University of Memphis

University of Memphis Digital Commons

Electronic Theses and Dissertations

11-17-2017

Predicting Speech Recognition using the Speech Intelligibility Index (SII) for Cochlear Implant Users and Listeners with Normal Hearing

Sung Min Lee

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

Recommended Citation

Lee, Sung Min, "Predicting Speech Recognition using the Speech Intelligibility Index (SII) for Cochlear Implant Users and Listeners with Normal Hearing" (2017). *Electronic Theses and Dissertations*. 1745.
https://digitalcommons.memphis.edu/etd/1745

This Dissertation is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.

PREDICTING SPEECH RECOGNITION USING THE SPEECH INTELLIGIBILITY INDEX (SII) FOR COCHLEAR IMPLANT USERS AND LISTENERS WITH NORMAL HEARING

by

Sungmin Lee

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Communication Sciences and Disorders

The University of Memphis

December 2017

Dedication

This dissertation is dedicated to my parents who provide all sorts of tangible and intangible support.

Acknowledgements

In acknowledging the various kinds of help and support I have received in preparation of my dissertation, I would first of all like to express my gratitude to my mentor, Dr. Lisa Lucks Mendel, whose perceptive criticism, words of encouragement, and willing assistance helped bring the project to a successful conclusion. Your consideration was a great help and encouragement to overcome difficulties during my doctoral program. It was the luckiest thing to be your student in my life.

I am also grateful to my dissertation committee. Invaluable help was furnished by Dr. Gavin Bidelman, Dr. Eugene Buder and Professor George Relyea. Your brilliant comments and insightful suggestions have contributed to improvements in the study. Not only your advice on research, but also your teaching has been a priceless lesson.

I would like to thank Dr. Chhayakant Patro and Monique Pousson. We have worked closely together on several projects in the SPAL. I believe we were a great team under Lisa Lucks Mendel for the laboratory projects.

I want to take this opportunity to thank my family, my parents, Jun-nyung Lee and Jungsook Back, and my sister, Arem Lee. They have assisted me in innumerable ways; whatever I might say here cannot do full justice to the extent and the value of their contribution. Your devout prayer for me was what sustained me so far.

I regret that I cannot acknowledge all the people who have given me help and encouragement over the period in which this dissertation was written. I will never forget your favour and will be eternally grateful to you for your help.

This study was financially supported in whole by a Student Investigator Research Grant supported by the American Academy of Audiology and American Academy of Audiology Foundation.

Preface

Chapter 2 is accepted for publication in *The Journal of the Acoustical Society of America*. Its authors are Sungmin Lee and Lisa Lucks Mendel.

Chapter 3 is in submission as a manuscript to *The Journal of the American Academy of Audiology.* Its authors are Sungmin Lee and Lisa Lucks Mendel.

Abstract

Lee, Sungmin. PhD. The University of Memphis. December 2017. Predicting Speech Recognition Using the Speech Intelligibility Index (SII) for Cochlear Implant Users and Listeners with Normal Hearing. Major Professor: Lisa Lucks Mendel, PhD.

Although the AzBio test is well validated, has effective standardization data available, and is highly recommended for Cochlear Implant (CI) evaluation, no attempt has been made to derive a Frequency Importance Function (FIF) for its stimuli. In the first phase of this dissertation, we derived FIFs for the AzBio sentence lists using listeners with normal hearing. Traditional procedures described in studies by Studebaker and Sherbecoe (1991) were applied for this purpose. Fifteen participants with normal hearing listened to a large number of AzBio sentences that were high- and low-pass filtered under speech-spectrum shaped noise at various signal-to-noise ratios. Frequency weights for the AzBio sentences were greatest in the 1.5 to 2 kHz frequency regions as is the case with other speech materials. A cross-procedure comparison was conducted between the traditional procedure (Studebaker and Sherbecoe, 1991) and the nonlinear optimization procedure (Kates, 2013). Consecutive data analyses provided speech recognition scores for the AzBio sentences in relation to the Speech Intelligibility Index (SII). Our findings provided empirically derived FIFs for the AzBio test that can be used for future studies. It is anticipated that the accuracy of predicting SIIs for CI patients will be improved when using these derived FIFs for the AzBio test.

In the second study, the SII for CI recipients was calculated to investigate whether the SII is an effective tool for predicting speech perception performance in a CI population. A total of fifteen CI adults participated. The FIFs obtained from the first study were used to compute the SII in these CI listeners. The obtained SIIs were compared with predicted SIIs using a transfer function curve derived from the first study. Due to the considerably poor hearing and large individual variability in performance in the CI population, the SII failed to predict speech perception performance for this CI group. Other predictive factors that have been associated with speech perception performance were also examined using a multiple regression analysis. Gap detection thresholds and duration of deafness were found to be significant predictive factors. These predictor factors and SIIs are discussed in relation to speech perception performance in CI users.

Table of Contents

List of Tables

List of Figures

Chapter 1

GENERAL INTRODUCTION

In the mid-20th century, the Articulation Index (AI) was developed by engineers at Bell Telephone Laboratories for the purpose of quantitatively evaluating speech intelligibility transmitted via their prototypes of telecommunication devices (French and Steinberg, 1947; Kryter, 1962). The AI was effective and efficient at predicting speech recognition by using this established mathematical concept that replaced effort and time in the actual testing procedure. After about half a century, the model was reviewed and updated by the American National Standards Institute (ANSI), and renamed the Speech Intelligibility Index (SII) (ANSI, 1997).

The SII is a value that quantifies the proportion of speech information available to listeners. The SII ranges between 0 (no speech information is available) and 1 (total speech information is available). Two critical components for frequency bands, audibility function and frequency importance function (FIF), are taken into account in the SII computational procedure. The audibility function is accounted for by the amount of speech energy available to the listener. Thus, levels of the speech spectrum, noise spectrum and the listeners' hearing thresholds are considered in the audibility calculation. The FIF refers to the importance of each frequency band and its relative weight for contributing to speech perception. The specific patterns of FIFs vary depending on the speech stimuli and the procedure used for deriving FIFs (Studebaker and Sherbecoe, 1991; DePaolis *et al.*, 1996). Most evidence from previous studies indicates that the most important frequency region is around 2 kHz where the greatest amount of speech information is centered (Studebaker and Sherbecoe, 1991; DePaolis *et al.*, 1996; Amlani *et al.*, 2002)

Cochlear implants (CIs) are prosthetic medical devices that electrically stimulate auditory nerve fibers to transmit acoustic outputs. Since their first approval by the Food and Drug Administration (FDA) in 1984, CIs have evolved and become a very successful option for people who are deaf or who have severe-to-profound sensorineural hearing loss. With advances in CI technology and continued success in speech perception performance for CI recipients, the FDA has lessened its eligibility criteria and extended accessibility to children as young as 12 months of age. In addition, the criteria for adult CI candidates have been expanded to include individuals with moderate-to-severe sensorineural hearing losses. These improvements in CI technology have restored audibility for many individuals with significant hearing loss. However, on closer inspection, CIs have not always provided satisfaction for all candidates, because there are still large numbers of individuals who have not benefited from CIs as much as others. This issue of variability in speech and language outcomes in the CI population is regarded as the most challenging problem that needs to be addressed in this population (Faulkner and Pisoni, 2013). These individual differences are not likely to be accounted for by a single factor, but in fact multiple parameters in different domains are likely interacting with each other, either positively or negatively, affecting these outcomes (Faulkner and Pisoni, 2013). In addition, many studies on CIs have investigated variables that contribute to individual differences in speech perception performance (Gordon *et al.*, 2000; Geers *et al.*, 2011). The factors scrutinized in many studies are limited to CI patients' demographics, surgical outcomes, auditory processing ability, and cognitive function. These contributing factors are also thought to be predictive factors, because both contributing and predictive factors can significantly affect speech perception outcomes associated with CIs.

The SII is used widely in hearing aid research, yet CI researchers pay scant attention to the SII as a predictive factor for estimating CI users' speech perception. Thus, the primary purpose of this dissertation was to explore the feasibility of the SII model for predicting speech perception performance for CI users.

This dissertation consists of two studies. In order to apply the SII model for CI evaluation, the FIF for the speech perception material that is often used with CI users needed to be known. Thus, in the first study, we derived FIFs for the AzBio sentences using a traditional procedure. The AzBio test was chosen because it is part of the Revised Minimum Speech Test Battery (MSTB) (Spahr *et al.*, 2012; Spahr *et al.*, 2014) which is considered as a standard test protocol in the field of CI evaluation. The derived FIF for the AzBio sentences was used to compute SIIs for CI users in the second phase of the dissertation.

The aim of the second study was to investigate whether the SII can be considered a useful tool for successfully predicting speech perception outcomes for CI patients. We hypothesized that conventional SII values could not predict speech perception performance for CI adults with a high degree of accuracy due to the enormous variability in speech perception outcomes. Other predictive factors that could possibly influence speech perception outcomes were further investigated.

Chapter 2

DERIVATION OF FREQUENCY IMPORTANCE FUNCTIONS FOR THE AZBIO SENTENCES

I. INTRODUCTION

The AzBio sentence test, named after the Arizona Biomedical Institute at Arizona State University, was first described and developed (Spahr and Dorman, 2004; Spahr *et al.*, 2012). The goals that the AzBio test pursued were to provide new test material that (1) minimized stimulus familiarization effects for listeners who were exposed to traditional test stimuli too often, (2) allowed a large number of test conditions, (3) had similar levels of difficulty for within subject comparisons, and (4) evaluated performance that reflects a high degree of correlation with the patient's everyday speech perception environments (Spahr et al. 2012). With a growing interest in cochlear implant (CI) studies, the AzBio sentence test has gained widespread use when assessing the speech perception ability of cochlear implant recipients. Yet, no attempt has been made to establish frequency importance functions (FIFs) for this sentence test which could be very useful in predicting the speech intelligibility index (SII) for these listeners. The spectral distribution of speech is important for estimating intelligibility; thus FIFs are critical for this process. These values reflect our understanding of the content of speech in each spectral band which contributes to a better understanding of speech processing. This study determined the frequency importance weights for the AzBio sentences for use in future SII studies that evaluate speech perception performance of CI patients.

A. Speech Intelligibility Index

Since the development of the model of articulation theory (French and Steinberg, 1947), the profession of audiology and related fields have made use of this concept, exploring extensive attempts to predict speech perception in an objective way. The underlying assumption of the articulation model is that intelligibility of speech can be quantitatively represented using weighted factors across the frequency bands of speech and corresponding audibility of listeners. Relying on this assumption, the Articulation Index (AI) (Kryter, 1962) in 1986, later named the Speech Intelligibility Index (SII) in 1997, has been used to establish the relationship between audible speech cues and the perception of speech.

Calculating SII in a traditional way is not a simple process, as it requires a fairly complicated procedure in its calculation (Amlani *et al.*, 2002). As a result, some researchers have attempted to simplify the calculation of the SII, while maintaining its accuracy (Mueller and Killion, 1990). The SII is a number between 0 and 1 with a value of 1 indicating that all speech cues were delivered to the listener, whereas a value of 0 suggests no speech cues were available to the listener. The SII is calculated by multiplying the audibility function (A_i) and FIF (I_i) which are summed across the total number of frequency bands [Eq. (1)]. Therefore, audibility functions (A_i) and FIFs (I_i) play an important role in determining the SII.

$$SII = \sum_{i=1}^{n} I_i A_i \, , \tag{1}$$

B. Audibility function and FIF

The audibility function defines the proportion of speech information delivered to the listener at frequency band *i*. The audibility function is typically represented by equation (2). The SNR_i denotes the SNR (or signal to hearing threshold ratio) in "i" frequency band; *K* is the level of the speech maxima above the long term average speech spectrum (LTASS); and *DR* is the dynamic range of the speech input. Despite some disagreement, it is generally assumed that a dynamic range of 30 dB (±15 dB relative to the LTASS in ANSI 1997) is a reasonable value in maximizing speech intelligibility (ANSI, 1969; Amlani *et al.*, 2002).

$$Audibility\ function\ (A_i) = \frac{SNR_i + K}{DR}, \tag{2}$$

The FIF, sometimes called Band Importance Function (BIF) or frequency weight, refers to defining the relative importance of the frequency band "i" in the speech spectrum in relation to speech intelligibility. In general, the greatest frequency weights are observed at approximately 2kHz (ANSI, 1969; Studebaker and Sherbecoe, 1991). The specific pattern of frequency importance weight, however, varies with specific stimuli, gender, equipment and procedures (Studebaker and Sherbecoe, 1991; DePaolis *et al.*, 1996; Sherbecoe and Studebaker, 2002; Chen *et al.*, 2016). Speech perception test materials are thought to be a major factor that contribute to the variability among FIFs due mostly to the different amounts of contextual information available in the various speech materials (DePaolis *et al.*, 1996). If there is more contextual information available in the material (e.g., discourse), then the peak of the FIF is closer to the lower frequencies. In contrast, higher frequency information becomes more informative when nonsense syllables, which do not have contextual cues, are recognized. Therefore, appropriate frequency weights need to be used to improve the accuracy in predicting speech recognition performance using the SII, and continuous efforts deriving FIFs should be made as new test materials become available. Currently, the FIFs for six speech tests [NNS (various nonsense syllable tests) (French and Steinberg, 1947), CID-W22 (Studebaker and Sherbecoe, 1991), NU-6 (Studebaker *et al.*, 1993) , DRT (Diagnostic Rhyme Test) (Duggirala *et al.*, 1988), short passages (Studebaker *et al.*, 1987), SPIN monosyllables (Speech Perception in Noise) (Bell *et al.*, 1992)] are included in the ANSI S3.5 (1997).

C. Prediction of speech intelligibility via transfer functions

Once the SII is calculated, it is typically used to predict speech recognition performance by means of a Transfer Function (TF). The TF is represented with an s-shaped curve to show a series of relationships between the SII and corresponding speech recognition scores. The equation for the TF is shown in Eq. (3), where *A* refers to the SII value, *P* is a proficiency factor, and *Q* and *N* are fitting consonants that determine the shape of the curve.

$$Proportion\ Correct = (1 - 10^{-(AP/Q)})^N , \tag{3}$$

Predicting speech intelligibility through the TF curve has been shown to be valid for listeners with normal hearing and for good performing listeners with mild-to-moderate hearing loss (French and Steinberg, 1947; Humes, 1986; Pavlovic *et al.*, 1986). The TF curves drawn from listeners with normal hearing, however, are likely to overestimate speech recognition performance for listeners with moderate-to-profound hearing loss having poor speech perception scores (Ching *et al.*, 1998). This deterioration in supra-threshold sound processing can be attributed to poor spectral and temporal resolution inherent in patients with sensorineural hearing loss (Pavlovic *et al.*, 1986). Therefore, when the SII serves as a predictor of speech perception performance for people with hearing loss, some correction factors (e.g., proficiency and hearing loss desensitization) are required (Sherbecoe and Studebaker, 2002; Scollie, 2008) to adjust the measured SII in proportion to the degree of hearing loss.

1. Proficiency factor

In 1950, Fletcher and Galt proposed a proficiency factor (*P*) as a means to reduce the predicted scores computed by the original TF. The proficiency factor accounts for talkers' and listeners' variation in proficiency with a maximum value of 1 when they use the same dialect. The SII, considering the proficiency factor (*P*), is represented by Equation 4.

$$Modified\ SII\ (considering\ proficiency\ factor) = P \sum_{i=1}^{n} I_i A_i, \tag{4}$$

2. Desensitization factor

A few decades later, a concept called the hearing loss desensitization factor was introduced to account for the effect of hearing loss on speech intelligibility (Pavlovic *et al.*, 1986). The desensitization factor was developed from the findings (Pavlovic, 1984) that supra-threshold sound processing is affected by hearing loss in a frequency specific way. Pavlovic *et al.* (1986) showed the superiority of the modified SII with desensitization factors to accurately predict the SII compared to using the unmodified SII without desensitization. The desensitization factor is computed by multiplying the hearing threshold by a number between 0 and 1. For hearing thresholds < 15 dB HL and > 95 dB HL, the desensitization factor is calculated by multiplying the threshold by 1 and 0, respectively. When the hearing threshold falls between 15 and 95 dB HL, the desensitization factor factor (D_i) is determined by the value derived from Equation 5.

$$D_i = 1.19 - 0.0127 \times hearing\ threshold(i), \tag{5}$$

In the SII calculation, the obtained desensitization factor (D_i) is multiplied by either A_i or I_i (Eq. 6). While the desensitization factor appears to be similar in concept to the proficiency factor, they are not identical. The desensitization factor reflects frequency specific deficits in hearing threshold, and thus is applied during the SII calculation. In contrast, the proficiency factor affects overall performance, and thus is applied after the SII calculation (Scollie, 2008). Even though desensitization factors improve the accuracy of TFs to some extent, they do not work perfectly for fitting TF curves depending on the degree of hearing loss and conditions (Humes, 2002; Scollie, 2008).

$$Modified\ SII\ (considering\ desensitization\ factor) = \sum_{i=1}^{n} I_i\, A_i D_i\ , \qquad (6)$$

D. Application of SIIs

The SII has been widely used with people with hearing loss in clinical and research environments. The SII is used in clinics to predict speech recognition ability for patients who have communication problems, on whom it is often difficult to obtain reliable speech perception scores. The SII is also typically applied in hearing aid evaluations by comparing aided and unaided performance. The count-the-dots audiogram is an example of a simplified version of the SII and is an effective tool for counseling patients with hearing impairments (Mueller and Killion, 1990). The SII is also used in probe microphone measurements when fitting appropriate gain to restore audibility. A high SII often results from amplifying speech signals so that the LTASS has up to 18 dB of sensational level which can facilitate speech perception (Humes, 1986; Amlani *et al.*, 2002). Finally, there have been attempts to develop hearing aid fitting formulas based on audibility across the frequencies and corresponding SIIs for hearing aid users (Dillon, 1999; Byrne *et al.*, 2001).

Despite a wide range of research on the SII associated with hearing aid outcomes, there is little published research available that has predicted SIIs for CI users. Mehr *et al.* (2001) attempted to develop and validate an estimation method to derive channel weights for multichannel CIs. In addition, some researchers have shown high correlations of several modified SII procedures with the intelligibility of vocoded speech (Chen and Loizou, 2011), and some proposed a couple of refinements to emulate CI auditory perception (Santos *et al.*, 2013). Neither of them, however, has shown traditional TFs that establish the relationship between SII and speech perception scores for CI users. Thus, global data related to the SII for CI listeners have not been collected using a standard approach, and comparative relationships to listeners with normal hearing have yet to be established. Significant variability in performance among CI populations has prevented the application of using the SII with CI users; however, at the very least, ways of measuring the SII with this population should be pursued.

E. Purpose of the study

The AzBio sentence test was first developed by Spahr and Dorman (2005) with the purpose of comparing speech perception performance of high-performing patients who used CIs from the various manufacturers (Spahr and Dorman, 2005; Spahr *et al.*, 2007). This test has become the gold standard for assessing CI users' performance, and is now included in the Revised Minimum Speech Test Battery used to evaluate pre- and post-implant speech perception abilities (Spahr *et al.*, 2012). As noted above, FIFs can differ considerably based on the specific speech test materials used (Sherbecoe and Studebaker, 2003). Although the AzBio is well known, has effective standardization data available, and was chosen as a gold standard for CI evaluation, no attempt has been made to derive a FIF for its stimuli. Establishing FIFs for the AzBio set would provide building blocks for future CI studies that employ the SII. Therefore, the purposes of this study were to: (1) derive FIFs for the AzBio sentences using a traditional approach (Studebaker and Sherbecoe, 1991) and (2) provide systematic comparisons of FIFs for the AzBio sentences with other speech perception materials.

II. METHOD

A. Participants

Fifteen native English speakers (4 males, 11 females) ranging in age from 21 to 51 years (M = 29, SD = 10.16) were recruited. Participants had normal hearing and reported negative history of cognitive deficits. Pure-tone audiometry was conducted to confirm air conduction thresholds < 20 dB HL at the octave frequencies from 250 to 8000 Hz. All participants underwent tympanometry to verify normal middle ear function as evidenced by Type A tympanograms. All of them received monetary compensation for their participation. This research was approved by The University of Memphis Institutional Review Board.

B. Materials

Recorded sentences from the AzBio lists were used which consist of 15 lists, each containing 20 sentences spoken by 2 male and 2 female talkers. The total number of possible words correct ranges from 133 to 154 depending on the list. Percent correct scores were computed by dividing the number of correctly identified words by the total number of words in the sentences in each list. In the present study, low- and high-pass filtered AzBio sentences were presented at various SNRs in sound field conditions.

C. Stimuli

The noise was designed to match the LTASS of the AzBio speech sounds. All silent pauses within and between sentences were eliminated and digitized at 44.1 kHz sampling frequency with 16-bit amplitude resolution using Adobe Audition 3.0 (Adobe Systems Inc., San Jose, CA, USA). All sentences in 14 of the AzBio lists were concatenated, with the exception of List 7, which was used as a practice list. The LTASS envelope of the concatenated speech was applied to white noise using Praat (Boersma, 2002). This process ensured that the noise and speech signals had the same spectral shape on average across the frequencies, preventing the effect of variation in SNRs across frequency bands. The SNRs across the frequencies, however, were slightly varied from sentence to sentence, even for the same talker.

As a large number of filtering/SNR conditions were required to be evenly applied to the limited number of AzBio lists, we generated a randomization table that randomly assigned each list to each stimulus condition. Following the randomization scheme, the AzBio lists were filtered through 18 high-pass (HP) and 18 low-pass (LP) conditions that were consistent with the 1/3 octave band calculation procedure (Table I). Linear phase FIR filters (Equiripple filter) with a slope of 96 dB/octave at the cutoff frequencies were used. The signal filtering was implemented using MATLAB 2016 (The Math-Works, Inc., Natick, MA). The speech-shaped white noise and the AzBio sentences were then mixed in separate channels using Adobe Audition to present a single stimulus at the desired SNR using an audiometer (GSI 61). Each channel was calibrated using a Bruel and Kjaer Type 2250 sound level meter. A total of 222 experimental stimuli were generated [(18 LP + 18 HP + 1 wide band) X 6 SNRs] with 30 unnecessary conditions based on the pilot study described below. Those stimuli were randomly assigned to one of the AzBio lists following the randomization table. All stimuli were presented using Adobe Audition from a laptop through a GSI 61 audiometer.

D. Procedures

1. Pilot study

First, it was necessary to determine the appropriate SNRs that would be used to draw a series of SNR curves for the curve bisectional procedure. The ideal scenario would include SNRs that generated maximum scores of 100% for the best condition (wideband frequency with the highest SNR), and the scores for the other conditions would gradually decrease with a decrease in either SNR or in the amount of speech energy in the cut-off frequency of the filters. Within our diverse filtering conditions (18 LP and 18 HP), those with a cut-off frequency at the extreme edge of the frequency bands were not intelligible at all, resulting in 0% correct. Thus, it was meaningless to conduct experiments in such extremely degraded conditions. As a result, we conducted a pilot study with 6 listeners who had normal hearing to determine appropriate SNRs and eliminate unnecessary filtering conditions.

The pilot study was performed using the procedures as in the main experiment (described below), except for the SNR/filtering conditions. Using wideband stimuli, we identified that a 4 dB SNR resulted in maximum scores of 100%. We then used an SNR range from -6 to 4 dB, with 2 dB steps. We further determined unnecessary conditions that resulted in 0% intelligibility by testing some cut-off frequency ranges at the edge of the frequency domain, and then eliminated those redundant conditions. For example, if LP450 (low-pass filter with a cut-off frequency at 450 Hz) at 4 dB SNR resulted in 0% correct, the other acoustically poorer conditions, such as LP450 at 2 dB SNR or LP335 at 4 dB SNR were also assumed to be 0%, as those conditions were acoustically more degraded. As a result, those unnecessary conditions (30) were removed, and 192 conditions remained from the initially planned 222 conditions [(18 LP + 18 HP + 1 WB) X 6 SNRs].

2. Primary study

Traditional procedures described by Studebaker and Sherbecoe (1991) were applied in order to derive the FIFs. For determining FIFs using the curve bisection technique, listeners' percent correct scores on the AzBio test were obtained in 192 filtering/SNR conditions that were determined in the pilot study. To avoid learning effects, each listener only heard each list of sentences one time. Fifteen AzBio lists were available to each listener. Each participant was randomly assigned to 14 different conditions using 14 different lists. According to Spahr *et al.* (2012), speech recognition performance in noise for AzBio List 7 was significantly poorer than for the other AzBio lists. Thus, List 7 was used as a practice list for familiarization of the procedure. A practice trial of the speech recognition test was conducted using List 7 under the unfiltered 2 dB SNR condition for each participant to get used to the experimental protocol.

Each participant was seated in the middle of a double-walled sound-treated room meeting permissible ambient noise levels (ANSI S3.1-1999 (R2013). The participants listened to the stimuli presented through a loudspeaker located 1m away (0° azimuth) from the listener. The stimuli were routed from a laptop computer to a GSI 61 audiometer. The noise level was set at 65 dB SPL, and the level of the speech signal was varied for the desired SNR conditions using the audiometer. Participants were instructed to listen carefully and repeat everything they heard, even if it was only part of a sentence. They were encouraged to guess. Responses were scored in percent based on the number of key words repeated correctly. The final perception score for each condition was determined by the average of two individuals' speech perception scores. Each participant required approximately one hour to complete their assigned test conditions. Table 1 shows the 192 filtering/SNR conditions used in the main experiment along with average speech recognition scores in percentage. The diagonal line boxes represent conditions that were eliminated based on the pilot study.

Table 1. The filtering/SNR conditions used in the present study. Average speech recognition scores are represented in percent for each condition. The cells with a diagonal line are the conditions that were eliminated.

SNR (dB)	Filter type	1/3 Octave cut-off frequency (Hz)																	
		180	224	280	355	450	560	710	900	1120	1400	1800	2240	2800	3550	4500	5600	7100	9000
-6	Low-pass filter					0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.03	4.52	7.98	8.38	12.47	16.32	18.04
	High-pass filter	11.43	13.04	16.42	13.41	18.35	12.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Wide band	19.00																	
-4	Low-pass filter					0.00	0.00	0.35	0.00	0.00	2.76	6.04	7.63	15.05	15.29	23.97	31.88	33.49	29.54
	High-pass filter	28.89	30.13	37.85	32.41	28.62	23.53	21.11	24.16	13.14	19.16	19.34	7.33	3.50	0.70	0.00	0.00	0.00	
	Wide band	33.00																	
-2	Low-pass filter					0.00	0.00	0.00	1.05	2.50	12.47	15.92	28.64	32.53	38.20	48.07	46.90	62.38	53.10
	High-pass filter	52.39	48.25	47.86	55.40	47.62	46.04	39.80	39.40	29.22	33.07	22.99	9.00	8.37	1.90	0.00	0.36	0.00	
	Wide band	55.00																	
0	Low-pass filter					0.00	0.00	2.05	7.10	14.14	30.07	29.51	36.50	48.85	72.72	62.42	72.42	74.74	68.37
	High-pass filter	74.96	68.01	67.59	63.81	67.38	62.09	50.37	48.45	39.79	40.93	30.90	13.79	4.34	9.62	0.00	0.00	0.00	
	Wide band	71.00																	
2	Low-pass filter					0.00	0.00	8.14	12.06	20.68	32.20	57.63	59.21	80.96	80.78	85.92	79.37	85.95	89.63
	High-pass filter	80.50	80.59	86.40	87.29	86.59	76.84	80.39	68.88	63.59	54.42	45.72	22.37	11.09	4.00	0.00	0.00	0.00	
	Wide band	85.40																	
4	Low-pass filter					0.00	0.97	10.37	15.06	31.57	44.20	70.36	73.10	88.82	95.22	92.17	91.42	95.91	97.80
	High-pass filter	95.20	91.55	96.00	99.29	87.66	82.31	82.33	77.72	82.39	60.61	55.23	26.08	17.27	9.07	2.17	0.00	0.00	
	Wide band	98.46																	

III. RESULTS

A. Curve smoothing

Multiple SNR curves were drawn plotting the speech recognition scores as a function of the cutoff frequencies of the HP and LP filters (Fig. 1). Unlike theoretical graph patterns that can show smoothed and even perception scores as a function of cut-off frequencies, our empirical graph patterns showed zigzag shapes for some frequency bands. This unsmoothed pattern has been observed in most studies, so smoothing curves were required prior to moving on to the next step of the curve bisection method. The smoothing method used is a technique that is conventionally used in most studies (Studebaker and Sherbecoe, 1991; Studebaker *et al.*, 1993; DePaolis *et al.*, 1996; Sherbecoe and Studebaker, 2002; Jin *et al.*, 2015). The curves were smoothed by drawing freehand curves following four rules demonstrated by Studebaker and Sherbecoe (1991): (1) two different SNR curves do not intersect if they are identical filter curves; (2) both HP and LP curves at any SNR terminate at the same score; (3) scores must increase or remain constant as the energy of the passband or SNR increases; (4) HP and LP curves at any SNR curves make only a single crossover point.

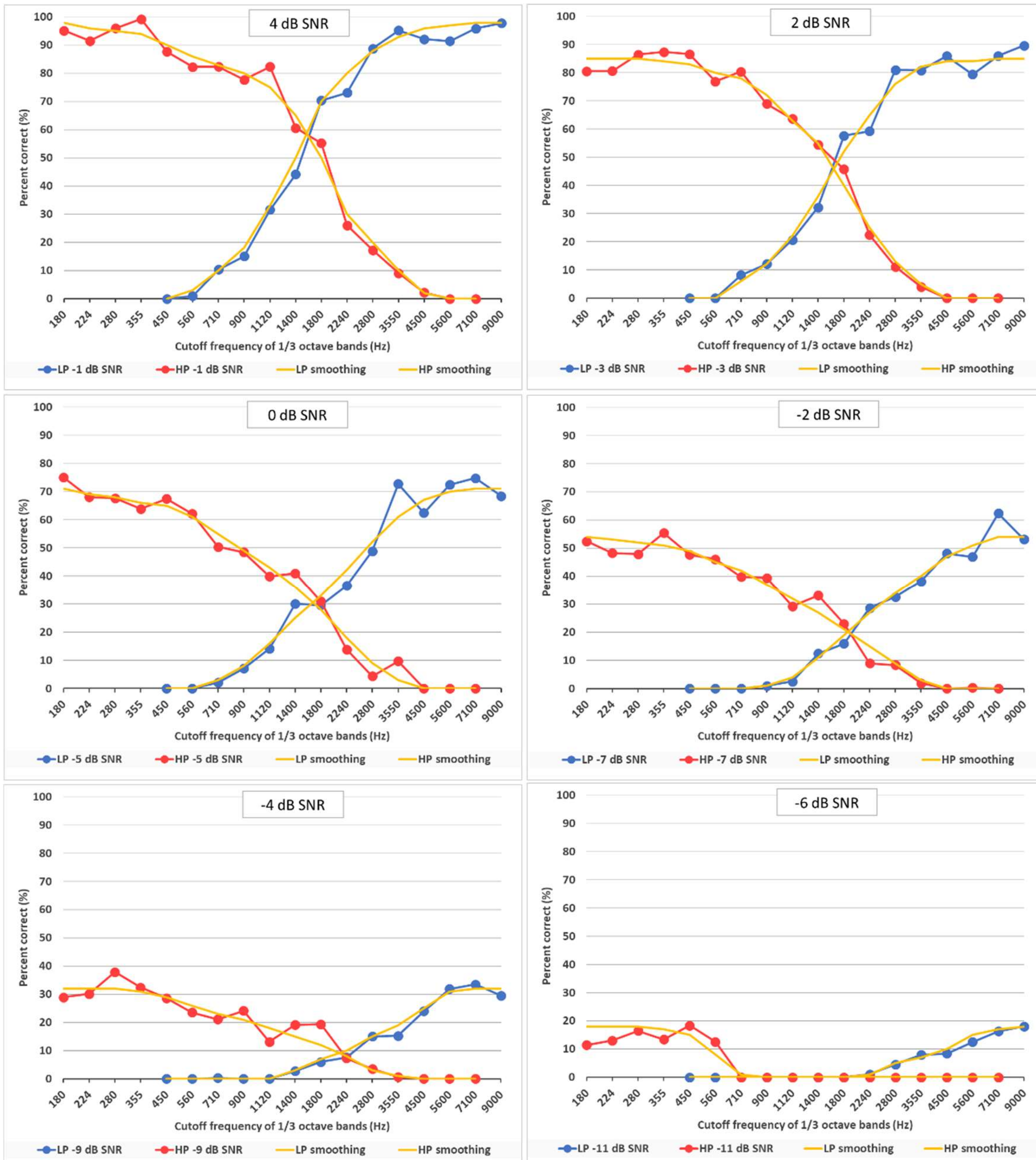

Figure 1. Group mean percent correct scores for AzBio lists plotted as a function of the 1/3 octave cutoff frequency bands. Each panel represents the curves in the order of different SNRs from 4 to -6 dB SNR in 2 dB steps. The curves with circles in red indicate HP, and the curves with circle in blue represent LP conditions. Smoothed curves drawn based on raw data are shown with yellow lines.

B. Curve bisection procedure

The smoothed curves were then analyzed using a sequence of procedures called the curve bisection procedure (Studebaker and Sherbecoe, 1991; Wong *et al.*, 2007; Chen *et al.*, 2016). The curve bisection method is a technique that is typically used to determine the relative transfer function as a basis for deriving a FIF and absolute transfer function.

Figure 2 shows the illustration of the curve bisection procedure for deriving relative transfer function curves. The procedure begins with an assumption that the total area for the 4 dB SNR curves, which have 100% maximum scores, is equal to an SII of 1.0. Thus, it can be assumed that the intersection point between the HP and LP curves for this SNR corresponds to an SII of 0.5, because half of the total auditory area is available below this point, and the other half of the auditory area is available above this point. In the same way, an intersection point for certain SNR curves having maximum scores that correspond to an SII of 0.5 is equal to an SII of 0.25, and another intersection point for certain SNR curves having maximum scores that correspond to an SII of 0.25 is equal to and SII of 0.125. Unfortunately, none of the obtained SNR curves in our procedure had maximum scores at the point of either an SII of 0.5 or 0.25. Therefore, we adopted interpolation methods to generate the curves that yielded the maximum scores corresponding to SIIs of 0.5 and 0.25. The interpolation curves were drawn on the basis of the two obtained curves that had the closest maximum scores to the 0.5 and 0.25 SIIs: 0 & -2 dB SNR curves for the interpolation curve having a maximum score at 0.5 SII and -4 & -6 dB SNR curves for the interpolation curve having a maximum score at 0.25 SII. The SII of 0.75 was obtained by extending two separate horizontal lines, left and right, from the 0.25 SII until they reached the HP and LP curve for the 4 dB SNR condition. Then, vertical lines were drawn up from those points until they touched the HP and LP curves for the 4 dB SNR condition. The average of those two points was 0.75 SII. A 0.875 SII point was derived in the same way using the SII of 0.125. Additional SII values were obtained with our speech perception score data at several different SNRs using this procedure.

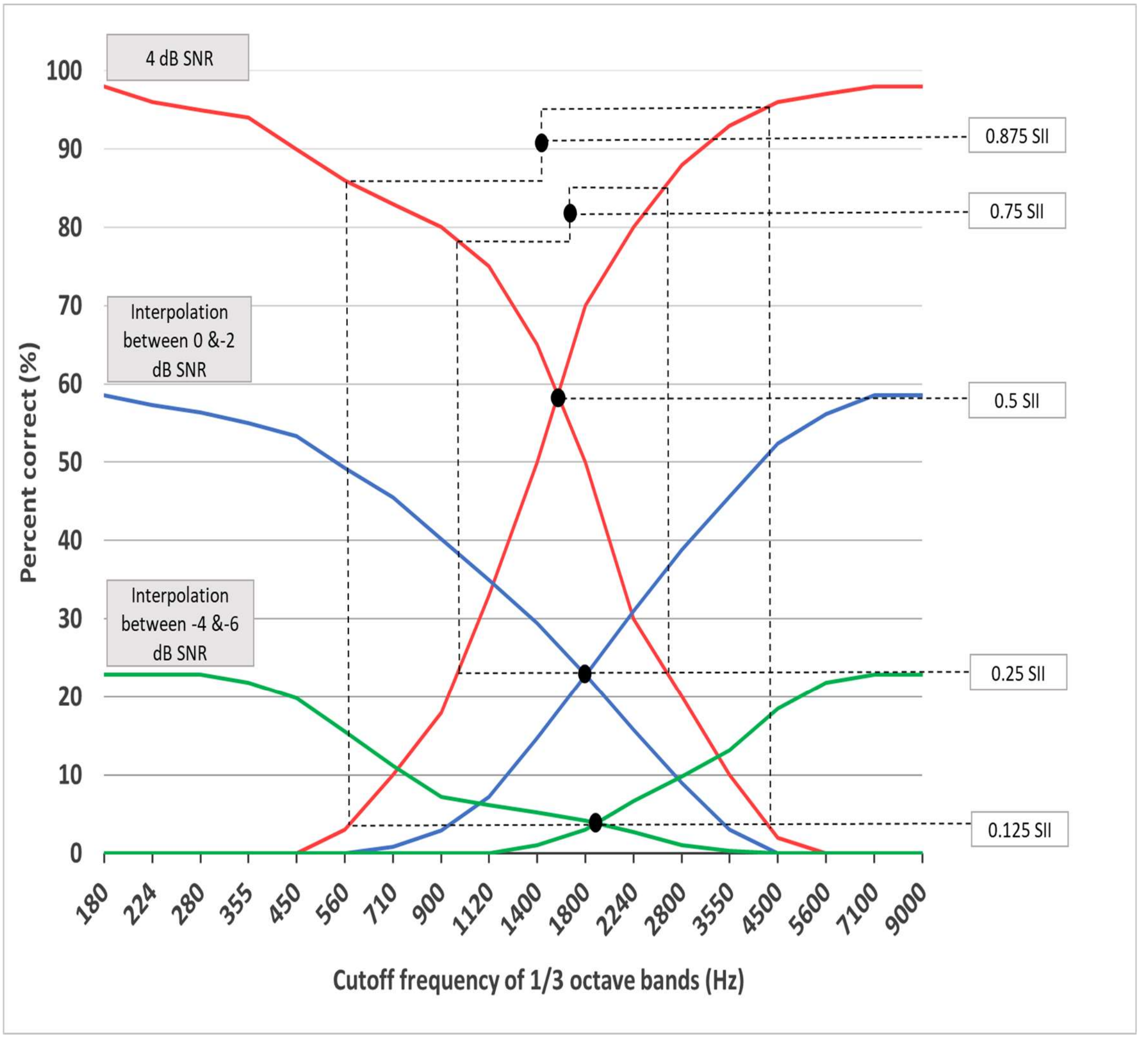

Figure 2. Illustration of the curve bisection procedure. Group mean percent correct scores obtained from speech perception tests at 4 dB SNR are represented with the smoothed curves (top two HP and LP curves) as a function of the 1/3 octave cutoff frequency bands. The following four curves were generated using an interpolation method based on 0 & -2 dB SNRs and -4 & -6 dB SNRs, respectively. The filled circles indicate relative SII points 0.875, 0.75, 0.5, 0.25, 0.125 from top to bottom.

C. Cross-over frequency

Cross-over frequencies are defined as the intersection points of HP and LP speech recognition curves. As shown in the previous step, these points account for equal intelligibility above and below the points at each SNR curve. Cross-over frequencies for the remaining five SNR conditions are shown in Table 2. The smoothed mean scores were used to determine the cross-over frequencies. The -6 dB SNR condition was not reported because its curve did not yield an intersection point due to significantly low intelligibility. To some extent, the cross-over frequencies tended to monotonically decrease with SNR conditions.

Table 2. Cross-over frequencies for the five SNR conditions.

SNR conditions (dB)	-4	-2	0	2	4	Average
Cross-over frequency (Hz)	2114	1863	1675	1645	1571	1774

D. Relative transfer function

Through the consecutive curve bisection procedures, 12 pairs of speech recognition scores and their associated SII values were obtained. The obtained pairs of SIIs and scores were used to calculate relative transfer functions yielding the curve fitting values Q and N in equation (3) (P will be assumed to be 1). The nonlinear regression using IBM SPSS (version 24) found that the fit between the SIIs and scores was excellent when $Q = 0.567$ and $N = 3.797$ (R^2 value of 0.991) for predicting scores (equation 3), and when $Q = 0.57$ and $N = 3.712$ (R^2 value of 0.995) for predicting SIIs (equation 7, an inverse of equation 3).

$$SII = -\frac{Q}{P} log_{10}(1 - S^{\frac{1}{N}}) , \qquad (7)$$

These values were also confirmed with the NLIN procedure in SAS program (version 9). The relative transfer function curve determined by Q (0.567) and N (3.797) based on the twelve pairs is shown in Figure 3.

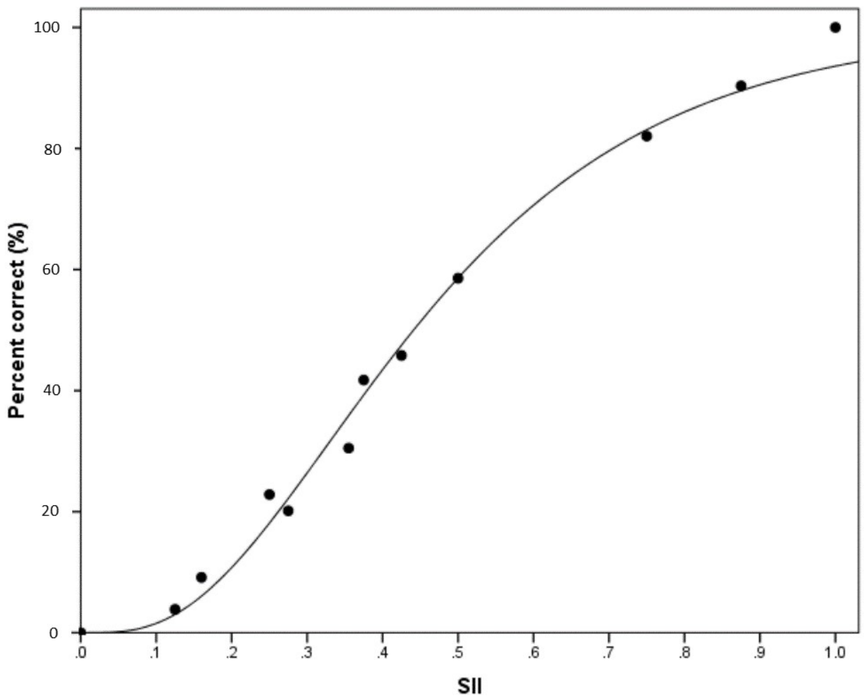

Figure 3. Relative transfer function plotted on the basis of the twelve pairs of speech recognition scores and corresponding SIIs (denoted with filled circles) obtained by the curve-bisection procedure.

E. Frequency importance functions

The FIFs were derived using the following procedures. First, all the smoothed mean HP and LP speech perception scores for each SNR were transformed into SII values using equation (7) using the Q and N values obtained for the relative transfer function. Then the difference in SII values between two adjacent bands was calculated to determine the SII value for individual bands. Specifically, for each HP condition, the SII value for the higher frequency band was subtracted from the SII value for the lower frequency band. In contrast, for each LP condition, the SII value for the lower frequency band was subtracted from the SII value for the higher frequency band. The averages of these two difference values were calculated for the six different SNRs, and then averages of the six SNR values were again determined. This procedure was repeated until all the mean values across the 1/3 octave bands were obtained. Eventually, FIFs for each frequency band were determined proportionally by dividing each SII value by the sum of all values over the frequency bands and multiplying by 100. Table 3 demonstrates the last computational stage for deriving FIFs, and Figure 4 reports the FIFs obtained from this study.

Table 3. Summary table of the FIF calculation. Differences in SII values between two adjacent bands are represented as a function of the 1/3 octave bands and the six SNR conditions. Final FIFs were derived by averaging the SII values for six SNRs at each frequency band and proportionally computing the value across the frequency bands.

NO.	1/3 Octave band (Hz)		Center frequency (Hz)	SNR (dB)						Average	FIF (%)
				-6	-4	-2	0	2	4		
1	0	180	160	0.006	0.006	0.007	0.000	0.007	0.065	0.015	2.203
2	180	224	200	0.000	0.000	0.003	0.009	0.000	0.087	0.017	2.406
3	224	280	250	0.000	0.000	0.007	0.009	0.000	0.056	0.012	1.738
4	280	355	315	0.006	0.006	0.003	0.009	0.009	0.023	0.009	1.347
5	355	450	400	0.013	0.012	0.013	0.008	0.016	0.065	0.021	3.076
6	450	560	500	0.052	0.017	0.013	0.016	0.022	0.105	0.037	5.420
7	560	710	630	0.090	0.018	0.018	0.083	0.091	0.060	0.060	8.708
8	710	900	800	0.085	0.012	0.057	0.046	0.058	0.049	0.051	7.426
9	900	1120	1000		0.019	0.040	0.048	0.072	0.075	0.051	7.354
10	1120	1400	1250		0.071	0.046	0.048	0.071	0.099	0.067	9.720
11	1400	1800	1600		0.032	0.045	0.047	0.097	0.131	0.070	10.208
12	1800	2240	2000	0.042	0.028	0.043	0.057	0.092	0.116	0.063	9.127
13	2240	2800	2500	0.062	0.044	0.042	0.063	0.091	0.097	0.067	9.647
14	2800	3550	3150	0.010	0.032	0.048	0.063	0.072	0.103	0.055	7.912
15	3550	4500	4000	0.025	0.018	0.040	0.043	0.036	0.113	0.046	6.652
16	4500	5600	5000	0.018	0.017	0.013	0.014	0.000	0.089	0.025	3.647
17	5600	7100	6300	0.013	0.006	0.010	0.010	0.017	0.051	0.018	2.565
18	7100	9000	8000	0.003	0.001	0.001	0.002	0.004	0.023	0.006	0.845
									SUM	0.690	100

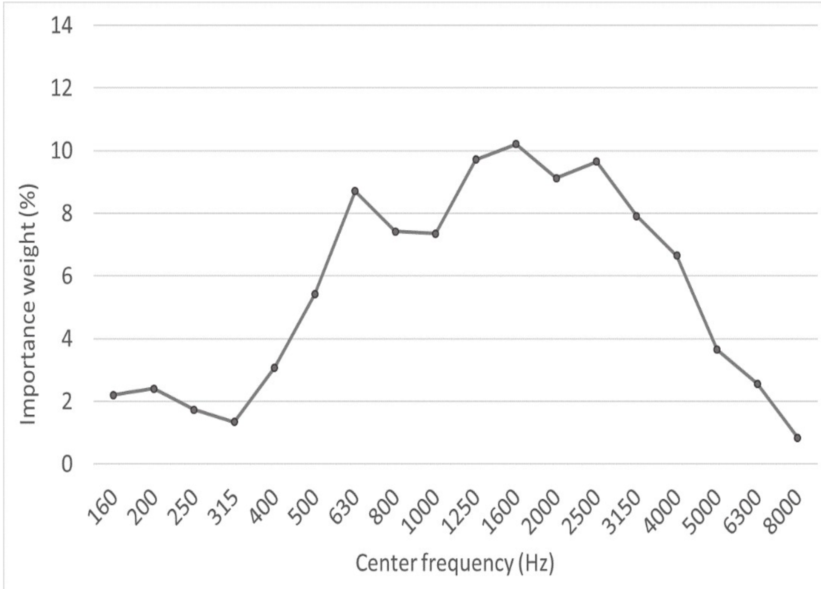

Figure 4. FIF plot for the AzBio test. Frequency weights for the AzBio sentences are displayed as a function of the 1/3 octave cutoff frequency bands.

F. Absolute transfer function

The last stage was to derive the absolute TF. As the relative TF was developed based only on the speech produced by the four speakers in the AzBio, it was necessary to adjust the TF curves to reflect variability in the speech spectrum using the LTASS. To identify the true absolute TF curves, the values *K* in the audibility formula (Eq. 2) needed to be determined and corresponding *Q* and *N* values were reestablished. An iterative process was used to search the *K* value that yielded the smallest mean square error between the predicted SII [calculated with Eq. 7) and the actual SII (calculated with Eq. 1)]. First, unsmoothed mean scores between 5% and 95% were plotted as a function of their SII values. Then, holding the DR value constant at 30 dB, *K* was varied starting from 10 dB in 1 dB steps in equation 1 to calculate the actual SII. The corresponding *Q* and *N* values to define *K* were used to calculate the predicted SII in equation 7. This comparison process was repeated until the smallest mean square error was identified. The corresponding *Q* and *N* values of the best fitting curve using Equation 3 and R^2 between raw scores and SII were obtained. The results showed that the mean square error was the least when K was 11. The corresponding Q and N values were 0.287 and 5.206, respectively, and R^2 was 0.923 for predicting speech intelligibility scores from SIIs using Equation 3. Figure 5 shows the TF curve plotted using the K, Q and N values obtained here. Q was 0.254 and N was 6.519, and R^2 was 0.914 for predicting SII from speech intelligibility scores using Equation 7. The obtained R^2 values in the current study are slightly lower than the values from other FIF studies: continuous discourse (Eq. 3: 0.984; Eq. 7: 0.977); CID W-22 monosyllabic word test (Eq. 3: 0.995; Eq. 7: 0.992). Comparatively lower correlations of our model are presumably attributed to some unknown potential methodological variables, such as a small sample size or unstable speech scores affecting the curve smoothing procedures. However, our R^2 is still thought to be high enough to show that the two models both provided an excellent fit to the data.

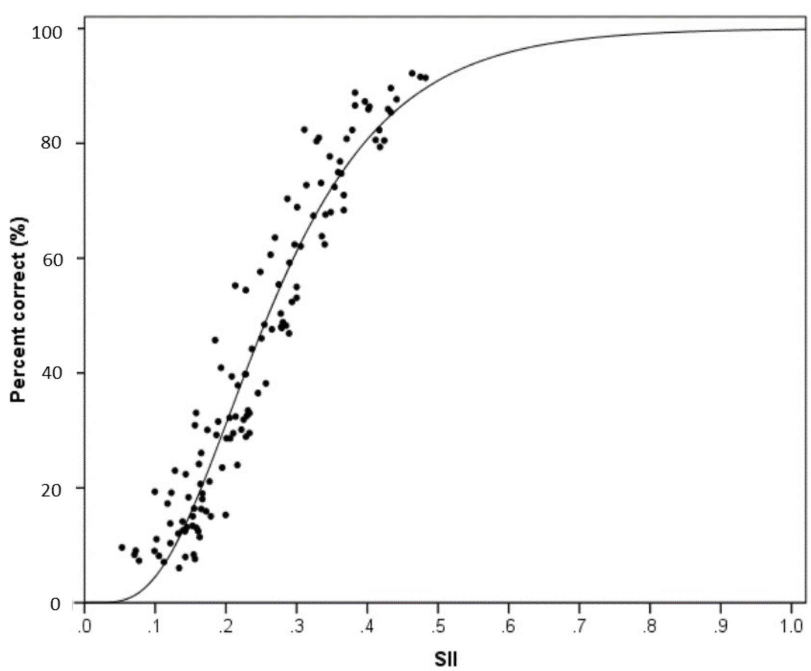

Figure 5. Absolute transfer function curve plotted on the basis of the unsmoothed speech recognition scores and corresponding SIIs (denoted with dots). The unsmoothed mean scores between 5% and 95% were plotted as a function of their SII values.

IV. DISCUSSION

A. Cross-over frequency

The geometric average of the cross-over frequencies for the five SNR curves was 1774 Hz. This value is somewhat higher in comparison with the cross-over frequencies of other English speech materials: Nonsense syllables (1660 Hz), HINT sentence test (1550 Hz), Connected Speech Test (1599 Hz), continuous discourse (1189 Hz) and CID W-22 monosyllabic word test (1314 Hz) (ANSI, 1969; Studebaker *et al.*, 1987; Studebaker and Sherbecoe, 1991; Eisenberg *et al.*, 1998; Sherbecoe and Studebaker, 2002). The higher cross-over frequency observed in this study may be associated with the gender of the talkers (Studebaker *et al.*, 1987). The two male and two female talkers used in the AzBio recordings probably resulted in higher cross-over frequencies than other studies that used only male talkers (ANSI, 1969; Studebaker and Sherbecoe, 1991; Eisenberg *et al.*, 1998). Some previous FIF studies have used either male and female talkers or only female talkers; however, they exhibited lower cross-over frequencies (Studebaker *et al.*, 1987; Sherbecoe and Studebaker, 2002). These studies used continuous discourse and connected speech as the speech materials. This implies that variance in cross-over frequency is probably accounted for by the redundancy effect of speech materials, with greater contextual cues associated with lower cross-over frequencies (Studebaker *et al.*, 1987). Interestingly, many of our participants unofficially reported that the female talkers were perceptually more intelligible than the male talkers. Thus, it seems reasonable that multiple variables interact to determine the cross-over frequency that eventually contributes to the FIFs.

As seen in Table 2, the cross-over frequencies in our study for the five usable SNRs decreased as speech intelligibility increased. Studebaker and Sherbecoe (1991) suggested that cross-over frequencies should be equal across different SNRs, and unequal cross-over frequencies may be caused by the adverse effect of spread of masking. They noted that spread of masking effects possibly occurs when the designated speech-shaped noise cannot completely cover 1% of the speech peaks causing a failure to mask the intensity variation in speech. However, the exact reason for the decrease in cross-over frequencies with SNR is unclear because this tendency has been observed not only in this study, but also in other studies (Kuo, 2013) that used validated speech-shaped noises.

B. Frequency importance function

Many articles regarding FIFs have reported that the primary peaks of speech are located around 2 kHz (Studebaker and Sherbecoe, 1991; DePaolis *et al.*, 1996; Henry *et al.*, 1998). The greatest amount of frequency weights at this frequency region is accounted for by the importance of the second and third formants in recognizing voicing in speech (Chen *et al.*, 2016). In most vowels, these formants show their dynamic trajectories at about 1 to 2 kHz. At first glance, our FIF for the AzBio test seems to have a broad mid-frequency peak extending from about 630 Hz to 2500 Hz. However, judging from the small valley between 800 and 1kHz, the shape appears to follow a bimodal pattern having two peaks at low and mid-high frequencies. There was a minor peak at 630 Hz and a major peak at 1600 Hz. Studebaker and Sherbecoe (1991) first proposed a possible association between the bimodal shape and perception of contextual cues. They provided some examples of highly redundant speech materials that produced a bimodal shape (Studebaker *et al.*, 1987; Duggirala *et al.*, 1988) as opposed to a nonsense syllable test that yielded unimodal configurations (ANSI, 1969). A similar view was expressed by DePaolis *et al.* (1996). They derived FIFs for words, sentences and continuous discourse under the same method and conditions, and suggested that highly contextual cues or low vocabulary size could be associated with a broad shape of frequency weights. Over the past few decades, highly contextual speech tests having unimodal shapes have been frequently reported (Bell *et al.*, 1992; DePaolis *et al.*, 1996; Eisenberg *et al.*, 1998). Therefore, it seems more reasonable to assume that context and linguistic information are associated with a more broadly shaped FIF, and further investigations on the origin of bimodal shapes are still necessary.

In Figure 6, we compared our FIF with those of other speech perception tests including importance weights for the R-SPIN test (Bilger, 1984) and the CNC words from Lehiste and Peterson (1959) along with average speech as presented in ANSI S3.5-1997. With the exception of the R-SPIN, the importance weights for the AzBio and the CNC words were nearly identical to those for average speech as provided in ANSI S3.5-1997. This implies that the FIFs provided by ANSI satisfactorily represent frequency weights for the AzBio sentences and CNC words in the SPIN test, and could be used for typical SII calculations for sentence intellegibility.

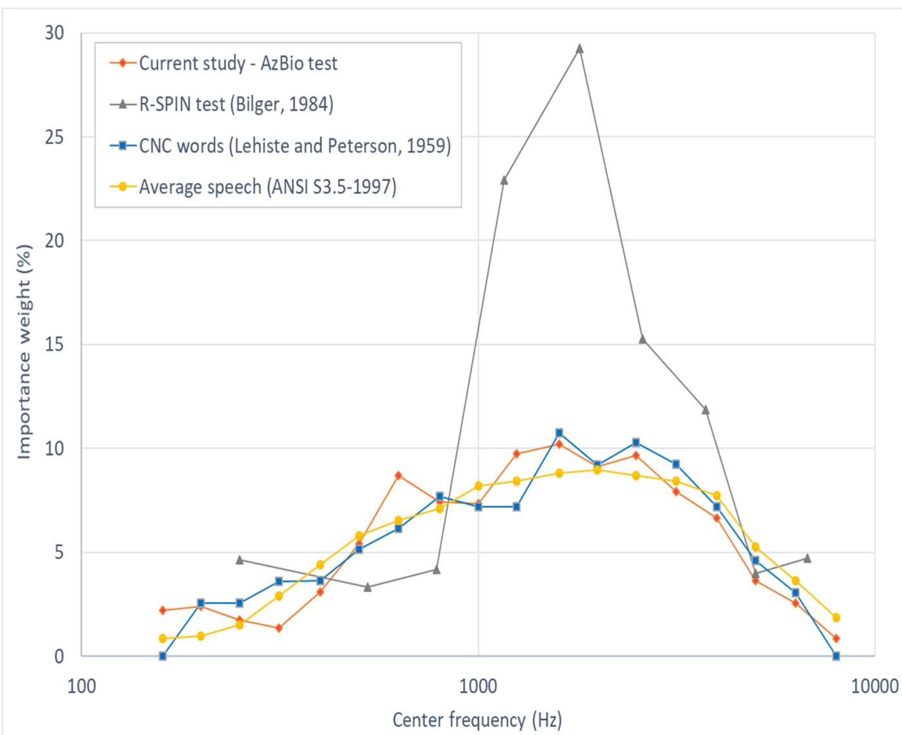

Figure 6. FIFs for four English stimuli were plotted as a function of frequency in logarithmic scale.

Interestingly, an extremely sharp and high peak at 2 kHz was observed in the FIF for the R-SPIN test. It is highly probable that the small number of frequency bands used gave rise to the extremely prominent peak because the distribution of weights across frequency was limited. Thus, caution should be taken regarding the number of bands used when interpreting comparative differences in FIFs for different studies. Using cumulative plots could be an option to eliminate the bias of the number of bands. Figure 7 shows the cumulative FIFs for the speech materials shown in Figure 6. Despite the different shape of the frequency curves, the abrupt change in frequency weights for the R-SPIN was also seen in the cumulative plot. DePaolis *et al.* (1996) assumed that this distinct shape was attributed to the degree of listeners' familiarity to the stimuli. In fact, methodological heterogeneity among studies has been noted as an obstacle for the accurate comparison of FIFs in many relevant studies (Studebaker *et al.*, 1993; DePaolis *et al.*, 1996; Sherbecoe and Studebaker, 2002; Kates, 2013). Neverthless, due to the complex nature of speech and technical procedures required for deriving FIFs, it has been challenging for studies to keep uniformity in their methods. In addition to speech stimuli, several other factors have been shown to cause variability in FIFs including gender of talkers, signal processing, stimulus familarization of listeners, curve smoothing methods, type of noise, data collection protocol, and vocabulary size (Bell *et al.*, 1992; DePaolis *et al.*, 1996; Sherbecoe and Studebaker, 2002; Kates, 2013).

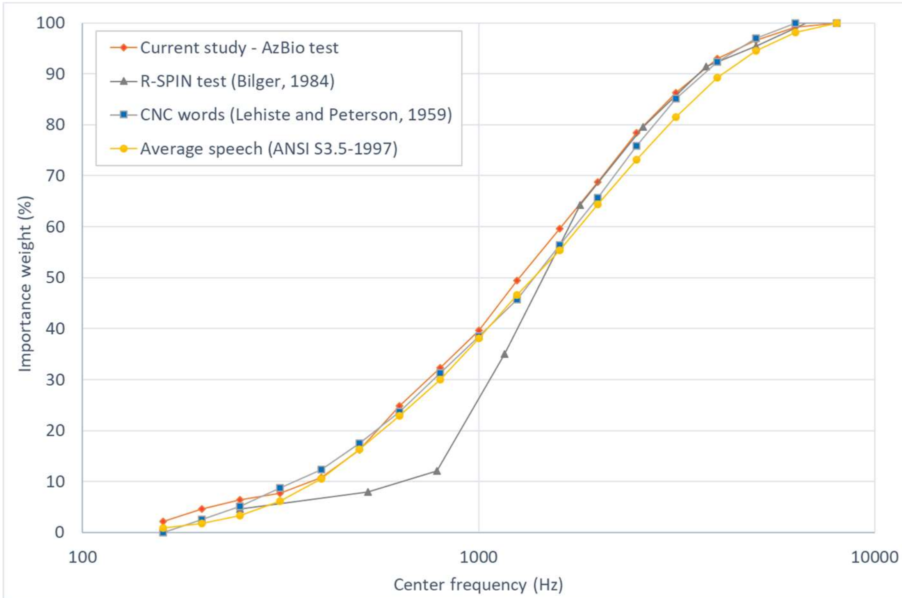

Figure 7. Cumulative FIFs for four English speech stimuli plotted as a function of frequency on a logarithmic scale.

An alternative way for FIF computation that uses a nonlinear optimization function has recently been released (Kates, 2013) highlighting its advantages of accuracy and simplicity over the more traditional procedure. The new procedure aims at minimizing RMS errors between speech perception scores observed from experiments and those predicted by the SII equation using MATLAB. An unconstrained minimization, *fminsearch* function, is used to find an approximation. Then the approximation is applied to the constrained minimization, *fmincon* function, to yield optimal values of variables.

The current study focused primarily on the traditional FIF derivation procedure, despite the laborious steps involved and its relatively lower accuracy as indicated by Kates (2013). We chose to use the traditional procedure because not only has it been used in most FIF studies, which makes it easier for comparisons, but also it provides more detailed information such as information about the cross-over frequency. Our FIF, derived by the traditional procedure, was compared to the FIF derived by the nonlinear optimization procedure as shown in Figure 8. The frequency weights for the traditional procedure and the nonlinear optimization procedure did not match up completely. The overall patterns were similar, but there were a few inconsistent frequency band weights between the two procedures (e.g., around the center frequency of 630 Hz and above the center frequency of 4 kHz). Some of the fluctuating FIF patterns seen across bands with the traditional procedure were not shown in the new procedure. It is likely that the five-point binomial smoothing filter (Marchand and Marmet, 1983) used in the new procedure removed undesirable fluctuations resulting in smoother morphology of the FIF for the nonlinear optimization procedure. The differences in the FIFs obtained in the current study are quite large compared to those observed in previous studies (Kates, 2013; Jin *et al.*, 2015) for unknown reasons.

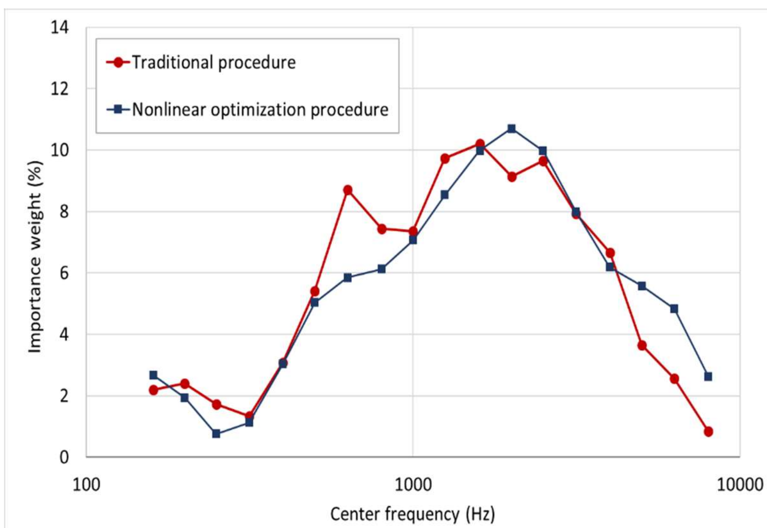

Figure 8. FIFs derived from the traditional procedure and the nonlinear optimization procedure.

The RMS error for the new procedure (0.044) was less than the RMS error for the traditional procedure (0.069). The Pearson correlation coefficients were 0.923 for the traditional procedure and 0.990 for the nonlinear optimization procedure ($p < 0.001$), implying that both procedures were accurate, but the new approach was slightly higher in accuracy. The new solution also produced the equation parameters that minimized the RMS error between the SIIs and the observed speech perception scores. The fitting parameters *Q, N*, and *K* for the two procedures are presented in Table 4.

Table 4. The fitting parameter values of Q, N, and K for the two procedures.

Parameter	Nonlinear Optimization procedure	Traditional procedure
Q	0.247	0.287
N	4.013	5.206
K	8.737	11

C. Transfer function (TF)

In this study, the TF for the AzBio test was derived to establish the relationship between SIIs and corresponding speech intelligibility. Figure 9 shows TFs that convert SII scores into speech recognition scores for our FIFs for the AzBio test compared to three other English (HINT, CST, and NU-6) speech materials (Studebaker *et al.*, 1993; Eisenberg *et al.*, 1998; Sherbecoe and Studebaker, 2002).

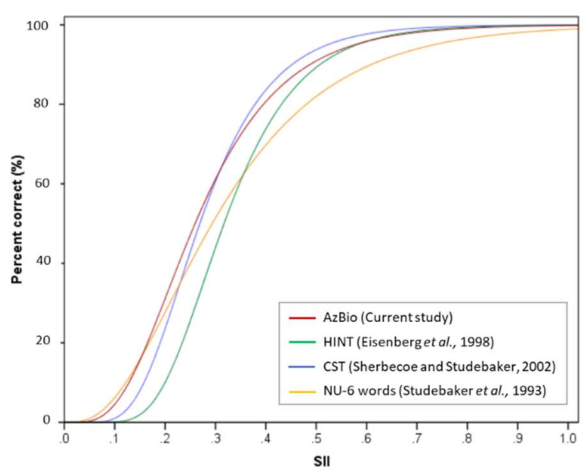

Figure 9. Comparison of transfer functions for the four English materials: (1) AzBio sentences (current study); (2) HINT (Eisenberg et al., 1998); (3) CST (Sherbecoe &Studebaker, 2002); (4) NU-6 words (Studebaker et al., 1993).

Examination of Figure 9 shows that the shape of the TFs differed depending on the amount of contextual cues embedded in the speech. The NU-6 words required a much higher SII value (approximately 1.0 SII) in order to achieve 100% correct compared to the other materials that required lower SIIs to reach 100%. This is presumably due to the limited linguistic context available in the NU-6 monosyllabic word lists (Sherbecoe and Studebaker, 2002). Notably, the TF for the AzBio test was positioned further to the left of the HINT and nearly the same as the CST indicating better performance on the AzBio test compared to the HINT. For example, an SII of 0.2 resulted in an AzBio score of about 40% compared to a HINT score of 10%, and at an SII of 0.5, performance on the AzBio and HINT converged. Even though the CST provides connected discourse compared to only sentences in the AzBio, our findings suggest that the AzBio sentences were comparable in difficulty.

Research has shown that performance on the AzBio test has been poorer than on the HINT test (Gifford *et al.*, 2008; Spahr *et al.*, 2012). However, our findings show the opposite results in the TF. It is possible this discrepancy is due mainly to the different experimental and mathematical methods used in this study and others. For example, in our experiments, listeners were highly encouraged to guess for correct responses. Also, we used a large number of filters, and outcomes were scored based on the key words correctly identified. In contrast, Eisenberg *et al.* (1998) used a limited number of filters for the purpose of establishing the TF rather than targeting FIFs. In addition, they scored the number of sentences, not words, that their listeners correctly answered. Further, in calculating the TF, we varied K (peak to RMS level), along with Q and N, to seek the best fitting curve, whereas Eisenberg *et al.* (1998) assumed K to be 12 dB. Thus, different methods used among the studies probably had some influence on the varying results. Table 5 reports the slopes for each of the four TF curves shown in Figure 9 based on the scores between 5 to 95% which eliminated floor and ceiling effects.

Table 5. Slopes of the TFs for the different speech materials.

	AzBio	HINT	CST	NU-6
Slope (%)	8.68	10.05	10.42	6.39

D. Limitations

There are several limitations to be noted in this study. The speech perception scores in each filtering/SNR condition were determined by limited experimental data (the average of two scores by two individuals) which some might characterize as insufficiently reliable. To be sure, testing more people and averaging more scores in each condition would have resulted in more reliable outcomes. However, we felt including a large number of filtering/SNR conditions in the experiments strengthened the reliability of this study.

For deriving the FIFs, we followed the traditional procedure that has been most widely used. As seen before, the traditional procedure calculates importance weights by comparing scores obtained with successive cut-off frequencies. As a result, when importance weights for each band are computed, speech energy either below or above the cut-off frequency are involved, while the other speech energy either above or below the cut-off frequency are excluded (Warren *et al.*, 2005; Healy *et al.*, 2013). Recently, some have posed the issue of redundancy and synergistic interactions among involved frequency bands, suggesting that traditional FIF procedures may not reflect independent importance weights for each band (Healy and Warren, 2003; Warren *et al.*, 2005). For instance, a target band presented with another adjacent band showed less importance in comparison to a target band presented with another band located some distance apart. To address this potential limitation, the “compound” method (Apoux and Healy, 2012) has been developed that derives FIFs by computing the performance differences between a target band with randomly selected bands and the same randomly selected band, but with the absence of the target band. Other alternative approaches also exist to determine the importance weights (Doherty and Turner, 1996; Turner *et al.*, 1998; Henry *et al.*, 2000; Mehr *et al.*, 2001). However, given the fact that almost all studies, even the ANSI standard for the six speech materials, have followed the traditional approach to FIF derivation, we opted to perform our research in this way as well. Thus, the possible effects of redundancy and synergistic interactions among bands needs to be taken into account when interpreting these results.

Lastly, the use of different data collection and analysis procedures is always problematic when comparing FIFs across studies. The accurate comparison of the results from our study with other studies is also limited due to the methodological heterogeneity. Further study will be required to identify the effect of variations in methodology. Validating our derived FIF for the AzBio sentences remains to be completed in future studies.

V. CONCLUSIONS

Advances in CI technology have seen rapid improvements in speech intelligibility among patients, prompting researchers to develop new speech perception materials, such as the AzBio test. In this study, we derived FIFs for the AzBio test because FIFs vary across different speech materials. The overall frequency weights for the AzBio test were similar to those for other English speech materials. For the cross-procedure comparison, the FIF derived from the nonlinear optimization procedure resulted in relatively higher accuracy compared to the FIF derived from the traditional procedure. The FIF shapes for the two procedures did not completely overlap each other. However, the results supported the globally accepted notion that speech cues in the 2 kHz region play a pivotal role in speech recognition. Our findings provide empirically derived FIFs for the AzBio test that can be used in future studies. The FIF will contribute to the interpretation of speech perception outcomes when using the AzBio test. It will be worthwhile to refer to the FIF when developing new signal processing strategies or providing optimal maps to the CI patients. The SII may not be a reliable tool to objectively estimate speech intelligibility for the CI population due to the tremendous variability in performance among CI recipients. The application of SII is challenged even more by the difference in FIF patterns between normal listeners and CI listeners, as well as the large cross-listener inconsistencies observed in FIFs among CI users (Bosen and Chatterjee, 2016). Nevertheless, the SII, which has a long history of widespread use, is certainly worth trying with CI patients. It is hoped that our obtained FIFs will contribute to improving the accuracy of the SII in predicting speech intelligibility for CI patients using the AzBio test.

Chapter 3

PREDICTING SPEECH RECOGNITION USING THE SPEECH INTELLIGIBILITY INDEX (SII) AND OTHER VARIABLES FOR COCHLEAR IMPLANT USERS

I. INTRODUCTION

A. Spectral/temporal resolution in cochlear implants

A cochlear implant (CI) is a prosthetic device that converts acoustic signals into electrical stimuli to excite surviving auditory nerve fibers. Over the past few decades, a growing number of people with severe-to-profound sensorineural hearing loss have benefited from CIs that can restore audibility. However, many patients still struggle with understanding speech not only due to the characteristics of electrical hearing, but also because the pathway through which sound travels is different in the CI compared to the normal hearing mechanism. Unlike the mechanisms of normal hearing that incorporate the peripheral (outer, middle, and inner ears) and central mechanisms, CIs bypass many of these auditory structures and directly stimulate auditory nerve fibers along the electrode array. This causes a loss of the significant roles that the outer and middle ears play as frequency-specific amplifiers. This loss of auditory processing ability is often compensated for by breaking the broad concept of auditory ability into spectral/temporal resolution. This segmentation strategy is especially useful when closely examining the subparts of auditory processing ability and coming up with a solution targeting the specific auditory ability that is diminished for people with hearing loss.

Spectral resolution refers to one's sensitivity in detecting fine acoustic changes in the frequency domain. CI users are known to have very poor spectral resolution for several reasons. Physiologically, neural excitation patterns of electrical hearing are broader than those of acoustic hearing (Macherey and Carlyon, 2014), resulting in poor frequency sensitivity caused by overlapping auditory filters for CI users. Functionally, a CI system primarily extracts and transfers temporal-envelope cues from band-pass filters, and temporal fine structure cues in the speech signal are typically lost. Although it has been established that envelope cues alone can transfer sufficient information for speech perception in quiet, the role of temporal fine structure cannot be disregarded considering its significant contribution to pitch perception (Smith *et al.*, 2002; Oxenham *et al.*, 2004) and speech perception in noise (Lorenzi *et al.*, 2006). In addition, up to 22 electrical channels in CI systems that substitute for thousands of inner and outer hair cells may not be enough to deliver fine frequency information. Increasing the number of electrodes could rather result in the adverse effect of channel interaction in certain circumstances such as with monopolar current. These physiological challenges combine with the technical impossibility of designing an equal number of the auditory filter bands typical of normal acoustic mechanism to cause poor spectral resolution for CI listeners. For this reason, spectral degradation is often thought to be a greater issue than deficiency in temporal resolution for CI users. That is, it is typically easier to restore normal-like temporal resolution by increasing stimulation rates (Shannon *et al.*, 2011). As a result, CI users often take advantage of temporal cues to compensate for their poor spectral resolution for phonetic perception (Winn *et al.*, 2012).

The ability to resolve or segregate temporal variances in a stream of sound is called temporal resolution. CIs use a train of biphasic pulses as the carrier of envelope cues to transmit acoustic information. Theoretically, higher stimulation rates of electrical pulses are beneficial, as fine temporal modulation information can be delivered to listeners. However, many cases have been reported where CI users cannot take advantage of these higher stimulation rates for improving speech perception (Fu and Shannon, 2000b; Vandali *et al.*, 2000; Shannon *et al.*, 2011), consistently resulting in poor temporal resolution. The reasons behind this are assumed to be due to the characteristics of auditory nerve firing patterns in response to electrical pulse trains used in CIs: (1) absolute refractory periods and resting potentials of neural firing patterns do not fit the fast rates of electrical stimulation and (2) a train of biphasic pulses is not appropriate to provide exact timing information because it consists of two opposite polarities that cause action potentials with different latencies (fixed rate stimuli primarily deliver envelope information) (Macherey and Carlyon, 2014).

These two aspects of auditory processing capacity are highly associated with speech perception performance. For this reason, many CI studies have used them in relation to speech recognition (Shannon *et al.*, 1995; Fu and Shannon, 2000a; Nie *et al.*, 2006; Xu and Zheng, 2007). Viewed in this light, the lack of such auditory processing capacities for CI users is thought to be an important contributing factor to the variance seen in their speech perception outcomes. Here, we included spectral and temporal resolution as predictive factors to examine the degree of contribution that these auditory processing abilities make for speech perception in CI users.

B. Variance in performance for CI users

One of the issues that always follows when we discuss listeners' performance with CIs is the enormous variability seen in their speech perception performance. CIs do not provide equal benefit in all users. Some CI recipients show nearly normal performance exceeding expectations, while the performance of others is so poor that some of them do not wear their CI. Those who cannot benefit from CIs challenge surgeons and clinicians to improve their procedures for better outcomes. Thus, determining, or even quantifying, factors that predict perceptual benefits from CI surgery is clinically crucial in establishing realistic expectations and rehabilitation strategies for CI recipients. In fact, a large number of studies has been conducted to address this issue by looking at the correlation between speech perception performance and surgical, demographic, psychophysical and cognitive variables. Overall, the same point has been made from previous studies that the duration of deafness is certainly one of the most critical factors determining performance with implantation (Blamey *et al.*, 1996; Daya *et al.*, 1999; van Dijk *et al.*, 1999; Gordon *et al.*, 2000; Green *et al.*, 2007; Holden *et al.*, 2013). Gordon *et al.* (2000) noted that all CI children who had extremely poor open-set word recognition scores had experienced deafness longer (implanted at ages beyond 5 years) than the control group. It is typically assumed that children implanted prior to two years of age have better speech and language performance compared to those who were implanted at a later age (Kirk *et al.*, 2000; Kral and Sharma, 2012; May-Mederake, 2012; O'Donoghue *et al.*, 2014). According to May-Mederake (2012), most children implanted at less than two years of age performed as well as or better than their control peers in speech and grammar development outcomes. The long duration of auditory deprivation causes delays in speech and language development and reconstruction of neural circuits in the brain (Faulkner and Pisoni, 2013).

Such pre-implant factors, however, cannot fully account for limited speech perception outcomes in CI individuals. Other factors, such as communication mode, audibility, etiology, habilitation and cognitive function also have been found to contribute a significant amount of variance in speech recognition performance in CI patients (Pisoni *et al.*, 1999; Collison *et al.*, 2004; Holden *et al.*, 2013; Schafer and Utrup, 2016). Geers *et al.* (2011) evaluated performance across a variety of domains for 112 CI teenagers by comparing outcomes obtained when they were in elementary grades with when they were in high school. They found that variability in speech/language outcomes was accounted for by neurocognitive processing measured with verbal rehearsal speed. They also emphasized the use of spoken language as a communication mode, suggesting that oral communication is positively correlated with verbal rehearsal skills and speech perception. In this study, CI users' audiologic/demographic variables were investigated to examine their effects on variability in speech perception performance.

C. Cognitive function as a contributing factor

Over the past few years, the effect of central cognitive function on CI users' speech perception has received much attention (Pisoni and Geers, 2000; Pisoni and Cleary, 2003b; Collison *et al.*, 2004; Burkholder and Pisoni, 2006). Cognitive function encompasses broad brain activities including memory, learning, judgment, reasoning, attention, language comprehension and production. In the field of cognitive science, a working memory model is frequently referenced to account for the cognitive information processing system. The notion of working memory stems from the concept of short-term memory (Atkinson and Shiffrin, 1968) that only refers to the short-term storage of information. From this simple point of view, short-term memory has been developed into working memory, extending its definition to include capacity to encode, store, and manipulate information. Among several models of working memory (Daneman and Carpenter, 1980; Hasher and Zacks, 1988; Caplan and Waters, 1999), the theory proposed by Baddeley and colleagues (Baddeley and Hitch, 1974; Baddeley, 1992) is considered as one of the most influential when discussing working memory. According to their multicomponent model, working memory functions by multiple components: central executive, phonological loop, and visuospatial sketchpad. As an active memory system, the central executive function is responsible for controlling attention to sensory inputs and regulating or coordinating "slave systems" that provide short-term storage of information. The visuospatial sketchpad is one of the slave systems responsible for storing visual and spatial information. The other slave system, the phonological loop that stores phonological information, is composed of two systems: a phonological input store and an articulatory rehearsal process. This model has been extended by an additional component, known as the episodic buffer, that links working memory and long-term memory (Baddeley, 2000). The episodic buffer integrates information from other slave systems and forms a combined unit such as the short story or a scene.

When assessing working memory associated with people with hearing loss, the concept of phonological loop can be frequently applied. In everyday life, linguistic information is encoded by sensory organs, and then the phonological information is rehearsed and stored in ones' memory. People with hearing loss whose auditory sensory functions are diminished may have a problem with making use of such phonological representations of input information. One of the more useful measures that has been used for assessing working memory and representing the phonological loop is digit span tests in which a participant is required to repeat a series of digits in designated order. A forward digit span test simply requires the participant to recall the series of digits, while a backward digit span test requires the participant to recall the digits in reverse order. Therefore, compared to the forward digit span test, backward digit span is thought to involve executive-attentional resources (Elliott *et al.*, 1997), reflecting more processing sources in working memory. Evidence has shown that outcome performance on digit span tests for CI subjects is poorer than that of their normal-hearing counterparts (Pisoni *et al.*, 2011; Geers *et al.*, 2013), and digit span tests are highly correlated with speech recognition in CI children (Pisoni and Geers, 2000; Pisoni and Cleary, 2003a). We included forward and backward digit span tests to examine working memory capacity for CI users in relation to their speech perception performance. The tests presented stimuli auditorily and visually to investigate working memory load with and without the detrimental effect of hearing loss.

D. Speech Intelligibility Index (SII) as a predictive factor

Previous studies have shown that better aided thresholds for CI users were significantly correlated with higher speech recognition performance, emphasizing the importance of a wider dynamic range that increases audibility in the CI map (Firszt *et al.*, 2004; Holden *et al.*, 2013). CIs are certainly more beneficial than hearing aids in terms of audibility. CIs allow clinicians to provide high frequency gains that sometimes are unavailable with hearing aids due to acoustic feedback or technological limitations. However, clinicians programming a CI speech processor frequently fit CIs based on the patient's loudness comfort rather than the consideration of audibility. This causes variability in aided thresholds across frequency as well as among CI users. Given that aided audibility is a contributing factor to speech perception outcomes, investigating the traditional speech intelligibility model that predicts speech perception outcomes is worth consideration.

Over 60 years ago, scholars working at a telephone laboratory explored a way to quantitatively represent a listeners' speech intelligibility, and developed the model of articulation theory (French and Steinberg, 1947). Over the years, the computation procedure has been enhanced and supplemented, such that the Articulation Index has been renamed the Speech Intelligibility Index (SII) (ANSI, 1997). This model considers audibility (Ai) and frequency importance functions (*Ii*) as key components to predict speech intelligibility [Eq. (1)].

$$SII = \sum_{i=1}^{n} IiAi \,, \tag{1}$$

The amount of speech energy available to listeners and the relative importance weights for each frequency band, respectively, are multiplied, and all outcome values are summed to calculate SIIs ranging from 0 to 1. The SII unit, however, does not solely account for speech recognition outcomes. To predict speech perception scores using SIIs, a transfer function that establishes the relationship between the SII and speech perception scores is required. The SII model typically displays high accuracy in prediction of speech perception scores in individuals with normal hearing (Pavlovic *et al.*, 1986; Sherbecoe and Studebaker, 2003), whereas incorporation of correction values, known as hearing loss desensitization factors, is recommended when calculating the SII for individuals with hearing impairment. In other words, there are additional factors that affect speech recognition beyond audibility for those with hearing loss. Several correction factors, known as Hearing Loss Desensitization (HLD) factors, that compensate for such supra-threshold deficits, have been developed and proposed (Fletcher and Galt, 1950; Pavlovic *et al.*, 1986; Studebaker *et al.*, 1997; Ching *et al.*, 1998; Studebaker *et al.*, 1999). Such correction factors associated with hearing thresholds of individuals with mild-to-moderate hearing loss have improved the accuracy of SII prediction to some extent. However, these modified and refined SIIs have not improved predictive accuracy for people with hearing loss greater than moderate impairment (Pavlovic *et al.*, 1986; Ludvigsen, 1987; Ching *et al.*, 1998). Given this limited success of the SII in the severe hearing loss group, it is reasonable to assume that applying an SII application to CI users may be met with limited success due to the considerable degree of hearing loss (nearly deaf) that typical CI users have. Furthermore, significantly deteriorated auditory processing and large individual variability in speech recognition outcomes among CI users could make it impossible to use such SIIs as a predictable tool for speech perception performance. Despite of these concerns, and in light of technological advancements seen in modern CIs that provide much more audibility across frequency, it is worth attempting to examine the feasibility of the SII to predict speech perception performance in CI users. This study examined the application of the SII for prediction of speech perception outcomes for CI users.

E. Aim of the study

Despite the widespread usage of SIIs for hearing aids and other related areas, little attention has been paid to the application of SIIs for CI patients. As noted above, the lack of SII studies with CI patients may be attributed to the significant hearing loss that typical CI individuals have and the provision of distorted electrical signals that CIs provide. Moreover, individual variability and heterogeneity typically observed in a CI population may also be a primary reason for this scant attention. As a result, this study attempted to use SIIs to predict CI users' speech perception outcomes. We designed this study in a way to partially replicate the methodology in a study by Humes (2002) that used two indirect approaches for predicting the aided and unaided speech perception ability of elderly hearing aid users: (1) calculation of the SII and (2) predictive factor analysis using a regression model. Thus, in addition to calculating SII, we also included other predictive factors that affect speech perception capabilities in a group of CI adults.

Few studies have examined the usefulness of the SII in predicting speech perception ability in CI users. Despite the likelihood that the SII would fall short of this goal, this study sought to provide evidence for the use of the SII with this population. First, a transfer function curve that established the relationship between SIIs and speech perception scores was used to determine if the SII could serve as an effective tool for predicting speech perception performance for this population. Then, we examined other predictive factors. Adult CI users' demographics, auditory processing ability, and working memory load were explored using a multiple regression analysis. Although the SII is primarily determined by the audibility and FIF, it has been improved with considerations of other psychoacoustic or perceptual effects (e.g., masking effects, level distortion effect, HLD, and age). Hence, examining such variables may provide us with some implications for a model for improving the accuracy of the SII.

II. METHOD

A. Participants

Fifteen CI adults ranging in age from 22 to 73 years (M = 53.13, SD = 17.27) were invited to participate in the current study. Younger than 80 years of age, experience with CI device(s) for at least 6 months, and American English as the first language were the inclusion criteria for participation. All participants had severe-to-profound sensorineural hearing loss with bilateral pure-tone averages (PTA, average loss at 0.5, 1, and 2 kHz) greater than 70 dB HL. The group mean PTA for left ears was 95 dB HL (*SD*: 7.2 dB HL) and right ears was 97 dB HL (*SD*: 4.8 dB HL).

The CI listeners signed informed consent forms, and all were paid for their participation after competing all procedures. The protocol employed in this research was approved by The University of Memphis Institutional Review Board. Participants completed a questionnaire addressing patient demographics and hearing history. Some of these demographics, such as duration of hearing loss, were used later for the regression analysis. Table 6 represents the demographic details of the CI participants.

Table 6. Demographic details of CI participants.

N	Gender	Age (years)	Onset of hearing loss (years)	Duration of deafness (years)	Etiology	CI Manufacturer and model	Level of education	Uni/ Bi-lateral CI	Communication mode
1	Female	41	13	0	Unknown	Cochlear Nucleus	Master’s degree	Uni	oral/speaking, lip reading
2	Male	73	40	0	Noise exposure	Cochlear Nucleus	Master’s degree	Uni	oral/speaking, lip reading
3	Male	63	46	13	Unknown	Cochlear Nucleus	High school	Uni	oral/speaking, lip reading, writing
4	Male	30	7	3	Unknown	Cochlear Nucleus	Doctoral degree	Uni	oral/speaking, lip reading
5	Female	56	41	7	Meniere’s	Cochlear Nucleus	Master’s degree	Uni	oral/speaking, lip reading
6	Female	65	51	10	Unknown	Cochlear Nucleus	Master’s degree	Uni	oral/speaking
7	Female	56	5	50	Nerve damage	Cochlear Kanso	Bachelor's degree	Uni	oral/speaking, lip reading
8	Male	73	41	31	Noise exposure	Cochlear Kanso	Master’s degree	Uni	oral/speaking, lip reading
9	Female	40	15	14	Brain tumor	Cochlear Nucleus	Master’s degree	Uni	oral/speaking
10	Female	59	0	55	Rubella	Medel Opus	Bachelor's degree	Uni	oral/speaking, lip reading, sign language
11	Female	65	56	2	Unknown	Cochlear Nucleus	Doctoral degree	Bi	oral/speaking, lip reading
12	Female	64	42	21	Meniere’s	Cochlear Nucleus	High school	Uni	oral/speaking, lip reading
13	Male	24	2	2	Illness	Advance Bionics Harmony	Master’s degree	Uni	oral/speaking, lip reading
14	Female	22	2	0	Meningitis	Advance Bionics Harmony	Bachelor's degree	Uni	Oral/speaking
15	Male	66	50	0	Meniere’s	Cochlear Nucleus	Bachelor's degree	Uni	Oral/speaking

B. Audiometric Testing

Aided and unaided audiometric tests were conducted to verify hearing thresholds and audibility with and without their CIs. Hearing thresholds were obtained at octave frequencies from 250 to 8000 Hz, and inter-octave frequencies (125, 750, 1500, 3000, and 6000 Hz) were also confirmed. Aided audiometry was carried out in a free-field condition with participants seated in the center of a double-walled sound booth meeting ANSI standard S.31-1999 (ANSI, R2013), facing the front speaker 1 meter away. In the case of bimodal participants who wore a CI on one ear and a hearing aid on the other ear, the hearing aid was removed during the test. This rule was applied to other experiments in the study as well. Unaided audiometry was conducted using pure tones presented through TDH-39 headphones.

C. Speech recognition test

CI listeners' speech recognition was measured using the AzBio Sentence Test. The AzBio test is one of the standardized speech perception tests in the Revised Minimum Speech Test Battery that was designed to be used with CI patients. The AzBio stimuli are produced by two male and two female speakers that can be presented in quiet or in noise (10-talker babble). Each participant listened to three AzBio sentence lists in three test conditions presented in the sound field. (1) one sentence list presented at a level of 65 dB SPL in quiet, (2) one sentence list presented at a level of 65 dB SPL with a +5 dB SNR, and (3) one sentence list presented at a level of 65 dB SPL with a +10 dB SNR. The level of the speech was fixed at 65 dB SPL, while noise levels were varied depending on the desired SNRs. Each CI listener was seated in the middle of the double-walled sound booth meeting ANSI standard S3.1-1999 (ANSI, R2013), 1m away from the speaker, wearing his/her CI device. The listener's job was to repeat the sentences or words they heard. Among 15 lists available in the AzBio test, lists 8, 9, 10, 11, 12 and 13 were chosen as they were equally difficult based on results from a previous study (Bush, 2016). Among these lists, three were randomly selected and presented in the different conditions. Each list consists of 20 sentences that contain a different number of target words per sentence. The tests were scored in percentage based on the number of correctly repeated target words by the listeners.

D. SII calculation

The SII was computed in the following way. For each CI patients' aided thresholds to be used in the SII calculation, equivalent hearing threshold levels needed to be established. The aided audiometric thresholds measured in dB HL were converted to dB SPL by adding the minimum audible field (MAF) values (Bentler and Pavlovic, 1989). Critical ratios (Pavlovic, 1987) and bandwidth adjustments (ANSI, 1997) were further used to transform the obtained thresholds into equivalent hearing threshold levels. The equivalent hearing threshold levels were eventually used in the SII equation. To yield 1/3 octave pure-tone thresholds that could not be obtained from the audiometric procedures, interpolation or extrapolation was used.

As noted earlier, two key components, audibility and frequency importance functions, need to be established for calculation of the SII. For computation of audibility in the three speech recognition task conditions, the Long Term Average Speech Spectrum (LTASS) of the AzBio lists and its noise were measured. To this end, the overall rms level of 65 dB SPL and the levels of the concatenated speech and noise were measured separately using a Bruel and Kjaer Type 2250 sound level meter. Figure 10 shows the band specific levels in LAeq across the 1/3 octave band frequencies. The shape of the speech and noise spectra reflect nearly identical patterns across the frequencies.

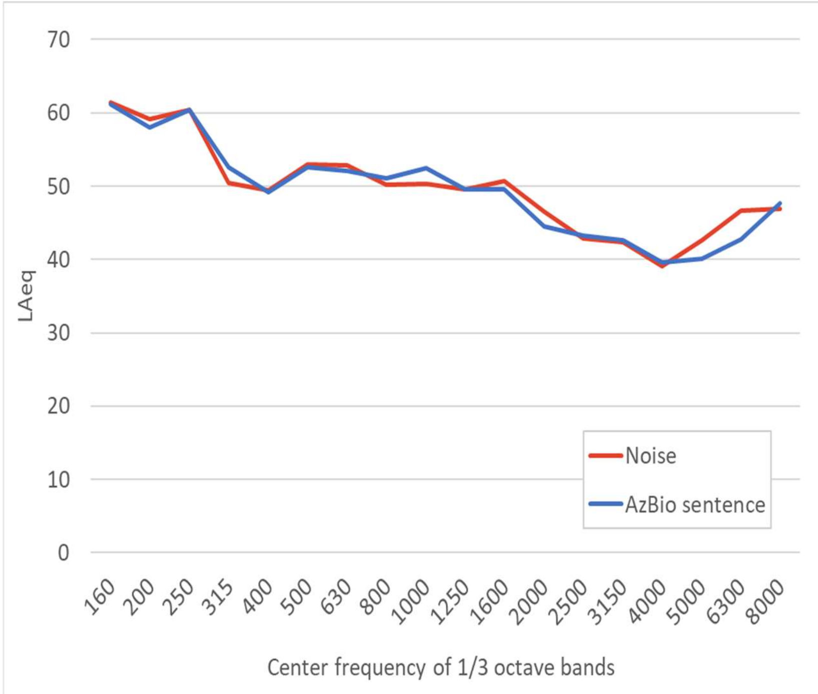

Figure 10. LTASS of AzBio sentences and noise across 1/3 octave band frequencies.

The audibility function (A*i*) was calculated by subtracting the larger one of either the long-term noise levels or the thresholds, from the speech peaks in each band, and dividing the difference by 30 dB. Table 7 shows the frequency importance function (FIF) for the AzBio sentences that were derived in our prior study (Lee et al., In Press). This band weight information of the AzBio sentences was applied in the SII calculation using Equation 1. The entire procedure of computing SII values followed the ANSI standard (ANSI S3.5-1997), which takes into consideration masking effects and a level distortion factor. After the SII calculation, an age correction factor proposed by Studebaker *et al.* (1997) was multiplied for those who were older than 70 years. This correction was applied to reflect the tendency of speech perception scores that decline with age. The SII algorithm was generated in an Excel program for all of the SII calculations.

Table 7. FIF across the 1/3 octave center frequencies (CF).

CF	160	200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000	6300	8000
FIF	2.20	2.41	1.74	1.35	3.08	5.42	8.71	7.43	7.35	9.72	10.21	9.13	9.65	7.91	6.65	3.65	2.56	0.84

E. Auditory processing tests

It is reasonable to assume that supra-threshold deficits associated with poor speech recognition in the hearing impaired are attributed to abnormal spectral and temporal resolution (Pavlovic et al., 1986). Assessing auditory processing ability that measures spectral and temporal resolution could be beneficial in terms of understanding these deficits in supra-threshold sound processing that cannot be accounted for by the SII. In addition, these psychophysical assessments could serve as predictive variables, along with the CI patients' demographics that potentially contribute to predicting speech perception performance for these listeners. For measuring auditory processing, a Gap Detection Test (GDT) and Spectral-temporally Modulated in Ripple Test (SMRT) were administered to determine each listener's temporal and spectral resolution, respectively. The auditory processing tests were administered twice for each participant, and the average of the two performances was used for subsequent statistical analysis.

PsyAcoustX (Bidelman *et al.*, 2015), which is a Matlab based platform allowing several auditory tests with 3 alternative forced choice (3AFC), was used to implement the GDT. Three successive broad band noises, 500 ms each, were presented at 65 dB SPL through a loud speaker located 1 meter and at 0° azimuth from the listener. One out of the three stimuli was designed to have a short silent gap, whereas the other two were continuous broad band noise. The durations of the short gap were varied depending on listener's response following a 2 down/1 up adaptive tracking rule to determine ones' GDT threshold with 71% consistent criterion performance level (Levitt, 1971). Starting gap duration was set to 10 ms. Each listener was asked to click a button on the screen to reflect a stimulus that sounded different from the other two stimuli, detecting a brief gap that divided two successive stimuli.

The Spectral Ripple Test, that uses stimuli containing a different number of spectral peaks at a particular modulation depth, is one of the most commonly used approaches to evaluate spectral resolution in modern CI studies (Henry and Turner, 2003; Henry *et al.*, 2005). Won *et al.* (2007) found that better spectral ripple discrimination was significantly correlated with better speech perception in noise and quiet for CI users. In our study, CI patients' spectral resolution was estimated using the SMRT software version 1.1 (Aronoff and Landsberger, 2013). The SMRT is a newly developed ripple test that was created to compensate for drawbacks found in the original test (local loudness cues and upper/lower frequency boundary cues). The SMRT was designed to seek the largest number of ripples per octave (RPO) that could be reliably detected by listeners. The test uses an adaptive procedure (1up/1down). Like the GDT, the SMRT was administered with 3AFC presenting stimuli at 65 dB SPL. The ripple density of reference stimuli was 20 RPO, and the target stimuli were adjusted starting from 0.5 RPO with a step size of 0.2 RPO. The trial ended when 10 reversals were found, and the mean of the last 6 reversals was reported as the RPO threshold. Listeners were instructed to click a button on the screen that reflected the stimulus that sounded different from the other two stimuli, discriminating the spectrally different sound.

F. Cognitive function tests

For the cognitive measure, the Digit Span Test (DST) designed to evaluate working memory function was administered. The DST is a subset of Wechsler Adult Intelligence Scale-Revised (WAIS-R), which assesses comprehensive cognitive ability for adults and consists of six verbal subtests (Information, Comprehension, Arithmetic, Digit Span, Similarities, and

Vocabulary) and five performance subtests (Picture Arrangement, Picture Completion, Block Design, Object Assembly, and Digit Symbol). In the DST, a participant is required to memorize a series of numbers presented either visually or auditorily, and repeat the correct numbers in the correct order. Outcomes from auditory DSTs may not reflect pure attention and memory deficits, as it is significantly influenced by hearing deficits for individuals with hearing loss. The DST can further be divided into two tasks depending on the answering method. The forward task asks participants to answer in the presented order, whereas the backward task requires listeners to answer in reverse order. We used both visual and auditory modalities and both forward and backward responses to compare the functional difference in short-term/working memory. To implement the test, Inquisit computer software (Draine, 1998) was used. For the visual DST, a sequence of numbers was shown on a computer screen and then disappeared. For the forward DST, participants were instructed to click the correct digits on the monitor in the correct order. For the backward DST, they were instructed to click the correct digits in the reverse order. The auditory DST was administered in the same way, but the sequence of numbers was presented from the front speaker at 65 dB SPL, instead of the monitor screen. The digit string was increased in length with each trial until the participant was unable to remember the correct numbers in the correct sequences. The maximum lengths of the correct numbers that the CI users remembered in a correct order were yielded from the tasks as a final outcome.

III. RESULTS

A. Prediction of speech perception scores using SIIs

Mean speech perception scores for three different SNR conditions (Quiet, SNR 5, and SNR 10) are shown in Figure 11. The CI listeners had particular difficulty under the noise conditions compared to the quiet condition. The scores were drastically decreased when background noise was presented along with the AzBio sentences. A one-way repeated measures Analysis of Variance (ANOVA) determined that mean speech perception scores differed significantly between the presentation conditions of the AzBio sentences [F (2, 25.717) = 112.893, $p < 0.001$]. Post hoc tests using the Bonferroni correction revealed that all pairs of conditions were significantly different from each other, indicating that an increase in the amount of background noise resulted in a statistically significant decrease in speech perception scores ($p < 0.001$). In addition, the extended boxes and whiskers in Figure 11 suggested that individual differences in speech perception performance were tremendously large.

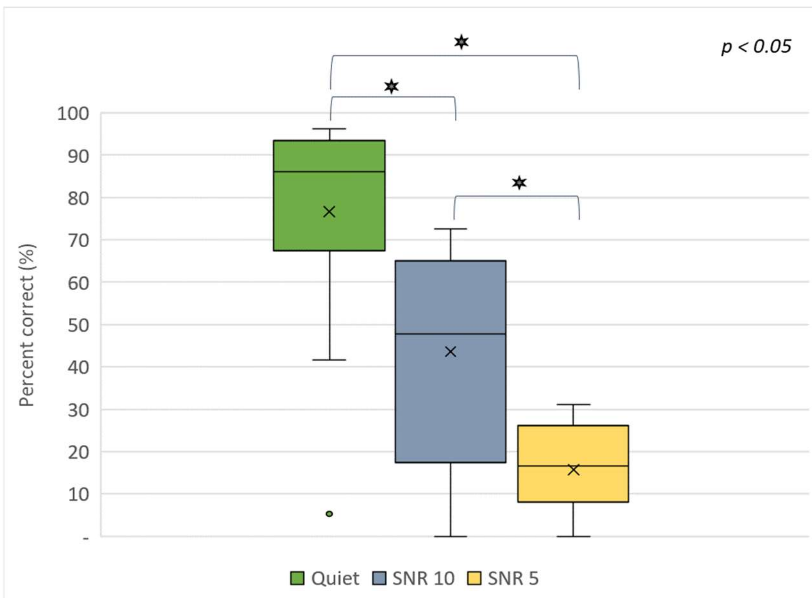

Figure 11. Mean speech perception scores (x) for the AzBio sentences in the three different SNR conditions.

To investigate the predictive role of the SII for speech perception performance, a transfer function curve that established the relationship between predicted scores using SIIs and observed scores needed to be considered. We made reference to the transfer function equation (Eq. 2) that was derived in our first study (Lee et al., In Press). The transfer function equation was obtained from listeners with normal hearing who were administered the AzBio test in a variety of filtering/SNR conditions. The fitting constants, Q (0.287) and N (5.206), in Equation 2 resulted in a good fit between the observed scores and predicted scores (RMS error = 0.069 and R^2 = 0.923). With the appropriate reference transfer function for normal hearing listeners as a normative point, our speech perception data and corresponding SIIs for CI listeners were examined.

$$Speech\ perception\ score = (1 - 10^{-(SII/0.287)})^{5.206}, \tag{2}$$

Figure 12A provides the scores-vs-SII transfer function for the AzBio test derived from listeners with normal hearing (solid line). Additionally, the SII values and corresponding speech perception scores obtained from this study are represented by circles in blue for the +10 dB SNR condition, gold for the +5 dB SNR condition, and green for quiet. Regardless of test condition, all obtained scores fell considerably below the predicted scores using the transfer function curve, suggesting that the transfer function curve for listeners with normal hearing is not capable of predicting speech perception scores for CI listeners using this SII model. In an attempt to address the issue of overestimation by the conventional SII calculation, we applied a HLD factor to the SII calculation. Among several HLD models, we adopted an equation similar to the one developed by Sherbecoe and Studebaker (2003) (Eq. 3). In the equation, the PTA is the average hearing loss of the better ear at 1, 2, and 4 kHz. This correction factor is applied by multiplying the SII values with the calculated correction factors.

$$Hearing\ loss\ desensitization = 1 - \left(\frac{PTA}{108.3072}\right)^3, \tag{3}$$

Paired-sample t-tests were conducted to compare the conventional SII values and SII values with HLD correction (HLD SII) for the three conditions. There were statistically significant differences in the SII values with and without HLD corrections for all conditions

[Quiet: t (14) = -21.958, $p < 0.001$, SNR 10: t (14) = -23.743, $p < 0.001$, and SNR 5: t (14) = -23.760, $p < 0.001$]. The comparisons showed that the HLD corrections significantly reduced the original SII values for all listening conditions.

In contrast to the predictions from the conventional SII that excluded the influence of the hearing loss, the SII calculations using the HLD corrections resulted in lower SII values compared to the normative transfer function predictions in most cases (Figure 12B). A considerable number of the observed SIIs were higher than the predicted SIIs, suggesting the normative transfer function curve was not suitable for predicting performance for CI listeners even when the influence of hearing loss was considered. As was the case for SIIs calculated without the HLD factor, the SIIs with the HLD factor in Figure 12B also displayed large individual variation in speech perception.

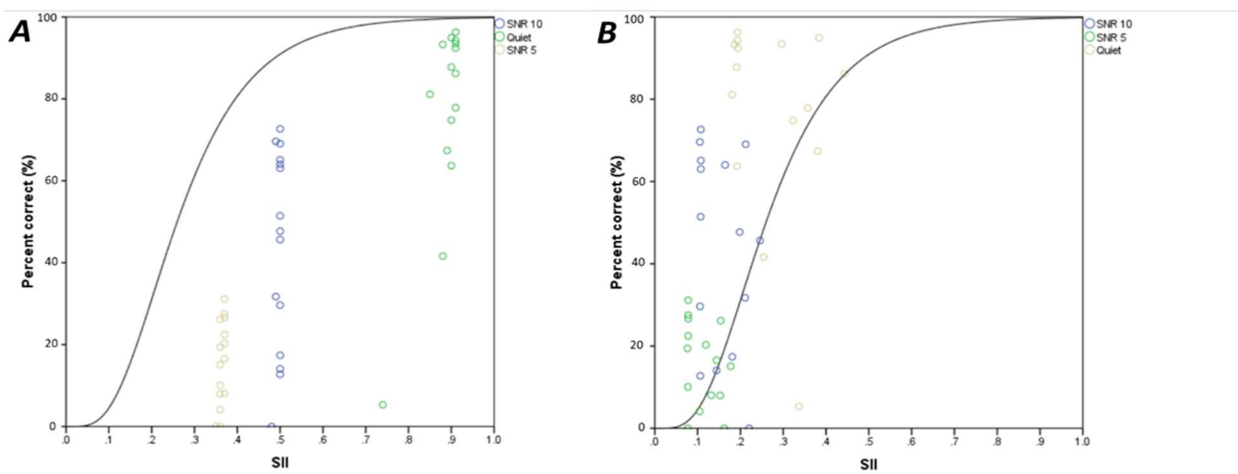

Figure 12. (A) Score-vs-SII distribution scatter-plots with the reference transfer function curve. (B) Score-vs-HLD SII distributions scatter-plots with the reference transfer function curve.

B. Prediction of speech perception scores using multiple variables

Prior to the multiple regression analysis, it was necessary for our dependent variable, speech perception scores, to be measured in three different conditions to be combined as a single value. To this end, the scores obtained from the three conditions were averaged, yielding one dependent variable. Furthermore, independent variables that were likely to be associated with speech perception scores were determined and defined clearly for the analysis. In the demographic data obtained from the questionnaires, age, onset of hearing loss, and duration of deafness were chosen as predictive variables. Onset of deafness was defined as the age at which hearing loss occurred. Duration of deafness was defined as the period between when hearing loss occurred and CIs were activated. In order to represent only the effects of hearing loss, ruling out the potential benefits of audibility from the better ear or hearing aids, these two demographic variables were primarily defined based on the history of the poorer ear. In our experimental data, the outcomes from the SMRT, GDT, visual DST, and unaided/aided audibility were selected as predictive variables. Only working memory outcomes from the DST for visual presentations were included to minimize the effects of hearing loss for auditory DSTs. The DST values used for the analysis were the mean of two visual DSTs (forward and backward DTS). Unaided and aided audibility values were determined by averaging auditory thresholds at 0.5, 1, 2, and 4 kHz in the better ear. If CI patients could not detect the pure tones at the highest level presented in the unaided condition, the thresholds were regarded as 100 dB at that frequency. Lastly, the HLD SII and SII were included as predictive variables. Like the speech perception scores, HLD SIIs and SIIs obtained in the three different conditions were averaged, and individual HLD SII and SII values were respectively obtained for the 15 CI participants. All of the variables used for the analysis are shown in Table 8.

Table 8. Nine predictive variables and one dependent variable (speech perception test scores) for the multiple regression analysis.

N	Onset of hearing loss (years)	Duration of deafness (years)	SMRT (RPO)	GDT (ms)	Unaided audibility (dB HL)	Aided audibility (dB HL)	Visual DST	Speech perception test scores (%)	HLD SII	SII
1	13	0	1.60	2.22	100.00	22.50	7	62.51	0.13	0.59
2	40	0	3.63	5.25	87.50	26.25	4.5	48.98	0.29	0.59
3	46	13	1.92	15.10	93.75	27.50	8	19.94	0.17	0.58
4	7	3	1.57	4.19	100.00	30.00	8.5	60.77	0.12	0.58
5	41	7	1.98	3.48	92.50	21.25	7	59.26	0.19	0.59
6	51	10	1.28	9.62	97.50	26.25	5.5	25.49	0.13	0.59
7	5	50	1.48	5.19	75.00	22.50	7	33.42	0.21	0.59
8	41	31	3.15	9.76	88.75	32.50	7	40.28	0.12	0.57
9	15	14	4.13	5.39	67.50	22.50	6	47.37	0.23	0.59
10	0	55	0.51	34.88	86.25	43.75	4.5	1.77	0.24	0.52
11	56	2	1.63	8.30	100.00	20.00	6	57.73	0.13	0.59
12	42	21	6.30	12.09	90.00	25.00	5	35.69	0.25	0.58
13	2	2	1.13	2.86	95.00	22.50	6	59.15	0.13	0.59
14	2	0	1.70	1.71	100.00	18.75	6	64.18	0.13	0.59
15	50	0	2.93	10.64	88.33	23.75	6	63.38	0.25	0.59

Pearson correlations and stepwise multiple linear regressions were conducted to investigate the relationship between the variables and the prediction model of speech perception scores for the CI users based on their demographic, auditory processing, and cognitive function variables. The software program IBM SPSS (version 24) was used. The inter-correlations of the variables are shown in Table 9. As can be seen, some correlations for several pairs of variables were statistically significant. There were strong, negative correlations between speech perception scores and duration of deafness, aided audibility, and GDT ($r > 0.7$). Increases in speech perception scores were correlated with decreases in duration of deafness, aided hearing threshold represented as unaided audibility, and GDT thresholds.

Table 9. Pearson correlation for 10 variables.

Variable	1	2	3	4	5	6	7	8	9	10
Speech perception score 1	1	-0.02	-0.767*	0.306	-0.713*	0.08	-0.815*	0.265	-0.272	0.697*
Onset of hearing loss 2		1	-0.29	0.167	-0.162	0.367	0.015	-0.08	0.118	0.281
Duration of deafness 3			1	-0.534*	0.626*	-0.088	0.647*	-0.144	0.261	-0.689*
Unaided audibility 4				1	-0.104	-0.362	-0.149	0.203	-0.624*	0.103
Aided audibility 5					1	-0.171	-0.836*	-0.169	0.175	-0.921*
SMRT 6						1	-0.126	-0.27	0.516*	0.22
GDT 7							1	-0.377	0.341	-0.907*
Visual DST 8								1	-0.567*	0.277
HLD SII 9									1	-0.189
SII 10										1

*$p < 0.05$

The stepwise multiple regression yielded a prediction model. The prediction model contained two out of the nine predictors with seven variables excluded. The significant regression equation was found [F (2, 12) = 19.343, $p < 0.001$], and this model accounted for approximately 76% of the variance of speech perception scores ($R^2 = 0.763$, adjusted $R^2 = 0.724$). Speech perception scores had significant negative correlations with GDT thresholds and duration of deafness, indicating that CI adults with better speech perception outcomes were expected to have lower GDT thresholds, and shorter duration of deafness. Table 10 shows the regression coefficient summary.

Table 10. Summary of regression coefficients.

Model	B	Std. Error	β	Sig.	Pearson-r
Constant	62.208	3.799			
GDT	-1.256	0.421	-0.550	0.011	-0.815
Duration of deafness	-0.428	0.192	-0.411	0.046	-0.767

The dependent variable was speech perception score, $R^2 = 0.763$, Adjusted $R^2 = 0.724$, $^*p < 0.05$

C. Cognitive function tests

The DSTs administered in four different ways (visual/auditory, forward/backward) were further analyzed to compare differences in performance. Group mean maximum lengths of numbers correctly answered by 15 CI users are shown in Figure 13 [visual-forward DST (*M*: 6.6, *SD*: 1.18); visual-backward DST (*M*: 5.93, *SD*: 1.27); auditory-forward DST (*M*: 5.4, *SD*: 1.05); auditory-backward DST (*M*: 5.06, *SD*: 1.53)]. A two-way repeated measures ANOVA was conducted with the two presentation modalities (visual and auditory) and two reproduction orders (forward and backward) as the two within-subject variables, and the maximum length of numbers as the dependent variable. Significant main effects were found for the presentation modalities [$F\ (1, 14) = 14.252$, $p < 0.05$] and reproduction orders [$F\ (1, 14) = 6.563$, $p < 0.05$] with no significant interaction effect between the two [$F\ (1, 14) = 1.094$, $p = 0.072$]. CI users performed better on forward DSTs and visual DSTs compared to backward DSTs and auditory DSTs, respectively.

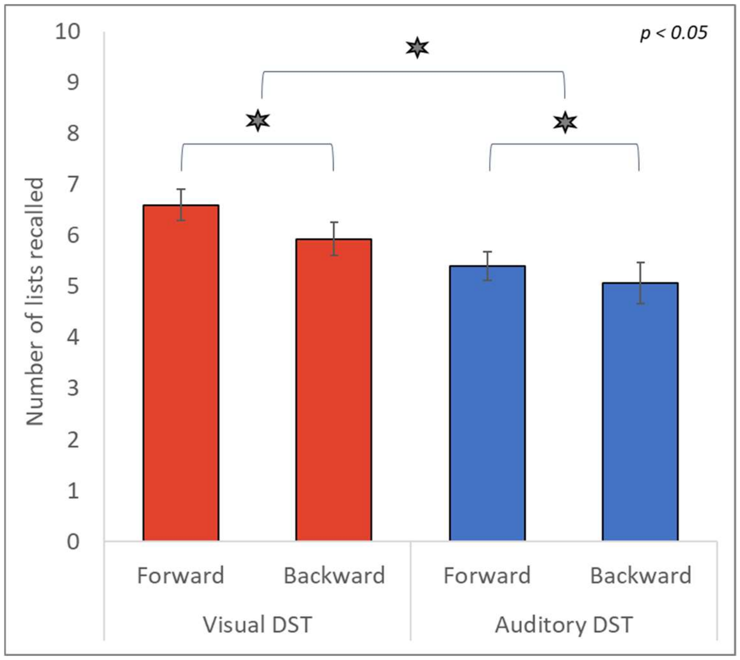

Figure 13. Mean number of lists correctly recalled for forward and backward DSTs presented with two different modalities (visual and auditory). Error bars denote ±1 SEM.

IV. DISCUSSION

A. Prediction of speech perception scores using SIIs

One of our primary goals was to find out whether the SII is an appropriate model to predict speech perception scores for adults who use CIs. We examined the observed SIIs with and without HLD for CI adults in comparison to the transfer function curve yielded by adults with normal hearing. Two trends were apparent from our results. First, the transfer function curve tended to overestimate speech perception scores for CI users when the HLD factor was not taken into account, whereas the curve tended to underestimate speech perception when the HLD factor was applied to the SII calculation. This finding suggests that speech perception ability of CI users cannot be accounted for adequately by the existing SII model based on the transfer function from listeners with normal hearing. It is likely that the severe-to-profound sensorineural hearing loss, the significantly damaged auditory system, and limitations posed by electrical hearing by CIs could be adverse factors affecting the SII prediction, beyond just audibility. When taking HLDs into account, a significant decrease in SII values with HLD corrections for the CI users appears to be the cause that the prediction failed. In general, HLD correction factors for modification of SII models are applied based on the hearing thresholds of the listeners. The amount of the correction factors increased with the greater hearing loss that these individual listeners have. This approach sometimes results in significant negative-weights on SII values for people having very poor hearing thresholds. For example, an HLD factor proposed by Pavlovic *et al.* (1986) provides a desensitization factor of zero when hearing thresholds are above 94 dB HL. This will eventually lead SIIs to 0 as a result of multiplication. This extreme application of correction factors has been criticized by arguing that individuals having thresholds of 94 dB HL would still be able to extract intelligible information from speech above 94 dB HL (Ching *et al.*, 1998). Given the fact that CI candidates have mostly severe to-profound hearing loss and the corresponding correction factors are substantially high, it is not surprising to see the underestimation effect observed in Figure 12B.

The other especially important observation is the significant variability seen in these listeners' speech perception scores. Given the consistent reporting of wide variation in individual speech perception outcomes (Kiefer *et al.*, 1998; Pisoni *et al.*, 1999), the limitation of the SII model to accurately predict speech perception was an anticipated result to some extent. Unless this issue of variability can be addressed, the applicability of SII models will likely not become an ideal tool. Given our findings and previous research, the general applicability of these SII models as a predictive tool for CI users appears impossible at this stage.

B. Prediction of speech perception scores using multiple variables

A total of nine independent variables considered as possible predictors of speech perception scores were included for a correlation and stepwise linear regression analysis. The correlation coefficient of the group data from 15 CI adults showed that duration of deafness, aided audibility, GDT, and SII were significantly correlated with speech perception scores. Unexpectedly, several variables that were thought to be associated with speech perception performance were excluded from the best regression model. Only GDT and duration of deafness were included as significant variables in the model. The GDT factor was the strongest predictor in the model, explaining more than 70% of the variance followed by duration of deafness (54%). Even though duration of deafness has been shown to be the strongest contributor to performance in CI users in other studies, the contribution accounted for by duration of deafness was less than that of GDT in our study. The majority of CI subjects in our study were post-lingually deafened CI users who had experienced hearing deficits after mastering oral communication skills. This CI population tends to show better performance in general and is less affected by duration of hearing loss compared to CI users whose hearing loss occurs prelingually. Thus, the lower significance of the duration of deafness factor that we found may be accounted for by such characteristics of those with post-lingual deafness.

The significant contribution of the GDT indicates that auditory processing abilities, especially temporal resolution, are highly associated with speech perception performance in CI adults. In contrast, the SMRT outcomes which represent spectral acuity did not show robust correlations with speech perception scores. The lack of any significant correlations between the SMRT outcomes and speech perception differs from previous studies that showed significant correlations between these two measures (Litvak *et al.*, 2007; Lawler *et al.*, 2017). It is hard to state reasonable reasons for this inconsistency, but possible explanations for this discrepancy may lie in the considerable degree of variability seen across the participants in this study. Individual differences in performance in CI populations are typically observed in psychoacoustic experiments as well as in speech intelligibility tasks. Even though an attempt was made to control for these effects by attempting random selection of CI participants for this study, the small sample size ($N = 15$) probably could not yield asymptotic performance on the SMRT tasks.

As noted above, it is often stated that temporal resolution is comparatively better than spectral resolution in CI users. Indeed, GDT performance for CI users was comparable to that of listeners with normal hearing (Shannon, 1989; Goldsworthy *et al.*, 2013). Better acuity of temporal cues in CI users probably allowed them to catch up to the rapid stream of speech cues, such as voice onset time, providing reliable data that systemically varied with speech scores. In addition, CI users' poorer spectral resolution likely made it difficult to produce meaningful scores on the SMRTs in relation to speech perception scores. Taken together, our auditory processing outcomes suggested that temporal resolution is more associated with speech perception than spectral resolution for CI users. This observation supports the notion that CI users who are exposed to only limited spectral information, rely heavily on temporal cures for speech perception (Kirby and Middlebrooks, 2010; Winn *et al.*, 2016).

C. Working memory capacity for CI users

To determine whether working memory capacity for CI users was affected by the deficits in auditory/phonological processing components, we administered the DST in two different modalities (auditory and visual presentations). Previous cognitive literature has shown mixed outcomes in terms of the superiority between visual and auditory presentation modalities for working memory tasks. Some studies on human memory have shown that visual memory is superior to auditory memory (Hilton, 2001; Cohen *et al.*, 2009). Hilton (2001) described the reasons that their subjects were better at visual learning for the memory task. She indicated that the visual stimuli, contrary to the auditory stimuli, were stored in two different forms, mental image and repletion, which makes it easier for the brain to reproduce. She also noted that auditory processing may cause more fatigue than visual processing. On the other hand, other studies have claimed an auditory superiority effect with the assumption of higher strength of association between successive auditory stimuli compared to successive visual stimuli (Penney, 1989; Kemtes and Allen, 2008). These researchers also argued that visual stimuli likely give rise to more attentional load relative to auditory stimuli. This controversy can be seen in DST studies using CI subjects who have significant hearing problems. AuBuchon *et al.* (2014) reported slightly higher performance for auditory DST than visual DST in a forward paradigm, but slightly lower performance for visual DST than auditory DST in a backward paradigm. This contradicts the results of Kronenberger *et al.* (2013) where visually presented stimuli resulted in slightly higher reproduction rates over auditorily presented stimuli in forward DSTs. Taken together, while there appears to be no definitive answer on this issue, it is clear that presentation modality plays a minor role in task performance.

Our DST results showed that CI users' performance on working memory tasks was significantly better when they perceived stimuli visually rather than auditorily. The heavy demand on working memory load for processing the auditory stimuli may make it difficult to store the stimuli into short-term memory, resulting in such variance in performance. The other possibility is that the poorer performance on auditory tasks might have been caused by an auditory perception issue, not auditory processing or memory demands. Some CI users having very poor speech perception might have misunderstood the auditory stimuli, substantially affecting the group mean performance on the auditory DST. However, since the DSTs for a normal hearing control group were not measured in this study, and previous studies showed mixed outcomes, it would be absurd to argue that auditory deficits in CI users led to such differences in working memory function. Indeed, studies that assessed listeners with normal hearing and those with CIs using both auditory and visual modalities indicated that the poorer working memory function for CI users compared to those with normal hearing are not solely accounted for by their auditory perception or speech production abilities (Cleary *et al.*, 2001; Kronenberger *et al.*, 2013; AuBuchon *et al.*, 2014). Therefore, further investigations are needed to examine the mechanism of these two modalities in relation to performance on working memory capacity.

D. Limitations

The primary shortcoming of this study is the small number of CI participants ($N = 15$). It is well known that large sample sizes are necessary for examining factors associated with experimental performance for CI users (Schafer and Utrup, 2016), yet it is often difficult to obtain large, random samples for such studies. Speech perception performance in CI individuals varies considerably from person to person, and a large number of variables is associated with such variability. A multiple regression analysis requires a proper sample size to have the desired statistical power needed for the number of predictors used. Our small sample size limited the range of predictive variables that could be included, as well as other potentially important variables, such as surgical- or device-relevant-variables that had to be excluded. Additionally, the small sample size may not have represented a typical CI population. Finding and recruiting CI users is certainly a practical challenge for studies such as these. Nevertheless, an effort should be made to include larger sample sizes to provide more reliable predictions in future studies.

V. CONCLUSION

This study investigated whether the SII could be a reliable predictor for speech perception performance in adults who use CIs. The speech perception scores for CI recipients obtained in three different SNR conditions yielded observed SIIs which were compared to the predicted SIIs based on the transfer function curve. Predictions of speech perception performance using the SII alone overestimated CI users' abilities, whereas SII calculations using HLD corrections underestimated performance. Furthermore, the large variability in speech perception performance across the CI users was shown to be a significant barrier for the SII to prove to be a reliable predictor. Other predictive factors that have been associated with speech perception performance were also examined using a correlation and stepwise multiple regression. Some of the demographic and experimental variables showed significant correlations with speech perception scores. Gap detection thresholds and duration of deafness were found to be significant predictable factors. These predictive factors and SIIs were discussed in relation to speech perception performance in CI users. We conclude that conventional SII models are not appropriate for predicting speech perception scores for CI users. CI users struggle with speech understanding not because of audibility, but because of other factors.

Future studies are required for developing a new SII model that reflects the characteristics typical of CI individuals. To improve the predictions made by SII calculations in a future SII model, more specific SII calculation methods need to be developed that consider such factors that contribute to the individual differences in speech perception ability found in this population. For example, our study found that GDT outcomes and duration of deafness are highly correlated with speech perception performance in CI users. As with some correction factors previously developed that consider variables associated with speech perception outcomes, such as the HLD with degree of hearing loss and age corrections, a model that takes into account GDT results or duration of deafness could contribute to improving SII prediction accuracy. The long term goal of future studies would include investigations that sophisticatedly quantify or control the enormous individual variability seen in CI populations and develop algorithms that incorporate such correction factors.

Chapter 4

GENERAL CONCLUSION

Having briefly demonstrated the history and fundamental concept of SII models, this dissertation began with raising the interesting question of whether the SII can be used as an adequate tool for predicting speech perception performance in CI recipients. Prior to implementing a study to answer this question, it was necessary to establish FIFs for the AzBio sentences, as the FIF is one of the critical components in the SII calculation that has never been established for the AzBio sentences. For this reason, we designed the dissertation to contain two consecutive studies.

The aim of the first study was to derive FIFs for the AzBio test. Traditional procedures described in studies by Studebaker and Sherbecoe (1991) were mainly applied for this purpose. Fifteen listeners with normal hearing underwent speech recognition testing using the AzBio sentences that were high- and low-pass filtered under speech-spectrum shaped noise at various SNRs. The results showed that frequency weights for the AzBio sentences were greatest in the 1.5 to 2 kHz region, supporting the globally accepted notion that speech cues in the 2 kHz region play an important role in speech perception. The frequency weights yielded with using traditional procedures (Studebaker and Sherbecoe, 1991) and those using a nonlinear optimization procedure (Kates, 2013) were compared. Both procedures were accurate, but the newer approach (Kates, 2013) was slightly more accurate. Consecutive data analyses provided speech recognition scores for the AzBio sentences in relation to the SII. Our findings provided empirically derived FIFs for the AzBio test that can be used for future CI studies. Specifically, the accuracy of predicting SIIs for CI patients should be improved when using these derived FIFs for the AzBio test.

In the second study, the SIIs for CI users were calculated to investigate whether the SII could be an effective tool for predicting speech perception performance in a CI population. Fifteen adults who were pre- and post-lingually deafened participated. CI users' speech recognition was measured using the AzBio sentence lists. The FIFs obtained from the first study were used to compute the SII in these CI listeners. The obtained SIIs were compared with predicted SIIs using a transfer function curve derived from the first study. Furthermore, CI users completed questionnaires and performed psychoacoustic and cognitive function tests. A multiple regression was conducted with predictive variables (demographics, cognition, and spectral/temporal resolution) to investigate which predictive factors could be associated with speech perception performance. Due to the considerably poor hearing loss and large individual variability in performance, the SII failed to predict speech performance for this CI group. Only gap detection thresholds and duration of deafness were found to be significant predictive factors. Although the applicability of the SII for predicting speech perception scores in CI users was not successful in this study, it is worth mentioning that this study represents the first study that devotes considerable attention to this area. Despite the unsuccessful SII predictions, we believe that SII models still have potential if we can strictly control confounding factors such as large individual variability among participants. To this end, development of correction models that compensate for such individual variability need to be developed. Therefore, future studies should focus particularly on increasing the sample size and predictive factors to provide more robust findings. The results from this study could contribute to future studies that aim to develop SII models for CI users by systematically controlling or quantifying confounding factors. Surely, prodigious efforts should be followed to enhance the possibility that SIIs can play are stronger role in predicting speech perception performance, resulting in practical benefits to CI patients. We hope this study represents an important breakthrough for that purpose.

REFERENCES

Amlani, A. M., Punch, J. L., and Ching, T. Y. (**2002**). "Methods and applications of the audibility index in hearing aid selection and fitting," Trends Amplif **6**, 81-129.

ANSI (**1969**). *American National Standard Methods for the Calculation of the Articulation Index (Acoustical Society of America, New York)*.

ANSI (**1997**). *American National Standard Methods for for the Calculation of the Speech Intelligibility Index (Acoustical Society of America, New York)*.

ANSI **(R 2013) S3.1-1999**. *American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms (Acoustical Society of America, New York)*.

Apoux, F., and Healy, E. W. (**2012**). "Use of a compound approach to derive auditory-filter-wide frequency-importance functions for vowels and consonants," J. Acoust. Soc. Am. **132**, 1078-1087.

Aronoff, J. M., and Landsberger, D. M. (**2013**). "The development of a modified spectral ripple test," J. Acoust. Soc. Am. **134**, EL217-EL222.

Atkinson, R. C., and Shiffrin, R. M. (**1968**). "Human memory: A proposed system and its control processes," Psychol Learn Motiv **2**, 89-195.

AuBuchon, A. M., Pisoni, D. B., and Kronenberger, W. G. (**2014**). "Short-Term and Working Memory Impairments in Early-Implanted, Long-Term Cochlear Implant Users Are Independent of Audibility and Speech Production," Ear Hear. **36**, 733-737.

Baddeley, A. (**1992**). "Working memory and conscious awareness," in *Theories of memory* (Lawrence Erlbaum Associates), pp. 11-20.

Baddeley, A. (**2000**). "The episodic buffer: a new component of working memory?," Trends Cogn. Sci. **4**, 417-423.

Baddeley, A. D., and Hitch, G. (**1974**). "Working memory," Psychol Learn Motiv **8**, 47-89.

Bell, T. S., Dirks, D. D., and Trine, T. D. (**1992**). "Frequency-importance functions for words in high-and low-context sentences," J. Speech. Lang. Hear. Res. **35**, 950-959.

Bentler, R. A., and Pavlovic, C. V. (**1989**). "Transfer functions and correction factors used in hearing aid evaluation and research," Ear Hear. **10**, 58-63.

Bidelman, G., Jennings, S., and Strickland, E. (**2015**). "PsyAcoustX: A flexible MATLAB® package for psychoacoustics research," Front. Psychol. **6**.

Bilger, R. (**1984**). "Speech recognition test development," Asha Reports **14**, 2-7.

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

Boersma, P. P. G. (**2002**). "Praat, a system for doing phonetics by computer," Glot international **5**.

Bosen, A. K., and Chatterjee, M. (**2016**). "Band importance functions of listeners with cochlear implants using clinical maps," J. Acoust. Soc. Am. **140**, 3718-3727.

Burkholder, R. A., and Pisoni, D. B. (**2006**). "Working memory capacity, verbal rehearsal speed, and scanning in deaf children with cochlear implants," Advances in the spoken language development of deaf and hard-of-hearing children, 328-357.

Bush, L. C. (**2016**). "List Equivalency of the AzBio Sentence Test," Honors Theses. 419.

Byrne, D., Dillon, H., Ching, T., Katsch, R., and Keidser, G. (**2001**). "NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures," J. Am. Acad. Audiol. **12**, 37-51.

Caplan, D., and Waters, G. S. (**1999**). "Verbal working memory and sentence comprehension," Behav. Brain Sci. **22**, 77-94; discussion 95-126.

Chen, F., and Loizou, P. C. (**2011**). "Predicting the intelligibility of vocoded speech," Ear Hear. **32**, 331.

Chen, J., Huang, Q., and Wu, X. (**2016**). "Frequency importance function of the speech intelligibility index for Mandarin Chinese," Speech Comm. **83**, 94-103.

Ching, T. Y., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128-1140.

Cleary, M., Pisoni, D. B., and Geers, A. E. (**2001**). "Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants," Ear Hear. **22**, 395-411.

Cohen, M. A., Horowitz, T. S., and Wolfe, J. M. (**2009**). "Auditory recognition memory is inferior to visual recognition memory," Proc. Natl. Acad. Sci. U. S. A. **106**, 6008-6010.

Collison, E. A., Munson, B., and Carney, A. E. (**2004**). "Relations among linguistic and cognitive skills and spoken word recognition in adults with cochlear implants," J. Speech. Lang. Hear. Res. **47**, 496-508.

Daneman, M., and Carpenter, P. A. (**1980**). "Individual differences in working memory and reading," J. Verbal Learn. Verbal Behav **19**, 450-466.

Daya, H., Figueirido, J. C., Gordon, K. A., Twitchell, K., Gysin, C., and Papsin, B. C. (**1999**). "The role of a graded profile analysis in determining candidacy and outcome for cochlear implantation in children," Int. J. Pediatr. Otorhinolaryngol. **49**, 135-142.

DePaolis, R. A., Janota, C. P., and Frank, T. (**1996**). "Frequency importance functions for words, sentences, and continuous discourse," J. Speech. Lang. Hear. Res. **39**, 714-723.

Dillon, H. (**1999**). "NAL-NL1: A new procedure for fitting non-linear hearing aids," Hear J **52**, 10-12.

Doherty, K. A., and Turner, C. W. (**1996**). "Use of a correlational method to estimate a listener's weighting function for speech," J. Acoust. Soc. Am. **100**, 3769-3773.

Draine, S. (**1998**). "Inquisit [computer software]," Seattle, WA: Millisecond Software.

Duggirala, V., Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (**1988**). "Frequency importance functions for a feature recognition test material," J. Acoust. Soc. Am. **83**, 2372-2382.

Eisenberg, L. S., Dirks, D. D., Takayanagi, S., and Martinez, A. S. (**1998**). "Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions," J. Speech. Lang. Hear. Res. **41**, 327-339.

Elliott, C. D., Smith, P., and McCulloch, K. (**1997**). *British ability scales II: technical manual* (NferNelson).

Faulkner, K. F., and Pisoni, D. B. (**2013**). "Some observations about cochlear implants: challenges and future directions," Neuroscience Discovery **1**.

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

Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89-151.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90-119.

Fu, Q.-J., and Shannon, R. V. (**2000a**). "Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589-597.

Fu, Q. J., and Shannon, R. V. (**2000b**). "Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589-597.

Geers, A. E., Pisoni, D. B., and Brenner, C. (**2013**). "Complex Working Memory Span in Cochlear Implanted and Normal Hearing Teenagers," Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology **34**, 396-401.

Geers, A. E., Strube, M. J., Tobey, E. A., Pisoni, D. B., and Moog, J. S. (**2011**). "Epilogue: factors contributing to long-term outcomes of cochlear implantation in early childhood," Ear Hear. **32**, 84s-92s.

Gifford, R. H., Shallop, J. K., and Peterson, A. M. (**2008**). "Speech recognition materials and ceiling effects: considerations for cochlear implant programs," Audiol. Neurootol. **13**, 193-205.

Goldsworthy, R. L., Delhorne, L. A., Braida, L. D., and Reed, C. M. (**2013**). "Psychoacoustic and Phoneme Identification Measures in Cochlear-Implant and Normal-Hearing Listeners," Trends Amplif **17**, 27-44.

Gordon, K. A., Daya, H., Harrison, R. V., and Papsin, B. C. (**2000**). "Factors contributing to limited open-set speech perception in children who use a cochlear implant," Int. J. Pediatr. Otorhinolaryngol. **56**, 101-111.

Green, K. M., Bhatt, Y., Mawman, D. J., O'Driscoll, M. P., Saeed, S. R., Ramsden, R. T., and Green, M. W. (**2007**). "Predictors of audiological outcome following cochlear implantation in adults," Cochlear Implants Int **8**, 1-11.

Hasher, L., and Zacks, R. T. (**1988**). "Working memory, comprehension, and aging: A review and a new view," Psychol Learn Motiv **22**, 193-225.

Healy, E. W., and Warren, R. M. (**2003**). "The role of contrasting temporal amplitude patterns in the perception of speech," J. Acoust. Soc. Am. **113**, 1676-1688.

Healy, E. W., Yoho, S. E., and Apoux, F. (**2013**). "Band importance for sentences and words reexamined," J. Acoust. Soc. Am. **133**, 463-473.

Henry, B. A., McDermott, H. J., McKay, C. M., James, C. J., and Clark, G. M. (**1998**). "A frequency importance function for a new monosyllabic word test," Scientific publications, vol. 11, 1998-1999, no. 1116.

Henry, B. A., McKay, C. M., McDermott, H. J., and Clark, G. M. (**2000**). "The relationship between speech perception and electrode discrimination in cochlear implantees," J. Acoust. Soc. Am. **108**, 1269-1280.

Henry, B. A., and Turner, C. W. (**2003**). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 2861-2873.

Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111-1121.

Hilton, E. (**2001**). "Differences in visual and auditory short-term memory," IU South Bend Undergraduate Research Journal **4**, 47-50.

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., Heydebrand, G., and Skinner, M. W. (**2013**). "Factors affecting open-set word recognition in adults with cochlear implants," Ear Hear. **34**, 342-360.

Humes, L. E. (**1986**). "An evaluation of several rationales for selecting hearing aid gain," J. Speech Hear. Disord. **51**, 272-281.

Humes, L. E. (**2002**). "Factors underlying the speech-recognition performance of elderly hearing-aid wearers," J. Acoust. Soc. Am. **112**, 1112-1132.

Jin, I. K., Kates, J. M., Lee, K., and Arehart, K. H. (**2015**). "Derivations of the band-importance function: A cross-procedure comparison," J. Acoust. Soc. Am. **138**, 938-941.

Kates, J. M. (**2013**). "Improved estimation of frequency importance functions," J. Acoust. Soc. Am. **134**, EL459-EL464.

Kemtes, K. A., and Allen, D. N. (**2008**). "Presentation modality influences WAIS Digit Span performance in younger and older adults," J. Clin. Exp. Neuropsychol. **30**, 661-665.

Kiefer, J., von Ilberg, C., Reimer, B., Knecht, R., Gall, V., Diller, G., Sturzebecher, E., Pfennigdorff, T., and Spelsberg, A. (**1998**). "Results of cochlear implantation in patients with severe to profound hearing loss--implications for patient selection," Audiology **37**, 382-395.

Kirby, A. E., and Middlebrooks, J. C. (**2010**). "Auditory temporal acuity probed with cochlear implant stimulation and cortical recording," J. Neurophysiol. **103**, 531-542.

Kirk, K. I., Miyamoto, R. T., Ying, E. A., Perdew, A. E., and Zuganelis, H. (**2000**). "Cochlear implantation in young children: effects of age at implantation and communication mode," Volta review **102**.

Kral, A., and Sharma, A. (**2012**). "Developmental neuroplasticity after cochlear implantation," Trends Neurosci. **35**, 111-122.

Kronenberger, W. G., Pisoni, D. B., Henning, S. C., and Colson, B. G. (**2013**). "Executive Functioning Skills in Long-Term Users of Cochlear Implants: A Case Control Study," J. Pediatr. Psychol. **38**, 902-914.

Kryter, K. D. (**1962**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689-1697.

Kuo, M.-W. (**2013**). "Frequency importance functions for words and sentences in Mandarin Chinese: implications for hearing aid prescriptions in tonal languages," Thesis (University of Western Sydney).

Lawler, M., Yu, J., and Aronoff, J. M. (**2017**). "Comparison of the Spectral-Temporally Modulated Ripple Test With the Arizona Biomedical Institute Sentence Test in Cochlear Implant Users," Ear Hear.

Lee, S., and Mendel, L. L. **(In Press).** "Derivation of frequency importance functions for the AzBio sentences," J. Acoust. Soc. Am.

Lehiste, I., and Peterson, G. E. (**1959**). "Linguistic considerations in the study of speech intelligibility," J. Acoust. Soc. Am. **31**, 280-286.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, Suppl 2:467+.

Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners," J. Acoust. Soc. Am. **122**, 982-991.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U. S. A. **103**, 18866-18869.

Ludvigsen, C. (**1987**). "Prediction of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners," J. Acoust. Soc. Am. **82**, 1162-1171.

Macherey, O., and Carlyon, R. P. (**2014**). "Cochlear implants," Curr. Biol. **24**, R878-R884.

Marchand, P., and Marmet, L. (**1983**). "Binomial smoothing filter: A way to avoid some pitfalls of least-squares polynomial smoothing," Rev. Sci. Instrum. **54**, 1034-1041.

May-Mederake, B. (**2012**). "Early intervention and assessment of speech and language development in young children with cochlear implants," Int. J. Pediatr. Otorhinolaryngol. **76**, 939-946.

Mehr, M. A., Turner, C. W., and Parkinson, A. (**2001**). "Channel weights for speech recognition in cochlear implant users," J. Acoust. Soc. Am. **109**, 359-366.

Mueller, H. G., and Killion, M. C. (**1990**). "An easy method for calculating the articulation index," Hearing Journal **43**, 14-17.

Nie, K., Barco, A., and Zeng, F.-G. (**2006**). "Spectral and temporal cues in cochlear implant speech perception," Ear Hear. **27**, 208-217.

O'Donoghue, G., Pisoni, D., Waltzman, S., and Roland, J. (**2014**). "Auditory and linguistic outcomes in pediatric cochlear implantation," Cochlear implants **3**, 158-166.

Oxenham, A. J., Bernstein, J. G., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U. S. A. **101**, 1421-1425.

Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253-1258.

Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413-422.

Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). "An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals," J. Acoust. Soc. Am. **80**, 50-57.

Penney, C. G. (**1989**). "Modality effects and the structure of short-term verbal memory," Mem. Cognit. **17**, 398-422.

Pisoni, D. B., and Cleary, M. (**2003a**). "Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation," Ear Hear. **24**, 106s-120s.

Pisoni, D. B., and Cleary, M. (**2003b**). "Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation," Ear Hear. **24**, 106S.

Pisoni, D. B., Cleary, M., Geers, A. E., and Tobey, E. A. (**1999**). "Individual Differences in Effectiveness of Cochlear Implants in Children Who Are Prelingually Deaf: New Process Measures of Performance," Volta Rev **101**, 111-164.

Pisoni, D. B., Kronenberger, W. G., Roman, A. S., and Geers, A. E. (**2011**). "Measures of digit span and verbal rehearsal speed in deaf children after more than 10 years of cochlear implantation," Ear Hear. **32**, 60s-74s.

Pisoni, D. D., and Geers, A. E. (**2000**). "Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language processing," Ann. Otol. Rhinol. Laryngol. **185**, 92.

Santos, J. F., Cosentino, S., Hazrati, O., Loizou, P. C., and Falk, T. H. (**2013**). "Objective speech intelligibility measurement for cochlear implant users in complex listening environments," Speech comm. **55**, 815-824.

Schafer, E., and Utrup, A. (**2016**). "The Effect of Age of Cochlear Implantation on Speech Intelligibility to Others," Journal of Educational, Pediatric & (Re) Habilitative Audiology **22**.

Scollie, S. D. (**2008**). "Children's speech recognition scores: The Speech Intelligibility Index and proficiency factors for age and hearing level," Ear Hear. **29**, 543-556.

Shannon, R. V. (**1989**). "Detection of gaps in sinusoids and pulse trains by patients with cochlear implants," J. Acoust. Soc. Am. **85**, 2587-2592.

Shannon, R. V., Cruz, R. J., and Galvin, J. J. (**2011**). "Effect of Stimulation Rate on Cochlear Implant Users' Phoneme, Word and Sentence Recognition in Quiet and in Noise," Audiology & Neurotology **16**, 113-123.

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

Sherbecoe, R. L., and Studebaker, G. A. (**2002**). "Audibility-index functions for the connected speech test," Ear Hear. **23**, 385-398.

Sherbecoe, R. L., and Studebaker, G. A. (**2003**). "Audibility-index predictions of normal-hearing and hearing-impaired listeners' performance on the connected speech test," Ear Hear. **24**, 71-88.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature **416**, 87-90.

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

Spahr, A. J., and Dorman, M. F. (**2005**). "Effects of minimum stimulation settings for the Med El Tempo+ speech processor on speech understanding," Ear Hear. **26**, 2S-6S.

Spahr, A. J., Dorman, M. F., Litvak, L. M., Cook, S., Loiselle, L. M., DeJong, M. D., Hedley-Williams, A., Sunderhaus, L. S., Hayes, C. A., and Gifford, R. H. (**2014**). "Development and validation of the Pediatric AzBio sentence lists," Ear Hear. **35**, 418.

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes, T., and Cook, S. (**2012**). "Development and validation of the AzBio sentence lists," Ear Hear. **33**, 112.

Spahr, A. J., Dorman, M. F., and Loiselle, L. H. (**2007**). "Performance of patients using different cochlear implant systems: effects of input dynamic range," Ear Hear. **28**, 260-275.

Studebaker, G. A., Gray, G. A., and Branch, W. E. (**1999**). "Prediction and statistical evaluation of speech recognition test scores," J. Am. Acad. Audiol. **10**, 355-370.

Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (**1987**). "A frequency importance function for continuous discourse," J. Acoust. Soc. Am. **81**, 1130-1138.

Studebaker, G. A., and Sherbecoe, R. L. (**1991**). "Frequency-importance and transfer functions for recorded CID W-22 word lists," J. Speech. Lang. Hear. Res. **34**, 427-438.

Studebaker, G. A., Sherbecoe, R. L., and Gilmore, C. (**1993**). "Frequency-importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word test," J. Speech. Lang. Hear. Res. **36**, 799-807.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gray, G. A. (**1997**). "Age-related changes in monosyllabic word recognition performance when audibility is held constant," J. Am. Acad. Audiol. **8**, 150-162.

Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., and Doherty, K. A. (**1998**). "Frequency-weighting functions for broadband speech as estimated by a correlational method," J. Acoust. Soc. Am. **104**, 1580-1585.

van Dijk, J. E., van Olphen, A. F., Langereis, M. C., Mens, L. H., Brokx, J. P., and Smoorenburg, G. F. (**1999**). "Predictors of cochlear implant performance," Audiology **38**, 109-116.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). "Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608-624.

Warren, R. M., Bashford Jr, J. A., and Lenz, P. W. (**2005**). "Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired," J. Acoust. Soc. Am. **118**, 3261-3266.

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

Winn, M. B., Won, J. H., and Moon, I. J. (**2016**). "Assessment of Spectral and Temporal Resolution in Cochlear Implant Users Using Psychoacoustic Discrimination and Speech Cue Categorization," Ear Hear. **37**, e377-e390.

Won, J. H., Drennan, W. R., and Rubinstein, J. T. (**2007**). "Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users," J Assoc Res Otolaryngol **8**, 384-392.

Wong, L. L., Ho, A. H., Chua, E. W., and Soli, S. D. (**2007**). "Development of the Cantonese speech intelligibility index," J. Acoust. Soc. Am. **121**, 2350-2361.

Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758-1764.

APPENDIX

IRB Approval

Institutional Review Board

Office of Sponsored Programs

University of Memphis

315 Admin Bldg, Memphis, TN 38152-3370

Mar 3, 2017

PI Name: Sung Min Lee

Co-Investigators:

Advisor and/or Co-PI: Lisa Mendel

Submission Type: Modification

Title: Predicting Speech Recognition using the Speech Intelligibility Index (SII) for Cochlear Implant Users and Listeners with Normal Hearing

IRB ID : #PRO-FY2017-246

Level of Review: Expedited

Approval: Mar 3, 2017

Expiration: *

*Modifications do not extend the expiration of the original approval

Approval of this project is given with the following obligations:

1. This IRB approval for modification has an expiration date, an approved renewal must be in effect to continue the project prior to that date. If approval is not obtained, the human consent form(s) and recruiting material(s) are no longer valid and any research activities involving human subjects must stop.

2. When the project is finished or terminated, a completion form must be submitted.

3. No change may be made in the approved protocol without prior board approval.

Thank you,

James P. Whelan, Ph.D.

Institutional Review Board Chair

The University of Memphis.