# Optimizing frequency channels for adults with cochlear implants

Research Thesis

Presented in partial fulfillment of the requirements for graduation

“with Research Distinction in Speech and Hearing Science” in the undergraduate colleges of

The Ohio State University

By

Lauren Boyce

The Ohio State University Departments of Speech and Hearing

and Otolaryngology-Head and Neck Surgery

May 2015

Project Advisors: Aaron Moberly, M.D., Department of Otolaryngology &

Rachael Frush Holt, Ph.D., Department of Speech and Hearing Science

# Abstract

Cochlear implants (CIs) are devices used by individuals with hearing loss to improve communication through the use of an electrode array that directly stimulates the auditory nerve. Existing signal processing strategies utilize a logarithmic frequency-to-electrode allocation, mimicking the representation of frequencies along the basilar membrane (high frequencies at the base and low frequencies at the apex). These strategies support some degree of open-set speech recognition for CI users; however, average speech recognition remains well below that of normal-hearing adults. To enhance speech recognition by adult CI users, this study examined one promising alternative to the standard logarithmic frequency-to-electrode allocation. The allocation map was modified to provide more refined representations of the first two (and most important) vowel formant frequencies (energy peaks vowel spectra that are critical to speech perception). Twelve participants were tested using two different CI maps: one based on existing clinical frequency-to-electrode allocation strategies (Standard) and one designed to improve the resolution of the first two formants, which should especially enhance vowel recognition (Speech). Alternating between these maps, participants listened to and repeated three kinds of stimulus materials: (1) highly meaningful five-word sentences, (2) syntactically correct but not meaningful four-word sentences, and (3) phonetically balanced consonant-vowel-consonant words in isolation. Analyses revealed that some participants benefitted from the Speech strategy. Moreover, an improvement in vowel recognition within words strongly predicted an improvement in recognition of words in sentences. These findings suggest that optimizing the representation of the first two formants could enhance speech recognition for CI users. Future efforts should focus on better representing this speech-specific information in modern-day signal processing strategies.

## Introduction

Considered one of the most successful neural prosthetics, cochlear implants (CIs) have improved the lives of almost 200,000 people to date (Cochlear, 2015). Advances in CI technology aid communication for profoundly deaf individuals; however, average speech recognition scores are still not equivalent to normal-hearing adults. Previous studies have tried to improve speech recognition scores by altering the way frequencies are allocated to the CI electrodes, with mixed results. This study attempts to improve speech recognition scores by examining whether the speech recognition performance of CI users might be improved by modifying their frequency-to-electrode allocation maps. Participants were tested while listening to two maps: one which focused on improving delivery of the acoustic information classically thought to underlie speech perception, and one representing a more typical clinical frequency-to-electrode allocation.

Speech sounds are generated during complex, coordinated patterns of articulatory activity. The resulting sounds are acoustic representations of gestures occurring in the oral, nasal, and pharyngeal cavities. These gestures are reflected in a number of properties of the acoustic signal, including acoustic structure in the frequency, duration, and intensity domains, which co-vary over time. These co-varying properties result in acoustic cues for speech perception, and some are more effective than others for a given language. Specifically, the general spectral structure of speech (the frequency-specific information) carries especially important acoustic and phonetic information that is used by a listener to understand speech (Studdert-Kennedy, 1983). For example, formant peaks (the resonant harmonic peaks that are formed as sound from the glottal source travels through the pharyngeal, oral, and nasal cavities) play a crucial role in the perception of vowels. Formant transitions, the relative changes over time of those formant peaks, underlie perception of many consonant-vowel and vowel-consonant transitions or glides.

The process of hearing and understanding these acoustic cues begins in the outer ear and ends in the brain. When a sound is produced, a wave is propagated through a medium and enters the external auditory canal. Subsequently, the tympanic membrane is set into motion, causing the ossicles to move and further propagate the sound energy through the oval window and into the fluid-filled cochlea. Together the outer and middle ear work to selectively transmit certain frequencies, as well as overcome the impedance mismatch from the air-filled middle ear to the fluid-filled cochlea. Within the cochlea, the fluid wave then stimulates movement of the basilar membrane in the cochlea, and triggers activation of hair cells. In general, these cells encode sounds according to frequency: hair cells at the base of the cochlea encode higher frequencies, and hair cells at the apex encode lower frequencies. Once the hair cells are stimulated, they produce electrical signals and send those signals to the brain via the auditory nerve (Yost, 2006).

In individuals with sensorineural hearing loss, certain parts of this process do not function as well as others. The waveform follows the same path as it does for normal-hearing individuals, but upon arriving at the cochlea, the hair cells are less able to transmit signals to the brain. The weakened signals can be due to damaged hair cells, as a result of age, disease, noise exposure, heredity, infections, trauma, or ototoxic drugs. Human hair cells do not regenerate, and any damage to these cells is irreversible. Upon the detection of a permanent hearing loss, a hearing aid may be recommended, depending on the degree and type of hearing loss (Yost, 2006). For individuals with more severe-to-profound hearing loss, a CI may be recommended.

A CI is a device that consists of two components: an external piece that is worn behind the ear and an internal piece that is surgically implanted. The external piece consists of a microphone and battery pack, as well as a sound processor, which are worn behind the ear, and an external transmitter that is held in place above the ear by an external magnet. The internal portion consists of a receiver/stimulator attached to an internal magnet and an electrode array. Sounds are detected by one or more microphones located on the external piece of the implant, after which signals are sent to the sound processor. This processor is a battery-powered digital signal processing unit that converts the acoustic sound signals into electrical stimulation. The electrical stimulation is transformed into radio-frequency signal, which is transmitted through the skin to the internal receiver/stimulator. The signals are then decoded by the internal receiver, which sends electrical stimulation to the surgically inserted electrode array in the cochlea. The electrode array has a length of about 2cm, which covers approximately the first one-and-a-half turns of the cochlea. Upon implant activation postoperatively, each electrode is assigned a certain range of frequencies to represent electrically. Electrical impulses are then delivered to stimulate auditory nerve fibers that exist near the electrodes, resulting in the production of action potentials that are delivered via the auditory nerve to the auditory cortex for processing (Macherey & Carlyon, 2014).

Although these complex devices are extremely helpful in many cases of hearing loss today, they are a relatively recent development. The first cochlear implantation in the United States occurred in 1961, by William House and John Doyle (Mudry & Mills, 2013). This original implant had a grounding electrode and one stimulating electrode (a single channel device), with a primary purpose of providing environmental sound awareness in deaf individuals (Moller, 2006). In terms of speech perception, the first few implants were primarily useful as aids to lip reading, because they provided little more than a sensation of sound and very gross envelope cues, with highly degraded spectral representations (Wilson & Dorman, 2008). However, even with these degraded signals, a few individuals were able to understand some speech. Because the single electrode was only capable of transmitting the temporal envelope (the broad amplitude structure of the entire frequency range of the acoustic signal), users had access to enough acoustic information to perceive voicing, but not much more (Clark, 2014). In order to improve speech understanding, including the perception of vowels and the glides connecting consonants and vowels, it was necessary to have better spectral resolution.

Improving spectral resolution of implants occurred through taking advantage of the tonotopicity of the cochlea. Advances in technology permitted the development of CIs with multiple stimulating electrodes and, therefore, multiple channels of processing and stimulation in the cochlea. Current multi-channel devices include up to 22 electrodes on a single electrode array (Macherey & Carlyton, 2014). This new technology has greatly improved speech recognition in CI users, and implants are no longer just for aiding lip reading and environmental awareness (Wilson & Dorman, 2008).

Although the number of electrodes inserted is now greater than with single-channel devices, spread of electrical excitation within the cochlea, as well as overlapping regions of neural stimulation, effectively lead to delivery of only about 6-8 independent channels of information (Friesen et al., 2001). Therefore, a large amount of spectral detail inherent in the speech signal is still lost. This signal degradation certainly plays a role in the imperfect speech recognition exhibited by CI users, especially under noisy conditions. It is still unclear which acoustic cues through an implant best support speech perception for CI users. Formants provide the necessary cues for people with normal hearing to identify articulator placement for vowels, while formant transitions into and out of the vowels help to identify the placement for consonants. Studies of normal-hearing individuals listening to degraded speech sound representations that preserve certain speech features have shed some light on how these cues are received by listeners using CIs. Remez et al. (1981) provided evidence that three sine waves replicating the time-varying spectral structure of the first three formants is enough for listeners to accurately repeat sentences, suggesting that this type of broad spectral structure can support accurate speech recognition. Studdert-Kennedy (1983) suggested that speech recognition is possible as long as the listener can identify instantaneous articulator placement, and claimed that the function of the speech signal is to specify articulation. In order to identify the instantaneous articulator placement, a user needs to accurately perceive not only formant information, but temporal patterns of spectral change, as well. The formants allow the listener to perceive exact articulator placement, and the spectral changes over time that underlie perception of connected speech.

Based on these classical premises of speech perception, early processing strategies of CIs attempted to provide representations of the formant frequencies of speech sounds. These processing strategies refer to the number and location of the electrodes stimulated, the frequency-to-electrode allocation, the type of stimulus, and the rate and amplitude of stimulation. An early processing strategy focused on representing just the fundamental frequency (F0), the second formant (F2), and the amplitude envelope of the speech. This strategy, known as F0F2, showed significant improvement of consonant recognition as well as overall speech recognition over the F0 strategy, which only represented the fundamental frequency (Clark, Tong & Dowell, 1984). The purpose of the F0F2 strategy was to improve lip reading, which it did by focusing on the fundamental frequency and second formant. Both the fundamental frequency and the second formant are very difficult or even impossible to see, because they are dictated by the vocal folds and tongue advancement, whereas the first formant is the easiest formant to see, because it is dictated by jaw height. The F0F2 strategy did not include the first formant because it is easily visible, but in an effort to further improve vowel perception, F1 was added to form the F0F1F2 strategy (Tye-Murray et al., 1990). Blamey et al. (1987) tested the F0F2 strategy against the F0F1F2 strategy and found that CI users preformed much better for prosodic and phonetic tasks with the F0F1F2 strategy. The authors also found that adding in F1 greatly improved vowel recognition, but consonant recognition was similar between the two programs (Blamey et al., 1987). To improve consonant recognition, MPEAK was created, which added 3 additional high frequency bands to the F0F1F2 processing strategy. This increase of consonant information greatly improved overall open set speech recognition (Skinner et al., 1991).

Cochlear implants utilize three main categories of processing strategies: feature extraction, waveform representation, and a mix of feature extraction and waveform. The strategies discussed earlier all were feature extraction strategies, in which the programmer decided what aspects of speech are most important and focuses on delivery of those aspects instead of the whole waveform. The previous strategies all focused on delivering formant information, which in turn nominally improved speech recognition.  No current processing strategies rely on feature extraction.  Waveform representation strategies attempt to maintain and deliver all of the waveform information found in the speech signal.  Some examples of waveform strategies include Compressed Analog, Simultaneous Analog Simulation (SAS) and Continuous Interleaved Sampling (CIS).  The first iteration of the Compressed Analog strategy delivered band-specific, compressed amplitude waveforms to different electrodes, and therefore different locations in the cochlea.  The main limitation of the Compressed Analog strategy was the cross-channel interactions, which was solved for the most part with the SAS strategy.  At the end of each band-pass channel, SAS uses a logarithmic mapping function, which allows each channel to be mapped individually (Zeng, 2004).  The CIS strategy is similar to the Compressed Analog strategy, but it uses more manipulations, a higher rate of stimulation, compression of envelopes occurs after filtering (as opposed to before filtering with the Compression Analog strategy), and it uses pulsatile stimulation (as opposed to analog).  CIS begins by increasing the amplitude of higher frequency sounds (where consonant information is located), which usually have lower intensities in speech.  Compression still occurs, but the amount of compression varies for low- and high-amplitude signals, while still preserving important changes in the temporal envelopes, resulting in high speech recognition in quiet (Shannon et al., 1995).  Two recent variations of the CIS strategy are the Paired Pulsatile Stimulation strategy and the HiRes strategy.  The PPS strategy pairs distant electrodes which are stimulated simultaneously, with each pair stimulated nonsimultaneously.  This doubles the rate of stimulation across all electrodes while minimizing interactions between simultaneously stimulated electrodes (Loizou et al., 2003).  The HiRes strategy is another variation of the CIS strategy but instead of pairing electrodes to increase stimulation, it increases the pulse rate and the number of channels used. The increase in pulse rate and channels led to improvement of speech recognition in noise for adults with cochlear implants (Firszt et al., 2009).

Two types of n-of-m strategies are the SPEAK strategy and the Advanced Combination Encoder strategy. The SPEAK strategy uses multiple band pass filters and then selects the filters with the largest output amplitudes, depending on the incoming signal. These amplitudes are then compressed into the listener's dynamic range, and the digitally encoded pulses get sent to the electrodes (Loizou, 1998). The ACE stimulation strategy is very similar to the SPEAK strategy, but the main difference is the ACE strategy uses a stimulation rate, about 800 to 1600 pps, which is much higher than SPEAK's stimulation rate, which is about 180 to 300 pps.

Regardless of which processing strategy is used, the sound processor is programmed through mapping, which allocates specific frequency ranges to each electrode on the electrode array. Roughly, current mapping techniques follow the tonotopic organization of the cochlea and represent high frequencies at more basal electrodes, whereas low frequencies are represented at the more apical electrodes. Moreover, frequency bands are assigned more widely at the base and more narrowly at the apex, based on a logarithmic scale and the known tonotopic arrangement of the cochlea. The frequency which produces the largest response with the smallest stimulation is the characteristic frequency, which is organized in a logarithmic scale in the cochlea (Gray, 1997). Although the clinical maps follow this logarithmic arrangement, the clinical technique does not specifically focus on the frequencies where speech sounds, and more specifically, formant peaks and transitions occur. Thus, current maps typically do not emphasize the representation of speech-specific information.

Previous studies have experimented with improving representation of the low-frequency speech structure, with mixed results. Skinner et al. (1995) extended the low frequency set of filters from a default filter set to one that provided one or two more filters to the area under 800 HZ, thus providing improved resolution in the F1 formant region of speech. After 3 weeks of experience listening to the device at home, participants were tested in the laboratory, and were found to have improved perception of vowels and nasality in consonants. These results showed that improving resolution of the lower frequencies, specifically the F1 region, could significantly improve vowel perception.

Fu and Shannon (1999a,b) carried out a series of experiments that examined relationships among frequency allocation, electrode location, and speech perception. In the first experiment, four electrode maps were tested with alterations of either electrode location or frequency allocation. Participants' consonant and vowel recognition was then assessed. They found that frequency range and electrode location had a significant effect on the perception of vowels and consonants. Vowel and consonant recognition was best when more filters and more electrodes were used to represent the lower frequencies. Vowel recognition and recognizing consonant place of articulation were strongly affected by the changes, whereas voicing and manner were not strongly affected by changes in electrode location and spacing, probably because they are largely based on temporal cues. The results of their second study (Fu & Shannon, 1999a) suggested that vowel discrimination was best when the frequency-to-electrode allocation was most similar to their clinical map.

One way to alter listeners' maps would be to allocate more electrodes to lower frequencies, which would allow more formant information to fall across adjacent channels in the low-to-mid frequency range, instead of within a single channel. Henry (2000) found that CI users' ability to discriminate between adjacent electrodes below 2.6 kHz was positively correlated with their speech recognition, suggesting that fine spectral discrimination is more important in vowel formant regions than higher frequency regions. McKay and Henshall (2002) altered the frequency information given to a CI user through the creation of two maps: one with evenly spaced allocations of frequency from 200-10,513 Hz across 10 electrodes; and one with 9 out of 10 electrodes allocated below 2,600 Hz. After wearing the new maps outside the clinic for two weeks, participants were tested using words in quiet and sentences in noise. The low-frequency allocation map led to improved sentence recognition in noise, as well as vowel recognition in quiet, but some users showed degraded consonant perception in quiet. Furthermore, users with poor speech perception to begin with did not show much of a difference between the two maps; the authors suggested that this might be due to the fact that those users might rely more on temporal cues than spectral cues. They concluded that vowel recognition could be improved by allocating more electrodes to lower frequencies without a detrimental effect on consonant recognition.

Similarly, Nittrouer (2014a) found that improving representation of low-to-mid frequencies might benefit CI users of all ages. Normal-hearing participants listened to vocoded (CI-simulation) stimuli that were processed to represent either a standard clinical CI map or an experimental map that represents the low-frequency formant information with greater resolution. Results demonstrated that speech recognition improved when the electrodes were organized so that the first and second formants were presented in separate channels (experimental map), as opposed to falling within the same channel (standard clinical map). This improved low-frequency resolution allowed representation of sound by CIs to specify instantaneous articulatory placement, as well as formant transitions, as deemed necessary by Studdert-Kennedy (1983).

The current study expands upon Nittrouer (2014b). The purpose was to determine whether changing the frequency allocation of a listener's map to better represent formant peaks (i.e., improve the resolution of low-frequency formants and their transitions) would improve speech recognition in adult, postlingually deafened patients with CIs. Two maps comprised of five electrodes were used. The first map, the "standard" map, was similar to a typical clinical map, with logarithmic frequency allocation. The second map, the "speech" map, consisted of allocating narrower low-frequency bands to the more apical electrodes in order to improve the resolution of the speech formant frequencies. Participants were tested for word and sentence recognition while listening using the standard and speech maps. It was hypothesized that speech recognition would be better with the speech map compared to the standard map. A second hypothesis was that better resolution of the formant information through the speech map, as evidenced by improved vowel recognition, should predict improved access to time-varying formant structure leading to improved sentence recognition.

## <u>Method</u>

### Participants

Twelve adults who wore CIs were recruited for this study from a pool of departmental patients. All of the participants were native speakers of English and were between the ages of

23 and 77 years, with various etiologies of hearing loss and ages of implantation. All participants had a progressive hearing loss and qualified for cochlear implantation at age 13 years or later (Table 1). Within the year prior to enrollment in the study, all participants were found to have CI-aided thresholds of 35 dB HL or better for all frequencies between .25 and 4 kHz, as measured by their clinical audiologists. All participants had at least one year of experience with their CIs, and all used Cochlear devices with an ACE processing strategy for normal everyday use. Six participants were implanted in their right ear, four in their left, and two were implanted bilaterally. Of the ten who had a single implant, five used a contralateral hearing aid. For the purpose of this study, all participants used a single Freedom processor on the ear which had been implanted first, and the contralateral ear was unaided. To ensure that none of the participants was cognitively impaired, all participants underwent a screening to rule out dementia using the Mini-Mental State Examination (MMSE). This test is a validated screening assessment of memory, attention, and the ability to follow instructions (Folstein et al., 1975). Results of this screening showed that none of the participants showed evidence of cognitive impairment.

## Equipment and Materials

All testing for this study was done in a sound proof booth with stimuli presented through a loudspeaker at 68 dB SPL, with the loudspeaker positioned one meter from the participant at zero degrees azimuth. All responses were video- and audio-recorded using a Sony video recorder and a Sony FM microphone to ensure good sound quality for later scoring. Participants wore specially designed vests that held the FM transmitters, which sent the speech signals to the receiver, which had direct input into the camera. All scoring was done at a later time by the first author.

All participants used the same processor that was programmed using an ACE strategy to stimulate only five electrodes during testing. All other electrodes were not stimulated. For each participant, a clinical audiologist programmed the processor to identify the T-level (the threshold) and the C-level (the maximum comfortable level) for electrodes 4, 8, 12, 16, and 20 (which were all inserted into the cochlea for all participants). The processor was programmed such that the first program was always designated as the speech program, and the second program was always designated as the standard program. The speech program was designed to split the frequency ranges for F1 and F2 across two electrodes each, and the last electrode allocated for the higher frequencies containing the consonant information. For the speech program, electrode 20 was allocated the frequency range of 250 to 550 Hz; for electrode 16, 550 to 936 Hz; for electrode 12, 936 to 1528 Hz; for electrode 8, 1528 to 2440 Hz; and for electrode 4, 2440 to 7938 Hz. For the standard program, electrode 20 was allocated the frequency range of 250 to 722 Hz; for electrode 16, 722 to 1528 Hz; for electrode 12, 1528 to 3066 Hz; for electrode 8, 3066 to 6000 Hz; and for electrode 4, 6000 to 7938 Hz. As such, the speech program divided the range of critical frequency information for formants one and two (250 to 1528 Hz) across four electrodes, whereas the standard program divided that same frequency range over only three electrodes. Thus, more refined information about formant energy should have been represented by the speech program compared to the standard program.

## Stimuli

Three types of speech stimuli were used: highly meaningful five-word sentences, syntactically correct but not meaningful four-word sentences, and phonetically balanced consonant-vowel-consonant words. Two different kinds of sentences were used to ensure that any effects observed were not due to the participant's understanding of the semantic context of the words within the sentences. This experiment used both words and sentences in order to see if any observed benefits might be greater in the sentences for the speech program, which would suggest that participants benefitted from receiving more information about the time-varying formant structure across word boundaries. Both meaningful and syntax-only sentences were used in this experiment to ensure that any results found were due to time-varying formant structure and not to context cues. The meaningful sentences were composed of 56 sentences from the Hearing in Noise Test (HINT) (Nilsson, Soli and Sullivan, 1994). All were five words in length, follow a subject-predicate structure, are syntactically correct, and semantically predictable. Half of the sentences were heard using the speech program, and the other half were heard using the standard program, with 3 for practice and 25 for testing using each program. Fifty-six syntax-only sentences, which were used in Nittrouer et al. (2014a), were also used in this experiment. These sentences are syntactically correct but semantically anomalous, and contained only monosyllabic content words (e.g., "Ducks teach sore camps"). Again, half of the sentences were heard using the speech program, and the other half were heard using the standard program, with 3 for practice and 25 for testing using each program. The last type of stimuli included was individual words, presented in lists, which were developed by Mackersie, Boothroyd, and Minnear (2001). Each list consisted of ten phonetically balanced CVC words.

Nine lists were heard using the speech program, and the other nine lists were heard using the standard program, with an extra five words for practice with each program. All sentences and words were recorded by an adult male talker of American English at 44.1 kHz sampling rate with 16 bit digitization.

## General Procedures

Testing took place at the Ear and Eye Institute (EEI) of The Ohio State University Wexner Medical Center. Approval was obtained from the Institutional Review Board of the Ohio State University, and informed written consent was obtained from all participants before beginning testing. Participants were tested during a single two-hour session.

Prior to testing, an audiologist programmed the Freedom processor for each individual participant. Of note, during this process, the participant had minimal exposure to speech using either program; that exposure was only sufficient to confirm T- and C-levels. The participant then entered the soundproof booth where testing was done.

Each stimulus was played only once, and the listener was asked to repeat what was heard. Measures were collected in the following order for each participant: (1) Meaningful five-word sentences (meaningful sentences), (2) words in isolation and (3) not meaningful but syntactically correct four-word sentences (syntax-only sentences). The order of programs was alternated based on enrollment in the study, such that half of the participants were tested with the speech program first and the other half with the standard program first.

Data included in analyses were collected from 5 pilot participants and 7 other participants. For the pilot participants, listeners heard the meaningful sentences, words, and syntax-only sentences, all presented in the same order, while switching between the speech and standard programs during testing. Following pilot testing, a randomization program was developed in MATLAB, so that sentences or words were presented in random fashion, but still divided into testing blocks of meaningful sentences, words, or syntax-only sentences. Again, participants were assigned to begin listening using either the speech program or the standard program first.

## **Results**

For both meaningful and syntax-only sentences, percent total (not key) words correct was computed, whereas percent phonemes correct and percent whole words correct were computed for the word lists. Percent phonemes correct for the words was further examined as percent correct first consonant, vowel, and final consonant to better examine the effect of the speech and standard programs on phoneme recognition, particularly vowel recognition, as the speech program was designed to benefit vowel recognition most strongly. Arc sine transformations of the percent correct scores were computed to give the data a normal distribution, and these arc sine transformed values were used as dependent measures in analyses.

To analyze the differences in mean scores between the speech and standard programs, paired t-tests were used. Difference scores for each of the transformed percent correct scores (speech minus standard) were also computed. A series of linear regression analyses was performed to examine whether differences in some scores between the two programs (i.e., correct phonemes, first consonant, vowel, or final consonant) would predict differences in sentence score (i.e., meaningful and syntax-only sentences).

Prior to analyzing the data for the group as a whole, data were analyzed and compared between the pilot participants and the other participants. Based on independent-samples *t*-tests, no significant differences were found in mean sentence or word recognition scores using either the speech or standard programs. Therefore, data for all 12 participants (pilot and other) were included together in further analyses.

A few general impressions can be made by inspection of the individual participant scores for the speech and standard programs, reported in Table 2. A wide range of scores was seen among participants for each task, in both conditions. It is difficult to determine the difference between the scores for the speech and standard program that would lead to a noticeable improvement in speech recognition in real life situations, but for the purposes of this study, participants who benefited more than five percentage points from either program were considered to have benefitted from that program. Five out of 12 participants had better scores using the speech program over the standard program for the meaningful sentences, with the largest benefit being 47 percentage points. Additionally, half of the participants benefitted from the speech program for vowel recognition from the word lists. It is interesting to note that these four are not the same four who benefitted from the speech program during meaningful sentence recognition, although one participant did improve by at least 5 percentage points for both the meaningful sentences and vowel stimuli. Out of the 9 participants who were able to complete the syntax-only sentences, 5 participants benefited by at least five percentage points using the speech program. For the word lists, only 1 participant benefited by at least 5 percentage points for the speech program; on the other hand, 5 participants benefited by more than 5 points for the standard program, and the rest did not have a 5-percentage point difference between scores.

The main question of interest for this study was whether a group benefit would be seen in speech recognition scores using the speech program over the standard program. Examination of the group mean scores in Table 2 revealed a larger mean score on meaningful sentences while using the speech program than using the standard program (69.2% versus 62.8%), but a paired *t*-test analysis revealed the difference did not reach statistical significance. Better scores for the speech program over the standard program were also seen for syntax-only sentences (20.5% versus 19.9%) and vowels (41.0% versus 37.5%); however the paired *t*-test analysis revealed that these differences were not statistically significant. Significant results favoring the standard program were found for words in isolation (19.6% versus 14.7%; *t* = 3.09, *p* = .01), and first consonant recognition (43.0% versus 37.2%; *t* = 2.44, *p* = .03). There also were higher scores for the standard program over the speech program for last consonant recognition (36.3% versus 33.3%), but those results were found not to be statistically significant.

The second question of interest was whether increased resolution of formant information (as measured by vowel recognition) would improve sentence recognition in adults with cochlear implants. To answer this question, linear regressions were used to determine if improved vowel recognition performance predicted improved scores on the meaningful and syntax-only sentence recognition. Analyses revealed that improvements in vowel scores strongly predicted improvements in meaningful sentence recognition, $\beta = .84$, $F = 23.38$, $p =$

.001, as well as syntax-only sentence recognition, *β = .67, F =*7.85*, p =* .026. These findings suggest that if improvements can be made in vowel recognition by CI users, this will likely result in improved sentence recognition as well.

## Discussion

The described experiment was undertaken to test two hypotheses: (1) Adults with cochlear implants would show improved speech recognition when using a modified frequency-to-electrode allocation program with improved resolution of formant information; and (2) Improved access to formant information, represented by improved vowel recognition, would predict better recognition of time-varying formant structure, as evidenced by better recognition of sentences.

The results of this study did not fully support the first hypothesis: the only significant improvement in speech recognition scores was seen for words in isolation using the standard program. It is important to note that one-third of participants showed substantial improvement using the speech program, suggesting that it is possible that there is a certain type of CI user that could benefit from the speech program. The group means for recognition scores of both meaningful and syntax-only sentences were higher for the speech program; however those results were not statistically significant. The higher group means using the speech program for the sentence stimuli suggest that the participants might benefit from the improved time-varying formant structure received through the speech program. Nonetheless, these results suggest that some adults with cochlear implants might benefit from increased resolution of information about time-varying formant structure across word boundaries.

The second hypothesis was supported by the results. Improvements in vowel recognition predicted improvements in syntax-only and meaningful sentence recognition, suggesting that improved access to formant information might support better information of the time-varying formant structure of running speech. These results stress the importance of formant information and time-varying spectral structure of running speech for speech recognition by CI users. This suggests that it might be worthwhile to focus future processing strategies on improved delivery of formant information in order to improve speech recognition.

A possible reason why some CI users did not show a benefit for the speech program over the standard program is the mismatch between the frequency allocation of the electrodes and the tonotopic arrangement of the cochlea. In an ideal situation, the frequency allocation of the electrodes would line up perfectly with the corresponding parts of the neural elements to be stimulated; however this is not yet possible. This mismatch is typically due to not knowing how deep the electrode array is inserted, or what part of the cochlea each electrode is stimulating.

A typical clinical map attempts to minimize this mismatch by allocating its electrode frequencies logarithmically to try and best match the tonotopicity of the basilar membrane, but this does not allocate many electrodes to the critical lower frequencies where formant frequencies are located. The speech map attempts to allocate more electrodes to the lower frequencies, but the frequency mismatch remains. Whitford et al. (1995) found improvements attributed to the experimental maps even when the experimental maps were substantially different from the clinical maps, suggesting that improvement with different maps is possible if the participants are given time to adjust. McKay and Hanshall (2002) also found some improvement with the clinical map when the participants were given a two-week adjustment period to get used to their new maps. Participants in this study did not have an adjustment period, and only had about 2 hours of total experience with the experimental maps, which might be the reason why there was not a significant improvement with the speech map.

An in-depth understanding of speech perception is an essential component in devising studies that will yield insight into the variable performance of CI users. Another consideration is the variability of the patients included. There were varying etiologies of hearing loss, ages of implantation, and degrees of residual hearing, as well as likely variability in CI electrode array placement within the cochlea; all of these factors might contribute to user performance. Future methods to improve apical stimulation and decrease frequency-to-place mismatch are needed. In the meantime, reallocation of frequencies to better represent the acoustic structure that underlies successful speech perception holds promise for patients with cochlear implants.

## Acknowledgements

I wish to thank Dr. Aaron Moberly for his tireless support throughout the entire process, and Dr. Rachael Frush Holt for her guidance and editing. I also wish to thank Dr. Susan Nittrouer for her guidance and support as well as her lab for the stimuli and modified software. Lastly, I wish to thank Audiologists Brenda Hall and Melissa Schnitzspahn for their assistance in device programming.

**References**

Blamey, P.J., Dowell R. C., Brown A.M., Clark G. M., and Seligman, P.M. (1987). "Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors,"J. Acoust. Soc. Am. 82(1), 48-57.

Clark, G. M. (2014). "The multi-channel cochlear implant: Multi-disciplinary development of electrical stimulation of the cochlea and the resulting clinical benefit," Hear. Res. 322, 4-13.

Clark, G. M., Tong Y. C., and Dowell, R. C. (1984). "Comparison of two cochlear implant speech-processing strategies." Ann. Otol. Rhinol. Laryngol. 2 Pt 1, 127-31

Cochlear celebrates 30 years of hearing revolution. (2015, January 1). Retrieved January 11, 2015.

Firszt, J. B., Holden, L. K., Reeder, R. M. and Skinner M. W. (2009). "Speech recognition in cochlear implant recipients: comparision of standard HiRes and HiRes 120 sound processing," Otol. Neurotol. 30(2), 146-52

Folstein, M. F., Folstein, S. E., McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12(3), 189-98

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

Fu, Q. J., Shannon, R. (1999a). Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant," Ear Hear. 20, 332-344

Fu, Q. J., Shannon, R. (1999b). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," J. Acoust. Soc. Am. 105, 1889-1900

Gray, L. (1997). Auditory system: Structure and function. *Neuroscience Online.* Retrieved from http://neuroscience.uth.tmc.edu/s2/chapter12.html.

Henry, B., Mckay, C., Mcdermott, H., & Clark, G. (2000). “The relationship between speech perception and electrode discrimination in cochlear implantees.” J. Acoust. Soc. Am., 108, 1269-1280.

Loizou, P. C., (1998). “Introduction to cochlear implants.” IEEE Signal Processing Magazine, September 1998, 101-130

Loizou, P.C., Stickney, G., Mishra, L., and Assmann, P. (2003). “Comparison od speech processing strategies used in the Clarion implant processor.” Ear Hear. 24(1), 12-9

Macherey, O. and Carlyton, R. P. (2014) “Cochlear implants,” Current Biology, 24, R878-R884.

Mackersie, C. L., Boothroyd, A., and Minnear, D. (2001). “Evaluation of the Computer-Assisted Speech Perception Assessment Test (CASPA),” J. Am. Acad. Audiol. 12, 390-396.

McKay, M. C., and Henshall, K. R. (2002). “Frequency-to-electrode allocation and speech perception with cochlear implants,” J. Acoust. Soc. Am. 111, 1036.

Moller, A.R. (2006). “History of cochlear and auditory brainstem implants,” Cochlear and brainstem implants 64, 1-10.

Mudry, A. and Mills, M. (2013). “The early history of the cochlear implant: A retrospective.” JAMA Otolaryngology-Head and Neck Surgery 139(5), 446-453.

Nilsson M., Soli, S. D., and Sullivan J. A. (1994). “Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise,” J. Acoust. Soc. Am. 95, 1085-1099.

Nittrouer S., Tarr, E., Bolster V., Caldwell-Tarr, A., Moberly, A. C., and Lowenstein J. H. (2014a). “Low frequency signals support perceptual organization of implant-simulated speech for adults and children,” Int. J. Audiol. 53, 270-284.

Nittrouer S., Lowenstein J. H., Wucinich, and T., Tarr, E. (2014b). “Benefits of preserving stationary and time-varying formant structure in alternative representations of speech: Implications for cochlear implants.” J. Acoust. Soc. Am. 136, 1845-1856.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). “Speech perception without traditional speech cues,: Science 212, 947-949.

Shannon R. V., Zeng, F. G. Kamath, V. Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues." Science 270(5234), 303-4

Skinner, M., Holden, L., and Holden, T. (1995). "Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy. Ann. Otol. Rhinol. Laryng. Suppl. 166, 307-311

Skinner M. W., Holden L. K., Holden T. A., Dowell R. C., Seligman, P.M., Brimacombe, J. A., Beiter, A. L. (1991). "Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus Multi-Electrode Cochlear Implant," Ear. Hear. 12(1), 3-22.

Studdert-Kennedy, M. (1983). "Limits on alternative auditory representations of speech," Ann. NY Acad. Sci. 405, 33-38.

Tye-Murray, N., Lowder, M., and Tyler,R. S. (1990). "Comparison of the F0F2 and F0F1F2 processing strategies for the Cochlear Corporation cochlear implant." Ear hear. 11(3), 195-200.

Vondrasek, M., Sovka, P., and Tichy, T. (2008). "ACE strategy with virtual channels," Radioengineering, 17(4), 55-61.

Whitford, L., Seligman, P., Everingham, C., Antognelli, T., Skok, M., Hollow, R., . . . Clark, G. (1995). "Evaluation of the Nucleus Spectra 22 Processor and New Speech Processing Strategy (SPEAK) in Postlinguistically Deafened Adults." *Acta Oto-Laryngologica, 115*(5), 629-637.

Wilson, B. S., Dorman, M. F., (2008). "Cochlear Implants: A remarkable past and brilliant future," Hear. Res. 242, 3-21.

Yost, W. A. (2006). Fundamentals of Hearing: An Introduction (5th ed.). *Structure of the Inner Ear and its Mechanical Response* (pp. 83-101*); The Abnormal Auditory System* (pp. 249-260). *Chicago*, IL: Emerald Group Publishing Limited

Zeng, F. (2004) "Trends in cochlear implants," Trends Amplif. 8(1), 1-34

## Tables

**Table 1.** Cochlear implant participant demographics. PTA: Unaided four-tone pure tone average at .5, 1, 2, and 4 kHz

| Participant | Gender | Age (years) | Implantation Age (years) | Side of Implant | Hearing Aid | Etiology of Hearing Loss | Better Ear PTA (dB HL) |
|---|---|---|---|---|---|---|---|
| 1 | F | 62 | 54 | B | N | Genetic | 105 |
| 2 | F | 64 | 62 | R | Y | Genetic, progressive as adult | 75 |
| 3 | M | 64 | 61 | L | N | Noise, Meniere's | 80 |
| 6 | M | 67 | 65 | R | N | Genetic, progressive as adult | 84 |
| 7 | M | 56 | 52 | B | N | Rubella, progressive | 105 |
| 8 | F | 54 | 48 | R | Y | Genetic, progressive | 105 |
| 9 | M | 77 | 67 | L | N | Genetic, progressive | 93 |
| 10 | M | 77 | 76 | R | Y | Progressive as adult, noise, sudden | 71 |
| 16 | F | 61 | 59 | R | N | Progressive as adult | 105 |
| 17 | M | 23 | 14 | L | Y | Congenital, progressive | 100 |
| 19 | F | 73 | 67 | L | N | Genetic, autoimmune | 105 |
| 25 | M | 57 | 56 | R | Y | Autoimmune, sudden | 76 |

**Table 2.** Individual participant scores for each task.

| Participant | Meaningful SPEECH | Meaningful STANDARD | Words SPEECH | Words STANDARD | Vowels SPEECH | Vowels STANDARD | Syntax-only SPEECH | Syntax-only STANDARD |
|---|---|---|---|---|---|---|---|---|
| 100001 | 99 | 94 | 48 | 48 | 72.2 | 66.7 | 60 | 52 |
| 100002 | 41 | 40 | 2.2 | 5.6 | 40 | 15.56 | 0 | 0 |
| 100003 | 75 | 28 | 13 | 6.7 | 47.8 | 56.7 | 0 | 0 |
| 100006 | 52 | 62 | 10 | 17 | 43.3 | 36.7 | 0 | 0 |
| 10007 | 79 | 77 | 16 | 17 | 27.8 | 20 | 28 | 22 |
| 100008 | 92 | 90 | 27 | 36 | 30 | 35.6 | 33 | 62 |
| 100009 | 86 | 65 | 17 | 20 | 25.56 | 23.3 | 27 | 16 |
| 100010 | 84 | 88 | 13 | 34 | 35.5 | 30 | 23 | 44 |
| 100016 | 15 | 8 | 0 | 3.3 | 48.9 | 35.6 | 24 | 4 |
| 100017 | 62 | 46 | 2.2 | 8.9 | 56.7 | 58.9 | 15 | 5 |
| 100019 | 90 | 88 | 27 | 33 | 14.4 | 14.4 | 33 | 31 |
| 100025 | 56 | 67 | 2.2 | 6.7 | 50 | 56.7 | 3 | 3 |
| Mean Performance | 69.23 | 62.8 | 14.73 | 19.63 | 41.01 | 37.51 | 20.5 | 19.92 |