown by black and white bars, respectively.

A multinomial regression model showed a significant difference in rating behavior between groups ($P < 0.001$), with CI users rating poorer speech understanding as less difficult than NH peers. This difference is especially stark for the rating of “very difficult,” where NH listeners report average speech understanding of 69% as “very difficult,” but CI users only assign this rating if performance drops to 14%, on average. In addition, relatively small drops in speech understanding (9%) corresponded to an increase in difficulty rating for NH listeners, whereas CI users tolerated much bigger dips in performance (23%) before assigning a different rating. It is also important to note that the frequency of selecting a given rating differed between the two groups. As shown in Panel A of Figure 2, NH listeners selected the rating “not difficult” significantly more often than CI users ($P < 0.001$), and CI users selected “very difficult” significantly more often than NH listeners ($P < 0.001$). This difference is due, at least in part, to the fact that overall speech understanding was significantly poorer for the CI group than the NH group at all SNRs tested ($P < 0.001$). Taken together, these results indicate that CI users as a group under-rate difficulty compared to NH listeners, even though overall performance for CI users is poorer.

CI users were also less able to tolerate background noise, as overall performance was poorer and difficulty rating was higher compared to NH listeners at the same SNR. This can most clearly be seen in Figure 1 where filled and open arrows denote average performance and difficulty rating at each SNR tested for the CI and NH groups, respectively. To give an example, average speech understanding for the NH group at an SNR of 5 was 91% and rated as “somewhat difficult,” whereas speech understanding for the CI group was 28% and rated as “very difficult” at the same SNR. This finding in particular has implications for speech understanding and social interactions in more realistic environments, which is explored further in Experiment 2.

6.3 Experiment 2: Ecological Momentary Assessments

6.3.1 Methods

6.3.1.1 Participants

The same 36 participants (18 CI users, 18 age-matched NH listeners) that completed Experiment 1 participated in this experiment.

6.3.1.2 Stimuli and materials

Participants completed one short and one longer survey multiple times over a two-week period using a smartphone app called Expimetrics. The questions and response options for the short survey completed by CI users are shown in Table 6.1. NH listeners provided responses for the first five questions of this survey, as well as an additional question asking whether they were wearing earbuds or headphones at the time of survey completion. This survey was meant to provide a snapshot of various social and listening environments participants found themselves in on a daily basis and participants provided in-the-moment responses about their activities.

Table 6.1

Questions and response options for CI users completing short survey administered via Expimetrics smartphone app. NH listeners completed the first five questions of this survey, as well as an additional question asking whether or not they were wearing earbuds at the time of survey completion.

Describe the setting you are in RIGHT NOW (check all that apply):
Home Work Leisure environment In transit Indoors Outdoors
Choose the option that best describes the social setting you are in RIGHT NOW:
I am alone/doing something independently. I am interacting with one other person. I am interacting with multiple people. I am around people but not engaging with them.
Choose the option that best describes what you are listening to RIGHT NOW:
I am not actively listening to anything right now. I am participating in conversation. I am listening to some type of media. I am listening to a live performer/speaker.
How difficult is it for you to hear what you are listening to RIGHT NOW?
Not difficult Somewhat difficult Difficult Very difficult
How much background sound is in your environment RIGHT NOW?
Very little background sound A moderate level of background sound A high level of background sound A very high level of background sound
How many cochlear implant processors are you wearing RIGHT NOW?
2 1 0
Are you using any other assistive listening devices RIGHT NOW? (Examples: direct audio cable, Mini Mic, headphones etc.)
Yes No

A second and slightly longer survey was also administered and participants answered questions in the evening, reflecting on the entirety of the day. The purpose of this survey was to probe attitudes and feelings of participants that may have influenced their social and listening behavior throughout the day. Some of these questions were modified from the Cochlear Implant Quality of Life Item Bank (McRackan, Hand, Velozo, Dubno, & Cochlear Implant Quality of Life Development Consortium, 2019) and others were created by the authors. A sample of the 28 questions and response options for this survey are shown in Table 6.2. Questions relating specifically to difficulties due to hearing impairment were modified or removed from surveys completed by NH listeners, but there was still significant overlap in questions completed by both groups of participants to facilitate comparative analysis.

Table 6.2

A sample of questions and response options for longer, reflective survey completed by participants in the evening on four different days over a two-week period.

Today I found it difficult to talk with staff in places such as shops, cafes, or banks.	**Today I used the following strategies to help me hear (check all that apply):**
Not at all	Watched mouth movements and facial expressions
Once or twice	Repositioned my body closer or further from the person speaking
Multiple times	Moved objects furniture etc in my environment
N/A	Adjusted the lighting
How much does mishearing or not hearing information when interacting with strangers bother you?	Selectively conversed with people who had clearer, louder voices None of the above
It's not a big deal It bothers me a little	**Today when I couldn't hear what was being said, I MOST OFTEN responded by:**
It bothers me quite a bit	Asking someone to repeat themselves
It bothers me a lot	Focusing on piecing together what I heard
Today I found it difficult to actively participate in casual conversation.	Pretending to hear what was said by smiling or nodding Withdrawing from the conversation
Not at all	I heard everything that was spoken to me today
Once or twice Multiple times	**Today I put effort into listening, even when it was difficult, because (check all that apply):**
N/A	What was being said was important
Today I felt anxious because of my inability to hear clearly.	I wanted to feel connected to others socially I felt social pressure to maintain appearances
Not at all	I didn't put effort into listening when it was difficult
Once or twice	Other
Multiple times N/A	**Today I stopped trying to listen to something because (check all that apply):**
Today I said the wrong thing in conversation, after mishearing what was said, and felt embarrassed.	The background sound was too loud The person talking was mumbling
Not at all	The person talking was speaking too softly
Once or twice	The person talking had a different accent
Multiple times	What was being said wasn't important
N/A	I was tired
Today I am mentally and/or physically tired from having to listen a lot.	I never stopped trying to listen to something Other
Not at all Somewhat tired	**Describe the MOST difficult listening situation you encountered today.**
Very tired I didn't actively listen much today	**Describe the LEAST difficult listening situation you encountered today.**

Participants received and completed both surveys on a smartphone app called Expimetrics, which is an integrated platform for building, scheduling and tracking surveys across mobile devices, developed by Dr. Louis Tay, an assistant professor of industrial-organizational psychology at Purdue University. All participants downloaded the app free of charge on their personal smartphones. Prior to downloading the app and starting the experiment, a schedule of the dates and times each participant would receive notifications (via the app) to complete surveys was created by the authors on the Expimetrics web platform. All participants were prompted to complete the same number of surveys on the same days of the week, but since participants did not all start the experiment on the same day, the actual calendar dates of the experiment varied across participants. Holiday weeks (i.e. Thanksgiving, Christmas, New Year's) were avoided to prevent the skewing of social engagement data between participants. To facilitate ease of data collection and analysis, individual survey schedules were created for each participant, which corresponded to a unique project code, generated by the Expimetrics software. After downloading the Expimetrics app, each participant entered their participant-specific project code into the app, which linked the survey schedule with the participant's smartphone. The "Settings" menu on each participant's smartphone was also checked to ensure that notifications to complete surveys, generated via the app, would be received throughout the experiment.

All data collected from participant smartphones could be viewed in real time (by the authors) on the Expimetrics web platform. The authors monitored the responsiveness of participants to scheduled surveys via the web portal to ensure adequate participation and detect technical issues that might be preventing participants from responding via the app. A few participants required further instruction on how to navigate the app and survey notifications, but these issues were remedied without significant loss of data. All participants were required to complete at least 30 short surveys and 2 reflective surveys for their data to be included in analysis. Two participants that completed less than 30 short surveys after the two-week experimental period were able to complete the required amount of surveys after authors scheduled two additional days of survey notifications be sent to their phones. Upon completion of the experiment, each participant's response data was downloaded from the Expimetrics web portal into an Excel spreadsheet and then imported into MATLAB for further analysis.

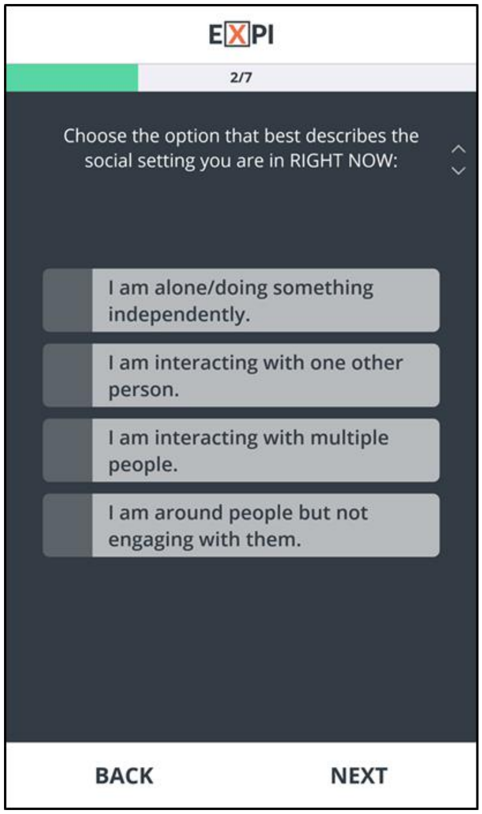

Figure 6.3

Display of sample survey question on Expimetrics smartphone app used in Experiment 2.

6.3.1.3 Procedure

Participants completed this experiment over a two-week period, beginning and ending with a visit to the lab. During the initial lab visit, the Expimetrics app was downloaded on participant's personal phones and each unique project code was entered into the app, which linked a schedule of survey times and dates to each participant's phone. Participants were told that they would be receiving notifications to complete short surveys on their phone several times a day and on multiple days over a two-week period, beginning the following day. Participants were also informed that they would not receive surveys every day, but to be mindful about keep their phones on their person throughout the day, so as not to miss notifications for surveys. The exact days and times of surveys were not given to participants to ensure participants did not alter their regular daily behavior during the study period.

Each participant received notifications to complete surveys on eight days (four week days and four weekend days) during the study period. On each of these eight days, participants received eight notifications to complete the short survey, every two hours starting at 7 am and ending at 9 pm. Participants had an hour window in which to complete each survey and received a reminder to complete the survey 15 minutes after the initial notification. On four of the eight "study days," participants also received a notification at 8 pm to complete the longer, more reflective survey. Participants had a two-hour window in which to complete this survey and received a reminder to complete the survey 15 minutes after the initial notification. After the two-week study period, participants returned to the lab to receive a $50 payment for participation in the study and were informed that they could remove the app from their phone, if desired.

6.3.1.4 Statistical analysis

All survey questions discussed in the Results section were analyzed using a one-way ANOVA with response options as within-subject dependent variables and group as a between-subjects factor, unless otherwise stated. Differences across response options were not tested, due to the number of possible comparisons and our focus on assessing differences in response behavior between groups. A p-value criterion of 0.05 was used to assess statistical significance. All statistical tests used to analyze results from Experiment 2 were performed using IBM SPSS Statistics software.

6.3.2 Results

6.3.2.1 In-the-Moment Survey Responses

The mean completion rate for the in-the-moment surveys was 72% for the CI group (46/64 surveys; SD = 9.8) and 79% (51/64 surveys; SD = 7.6) for the NH group. Survey responses were normalized for each individual participant before being included in further analysis.

The average percentage of time each response option was selected for the first five questions of the in-the-moment survey is shown in Figure 6.4. The white speckled bars represent group averages for the NH listeners, black bars show averages for high-performing CI users and gray bars show averages for lower-performing CI users. The CI group was split into two groups of 9 users each, based on average speech perception scores from Experiment 1. Average speech understanding for the "good CI users" was greater than 40% across all SNRs tested (0, 5, 10, 15, quiet), and "poor CI users" had average scores below 40%.

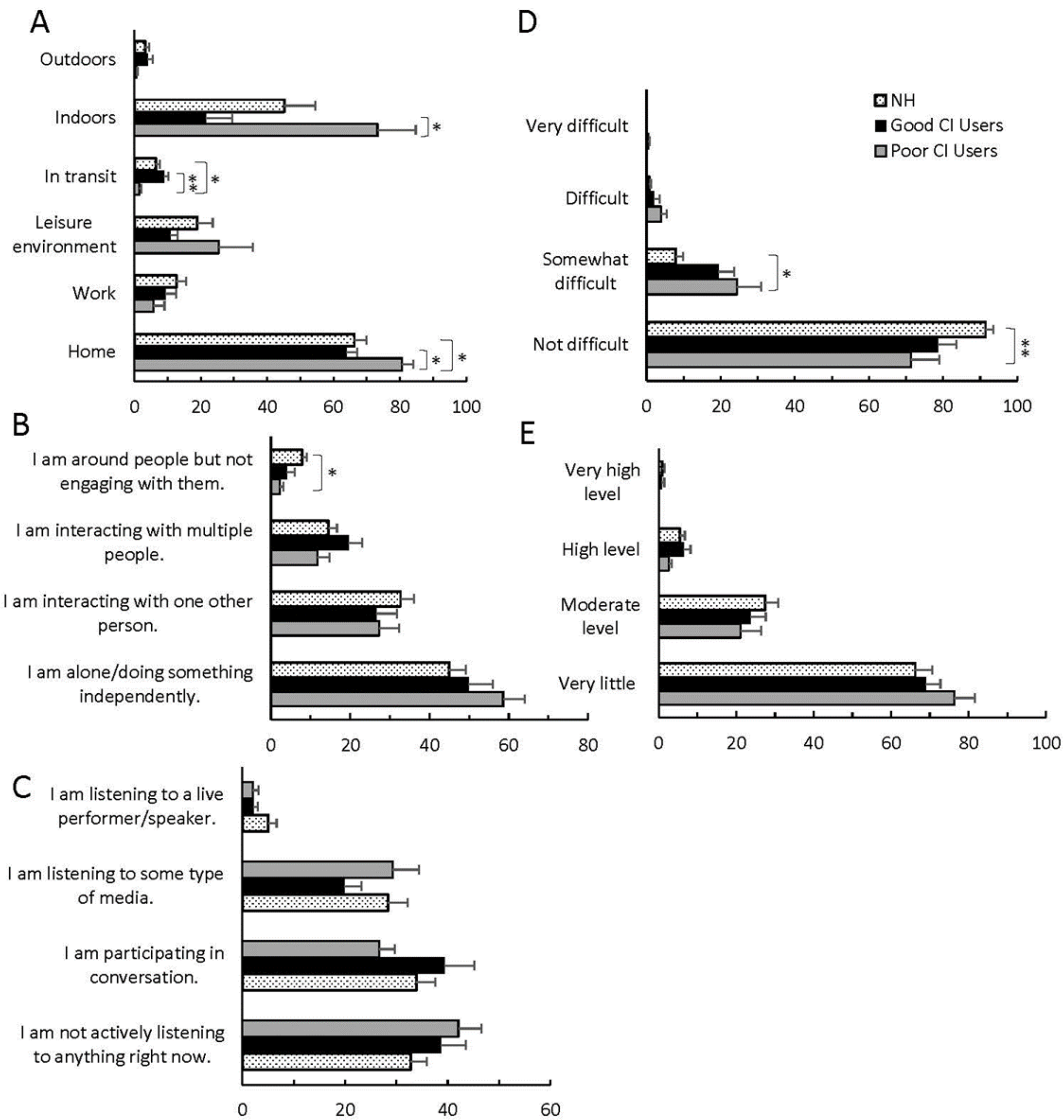

Figure 6.4

Group average responses for NH listeners, good CI users and poor CI users to the first five questions from the in-the-moment survey. Error bars represent ± 1 standard error and significance is depicted as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Panel A of Figure 6.4 shows group average responses for the first question of the survey which asked participants to describe the setting they were in at the time they completed the survey. Participants were allowed to pick more than one response option for this question. In general, all groups of participants spent more time at home than at work, and more time indoors than outdoors. Since many participants were not employed, due to age, retirement or hearing status, they were more often at home than in any other environment. The average amount of time participants reported being indoors differed between groups, but upon inspection of individual responses, often times participants would indicate they were at home, but fail to report whether they were indoors or outdoors. Thus, group differences in the average time spent indoors was not analyzed in further detail. One interesting difference between groups was that poor CI users spent significantly more time at home than both good CI users (P = 0.032) and NH peers (P = 0.035). Poor CI users also spent significantly less time in transit than good CI users (P = 0.003) and NH listeners (P = 0.025). Taken together, these results may indicate a hesitancy for CI users with poorer hearing outcomes to interact with the greater outside world in daily life.

Social engagement was assessed more directly in the second question of the in-the-moment survey, shown in Panel B of Figure 6.4, which asked participants to describe the social setting they were in at the time of the survey. On average, NH participants spent 45% of their time alone or doing something independently and 55% of their time either interacting with or around other people. In contrast, poor CI users were alone 60% of the time and only interacting or around others about 40% of the time, on average. Good CI users were somewhere in the middle, reporting being alone 50% of the time and with others 50% of the time. Among CI users, good and poor performers spent about the same amount of time interacting with one other person (27%), but poor performers spent less time interacting with multiple people (12%) than good performers (20%), though this difference was not significant. Poor performers also spent significantly less time around other people but not engaging (2%) than their NH peers (8%, P = 0.025).

Panel C of Figure 6.4 shows results from a question probing listening behavior in the daily life of participants and, while there are not significant differences in responses between groups, trends in listening behavior are in line with previous questions indicating less engagement among poor CI users. On average, poor CI users spend more time not actively listening (42%) than good CI users (39%) and also less time participating in conversation (27% vs. 39%). Compared to NH participants, both groups of CI users spent more time not actively listening to anything and also less time listening to a live performer or speaker.

Panels D and E of Figure 6.4 show responses to survey questions asking about the perceived difficulty of understanding speech and the level of background sound in the everyday environments of participants. Not surprisingly, NH participants reported having no difficulty understanding speech 91% of the time and speech being somewhat difficult to understand 8% of the time, on average. Poor-performing CI users reported understanding speech in their environment as "not difficult" significantly less of the time ($P = 0.008$) and "somewhat difficult" significantly more of the time ($P = 0.011$) than their NH peers. Good CI users again fell between poor CI users and NH listeners when reporting perceived difficulty in understanding speech. Participants in all groups seem to avoid environments where speech is difficult or very difficult to understand, with NH listeners, good CI users and poor CI users spending 1%, 2% and 4% in these environments, respectively. Though it should be noted that the types of environments and social settings that would be perceived as difficult, as well as the ways in which listening difficulty is perceived, varies between groups. In terms of level of background sound in everyday environments, there were no significant differences between groups. In general, all groups spent more time in environments with very little or moderate levels of background sound than environments with high or very high levels of background sound.

Question 6 of the in-the-moment survey was asked only of CI users as it probed the number of processors participants were wearing at the time of the survey. In general, unilateral CI users reported wearing their processor the vast majority of the time (95%) and bilateral users most often wore both processors (72%). However, 28% of the time, bilateral users wore only one (18%) or neither of their processors (10%). Notably, there were no significant differences in device usage between good and poor CI users, indicating that poorer hearing outcomes are not a result of less time wearing the processors themselves.

The final question of the in-the-moment survey asked CI users to report whether or not they were using assistive listening devices and NH listeners to report whether or not they were wearing headphones or earbuds at the time of the survey. Overall, the use of assistive technology among CI users was very low, with good and poor CI users utilizing assistive listening devices 2% and 8% of the time, respectively. The use of headphones or earbuds was also very rare among NH listeners, who reported wearing them just 1% of the time, on average. Poor CI users did use assistive listening devices significantly more often than NH listeners wore headphones ($P = 0.041$), indicating that, although the use of listening technology in general may be low among older people, poor CI users feel the need to seek out assistive technology to improve their hearing outcomes.

6.3.2.2 Correlations between survey responses and speech perception

To further assess our hypothesis that an association may exist between hearing outcomes in CI users and social engagement, correlations were calculated between average speech understanding (from Experiment 1) and a subset of responses from the in-the-moment survey. The *R* and *P*-values from these correlations are shown in Table 6.3.

Table 6.3

Correlations between average speech perception and responses to questions from the in-the-moment survey for CI users. Significance is depicted as: *p < 0.05.

	R	p-value
Q1 - Home	-0.542	0.02 *
Q1 - Work	0.092	0.717
Q2 - I am alone/doing something independently.	-0.323	0.191
Q2 - I am interacting with multiple people.	0.462	0.054
Q3 - I am not actively listening to anything right now.	-0.364	0.138
Q3 - I am participating in conversation.	0.49	0.039 *
Q4 - Not difficult	0.352	0.152
Q4 - Difficult	-0.412	0.089
Q5 - Very little background sound	-0.331	0.179
Q5 - A high level of background sound	0.367	0.134
Q7 - Yes	-0.512	0.03 *

There was a significant correlation between speech understanding and time spent at home, with better CI users spending less time at home and poorer users spending more time at home (R = -0.54, P = 0.02). There was not a significant correlation between speech understanding and time spent at work (R = 0.09, P = 0.72), but as many of the CI participants were retired, this association was hard to assess in this sample. The amount of time CI users spent conversing with others was also correlated with speech understanding, with better performers spending more time in conversation and poorer performers spending less time conversation (R = 0.49, P = 0.039). Though not significant, there was also a trend for better CI users to spend more time conversing with others than poorer CI users (R = 0.46, P = 0.054). The perceived difficulty of understanding speech in daily environments was not significantly correlated with speech understanding in CI users. This is not entirely surprising given the vast range of speech understanding scores associated with individual difficulty ratings, found in Experiment 1. There was not a significant correlation between the level of background sound in daily environments frequented by CI users and speech understanding, indicating that all CI users are in environments with similar levels of noise. Finally, the use of assistive listening technology (i.e. Mini Mic, Roger Pen, audio cables) was significantly correlated with speech understanding, with poorer CI users utilizing additional assistive listening technology more often than better CI users (R = -0.51, P = 0.03).

6.3.2.3 Reflective survey responses

Both the CI group and the NH group completed 3 out of 4 (75%; SD = 0.7 and 0.8) reflective surveys, on average. Survey responses were normalized for each individual participant before being included in further analysis. Since the reflective survey was exploratory in nature and more than 20 questions were included for both CI and NH participants, only a subset of the most interesting and informative questions and responses are included in this section. Response data from several of the survey questions was hard to analyze and interpret, due to the fact that the response options “not at all” and “N/A” were frequently misunderstood and conflated in participant responses.

Questions from the reflective survey probing listening behavior and coping strategies are shown in Figure 6.5. The average percentage of time each response option was selected for each question is represented by white speckled bars for the NH group, black bars for high-performing CI users and gray bars for lower-performing CI users.

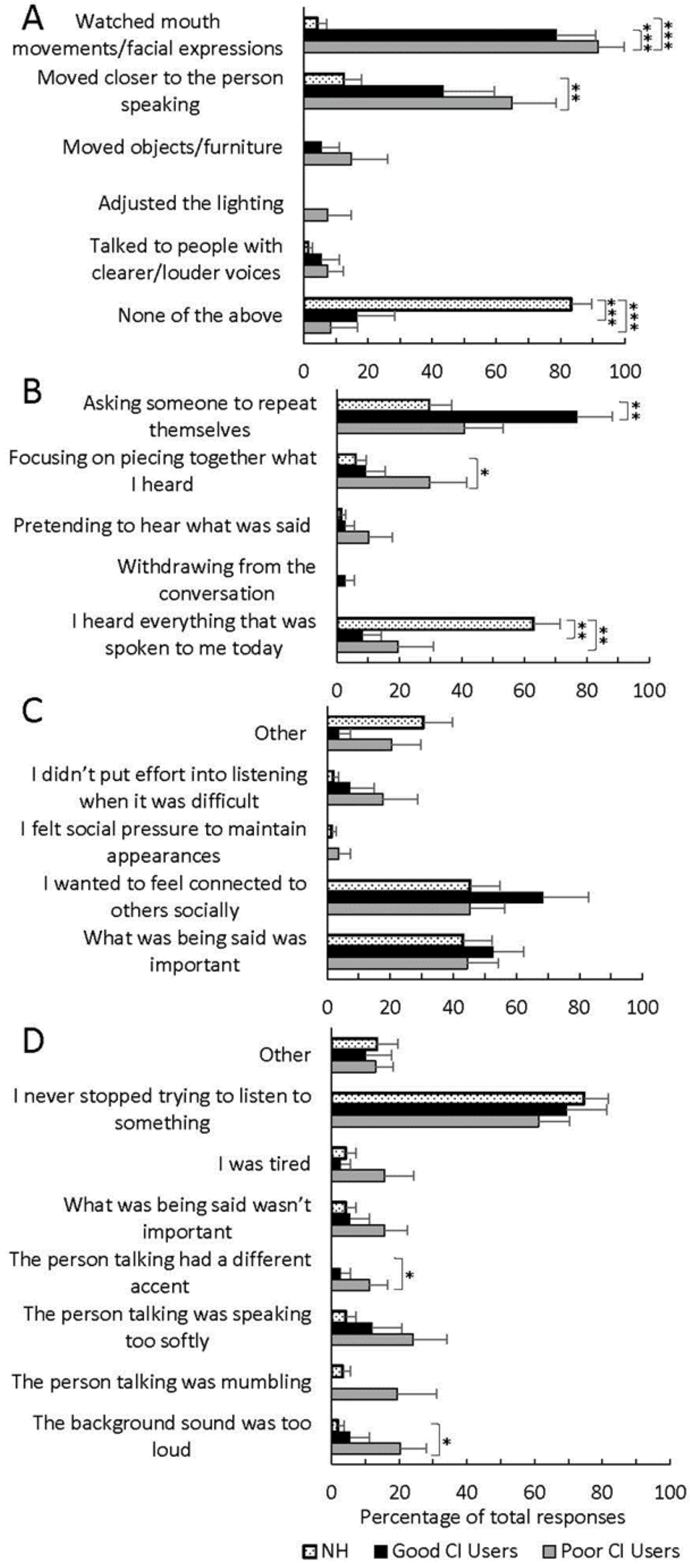

Figure 6.5

Group average responses for NH listeners, good CI users and poor CI users to four questions from the reflective survey. Error bars represent ± 1 standard error and significance is depicted as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Panel A of Figure 6.5 shows group average responses to the prompt "Today I used the following strategies to help me hear," and participants could choose multiple response options. As expected, NH listeners most often selected the option "none of the above," indicating that listening strategies are not required on a daily basis for individuals with normal hearing. In contrast, both groups of CI users reported not having to employ a listening strategy significantly less often, with good CI users *not* using some sort of coping strategy 17% of the time ($P < 0.001$) and poor CI users just 8% of the time ($P < 0.001$). The coping strategy used most often by both good and poor CI users was watching mouth movements and/or facial expressions, followed by moving closer to whoever was speaking. Both good and poor CI users used visual facial cues significantly more than NH listeners, reporting watching mouth movements and/or facial expressions 92% ($P < 0.001$) and 79% ($P < 0.001$) of the time, compared to 4% of the time by NH listeners. Poor CI users moved closer to whoever was speaking significantly more often than NH listeners ($P = 0.003$) as well, indicating this behavior was necessary 65% of the time. Not surprisingly, poor CI users also employed multiple coping strategies on a daily basis and used those strategies more often than either good CI users or NH listeners, including moving objects or furniture, adjusting the lighting or selectively talking to people with louder or clearer voices.

The use of specific listening and social strategies was addressed more directly in a subsequent survey question where participants had to select the strategy they used most often when they could not hear what was being said. Group average responses are shown in Panel B of Figure 6.5. As expected, NH listeners most often reported hearing everything that was spoken to them (the day the survey was completed) and almost always asked someone to repeat themselves if they failed to hear what was said. In contrast, both groups of CI users reported hearing everything spoken to them significantly less often, with good CI users selecting this response 8% of the time ($P = 0.001$) and poor CI users 19% of the time ($P = 0.006$). Although asking someone to repeat themselves was the most popular strategy (when necessary) for NH listeners, good CI users actually employed this strategy significantly more often (P = 0.004), having to ask for repetition 77% of the time as opposed to 30% of the time. Interestingly, poor CI users did not ask people to repeat themselves significantly more than NH listeners (P = 1.000), but instead reported focusing on piecing together what was said (30%) or pretending to hear what was said (10%) almost as often. As a result, poor CI users reported mentally trying to piece together what was said significantly more often than NH listeners (P = 0.045). In general, good CI users seemed to almost universally ask people to repeat themselves when they could not hear what was said, whereas poor CI users were much more varied in their coping strategies.

Panels C and D of Figure 6.5 show responses to questions focusing on especially difficult listening situations. Panel C shows responses for why participants decided to put effort into listening, even when it was difficult, and Panel D details the types of situations and settings that caused participants to stop putting effort into listening. The response option "other" in Panels C and D indicates that participants put effort into listening even when it was difficult or stopped putting effort into listening for a reason other than the options listed in the survey. Perhaps not surprisingly, there were no significant differences between groups in regards to what motivated participants to put effort into listening. Overall, all three groups put effort into listening because they felt what was being said was important or they wanted to feel connected to others socially. Good CI users seemed to be more motivated by wanting to feel socially connected than poor CI users or NH listeners, selecting this response 69% of the time, as opposed to 45% of the time, but these differences were not significant (P = 0.663, P = 0.477). It is also worth noting that both good and poor CI users reported not putting effort into listening only 7% and 18% of the time, respectively. This determination is reflected in responses shown in Panel D where participants reflected on why they decided to abandon listening, in a given situation. All three groups most often reported that they never stopped trying to listen to something (the day the survey was completed), with good and poor CI users selecting this response 69% and 61% of the time, respectively. Once again, poor CI users reported more reasons for abandoning listening and doing so more often than good CI users. Poor CI users gave background sound and talkers with accents as reasons they, at times, stopped putting effort into listening significantly more often than NH listeners (*P* = 0.015, *P* = 0.017). The most common reason both good and poor CI users stopped listening was because the person they were conversing with was speaking too softly, selecting this response 12% and 24% of the time, respectively.

Responses from three questions from the reflective survey probing how much CI users were frustrated by mishearing or not hearing speech in different social settings are shown in Figure 6.6. Participants were asked how much mishearing or not hearing information bothered them when 1) interacting with strangers, 2) talking in a group setting or at work, and 3) in casual conversation with friends or family. Group average responses for questions relating to social settings 1, 2 and 3 are shown in light gray, medium gray and black bars, respectively. Filled bars represent data for good CI users and unfilled bars represent data for poor CI users.

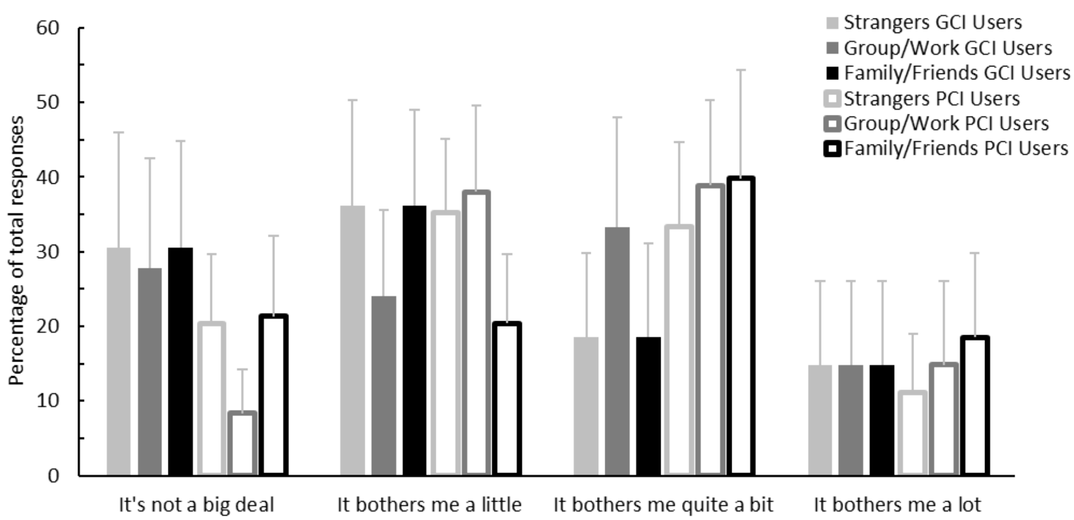

Figure 6.6

Group average responses for good and poor CI users to three questions from the reflective survey, each with the same four response options. The three questions ask "how much does mishearing or not hearing information when 1) interacting with strangers, 2) talking in a group setting or at work, and 3) in casual conversation with friends or family, bother you? Light gray bars show responses to the first question, medium gray bars show responses to the second question and black bars show responses for the third question. Filled and unfilled bars represent group averages for good and poor CI users, respectively. Error bars represent ± 1 standard error.

In general, the frustration felt from mishearing or not hearing information varies widely across individual participants and is not modulated by actual speech perception abilities. No significant difference was found between good and poor CI users in how bothered they were by their hearing loss across social settings (all P-values > 0.15). The *type* of social interaction (i.e. strangers vs. group setting vs. family) also did not influence how much mishearing or not hearing speech bothered participants (all P-values > 0.7). Overall, CI users more often reported being a little bothered or bothered quite a bit when not being able to hear in social situations, as opposed to not at all bothered or bothered a lot.

6.4 Discussion

The aim of this study was to explore social engagement and real-world listening behavior in adult CI users and age-matched NH listeners, and assess to what extent, if any, speech perception abilities relate to social decisions made in everyday life. We used measures of speech understanding, subjective ratings of listening difficulty and EMAs administered via a smartphone app, to assess social engagement and listening behavior. The main findings and their implications are discussed below.

6.4.1 Perception of listening difficulty and real world implications

Understanding the extent to which CI users struggle to understand speech and navigate social situations in everyday life is crucial to assessing the efficacy of CIs in positively influencing quality of life. When deciding whether or not to engage with others socially or allocate effort to listening, a CI user's perception of listening difficulty in a given situation will influence social behavior (Pichora-Fuller, 2016; Pichora-Fuller et al., 2016). One surprising finding of this study was how often CI users indicated that understanding speech in their daily lives was "not difficult" or only "somewhat difficult." Even poorer-performing CI users reported speech understanding in their daily lives as "not difficult" 71% of the time and "difficult" or "very difficult" less than 5% of the time. One explanation for this for this apparent lack of difficulty navigating conversation in daily life is the differences in difficulty perception between individual CI users, and also between NH listeners and CI users, as a group (see Figure 6.1). In general, CI users rate a much lower level of speech understanding as less difficult than their NH peers. So a "not difficult" rating for a CI user in daily life might correspond to understanding 80% of speech, whereas the same rating for a NH participant might indicate 100% speech understanding. However, based on results from Experiment 1, most CI users will assign a rating of "difficult" or "very difficult" when perceiving 50% of speech or less. This would seem to indicate that over 95% of the time, CI users are frequented social environments where they can understand at least 50% of what is being said. Since many of the poor CI users could not achieve 50% speech understanding on PRESTO sentences, even in quiet, this seems to suggest that CI users are relying heavily on visual or other non-auditory cues and the familiarity of speakers to understand speech in daily life.

Responses to questions from the reflective survey seem to support this assertion, where CI users indicated watching mouth movements and facial expressions in more than 75% of daily social situations (Panel A, Figure 6.5). When asked to describe the most difficult and least difficult listening situations encountered during survey days, CI users consistently mentioned the presence of visual information (face-to-face and one-on-one conversation) as promoting ease of listening and the lack of visual information (speaker was turned away, tracking speaker in a group setting) as making listening much more difficult. Familiarity with both the environment (home setting) and the talker (family member or friend) were also recurring themes reported by CI users as facilitating hearing success. For example, one participant described the least difficult listening situation encountered one day as "*talking face to face with [my] son, no background noise, in [the] kitchen*." Another participant reported an easy listening situation as when she "*was sitting across from one person talking directly to that person and that person looking at me while speaking directly to me."* Conversely, unfamiliar spaces, unfamiliar talkers and multiple talkers or background noise were reported as making listening more difficult. One participant described understanding important information given by medical personnel at a local blood drive as especially difficult: "*Giving blood today was difficult as they were playing loud music in the background in order to protect our privacy during questioning. The nurse was also very soft spoken and kept forgetting to look at me while speaking."* Another CI participant related her struggles and frustrations while trying to take part in a workout class at the gym: *"[On the] tread mill, meeting [the] gal next to me* *in class. [I] wanted to know what she said about herself. Also [the] trainer was indicating time and resistance and I strained to see her around [my] screen!"* So while the apparent lack of listening difficulty in everyday life among CI users is a positive finding, this result is likely a reflection of CI users under-rating difficulty in understanding speech when compared to NH listeners, reliance on visual information and spending most of their time in quiet, familiar environments with familiar talkers.

6.4.2 Speech understanding is related to time spent at home, in conversation and use of assistive technology in CI users

Our hypothesis that poorer-performing CI users would be less socially engaged than better-performing CI users was supported by our data, overall. Speech perception abilities of CI users were significantly correlated with the amount of time spent at home, with better CI users spending less time at home and poorer CI users spending more time at home. Speech understanding was also positively correlated with the amount of time spent conversing with others, with poorer CI users spending less time in conversation than better CI users. Across multiple questions, poor CI users reported more difficulty in everyday environments and were less socially active than their NH peers. Poor CI users also used assistive listening technology (i.e. Mini Mic, Roger Pen, etc.) more often than NH listeners used earbuds. While this is not a direct comparison, it does indicate more use of hearing technology, in general, among poor CI users than NH peers would seek out, by choice. It is also worth noting that, despite the time and energy that has gone into creating assistive listening technology, even the poor CI users were only actually using it 8% of the time, on average. Non-technological strategies, such as using visual cues, moving closer to a speaker, moving objects or furniture, asking someone to repeat themselves, or piecing together what was said, seem to be more common coping strategies among both good and poor CI users.

In contrast to the differences between poor CI users and NH peers, there were no significant differences in responses to questions probing social engagement between good CI users and NH participants. This is very encouraging, as it seems to indicate that, for some individuals, CIs enable the same level of social engagement as found in people with normal hearing. However, this conclusion should be interpreted with caution. First, the average age of participants in this study was 62 years old and, as a result, many participants were retired and less socially active overall than a younger demographic might be. Secondly, there are still significant differences in *how* good CI users and NH participants successfully navigated social situations. While NH listeners usually heard everything spoken to them and thus rarely used any listening coping strategies, good CI users relied on visual cues, asking for repetition and moving closer to the speaker to facilitate social interactions.

6.4.3 Speech understanding not related to time spent at work, time spent in noisy environments or CI usage

An unexpected finding from the EMA data was that CI users (poor or good) did not spend less time in noisy environments than NH peers, and time spent in environments with high or low levels of background noise did not correlate with speech understanding in CI users. This is surprising, given that a frequent complaint of CI users is difficulty in understanding speech in background noise. However, it is important to note that CI users spent over 70% of their time in environments with very little background sound. NH listeners also spent most of their time in quiet environments (over 65%), but it would be interesting to see how these results might change with younger adults or children who might spend more time in noisy bars or gymnasiums.

There was also no difference in the amount of time spent at work between NH and CI participants and speech understanding was not correlated with time spent working, for CI users. However, since the majority of participants were retired, we could not realistically assess how hearing impairment affects employment, in this study. Since hearing impairment and other disabilities are often associated with lower levels of employment and socioeconomic status (Emmett & Francis, 2015; He et al., 2018), this would be a relevant factor to explore in a younger population of CI users and NH individuals.

The amount of time CI users spend actually wearing their devices in everyday life and its influence on CI outcomes is still not fully understood. Clinicians often counsel patients on the importance of wearing their CIs as much as possible, especially in the weeks and months immediately following implantation. A recent study by Holder et al.

(2020) found a correlation between daily CI use and speech perception outcomes, but since the participants were experienced CI users it is hard to know if more CI usage facilitates better hearing outcomes or if poor hearing outcomes lead to less usage of the device. Interestingly, in this study, there were no significant differences in device usage between good and poor CI users. This seems to indicate that some CI users are not achieving poorer speech understanding outcomes simply because they are not wearing their CIs as much or as often as better users. Bilateral CI users wore both CIs more than 70% of the time and unilateral users wore their CI more than 90% of the time. It is interesting to note that both unilateral and bilateral users wore no CI processors about 10% of the time. This may indicate a need for listening breaks on the part of both good and poor CI users (see Panel F of Figure 6.5).

6.4.4 Individual differences in level of frustration felt due to hearing impairment in everyday life

Mishearing or not hearing information in conversation is a daily reality of CI users. Despite this social barrier, the level of frustration felt by individual CI users when this occurred was not modulated by speech understanding. There was a huge amount of variability in responses across users when asked how much mishearing or not hearing information bothered them (see Figure 6.6) that does not seem to be related to overall hearing outcomes or even the type of social setting. Some star performers are not bothered at all by missing spoken information, some poor performers are very bothered, and vice versa. Likewise, some CI users seem to be bothered more by not hearing information when speaking with friends and family, whereas others are more affected by difficulties encountered when talking to strangers. It is likely that frustration felt by CI users is more closely related to differences in personality or expectations, but this cannot be confirmed from our data, as these dimensions were not assessed. However, the lack of correlation between hearing outcomes and frustration in daily life is important for clinicians to understand when counseling patients.

6.5 Conclusions

Social engagement, listening behaviors and perception of listening difficulty were measured in 18 adult CI users and 18 age-matched NH listeners and related to speech perception abilities. The main findings can be summarized as follows:

- The perception of listening difficulty differs greatly between CI users and NH listeners, with CI users consistently rating poorer levels of speech understanding as less difficult than NH peers.
- CI users spend most of their time in listening and social situations they find to be "not difficult." The use of visual cues, such as lipreading, and social engagement with familiar people in familiar settings most likely facilitates this ease of listening in daily life.
- Poorer CI users spend more time at home and less time in conversation than higher-performing CI users or age-matched NH listeners. Thus, there is an association between hearing outcomes and social engagement in CI users.
- Good CI users did not spend more time in noisy environments or wear their CIs more often in daily life than poor CI users. This is most likely due to the fact that all CI users in this study wore their devices the vast majority of the time and also spent lots of time in quiet environments.
- The frustration experienced by individual CI users as a result of their hearing impairment is not mediated by speech perception abilities or differences in social settings. This is important for clinicians to understand when counseling patients.

CHAPTER 7: CONCLUSIONS AND FUTURE DIRECTIONS

The goal of this dissertation was to better understand the peripheral and non-peripheral factors that contribute to the variability in speech perception outcomes of post-lingual adult CI users. We examined cognitive aspects of working memory and intellectual efficiency, listening strategies related to the use of semantic context and talker differences, patterns related to social engagement and perceived listening difficulty, as well as peripheral spectral resolution, to form a more comprehensive picture of less-understood variables in the CI population related to hearing outcomes. We also compared these results to carefully age-matched NH control groups to isolate those factors that are indeed unique to the CI population from those that are simply inherent to difficult listening. Our findings offer a fresh perspective on the long-standing question of variability in outcomes of CI users, which has traditionally been dominated by peripheral explanations, and open up new avenues for future research.

7.1 Cognitive, Social, and Peripheral Sources of Variability

There is a general consensus in the field of CI research that a portion of variability in speech perception outcomes can be attributed to differences in spectral resolution. Perhaps the most convincing evidence to date was found by Gifford et al. (2018), where results showed that 25% of overall variance for speech-in-noise perception in 477 post-lingual adult CI users was associated with differences in spectral-modulation detection thresholds, a recognized measure of spectral resolution. Findings from our studies support this modest, but significant, relationship between spectral resolution and speech perception, with ripple discrimination and detection thresholds accounting for 28-41% of variance across studies. While informative, the remaining 60-75% of variance within the CI population cannot be explained by differences in peripheral factors. In addition, our results also suggest that spectral-ripple detection thresholds reflect more than just peripheral factors.

Our findings indicate that non-peripheral factors, both cognitive and social in nature, also contribute to varied outcomes seen in adult CI users. Specifically, visual, verbal working memory accounted for 12-18% of variability in performance of CI users across studies. Results from a recent study by Moberly et al. (2017) indicate that this relationship is even stronger for auditory working memory and is modulated by the difficulty of the speech material, with longer and more complex sentence perception requiring even better working memory. Findings from our final study also suggest that social engagement post-implantation is correlated with speech perception outcomes in adult CI users. The amount of time individual users spent at home was negatively correlated with hearing outcomes, accounting for 29% of variance in speech-in-noise understanding. In addition, the amount of time CI users spent conversing with others in everyday life was positively associated with speech perception outcomes, accounting for 24% of variance across individuals. While the casual relationship between these variables is yet to be determined, these results highlight the need for more research into social dynamics that could have a significant impact on overall hearing success post-implantation.

7.2 Variability Inherent to Difficulty of Auditory Task

The high variability of speech perception outcomes in CI users has been extensively studied and reported. To our knowledge, however, the studies in this thesis were the first to quantify the variability inherent in listening to degraded speech in general, versus variability that is a result of factors specific to the CI population. Our interest in exploring this topic originated while collecting data for our first study, reported in Chapter 2. Three control groups were tested, comprising of young NH listeners attempting to understand speech presented through three different vocoder simulations that degraded speech to varying degrees. Through interacting with these participants in testing sessions and monitoring their speech perception results, we noticed a huge range of abilities among young NH individuals in both how well they perceived degraded speech and how quickly they learned to interpret the degraded signal. Some participants never achieved speech understanding in the most degraded condition, even after extensive training, while others understood 50% of sentences in quiet after relatively little exposure to the stimulus. This lead us to wonder just how much of the immense variability in the CI population is unrelated to spectral resolution and other patient-specific factors.

Results from our second study, reported in Chapter 3, revealed that context and nonsense speech perception, in quiet and in noise, was significantly more variable for CI users than NH participants listening to sentences degraded by a vocoder with spectral spread characterized by 12 dB/oct filter slopes. Interestingly, ripple detection thresholds across multiple ripple rates were *not* more variable in CI users than in young or age-matched NH listeners. This was surprising, given that spectral ripple detection is thought to measure differences in spectral resolution in CI users, which should not be nearly as variable in listeners with normal hearing, as resolution was limited by the vocoder. However, ripple detection thresholds correlated significantly with working memory in age-matched NH listeners, indicating that this measure may be assessing more than just peripheral differences in CI users. Our second study looking at this question, reported in Chapter 5, again showed similar levels of variance in ripple detection thresholds for CI and age-matched NH groups, but also increased variability in working memory in the CI group. Most notably, both the age-matched and young NH listeners showed similar variance for the perception of context and nonsense sentences in quiet, when compared to CI users. In that study, the vocoder was further degraded (by using filter slopes of 8 dB/oct) to match performance of CI users on nonsense sentences and perhaps better reflect the difficulty of speech understanding characteristic of CI use. However, perception of context and nonsense sentences in noise was still more variable for CI users.

Another consistent and somewhat frustrating finding across our studies was how the extent to which vocoders could accurately simulate or reflect performance in CI users was limited. For example, while a 12 dB/oct spread vocoder approximated performance for CI users on context sentences, it appeared to overestimate performance for nonsense sentences and underestimate performance for measures of spectral resolution. So while we can be fairly confident that a significant amount of variability can be attributed to the difficulty of understanding degraded speech alone, the inability of vocoders to account for factors like differences in interpreting electric versus acoustic stimulation or the electrode-neural interface, make it hard to extrapolate these findings.

7.3 Use of Semantic Context and Cognitive Flexibility

Traditionally, and especially among clinicians, speech perception in CI users has been measured using isolated words in quiet (Nilsson et al., 1996) or highly contextual sentences in quiet and in noise (Bench et al., 1979; Killion et al., 2004; Nilsson et al., 1994; Spahr et al., 2012). The argument could be made that these types of measures approximate performance in real world situations, but they also conflate what a CI user is actually hearing with what they are *inferring*, based on semantic context. This process of "filling in the blanks" is so common among hearing-impaired listeners that there have been anecdotal reports of individuals with extended periods of hearing loss failing to qualify for CIs based on speech perception testing because they were too skilled at piecing together basic sentences during clinical testing. As an alternative to this type of testing, we developed a corpus of semantically sparse nonsense sentences that can be directly compared to performance on context sentences with the same vocabulary, length, and grammatical structure. One revealing finding from our studies using both nonsense and context sentences as measures of speech perception, was that performance for nonsense sentences in quiet was 21-27 percentage points (after RAU transformation) worse, on average, than performance on context sentences for CI users across studies. This finding implies that a significant portion of speech that is understood by CI users, may not actually be *heard* by these individuals, but is instead likely inferred based on available semantic context. Clinically, it is widely reported that adult CI users understand 60% of speech in quiet, on average. However, CI users might actually only be hearing around 40% of words in sentences, a finding which highlights both the need to improve the device itself and the importance of the human brain in facilitating the hearing success of CI users, in general.

The study presented in Chapter 5 took this nuanced approach to speech testing one step further, by measuring speech understanding of context and nonsense sentences spoken by multiple talkers and at different SNRs. Results from this study revealed the dynamic and complex nature of speech perception for CI users, showing that the amount of benefit that can be gained from semantic context is not fixed, but is modulated by differences in talkers and levels of background noise. More specifically, overall intelligibility can alter the benefit from context, with CI users seeing less benefit from context when overall performance was very poor, due to either the delivery of speech by a given talker or too much background noise. While not altogether surprising, this result is sobering in that it indicates that one of the main listening strategies employed by CI users cannot be leveraged in very difficult listening situations encountered in the real world. Results from our final study indicate that CI users may actually avoid these types of listening environments in daily life, perhaps due to the expectation that speech understanding will not be adequately facilitated. Intelligibility accounted for some of the differential benefits of semantic context in various listening situations, but not all. There seems to be somewhat of a “sweet spot” for maximal benefit gained from semantic context but that the way in which speech is delivered by different talkers modulates the benefit of context even within this intelligibility range. Though further research is needed to completely untangle the interactions between semantic context, talker differences, and SNR, it is clear that the dynamic nature of speech perception characteristic of everyday life requires rapid cognitive flexibility in CI users.

7.4 Future Directions

Findings from this dissertation have expanded on existing CI research related to the role of spectral resolution in hearing outcomes and have pushed the question of variability into cognitive and social realms. Decades of research have gone into understanding the intricate interactions between a myriad of peripheral factors and speech understanding in CI users, and the same is necessary to get a fuller understanding of cognitive and social factors affecting these outcomes. However, results from these studies give a fresh perspective on this topic and will hopefully help push the field as a whole towards exploring non-peripheral aspects of degraded speech perception. A better understanding of these factors may lead to both better prediction of CI outcomes pre-implantation, but also the development of rehabilitation and lifestyle strategies that may improve results for struggling CI users. Tools like our multi-talker nonsense and context sentence recordings may help better assess differences in speech understanding early on as well as progress over time in adult CI users, which can be used to tailor counseling by audiologists and other medical professionals. A longitudinal study would be the logical next step in understanding how peripheral, cognitive, and social factors interact from pre-implantation to activation, activation to progress at one year, and eventually determine peak performance over time. Findings from these studies also highlight both the dynamic nature of speech understanding in everyday environments and the diversity of strategies used by CI users to compensate. Now that a general understanding of these topics has been established in these studies, additional, more focused research is needed to answer more specific questions related to these topics.

BIBLIOGRAPHY

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47 Suppl 2*, S53-71. https://doi.org/10.1080/14992020802301142

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, *135*(3), 1371–1379. https://doi.org/10.1121/1.4863269

Amichetti, N. M., Atagi, E., Kong, Y.-Y., & Wingfield, A. (2018). Linguistic Context Versus Semantic Competition in Word Recognition by Younger and Older Adults With Cochlear Implants. *Ear and Hearing*, *39*(1), 101–109. https://doi.org/10.1097/AUD.0000000000000469

Anderson, C. A., Wiggins, I. M., Kitterick, P. T., & Hartley, D. E. H. (2017). Adaptive benefit of cross-modal plasticity following cochlear implantation in deaf adults. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(38), 10256–10261. https://doi.org/10.1073/pnas.1704785114

Anderson, E. S., Nelson, D. A., Kreft, H., Nelson, P. B., & Oxenham, A. J. (2011). Comparing spatial tuning curves, spectral ripple resolution, and speech perception in cochlear implant users. *The Journal of the Acoustical Society of America*, *130*(1), 364–375. https://doi.org/10.1121/1.3589255

Anderson, E. S., Oxenham, A. J., Nelson, P. B., & Nelson, D. A. (2012). Assessing the role of spectral and intensity cues in spectral ripple detection and discrimination in cochlear-implant users. *The Journal of the Acoustical Society of America*, *132*(6), 3925–3934. https://doi.org/10.1121/1.4763999

Apoux, F., & Healy, E. W. (2010). Relative contribution of off- and on-frequency spectral components of background noise to the masking of unprocessed and vocoded speech. *The Journal of the Acoustical Society of America*, *128*(4), 2075–2084. https://doi.org/10.1121/1.3478845

Araneda, R., De Volder, A. G., Deggouj, N., Philippot, P., Heeren, A., Lacroix, E., … Renier, L. (2015). Altered top-down cognitive control and auditory processing in tinnitus: evidences from auditory and visual spatial stroop. *Restorative Neurology* *and Neuroscience*, *33*(1), 67–80. https://doi.org/10.3233/RNN-140433

Arehart, K., Souza, P., Baca, R., & Kates, J. (2013). Working memory, age, and hearing loss: susceptibility to hearing aid distortion. *Ear and Hearing*, *34*(3), 251–260. https://doi.org/10.1097/AUD.0b013e318271aa5e

Aronoff, J. M., & Landsberger, D. M. (2013). The development of a modified spectral ripple test. *The Journal of the Acoustical Society of America*, *134*(2), EL217-22. https://doi.org/10.1121/1.4813802

Aschendorff, A., Kromeier, J., Klenzner, T., & Laszig, R. (2007). Quality control after insertion of the nucleus contour and contour advance electrode in adults. *Ear and Hearing*, *28*(2 Suppl), 75S-79S. https://doi.org/10.1097/AUD.0b013e318031542e

Azadpour, M., & Mckay, C. M. (2012). A Psychophysical Method for Measuring Spatial Resolution in Cochlear Implants. *Journal of the Association for Research in Otolaryngology*, *13*(1), 14–157. https://doi.org/10.1007/s10162-011-0294-z

Bacon, S. P., & Brandt, J. F. (1982). Auditory processing of vowels by normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Research*, *25*(3), 339–347. https://doi.org/10.1044/jshr.2503.339

Baskent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., … Gaudrain, E. (2016). Cognitive Compensation of Speech Perception With Hearing Impairment, Cochlear Implants, and Aging: How and to What Degree Can It Be Achieved? *Trends in Hearing*, *20*, 1–16. https://doi.org/10.1177/2331216516670279

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Beer, J., Kronenberger, W. G., Castellanos, I., Colson, B. G., Henning, S. C., & Pisoni, D. B. (2014). Executive functioning skills in preschool-age children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1521–1534. https://doi.org/10.1044/2014_JSLHR-H-13-0054

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108–112. https://doi.org/10.3109/03005367909078884

Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., & Leek,

M. R. (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *Journal of the American Academy of Audiology*, *24*(4), 293–306. https://doi.org/10.3766/jaaa.24.4.5

Bernstein, L. R., & Green, D. M. (1988). Detection of changes in spectral shape: uniform vs. non-uniform background spectra. *Hearing Research*, *34*(2), 157–165. https://doi.org/10.1016/0378-5955(88)90103-7

Beyea, J. A., McMullen, K. P., Harris, M. S., Houston, D. M., Martin, J. M., Bolster, V. A., … Moberly, A. C. (2016). Cochlear Implants in Adults: Effects of Age and Duration of Deafness on Speech Recognition. *Otology & Neurotology*, *37*(9), 1238–1245. https://doi.org/10.1097/MAO.0000000000001162

Bharadwaj, S. V, Maricle, D., Green, L., & Allman, T. (2015). Working memory, short-term memory and reading proficiency in school-age children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, *79*(10), 1647–1653. https://doi.org/10.1016/j.ijporl.2015.07.006

Bhargava, P., Gaudrain, E., & Başkent, D. (2014). Top-down restoration of speech in cochlear-implant users. *Hearing Research*, *309*, 113–123. https://doi.org/10.1016/j.heares.2013.12.003

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, *27*(1), 32–48. https://doi.org/10.1044/jshr.2701.32

Bingabr, M., Espinoza-Varas, B., & Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants. *Hearing Research*, *241*(1–2), 73–79. https://doi.org/10.1016/j.heares.2008.04.012

Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., … Whitford, L. (1996). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants. *Audiology & Neuro-Otology*, *1*(5), 293–306. https://doi.org/10.1159/000259212

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., … Lazard, D. S. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. *Audiology & Neuro-Otology*, *18*(1), 36–47. https://doi.org/10.1159/000343189

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065–1066. https://doi.org/10.1121/1.428288

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America*, *84*(1), 101–114. https://doi.org/10.1121/1.396976

Buckley, K. A., & Tobey, E. A. (2011). Cross-modal plasticity and speech perception in pre- and postlingually deaf cochlear implant users. *Ear and Hearing*, *32*(1), 2–15. https://doi.org/10.1097/AUD.0b013e3181e8534c

Calandruccio, L., & Smiljanic, R. (2012). New Sentence Recognition Materials Developed Using a Basic Non-Native English Lexicon. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1342–1355. https://doi.org/10.1044/1092-4388(2012/11-0260)

Carbonell, K. M. (2017). Reliability of individual differences in degraded speech perception. *The Journal of the Acoustical Society of America*, *142*(5), EL461. https://doi.org/10.1121/1.5010148

Carroll, L. (1883). *Through the Looking-Glass.* New York: Macmillan and Co.

Castiglione, A., Benatti, A., Velardita, C., Favaro, D., Padoan, E., Severi, D., … Martini, A. (2016). Aging, Cognitive Decline and Hearing Loss: Effects of Auditory Rehabilitation and Training with Hearing Aids and Cochlear Implants on Cognitive Function and Depression among Older Adults. *Audiology & Neuro-Otology*, *21 Suppl 1*, 21–28. https://doi.org/10.1159/000448350

Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, *106*(5), 2719–2732. https://doi.org/10.1121/1.428100

Chia, E.-M., Wang, J. J., Rochtchina, E., Cumming, R. R., Newall, P., & Mitchell, P. (2007). Hearing impairment and health-related quality of life: the Blue Mountains Hearing Study. *Ear and Hearing*, *28*(2), 187–195. https://doi.org/10.1097/AUD.0b013e31803126b6

Choi, J. E., Won, J. H., Kim, C. H., Cho, Y.-S., Hong, S. H., & Moon, I. J. (2018). Relationship between spectrotemporal modulation detection and music perception in normal-hearing, hearing-impaired, and cochlear implant listeners. *Scientific Reports, 8*(1), 800. https://doi.org/10.1038/s41598-017-17350-w

Chung, J., Chueng, K., Shipp, D., Friesen, L., Chen, J. M., Nedzelski, J. M., & Lin, V. Y. W. (2012). Unilateral multi-channel cochlear implantation results in significant improvement in quality of life. *Otology & Neurotology*, *33*(4), 566–571. https://doi.org/10.1097/MAO.0b013e3182536dc2

Claes, A. J., Van de Heyning, P., Gilles, A., Hofkens-Van den Brandt, A., Van Rompaey, V., & Mertens, G. (2018). Impaired Cognitive Functioning in Cochlear Implant Recipients Over the Age of 55 Years: A Cross-Sectional Study Using the Repeatable Battery for the Assessment of Neuropsychological Status for Hearing-Impaired Individuals (RBANS-H). *Frontiers in Neuroscience*, *12*, 580. https://doi.org/10.3389/fnins.2018.00580

Collison, E. A., Munson, B., & Carney, A. E. (2004). Relations among linguistic and cognitive skills and spoken word recognition in adults with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *47*(3), 496–508. https://doi.org/10.1044/1092-4388(2004/039)

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Conway, C. M., Deocampo, J. A., Walk, A. M., Anaya, E. M., & Pisoni, D. B. (2014). Deaf Children With Cochlear Implants Do Not Appear to Use Sentence Context to Help Recognize Spoken Words. *Journal of Speech Language and Hearing Research*, *57*(6), 2174. https://doi.org/10.1044/2014_JSLHR-L-13-0236

Cosetti, M. K., & Waltzman, S. B. (2012). Outcomes in cochlear implantation: variables affecting performance in adults and children. *Otolaryngologic Clinics of North America*, *45*(1), 155–171. https://doi.org/10.1016/j.otc.2011.08.023

Crew, J. D., Galvin, J. J., & Fu, Q.-J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *The Journal of the Acoustical Society of America*, *132*(4), EL429. https://doi.org/10.1121/1.4758770

Crowson, M. G., Semenov, Y. R., Tucci, D. L., & Niparko, J. K. (2017). Quality of Life and Cost-Effectiveness of Cochlear Implants: A Narrative Review. *Audiology & Neuro-Otology*, *22*(4–5), 236–258. https://doi.org/10.1159/000481767

Cusumano, C., Friedmann, D. R., Fang, Y., Wang, B., Roland, J. T., & Waltzman, S. B. (2017). Performance Plateau in Prelingually and Postlingually Deafened Adult Cochlear Implant Recipients. *Otology & Neurotology*, *38*(3), 334–338. https://doi.org/10.1097/MAO.0000000000001322

Dawes, P., Emsley, R., Cruickshanks, K. J., Moore, D. R., Fortnum, H., Edmondson-Jones, M., … Munro, K. J. (2015). Hearing loss and cognition: the role of hearing aids, social isolation and depression. *PloS One*, *10*(3), e0119616. https://doi.org/10.1371/journal.pone.0119616

Dingemanse, J. G., & Goedegebure, A. (2019). The Important Role of Contextual Information in Speech Perception in Cochlear Implant Users and Its Consequences in Speech Tests. *Trends in Hearing*, *23*, 1–16. https://doi.org/10.1177/2331216519838672

Dorman, M F, Loizou, P. C., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *The Journal of the Acoustical Society of America*, *104*(6), 3583–3585. https://doi.org/10.1121/1.423940

Dorman, M. F., Dankowski, K., McCandless, G., Parkin, J. L., & Smith, L. (1990). Longitudinal changes in word recognition by patients who use the ineraid cochlear implant. *Ear and Hearing*, (11), 455–459. https://doi.org/10.1097/00003446-199012000-00008

Dorman, M. F., & Wilson, B. S. (2004). The design and function of cochlear implants. *American Scientist*, *92*(5), 436–445. https://doi.org/10.1511/2004.5.436

Dowell, R. C., Mecklenburg, D. J., & Clark, G. M. (1986). Speech Recognition for 40 Patients Receiving Multichannel Cochlear Implants. *Archives of Otolaryngology--Head and Neck Surgery*, (112), 1054–1059. https://doi.org/10.1001/archotol.1986.03780100042005

Drennan, W. R., Anderson, E. S., Won, J. H., & Rubinstein, J. T. (2014). Validation of a Clinical Assessment of Spectral-Ripple Resolution for Cochlear Implant Users. *Ear and Hearing*, *35*(3), e92–e98. https://doi.org/10.1097/AUD.0000000000000009

Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in Hearing*, *21*, 1–21. https://doi.org/10.1177/2331216517744675

Eddins, D. A., & Bero, E. M. (2007). Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region. *The Journal of the Acoustical Society of America*, *121*(1), 363–372. https://doi.org/10.1121/1.2382347

Elhilali, M., Chi, T., & Shamma, S. A. (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication*, *41*(2–3), 331–348. https://doi.org/10.1016/S0167-6393(02)00134-6

Emmett, S. D., & Francis, H. W. (2015). The socioeconomic impact of hearing loss in U.S. adults. *Otology & Neurotology*, *36*(3), 545–550. https://doi.org/10.1097/MAO.0000000000000562

Finley, C. C., Holden, T. A., Holden, L. K., Whiting, B. R., Chole, R. A., Neely, G. J., … Skinner, M. W. (2008). Role of electrode placement as a contributor to variability in cochlear implant outcomes. *Otology and Neurotology*, *29*(7), 920–928. https://doi.org/10.1097/MAO.0b013e318184f492

Foo, C., Rudner, M., Rönnberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology*, *18*(7), 618–631. https://doi.org/10.3766/jaaa.18.7.8

Freyman, R. L., Griffin, A. M., & Macmillan, N. A. (2013). Priming of lowpass-filtered speech affects response bias, not sensitivity, in a bandwidth discrimination task. *The Journal of the Acoustical Society of America*, *134*(2), 1183–1192. https://doi.org/10.1121/1.4807824

Freyman, R. L., Griffin, A. M., & Oxenham, A. J. (2012). Intelligibility of whispered speech in stationary and modulated noise maskers. *The Journal of the Acoustical Society of America*, *132*(4), 2514–2523. https://doi.org/10.1121/1.4747614

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. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163. https://doi.org/10.1121/1.1381538

Fu, Q.-J. (2002). Temporal processing and speech recognition in cochlear implant users. *Neuroreport*, *13*(13), 1635–1639. https://doi.org/10.1097/00001756-200209160-00013

Fu, Q.-J., Shannon, R. V, & Wang, X. (2013). Effects of noise and spectral resolution on vowel and consonant recognition : Acoustic and electric hearing. *The Journal of the Acoustical Society of America*, *104*(6), 3586–3596. https://doi.org/10.1121/1.423941

Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, *6*(1), 19–27. https://doi.org/10.1007/s10162-004-5024-3

Füllgrabe, C., & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Frontiers in Psychology*, *7*, 1268. https://doi.org/10.3389/fpsyg.2016.01268

Gifford, R. H., Noble, J. H., Camarata, S. M., Sunderhaus, L. W., Dwyer, R. T., Dawant, B. M., … Labadie, R. F. (2018). The Relationship Between Spectral Modulation Detection and Speech Recognition: Adult Versus Pediatric Cochlear Implant Recipients. *Trends in Hearing*, *22*, 1–14. https://doi.org/10.1177/2331216518771176

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech Recognition Materials and Ceiling Effects: Considerations for Cochlear Implant Programs. *Audiology and Neurotology*, *13*(3), 193–205. https://doi.org/10.1159/000113510

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. *Journal of the American Academy of Audiology*, *24*(1), 26–36. https://doi.org/10.3766/jaaa.24.1.4

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138. https://doi.org/10.1016/0378-5955(90)90170-T

Glick, H., & Sharma, A. (2017). Cross-modal plasticity in developmental and age-related hearing loss: Clinical implications. *Hearing Research*, *343*, 191–201. https://doi.org/10.1016/j.heares.2016.08.012

Grange, J. A., Culling, J. F., Harris, N. S. L., & Bergfeld, S. (2017). Cochlear implant simulator with independent representation of the full spiral ganglion. *The Journal of the Acoustical Society of America*, *142*(5), EL484. https://doi.org/10.1121/1.5009602

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology*, *11*(2), 79–87. https://doi.org/10.3109/01050398209076203

Harris, R. W. (1991). *Speech Audiometry Materials Compact Disk*. Provo, Utah: Brigham Young University.

Hast, A., Schlücker, L., Digeser, F., Liebscher, T., & Hoppe, U. (2015). Speech Perception of Elderly Cochlear Implant Users Under Different Noise Conditions. *Otology & Neurotology*, *36*(10), 1638–1643. https://doi.org/10.1097/MAO.0000000000000883

He, P., Luo, Y., Hu, X., Gong, R., Wen, X., & Zheng, X. (2018). Association of socioeconomic status with hearing loss in Chinese working-aged adults: A population-based study. *PloS One*, *13*(3), e0195227. https://doi.org/10.1371/journal.pone.0195227

Helfer, K. S. (1997). Auditory and Auditory-Visual Perception of Clear and Conversational Speech. *Journal of Speech Language and Hearing Research*, *40*(2), 432. https://doi.org/10.1044/jslhr.4002.432

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *The Journal of the Acoustical Society of America*, *113*(5), 2861–2873. https://doi.org/10.1121/1.1561900

Henry, B. A., Turner, C. W., & Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. *The Journal of the Acoustical Society of America*, *118*(2), 1111–1121. https://doi.org/10.1121/1.1944567

Heydebrand, G., Hale, S., Potts, L., Gotter, B., & Skinner, M. (2007). Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation. *Audiology & Neuro-Otology*, *12*(4), 254–264. https://doi.org/10.1159/000101473

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., …

Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear and Hearing*. https://doi.org/10.1097/AUD.0b013e3182741aa7

Holden, L. K., Firszt, J. B., Reeder, R. M., Uchanski, R. M., Dwyer, N. Y., & Holden, T. A. (2016). Factors affecting outcomes in cochlear implant recipients implanted with a perimodiolar electrode array located in scala tympani. *Otology and Neurotology*, *37*(10), 1662–1668. https://doi.org/10.1097/MAO.0000000000001241

Holder, J. T., Dwyer, N. C., & Gifford, R. H. (2020). Duration of Processor Use Per Day Is Significantly Correlated With Speech Recognition Abilities in Adults With Cochlear Implants. *Otology & Neurotology*, *41*(2), e227–e231. https://doi.org/10.1097/MAO.0000000000002477

Holder, J. T., Reynolds, S. M., Sunderhaus, L. W., & Gifford, R. H. (2018). Current Profile of Adults Presenting for Preoperative Cochlear Implant Evaluation. *Trends in Hearing*, *22*, 1–16. https://doi.org/10.1177/2331216518755288

Houtgast, T. (1977). Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker. *The Journal of the Acoustical Society of America*, *62*(2), 409–415. https://doi.org/10.1121/1.381541

Hua, H., Johansson, B., Magnusson, L., Lyxell, B., & Ellis, R. J. (2017). Speech Recognition and Cognitive Skills in Bimodal Cochlear Implant Users. *Journal of Speech, Language and Hearing Research*, *60*(9), 2752–2763. https://doi.org/10.1044/2017_JSLHR-H-16-0276

Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., & Boisvert, I. (2018). Social Connectedness and Perceived Listening Effort in Adult Cochlear Implant Users: A Grounded Theory to Establish Content Validity for a New Patient-Reported Outcome Measure. *Ear and Hearing*, *39*(5), 922–934. https://doi.org/10.1097/AUD.0000000000000553

Humes, L. E., Kidd, G. R., & Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Frontiers in Systems Neuroscience*, *7*, 55. https://doi.org/10.3389/fnsys.2013.00055

IEEE. (1969). IEEE Recommended Practices for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*, 227–246.

Jeon, E. K., Turner, C. W., Karsten, S. A., Henry, B. A., & Gantz, B. J. (2015). Cochlear implant users' spectral ripple resolution. *The Journal of the Acoustical Society of America*, *138*(4), 2350–2358. https://doi.org/10.1121/1.4932020

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kidd Jr, G., Mason, C. R., & Gallun, F. J. (2005). Combining energetic and informational masking for speech identification. *The Journal of the Acoustical Society of America*, *118*(2), 982–992. https://doi.org/10.1121/1.1953167

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4 Pt 1), 2395–2405. https://doi.org/10.1121/1.1784440

Kim, S. Y., Kim, H.-J., Park, E.-K., Joe, J., Sim, S., & Choi, H. G. (2017). Severe hearing impairment and risk of depression: A national cohort study. *PloS One*, *12*(6), e0179973. https://doi.org/10.1371/journal.pone.0179973

Kitterick, P. T., & Ferguson, M. A. (2018). Hearing Aids and Health-Related Quality of Life in Adults With Hearing Loss. *The Journal of the American Medical Association*, *319*(21), 2225–2226. https://doi.org/10.1001/jama.2018.5567

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, *54 Suppl 2*, 3–16. https://doi.org/10.3109/14992027.2015.1020971

Landsberger, D. M., Padilla, M., Martinez, A. S., & Eisenberg, L. S. (2017). Spectral-Temporal Modulated Ripple Discrimination by Children With Cochlear Implants. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000463

Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013a). Expectation and entropy in spoken word recognition: effects of age and hearing acuity. *Experimental Aging Research*, *39*(3), 235–253. https://doi.org/10.1080/0361073X.2013.779175

Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013b). Expectation and entropy in spoken word recognition: Effects of age and hearing acuity. *Experimental Aging Research*, *39*(3), 235–253. https://doi.org/10.1080/0361073X.2013.779175

Lash, A., & Wingfield, A. (2014). A Bruner-Potter effect in audition? Spoken word recognition in adult aging. *Psychology and Aging*, *29*(4), 907–912. https://doi.org/10.1037/a0037829

Lazard, D. S., Vincent, C., Venail, F., Van de Heyning, P., Truy, E., Sterkers, O., … Blamey, P. J. (2012). Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: a new conceptual model over time. *PloS One*, *7*(11), e48739. https://doi.org/10.1371/journal.pone.0048739

Lee, J. S., Lee, D. S., Oh, S. H., Kim, C. S., Kim, J.-W., Hwang, C. H., … Lee, M. C. (2003). PET evidence of neuroplasticity in adult auditory cortex of postlingual deafness. *Journal of Nuclear Medicine*, *44*(9), 1435–1439.

Lenarz, M., Sönmez, H., Joseph, G., Büchner, A., & Lenarz, T. (2012). Long-Term Performance of Cochlear Implants in Postlingually Deafened Adults. *Otolaryngology–Head and Neck Surgery*, *147*(1), 112–118. https://doi.org/10.1177/0194599812438041

Lin, F. R., Ph, M. D. D., Yaffe, K., Xia, J., Xue, Q., Ph, D., … Simonsick, E. M. (2013). Hearing Loss and Cognitive Decline in Older Adults. *The Journal of the American Medical Association Internal Medicine*, *173*(4), 293–299. https://doi.org/10.1001/jamainternmed.2013.1868.Hearing

Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., … Health ABC Study Group. (2013). Hearing loss and cognitive decline in older adults. *The Journal of the American Medical Association Internal Medicine*, *173*(4), 293–299. https://doi.org/10.1001/jamainternmed.2013.1868

Lin, F. R., Yaffe, K., Xia, J., Xue, Q. L., Harris, T. B., Purchase-Helzner, E., … Simonsick, E. M. (2013). Hearing loss and cognitive decline in older adults. *The Journal of the American Medical Association Internal Medicine*. https://doi.org/10.1001/jamainternmed.2013.1868

Litvak, L. M., Spahr, A. J., Saoji, A. A., & Fridman, G. Y. (2007). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *The Journal of the Acoustical Society of America*, *122*(2), 982–991. https://doi.org/10.1121/1.2749413

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., …

Gitlin, N. (2017). The Lancet Commissions Dementia prevention, intervention, and care. *Lancet*, *390*, 2673–2734. https://doi.org/10.1016/S0140-6736(17)31363-6

Loughrey, D. G., Kelly, M. E., Kelley, G. A., Brennan, S., & Lawlor, B. A. (2018). Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis. *The Journal of the American Medical Association Otolaryngology-- Head & Neck Surgery*, *144*(2), 115–126. https://doi.org/10.1001/jamaoto.2017.2513

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, *42*(sup1), 49–58. https://doi.org/10.3109/14992020309074624

Lyxell, B., Andersson, J., Andersson, U., Arlinger, S., Bredberg, G., & Harder, H. (1998). Phonological representation and speech understanding with cochlear implants in deafened adults. *Scandinavian Journal of Psychology*, *39*(3), 175–179. https://doi.org/10.1111/1467-9450.393075

Mahmoud, A. F., & Ruckenstein, M. J. (2014). Speech perception performance as a function of age at implantation among postlingually deaf adult cochlear implant recipients. *Otology & Neurotology*, *35*(10), e286-291. https://doi.org/10.1097/MAO.0000000000000581

Mäki-Torkko, E. M., Vestergren, S., Harder, H., & Lyxell, B. (2015). From isolation and dependence to autonomy - expectations before and experiences after cochlear implantation in adult cochlear implant users and their significant others. *Disability and Rehabilitation*, *37*(6), 541–547. https://doi.org/10.3109/09638288.2014.935490

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477

McGettigan, C., Rosen, S., & Scott, S. K. (2008). Investigating the perception of noise-vocoded speech - an individual differences approach. *The Journal of the Acoustical Society of America*, *123*(5), 3330–3330. https://doi.org/10.1121/1.2933839

McKay, C. M., Rickard, N., & Henshall, K. (2018). Intensity Discrimination and Speech Recognition of Cochlear Implant Users. *Journal of the Association for Research in Otolaryngology*, *19*(5), 589–600. https://doi.org/10.1007/s10162-018-0675-7

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., & Dubno, J. R. (2018). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *The Laryngoscope*, *128*(4), 982–990. https://doi.org/10.1002/lary.26738

McRackan, T. R., Hand, B. N., Velozo, C. A., Dubno, J. R., & Cochlear Implant Quality of Life Development Consortium. (2019). Development of the Cochlear Implant Quality of Life Item Bank. *Ear and Hearing*, *40*(4), 1016–1024. https://doi.org/10.1097/AUD.0000000000000684

Mesnildrey, Q., & Macherey, O. (2015). Simulating the dual-peak excitation pattern produced by bipolar stimulation of a cochlear implant: Effects on speech intelligibility. *Hearing Research*, *319*, 32–47. https://doi.org/10.1016/j.heares.2014.11.001

Mick, P., Kawachi, I., & Lin, F. R. (2014). The association between hearing loss and social isolation in older adults. *Otolaryngology--Head and Neck Surgery*, *150*(3), 378–384. https://doi.org/10.1177/0194599813518021

Mikkola, T. M., Polku, H., Portegijs, E., Rantakokko, M., Tsai, L.-T., Rantanen, T., & Viljanen, A. (2016). Self-reported hearing is associated with time spent out-of-home and withdrawal from leisure activities in older community-dwelling adults. *Aging Clinical and Experimental Research*, *28*(2), 297–302. https://doi.org/10.1007/s40520-015-0389-1

Miller, J. B., Tong, Y. C., Blamey, P. J., Clark, G. M., Dowell, R. C., & Seligman, P. M. (1986). Speech processing for electrical stimulation of the auditory nerve. *In Proceedings of the International Conference on Speech Input/Output: Techniques and Applications, London*.

Moberly, A. C., Harris, M. S., Boyce, L., & Nittrouer, S. (2017). Speech Recognition in Adults With Cochlear Implants: The Effects of Working Memory, Phonological Sensitivity, and Aging. *Journal of Speech Language and Hearing Research*, *60*(4), 1046–1061. https://doi.org/10.1044/2016_JSLHR-H-16-0119

Moberly, A. C., Houston, D. M., Harris, M. S., Adunka, O. F., & Castellanos, I. (2017). Verbal working memory and inhibition-concentration in adults with cochlear implants. *Laryngoscope Investigative Otolaryngology*, *2*(5), 254–261. https://doi.org/10.1002/lio2.90

Moberly, A. C., Vasil, K. J., Wucinich, T. L., Safdar, N., Boyce, L., Roup, C., … Pisoni, D. B. (2018). How does aging affect recognition of spectrally degraded speech? *The Laryngoscope*, *128 Suppl*. https://doi.org/10.1002/lary.27457

Moon, I. S., Park, S., Kim, H.-N., Lee, W.-S., Kim, S. H., Kim, J.-H., & Choi, J. Y. (2014). Is there a deafness duration limit for cochlear implants in post-lingual deaf adults? *Acta Oto-Laryngologica*, *134*(2), 173–180. https://doi.org/10.3109/00016489.2013.850175

Moradi, S., Lidestam, B., Hällgren, M., & Rönnberg, J. (2014). Gated Auditory Speech Perception in Elderly Hearing Aid Users and Elderly Normal-Hearing Individuals: Effects of Hearing Impairment and Cognitive Capacity. *Trends in Hearing*, *18*, 1–12. https://doi.org/10.1177/2331216514545406

Naylor, G., Koelewijn, T., Zekveld, A. A., & Kramer, S. E. (2018). The Application of Pupillometry in Hearing Science to Assess Listening Effort. *Trends in Hearing*, *22*, 1–3. https://doi.org/10.1177/2331216518799437

Nechaev, D. I., Milekhina, O. N., & Supin, A. Y. (2019). Estimates of Ripple-Density Resolution Based on the Discrimination From Rippled and Nonrippled Reference Signals. *Trends in Hearing*, *23*, 2331216518824435. https://doi.org/10.1177/2331216518824435

Neher, T., Lunner, T., Hopkins, K., & Moore, B. C. J. (2012). Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L). *The Journal of the Acoustical Society of America*, *131*(4), 2561–2564. https://doi.org/10.1121/1.3689850

Nelson, D. A., Donaldson, G. S., & Kreft, H. (2008). Forward-masked spatial tuning curves in cochlear implant users. *The Journal of the Acoustical Society of America*, *123*(3), 1522–1543. https://doi.org/10.1121/1.2836786

NIDCD. (2016). Retrieved from National Institute on Deafness and other Communication Disorders (NIDCD) website: https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing

Nilsson, M, Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Nilsson, M, McCaw, V., & Soli, S. (1996). *Minimum Speech Test Battery for adult Cochlear Implant Users: User Manual*. Los Angeles: House Ear Institute.

Nittrouer, S., Caldwell-Tarr, A., Low, K. E., & Lowenstein, J. H. (2017). Verbal Working Memory in Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3342–3364. https://doi.org/10.1044/2017_JSLHR-H-16-0474

Nuesse, T., Steenken, R., Neher, T., & Holube, I. (2018). Exploring the link between cognitive abilities and speech recognition in the elderly under different listening conditions. *Frontiers in Psychology*, *9*, 678. https://doi.org/10.3389/fpsyg.2018.00678

O'Connell, B. P., Hunter, J. B., & Wanna, G. B. (2016). The importance of electrode location in cochlear implantation. *Laryngoscope Investigative Otolaryngology*, *1*(6), 169–174. https://doi.org/10.1002/lio2.42

O'Neill, E. R., Kreft, H. A., & Oxenham, A. J. (2019a). Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions. *The Journal of the Acoustical Society of America*, *146*(1), 195–210. https://doi.org/10.1121/1.5116009

O'Neill, E. R., Kreft, H. A., & Oxenham, A. J. (2019b). Speech Perception with Spectrally Non-overlapping Maskers as Measure of Spectral Resolution in Cochlear Implant Users. *Journal of the Association for Research in Otolaryngology*, *20*(2), 151–167. https://doi.org/10.1007/s10162-018-00702-2

Oxenham, A. J., & Kreft, H. A. (2014). Speech Perception in Tones and Noise via Cochlear Implants Reveals Influence of Spectral Resolution on Temporal Processing. *Trends in Hearing*, *18*, 1–14. https://doi.org/10.1177/2331216514553783

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of* *the Acoustical Society of America*, *59*(3), 640–654. https://doi.org/10.1121/1.380914

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, *72*(6), 1788–1803. https://doi.org/10.1121/1.388652

Pichora-Fuller, M. K. (2016). How Social Psychological Factors May Modulate Auditory and Cognitive Functioning During Listening. *Ear and Hearing*, *37 Suppl 1*, 92S-100S. https://doi.org/10.1097/AUD.0000000000000323

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., … Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pronk, M., Deeg, D. J. H., Smits, C., van Tilburg, T. G., Kuik, D. J., Festen, J. M., & Kramer, S. E. (2011). Prospective effects of hearing status on loneliness and depression in older persons: identification of subgroups. *International Journal of Audiology*, *50*(12), 887–896. https://doi.org/10.3109/14992027.2011.599871

Raven, J., Raven, J., & Court, J. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. NCS Pearson, Inc.

Ray, J., Popli, G., & Fell, G. (2018). Association of Cognition and Age-Related Hearing Impairment in the English Longitudinal Study of Ageing. *The Journal of the American Medical Association Otolaryngology-- Head & Neck Surgery*, *144*(10), 876–882. https://doi.org/10.1001/jamaoto.2018.1656

Resnick, J. M., Horn, D. L., Noble, A. R., & Rubinstein, J. T. (2020). Spectral aliasing in an acoustic spectral ripple discrimination task. *The Journal of the Acoustical Society of America*, *147*(2), 1054–1058. https://doi.org/10.1121/10.0000608

Rimikis, S., Smiljanic, R., & Calandruccio, L. (2013). Nonnative English Speaker Performance on the Basic English Lexicon (BEL) Sentences. *Journal of Speech, Language, and Hearing Research*, *56*(3), 792–804. https://doi.org/10.1044/1092-4388(2012/12-0178)

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., … Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(31), 1–17. https://doi.org/10.3389/fnsys.2013.00031

Rothauser, E. H., Chapman, W. D., Guttman, N., Silbiger, H. R., Hecker, M. H. L., Urbanek, G. E., … Weinstock, M. (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225–246. https://doi.org/10.1109/TAU.1969.1162058

Rouger, J., Lagleyre, S., Démonet, J.-F., Fraysse, B., Deguine, O., & Barone, P. (2012). Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Human Brain Mapping*, *33*(8), 1929–1940. https://doi.org/10.1002/hbm.21331

Rudner, M., Foo, C., Rönnberg, J., & Lunner, T. (2007). Phonological mismatch makes aided speech recognition in noise cognitively taxing. *Ear and Hearing*, *28*(6), 879–892. https://doi.org/10.1097/AUD.0b013e3181576c9c

Ruffin, C. V., Tyler, R. S., Witt, S. A., Dunn, C. C., Gantz, B. J., & Rubinstein, J. T. (2007). Long-Term Performance of Clarion 1.0 Cochlear Implant Users. *The Laryngoscope*, *117*(7), 1183–1190. https://doi.org/10.1097/MLG.0b013e318058191a

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PloS One*, *9*(1), e86980. https://doi.org/10.1371/journal.pone.0086980

Sandmann, P., Dillier, N., Eichele, T., Meyer, M., Kegel, A., Pascual-Marqui, R. D., … Debener, S. (2012). Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. *Brain*, *135*(Pt 2), 555–568. https://doi.org/10.1093/brain/awr329

Saoji, A. A., & Eddins, D. A. (2007). Spectral modulation masking patterns reveal tuning to spectral envelope frequency. *The Journal of the Acoustical Society of America*, *122*(2), 1004–1013. https://doi.org/10.1121/1.2751267

Saoji, A. A., Litvak, L., Spahr, A. J., & Eddins, D. A. (2009a). Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners. *The Journal of the Acoustical Society of America*, *126*(3), 955–958. https://doi.org/10.1121/1.3179670

Saoji, A. A., Litvak, L., Spahr, A. J., & Eddins, D. A. (2009b). Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3179670

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230–1240. https://doi.org/10.1044/1092-4388(2009/08-0111)

Shafiro, V., Sheft, S., Norris, M., Spanos, G., Radasevich, K., Formsma, P., & Gygi, B. (2016). Toward a Nonspeech Test of Auditory Cognition: Semantic Context Effects in Environmental Sound Identification in Adults of Varying Age and Hearing Abilities. *PloS One*, *11*(11), e0167030. https://doi.org/10.1371/journal.pone.0167030

Shankar, A., McMunn, A., Demakakos, P., Hamer, M., & Steptoe, A. (2017). Social isolation and loneliness: Prospective associations with functional status in older adults. *Health Psychology*, *36*(2), 179–187. https://doi.org/10.1037/hea0000437

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Signoret, C., Johnsrude, I., Classon, E., & Rudner, M. (2018). Combined effects of form- and meaning-based predictability on perceived clarity of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 277–285. https://doi.org/10.1037/xhp0000442

Skinner, M. W., Holden, T. A., Whiting, B. R., Voie, A. H., Brunsden, B., Neely, J. G., … Finley, C. C. (2007). In vivo estimates of the position of advanced bionics electrode arrays in the human cochlea. *The Annals of Otology, Rhinology & Laryngology. Supplement*, *197*, 2–24. https://doi.org/10.1177/000348940711600401

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., … Cook, S. (2012). Development and Validation of the AzBio Sentence Lists. *Ear and Hearing*, *33*(1), 112–117. https://doi.org/10.1097/AUD.0b013e31822c2549

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201. https://doi.org/10.2307/1412107

Stone, M. A., Füllgrabe, C., Mackinnon, R. C., & Moore, B. C. J. (2011). The importance for speech intelligibility of random fluctuations in “steady” background noise. *The Journal of the Acoustical Society of America*, *130*(5), 2874–2881. https://doi.org/10.1121/1.3641371

Stone, M. A., Füllgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, *132*(1), 317. https://doi.org/10.1121/1.4725766

Stone, M. A., & Moore, B. C. J. (2014). On the near non-existence of “pure” energetic masking release for speech. *The Journal of the Acoustical Society of America*, *135*(4), 1967–1977. https://doi.org/10.1121/1.4868392

Strelnikov, K., Rouger, J., Demonet, J.-F., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2013). Visual activity predicts auditory recovery from deafness after adult cochlear implantation. *Brain*, *136*(Pt 12), 3682–3695. https://doi.org/10.1093/brain/awt274

Stubbs, B., Vancampfort, D., Veronese, N., Kahl, K. G., Mitchell, A. J., Lin, P.-Y., … Koyanagi, A. (2017). Depression and physical health multimorbidity: primary data and country-wide meta-analysis of population data from 190 593 people across 43 low- and middle-income countries. *Psychological Medicine*, *47*(12), 2107–2117. https://doi.org/10.1017/S0033291717000551

Studebaker, G. A. (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*, *28*(3), 455–462. https://doi.org/10.1044/jshr.2803.455

Supin, A. Y., Popov, V. V., Milekhina, O. N., & Tarakanov, M. B. (1994). Frequency resolving power measured by rippled noise. *Hearing Research*, *78*(1), 31–40. https://doi.org/10.1016/0378-5955(94)90041-8

Taljaard, D. S., Olaithe, M., Brennan-Jones, C. G., Eikelboom, R. H., & Bucks, R. S. (2016). The relationship between hearing impairment and cognitive function: a meta-analysis in adults. *Clinical Otolaryngology : Official Journal of ENT-UK ; Official Journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery*, *41*(6), 718–729. https://doi.org/10.1111/coa.12607

Tao, D., Deng, R., Jiang, Y., Galvin, J. J., Fu, Q.-J., & Chen, B. (2014). Contribution of auditory working memory to speech understanding in mandarin-speaking cochlear implant users. *PloS One*, *9*(6), e99096. https://doi.org/10.1371/journal.pone.0099096

Teoh, S. W., Pisoni, D. B., & Miyamoto, R. T. (2004). Cochlear implantation in adults with prelingual deafness. Part I. Clinical results. *The Laryngoscope*, *114*(9), 1536–1540. https://doi.org/10.1097/00005537-200409000-00006

Thorndike, E. L., & Lorge, I. (1945). The teacher's word book of 30,000 words. *Journal of Consulting Psychology*, Vol. 9, pp. 306–307. https://doi.org/10.1037/h0050354

Tobey, E. A., Thal, D., Niparko, J. K., Eisenberg, L. S., Quittner, A. L., Wang, N.-Y., & CDaCI Investigative Team. (2013). Influence of implantation age on school-age language performance in pediatric cochlear implant users. *International Journal of Audiology*, *52*(4), 219–229. https://doi.org/10.3109/14992027.2012.759666

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., & Henry, B. A. (2004). Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. *The Journal of the Acoustical Society of America*, *115*(4), 1729–1735. https://doi.org/10.1121/1.1687425

Van Der Marel, K. S., Briaire, J. J., Verbist, B. M., Muurling, T. J., & Frijns, J. H. M. (2015). The influence of cochlear implant electrode position on performance. *Audiology and Neurotology*, *20*(3), 202–211. https://doi.org/10.1159/000377616

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *The Journal of the Acoustical Society of America*, *82*(4), 1152–1161. https://doi.org/10.1121/1.395251

Wagner, A., Pals, C., de Blecourt, C. M., Sarampalis, A., & Başkent, D. (2016). Does signal degradation affect top–down processing of speech? *Advances in Experimental Medicine and Biology*, *894*, 297–306. https://doi.org/10.1007/978-3-319-25474-6_31

Wanna, G. B., Noble, J. H., Carlson, M. L., Gifford, R. H., Dietrich, M. S., Haynes, D. S., … Labadie, R. F. (2014). Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes. *The Laryngoscope*, *124 Suppl*, S1–S7. https://doi.org/10.1002/lary.24728

Wechsler, D. (2011). Wechsler Abbreviated Scale of Intelligence, Second edition. *Pearson*, (Bloomington, MN).

Whitmal, N. A., Poissant, S. F., Freyman, R. L., & Helfer, K. S. (2007). Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *The Journal of the Acoustical Society of America*, *122*(May 2013), 2376–2388. https://doi.org/10.1121/1.2773993

Winn, M. B. (2016). Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants. *Trends in Hearing*, *20*, 1–17. https://doi.org/10.1177/2331216516669723

Won, J. H., Clinard, C. G., Kwon, S., Dasika, V. K., Nie, K., Drennan, W. R., … Rubinstein, J. T. (2011). Relationship between behavioral and physiological spectral-ripple discrimination. *Journal of the Association for Research in Otolaryngology*, *12*(3), 375–393. https://doi.org/10.1007/s10162-011-0257-4

Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, *8*(3), 384–392. https://doi.org/10.1007/s10162-007-0085-8

Won, J. H., Moon, I. J., Jin, S., Park, H., Woo, J., Cho, Y.-S., … Hong, S. H. (2015). Spectrotemporal Modulation Detection and Speech Perception by Cochlear Implant Users. *PloS One*, *10*(10), e0140920. https://doi.org/10.1371/journal.pone.0140920

Yamada, Y., & Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Research*, *1130*(1), 167–180. https://doi.org/10.1016/j.brainres.2006.10.052

Yeend, I., Beach, E. F., & Sharma, M. (2019). Working Memory and Extended High-Frequency Hearing in Adults: Diagnostic Predictors of Speech-in-Noise Perception. *Ear and Hearing*, *40*(3), 458–467. https://doi.org/10.1097/AUD.0000000000000640

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480–490. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Heslenfeld, D. J., & Rönnberg, J. (2012). Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility. *Brain and Language*, *122*(2), 103–113. https://doi.org/10.1016/j.bandl.2012.05.006

Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., & Feng, H. (2008). Cochlear implants:

system design, integration, and evaluation. *IEEE Reviews in Biomedical Engineering*, Vol. 1, pp. 115–142. https://doi.org/10.1109/RBME.2008.2008250

Zhao, E. E., Dornhoffer, J. R., Loftus, C., Nguyen, S. A., Meyer, T. A., Dubno, J. R., & McRackan, T. R. (2020). Association of Patient-Related Factors With Adult Cochlear Implant Speech Recognition Outcomes. *The Journal of the American Medical Association Otolaryngology–Head & Neck Surgery*. https://doi.org/10.1001/jamaoto.2020.0662

Zhou, N. (2017). Deactivating stimulation sites based on low-rate thresholds improves spectral ripple and speech reception thresholds in cochlear implant users. *The Journal of the Acoustical Society of America*, *141*(3), EL243. https://doi.org/10.1121/1.4977235

Zhou, X., Seghouane, A.-K., Shah, A., Innes-Brown, H., Cross, W., Litovsky, R., & McKay, C. M. (2018). Cortical Speech Processing in Postlingually Deaf Adult Cochlear Implant Users, as Revealed by Functional Near-Infrared Spectroscopy. *Trends in Hearing*, *22*, 1–18. https://doi.org/10.1177/2331216518786850

APPENDICES

Appendix A

Context Sentences	Key	Participant responses (Synonym)
The milk and cheese smelled ___________.	horrible	stinky
My strong ___________ carried my brother.	father	dad
The family ____________ in an expensive restaurant.	ate	dined
This new computer is _____________ useful.	quite	very
The ____________ broke a glass window.	baseball	ball
My grandmother ___________ a chocolate cake.	baked	made
The party game was __________ easy.	really	pretty
That ___________ sells cheap clothes.	store	shop
The kitchen garbage smelled _____________.	terrible	horrible, bad
The _____________ write simple problems.	professor	teacher
The trees _____________ sweet apples.	grow	produced (2)
Some _____________ drink black coffee.	people	folks
Her grandparents are serious and sometimes _____________.	cruel	stern
The last year was ___________ and peaceful.	calm	quiet, relaxing
The popular club is often ___________.	full	crowded
My neighbor sings country ___________.	songs	music
The tiny fly ___________ everyone.	bothered	annoys
The fish swam ___________ in the lake.	slowly	quickly
A ___________ country is exciting to visit.	foreign	new
The church ___________ inspired the community.	group	congregation
The news was on TV ___________ morning.	every	each
The wind damaged the ___________ boat.	tiny	little
The stars ___________ the night sky.	lit	illuminated
Her apartment was near the private ___________.	school	campus
The ___________ and trees look beautiful.	plants	flowers
The restaurant ___________ red wine.	sells	served, serves
She drove the bus down the ___________.	street	road
The spicy carrots were her favorite ___________.	dish	food
The strange animal scared the ___________.	baby	kid
My brother sleeps until ___________ morning.	late	mid
The small boy seemed ___________.	sad	scared (2)
Our ___________ drinks orange juice.	mother	mom
A ___________ kitten climbed over the fence.	little	young
The chicken soup was a ___________ meal.	tasty	good

That ___________ animal is cute but dangerous.	tiny	small
A summer vacation is ___________ relaxing.	always	very, quite
The ___________ learned about earth science.	class	students (2)
The boss ___________ the lazy waiter.	fired	scolds
Some writers ___________ interesting stories.	tell	write
They played fast ___________ on the radio.	music	songs
The best worker went on the ___________.	trip	vacation
Their ___________ son danced well.	famous	talented
The ___________ destroyed some plants.	rain	storm
Those cute animals ___________ the plants.	chewed	eat (2)
His grandma and grandpa helped the ___________.	kids	child
The big ___________ felt empty.	room	house
The honest mother is loving and ___________.	nice	caring
The lonely ducks swims in the ___________.	lake	pond (2)
Her uncle ___________ quietly for the answer.	waits	listens
The instructor ___________ books to her class.	gave	brought, brings
These colleges offer many ___________.	courses	classes
The ___________ used soft brushes.	painter	artist
The Christmas ___________ interested my son.	show	play
The ___________ lays tiny eggs.	bird	chicken
The chef ___________ pasta every day.	cooks	made
Her youngest son was always ___________.	upset	crying
They ___________ salty meat in the pan.	roasted	cooked (2)
That ___________ carried six bags.	boy	man (2)
A lonely person is usually ___________.	unhappy	sad
They loved the French ___________ and dessert.	food	meal, dinner
The party ___________ for three hours.	lasted	continued
The teacher ___________ homework daily.	gave	assigned (2)
The old shirt was warm and ___________.	soft	fuzzy, comfortable
That fast ___________ chased a mouse.	kitten	cat (2)
The student studies in the quiet ___________.	room	library
The ___________ serves dinner and drinks.	restaurant	diner
My cousin ___________ my a birthday cake.	baked	made
The tropical ___________ had many trees.	forest	rainforest
He ___________ loudly in the crowded room.	screamed	yelled
The green ___________ look healthy.	plants	grass
The artist took a ___________ picture.	beautiful	pretty
The performer worked for little ___________.	money	pay
The ___________ looked perfectly calm.	ocean	sea
The boss tells horrible ___________.	jokes	stories (2)

The Catholic priest sang ___________.	songs	hymns
The slow computer had many ___________.	problems	viruses
That new book is ___________ helpful.	really	very
The ___________ hung above the door.	picture	sign
That goat ___________ in the deep hole.	fell	jumped (2)
The talented artist drew a ___________.	picture	portrait
The jazz ___________ sounded great.	singer	concert, ensemble
The lesson seemed too ___________.	difficult	complicated
The bird flew ___________ the sea.	over	above
The company ___________ foreign cars.	buys	purchased
The young woman is ___________ smart.	very	incredibly
A pretty boat went down the ___________.	river	stream
The two friend ___________ up the mountain.	hiked	climbed
The nurses work hard every ___________.	day	shift
The nurses work hard every ___________.	day	shift
The couple sang the ___________ well.	song	anthem
The hot sun ___________ the pool.	warmed	heated
The divorced couple sat at the ___________.	table	desk
Our business paid for the daily ___________.	newspaper	paper
The vegetables grew in the green ___________.	garden	yard
The ___________ wrote thirty books.	author	writer
The helpful nanny cleaned the ___________.	house	room (2)
The ___________ lady give advice.	kind	nice
The team ___________ goals easily.	scored	shoots
Context Sentences	**Key**	**(Same category)**
The __________ loves sweet candy.	girl	child, boy
The park opens in ___________ months.	eleven	two
The worried adult _____________ home.	ran	walks, came
The _____________ bears eat fruit.	brown	black (2)
The woman met her favorite _____________.	actor	celebrity
The glass _____________ broke in the kitchen.	dish	cup
The twins live with their _____________.	grandparents	grandmother
The sun sets in the late ___________.	afternoon	evening (2)
The hungry ___________ made a sandwich.	girl	boy, man
Her black ___________ looked funny.	sweater	shirt
His ___________ tell boring stories.	parents	grandparents
Her thoughtful ___________ sent flowers.	boyfriend	husband (2)
The kind ___________ helps strangers.	girl	person, woman
The fat pig ___________ on the floor.	slept	lay, laid

The talented ___________ received an award.	writer	musician
They bought three ___________ cars.	blue	red
The ___________ juice spilled on the floor.	grape	cranberry
Our ___________ drinks orange juice.	mother	father
The snowman had ___________ green gloves.	two	four
The dark cloud covered the ___________.	sky	sun
The meeting starts in ___________ minutes.	twenty	five, ten
That ___________ test was really difficult.	English	math (2)
The three ___________ did their math homework.	cousins	kids, siblings
His grandma and grandpa helped the ___________.	kids	son
The happy ___________ laugh at the story.	children	teenagers, people
They took a ___________ picture every year.	school	family (2)
The shopper bought ___________ things.	many	ten
My grandmother read the ___________ quickly.	newspaper	book (2)
The ___________ team practiced at night.	tennis	soccer
The fans watched ___________ games.	football	baseball
Her ___________ watched movies with another girl.	boyfriend	friend
The artist studies ___________ and French.	Italian	spanish
The ___________ ring fit her finger.	gold	diamond
The ___________ sandwich came with salad.	chicken	turkey
The ___________ sandwich came with salad.	chicken	beef
The dedicated ___________ help patients.	nurses	doctors, surgeon
My ___________ made wooden chairs.	grandfather	father
Their oldest ___________ plays with toys.	daughter	child, son
A little ___________ runs through the forest.	rabbit	fox
The player ___________ the soccer ball.	kicked	hits
Her right arm and ___________ were broken.	leg	hip
The hungry teenagers eat ___________.	snacks	pizza
Context Sentences	**Key**	**(Antonyms)**
The wooden door was hard to ___________.	close	open (2)
Our team practices every ___________.	night	day
The ___________ night was comfortable and relaxing.	cool	warm
The customers ___________ black tea.	hate	love, loves
The city bus is usually ___________.	early	late
The show ___________ early today.	ended	started
The worker hurt his ___________ hand.	left	right (2)
Nonsense Sentences	**Key**	**(Synonyms)**
The weekend is expensive and ___________ sad.	not	never
A child ___________ down the office.	chased	ran
The hot horse is ___________ worried.	too	very

A dangerous weekend is _____________ thirsty.	really	very
The painter cut milk ___________ day.	every	all
The ___________ travelled some room.	trip	tour
The math store was ___________ relaxing.	always	very
The sweet group is ___________ slow.	very	really
Nonsense Sentences	**Key**	**(Same category)**
My ___________ ran every sun.	girl	son
The ___________ and cloud screamed their mother.	sister	child, boy
They wrote cake in the ___________ police.	green	blue