CUSTOMIZED SOUND PROCESSING AND FITTING PARADIGMS FOR COCHLEAR IMPLANT USERS

by

Hussnain Ali

APPROVED BY SUPERVISORY COMMITTEE:

____
Dr. John H. L. Hansen, Chair

____
Dr. Blake Wilson

____
Dr. Peter Assmann

____
Dr. P. K. Rajasekaran

____
Dr. Carlos Busso

To my beloved country – Pakistan,

my teachers – Dr. Shoab Khan and Dr. Philipos Loizou,

and my parents.

CUSTOMIZED SOUND PROCESSING AND FITTING PARADIGMS FOR COCHLEAR IMPLANT USERS

by

HUSSNAIN ALI, BE, MSEE

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

May 2016

ACKNOWLEDGMENTS

I would like to begin by expressing my gratitude to all my teachers who have imparted knowledge to me over the years. Thank you, Baji Rabia, Ms. Gulcha, Ms. Benedicta, Ms. Angelina, Ms. Sabookhi, Ms. Dulsee Isaac, Ms. Yaseen, Ms. Aamra, Ms. Shagufta, Mr. Taher, Mr. Shaukat, Mrs. Nasreen Iftikhar, Ms. Sajda, Mr. Shehryar, Mr. Taji, Dr. Khalid Munawer, Dr. Shoab Khan, Dr. Rajasekaran, Dr. Fonseka, Dr. John Hansen, and the late Dr. Philip Loizou. I owe it to every single of them for this achievement and to every success I may achieve in the future, for which I will always be indebted to them.

Many thanks to my committee members, Dr. Blake Wilson, Dr. Peter Assmann, Dr. P. K. Rajasekaran, and Dr. Carlos Busso for their precious time and invaluable feedback and guidance on my research. I would like to particularly like to thank my doctoral advisor, Dr. John H. L. Hansen, for his constant support and patience during the course of my PhD. I am indefinitely indebted to him for sharing with me his knowledge and expertise in speech signal processing, hearing science. I am immensely grateful to him for providing me with freedom in research that allowed me to explore many fruitful directions during my graduate studies.

My sincere gratitude to Dr. René Gifford and Dr. Jack Noble at the Vanderbilt University, who opened up their lab and resources to help me with my research. I would also like to take this opportunity to thank Dr. Emily Tobey, Dr. Michael Dorman, and all the researchers working in the field of speech, hearing sciences, hearing impairment, and auditory prostheses.

This acknowledgement would be incomplete without the mention of my lab members, and cochlear implant users who participated in my research studies. I am grateful to them for all the difficult listening tests they willingly participated in just to help the cochlear implant community in the future. Their dedication and hard work will pay off and impact society.

Last but not the least, I am grateful to my parents and family for their love and unconditional support. Thank you for giving me the opportunity to pursue my dreams. My *abbu* (dad) for ALL you have done for me, which I can't even begin to describe. My sweet *ammi* (mom) for being the blessing in my life. You both deserve all the credit, I hope I did you proud. My brother for buying me that chemistry book. My sisters for their unconditional love. My nephew and nieces for filling my life with color and joy. (I'm sorry for missing out your precious childhood years. I hope you would put pursuit of knowledge and education as big priorities in your lives.) My dearest friend and colleague, Talha, for being with me in spirit in this arduous journey to uplift my energy and motivation. Sadaf, my wife, for coping with my schedule and standing with me to the last minute. Thank you, all family and friends for your best wishes, prayers, and support.

I hope this work will advance science and technology related to cochlear implants and will be able to create a positive impact on the quality of life of hearing-impaired people around the world. I hope that I continue to spread the knowledge that has been imparted to me by my teachers and I help others as they have helped me.

The work presented in this dissertation was supported by Grant No. R01 DC010494 from the National Institute on Deafness and Other Communication Disorders (NIDCD), National Institutes of Health (NIH).

November 2015

PREFACE

This dissertation was produced in accordance with guidelines which permit the inclusion as part of the dissertation the text of an original paper or papers submitted for publication. The dissertation must still conform to all other requirements explained in the "Guide for the Preparation of Master's Theses and Doctoral Dissertations at The University of Texas at Dallas." It must include a comprehensive abstract, a full introduction and literature review, and a final overall conclusion. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported.

It is acceptable for this dissertation to include as chapters authentic copies of papers already published, provided these meet type size, margin, and legibility requirements. In such cases, connecting texts which provide logical bridges between different manuscripts are mandatory. Where the student is not the sole author of a manuscript, the student is required to make an explicit statement in the introductory material to that manuscript describing the student's contribution to the work and acknowledging the contribution of the other author(s). The signatures of the Supervising Committee which precede all other material in the dissertation attest to the accuracy of this statement.

CUSTOMIZED SOUND PROCESSING AND FITTING PARADIGMS FOR

COCHLEAR IMPLANT USERS

Publication No. ___________________

Hussnain Ali, PhD
The University of Texas at Dallas, 2016

Supervising Professor: Dr. John H. L. Hansen

The peripheral sensory auditory system is an intricate synergy between mechanical, chemical, and electrical bio-systems that generate complex temporal-spectral patterns of neural activity for sound perception. Artificial electrical stimulation provided by contemporary multichannel cochlear implants (CIs) aim to mimic the natural electrical stimulation patterns that occur at the spiral ganglion nerves; however, sparsity of information provided by the CIs is beyond comparison to the exquisite intricacies of a normal auditory system. This, in part is due to the current limitations inherent in the design of CIs, particularly the electrode-nerve interface, which may be a bottleneck for optimal performance. Spectral mismatch, poor temporal and spectral resolution, and current spread combined with patient-specific physiological, audiological, and cognitive factors are some of the key challenges that may impact performance and could potentially be responsible for large variations in CI performance outcomes.

Although electrode placements relative to the spiral ganglion are generally unknown, conventional sound coding algorithms and clinical fitting procedures follow a *one-size-fits-all* strategy. This dissertation aims to customize/personalize the sound coding and fitting procedures according to an individual's cochlear physiology. This is achieved by utilizing novel imaging procedures, more specifically CT images of recipients' cochleae, to determine the precise spatial location and orientation of the electrode contacts and the corresponding neural stimulation sites to produce a tailor-fit frequency-to-place function. In addition, patient-specific channel selection optimization techniques have been presented that aim to optimize presentation of electrical stimulation patterns. The proposed schemes have been evaluated in groups of normal hearing individuals using CI simulations as well as cochlear implant recipients. The data from the experiments suggest that patient-specific sound coding and fitting schemes may potentially aid in achieving higher asymptotic performance and possibly faster adaptation to electric hearing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The Cochlear Implant (CI) is one the greatest engineering marvels of modern medicine and can be considered a benchmark in neural prostheses for auditory restoration and rehabilitation. Although the history of auditory prostheses is relatively short, electrical stimulation of the auditory pathways has its roots across the globe which spans several decades. Our understanding of sound and the science of hearing dates as far back as the sixth century B.C., with fundamental interpretation of sound as vibrations by Pythagoras, followed by Galen's discovery on nerves as the transmission mechanism of sound sensation to the brain in 175 A.D. The mechanics of the middle ear were not understood until the mid-sixteenth century when Andreas Vesalius discovered the malleus and incus, two of the three middle ear bones. This was followed by Gabriello Fallopio's discovery of the cochlea in 1561, and finally the discovery of the organ of hearing – termed organ of corti by Alfonso Corti in 1851. The first demonstration of electrical stimulation came from none other than Alessandro Volta, the inventor of the electrical battery, in 1800 when he placed the ends of a 50-volt battery in his each ear and experienced sensation of *noise-like crackling* sound [1]. Many years later, the first attempt towards an auditory prosthesis was reported by Drs. André Djourno and Charles Euriès in Paris in1957, which spurred intensive activities in the 1960s and 1970s in the United States, central Europe, and Australia towards the modern era cochlear implant. Pioneering scientists in the Unites States were Drs. William F. House, Doyle brothers – John and James Doyle, Francis B. Simmons, Robin Michelson, Michael

Merzenich, and Blake Wilson. In Australia, Dr. Graeme Clark led the research and development activities by founding the Bionic Ear Institute. Parallel research efforts across Europe were led by Drs. C. H. Chouard, Stefaan Peeters, F. E. Offeciers, Kurt Burian, Erwin Hochmair, and Ingeborg Hochmair. In these early years of struggle towards commercial CI development, technical and safety concerns raised skepticism among critics, mainly in the basic science community, who adamantly opposed the idea of a cochlear prosthesis. They believed that it was impossible to mimic the natural exquisite machinery of the inner ear and that cochlear implantation would not yield any useful hearing beyond environmental awareness or possibly speech cadences, at best. The pioneering and persistent spirit of the above scholars laid the foundation of the modern-day cochlear implant. The famous Bilger report commissioned by NIH in 1977 finally paved the way towards commercial FDA approved devices in the United States with the House-3M single-electrode implant being the first to get FDA approval in 1984, shortly followed by Nucleus multichannel implant in 1985. For further details on the inspirational history of cochlear implants, please see [2-5].

The last three decades have seen exponential growth in cochlear implantation, both in terms of the number of implant recipients, as well as scientific advancements and publications across the globe. Figure 1.1 depicts the number of implant recipients which has increased at least 5 fold in the last decade from 60, 000 implant recipients in 2001 to 324, 000 by the end of 2012 [6]. The number of scientific publications as indexed by PubMed mirrors the same exponential trend, from on average 141 publications per year in the 1990s to 284 per year in the years 2000 – 2010 (see Figure 1.2). Year 2015 alone has produced 630 publications in the field of cochlear implants. These trends indicate both a maturity of the field as well as reflection of the confidence

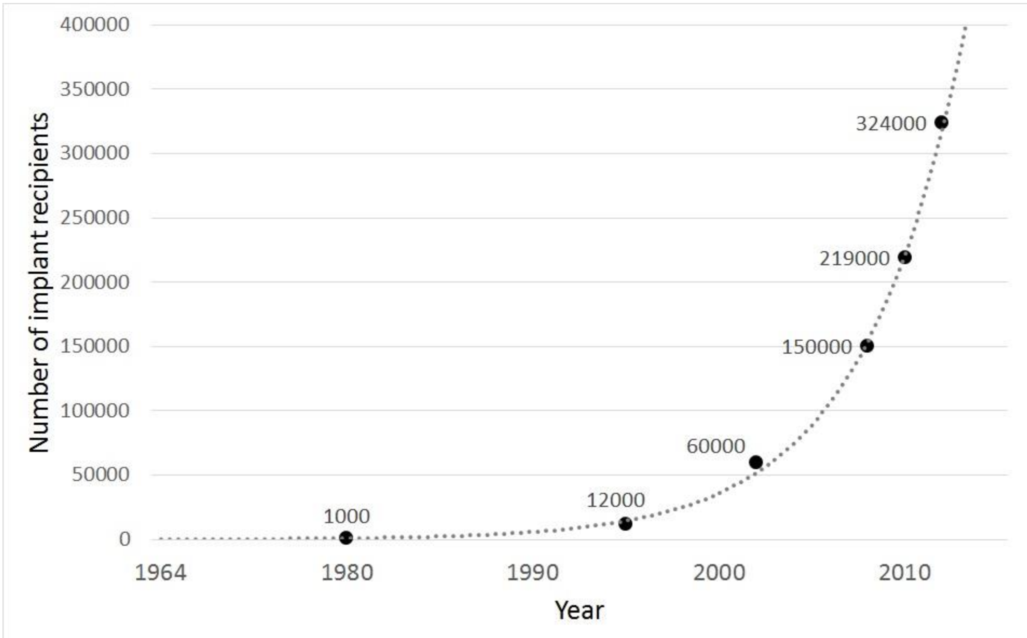


Figure 1.1. Number of implant recipients over time [3, 6-8].

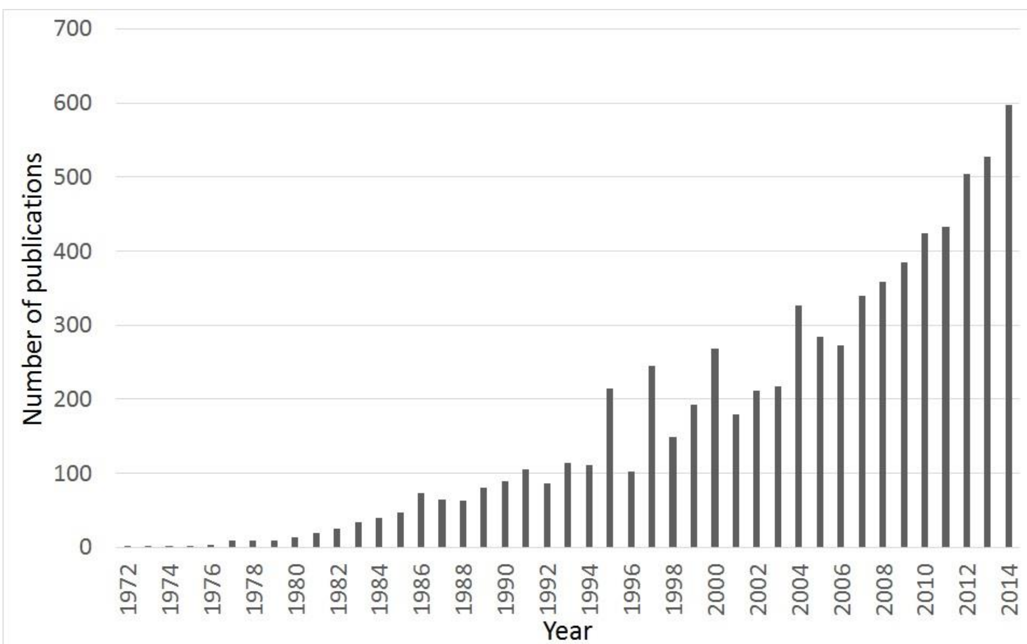


Figure 1.2. Annual number of scientific publications in the field of cochlear implants as indexed by PUBMED. The data were retrieved on October 17, 2015 from http://www.pubmed.gov [9].

in implant technology and performance that has led to a growing market share of CIs as an effective auditory prosthesis.

Contemporary cochlear implants can provide a high level of speech recognition and aural ability to a majority of implant recipients that are on par with normal hearing individuals, at least in quiet environments. In general, CI outcomes are very encouraging and to a much greater extent very surprising because implants provide only a very crude mimicking of limited aspects of the normal hearing physiology. In fact, it is the remarkable ability of the brain to decipher the artificial electrical stimuli of sparse inputs that has led to the ultimate success of the cochlear implant. Scientific studies suggest that implant outcomes largely depend on auditory pathways, central decoding, and neural brain plasticity. The importance of the brain in learning to classify the neural activity evoked by cochlear implants is so crucial, Drs. Wilson and Dorman describe the brain as *"the tail that wags the dog"* [4, 10, 11]. Therefore, in recent years there has been a strong push towards understanding the cortical function and brain plasticity in response to synthetic electrical stimulation of the auditory pathways. While the focus of this dissertation is not on top down processing, but instead to leverage the physiological and cognitive factors as blueprints when devising new signal processing strategies to improve the bottom-up peripheral processes in cochlear prosthesis.

1.1 Problem Statement

Cochlear implants, like other neural prosthetic devices, face two key challenges that have both theoretical and practical implications, namely analysis issues and presentation issues [1 - 3]. The analysis problem is concerned with how to effectively and efficiently encode the sound information given the limited number of channels. The presentation problem focuses on how to optimally map the coded information into a perceptually meaningful neural stimulation via the prosthesis. At one end, current and emerging sound coding strategies along with novel speech processing algorithms aim to address the analysis issues; however, the presentation challenge is increasingly emerging as a bottleneck in limiting the possibilities and scope of what signal processing engineers can achieve with current technology. Shortcomings inherent in the design of contemporary CIs such as the limited number of electrodes and the associated current spread due to spatial channel interactions limit the number of perceptually discriminable stimulation sites and hence limit spectral resolution. The presentation issue, in part, may be one of the likely factors that results in large variability in implant outcomes – a question that continues to puzzle the CI research community. In essence, the presentation issue forms a bridge between the peripheral processing and the central auditory perception.

Bridging the peripheral and central auditory systems are the neurons of the spiral ganglion that reside in the organ of corti. In natural hearing, sound creates a sequence of mechanical, chemical, and electrical bio-processes to stimulate these auditory nerve fibers to induce the sensation of hearing. Auditory nerve fibers have intrinsic “characteristic frequencies” (CFs) which are tonotopically organized in the cochlea, (e.g., nerve fibers located at deeper sites along the length of the cochlea have lower CFs and thus when they are simulated, lower pitched sounds are perceived). CIs exploit this natural phenomenon by providing electrical stimulation across the length of the cochlea via an electrode array, which is threaded into the cochlear bony labyrinth during surgery. Insertion depth of the electrode array, number of electrodes, electrode configuration, degree of neuronal survival, positioning and proximity of electrodes to the auditory nerve fibers largely determine which tonotopically mapped groups of nerve fibers are stimulated by each electrode and thus form the core of the presentation challenge. Deeper electrode insertions generally favor improved speech recognition in CIs [4-9]. This is due to the accessibility of the apical regions of the cochlea which correspond to lower frequencies and hence theoretically more low-frequency speech information (e.g., location of fundamental frequency and the first formant frequency) which can be provided without spectral distortion. However, deeper electrode insertion has challenges of its own, including insertion trauma and optimal electrode placement (electrode array migrating to neighboring cochlear canals). Variations in electrode insertion depth result in differences in the accessible tonotopic range among implant recipients, (i.e., the range of CF stimulated at the corresponding electrode locations). For example, according to the Greenwood function, with an insertion depth of 30 mm from the round window, the most apical electrode would correspond to CF of approximately 185 Hz, while a shallower insertion of 20 mm would correspond to 1170 Hz. This offset/mismatch requires user adaptation for effective decoding of distorted representation of speech through the CI. Despite the variations in electrode insertion depth and the fact that the final positions of the electrodes in relation to the nerve fibers are generally unknown and unique for each patient, contemporary CI sound processors use a standard mapping strategy which assumes that electrodes are optimally situated in the cochlea. Each electrode is programmed by the CI processor to stimulate nerve fibers corresponding to a predefined frequency-bandwidth. The standard mapping assigns default frequency allocation to all implantees and maps the full acoustic range of speech (approximately 100 - 200 up to 8500 Hz) to the tonotopic location of the electrode array with the expectation that CI users will learn to adapt to the modified map over time. Such a mismatch between frequency analysis bands of the CI sound processor and the CF of the nerve fibers that are stimulated can result in frequency-place shifting (frequency offset), frequency compression, expansion, warping, or a combination of the above. These factors not only deteriorate spectral characteristics of the perceived speech but may degrade overall user performance.

In addition to frequency-place mismatch, electrode channel-interference caused by the overlap of electric fields within the cochlea is a serious concern in present scala tympani implants. While the number of electrode contacts in a cochlear array may be up to 22 independent electrodes, clinical evidence suggests that cochlear implant users are not able to utilize more than about 4 - 8 effective channels of stimulation when stimuli are rapidly sequenced across electrodes in a real-time, speech processor context [10]. While speech recognition in quiet can be achieved by only 3 – 4 channels of spectral information, listening in complex acoustic environments can require as much as 30 or more effective channels for an equivalent level of performance [12]. Scientific evidence suggests that poor performers generally do not have more than four effective channels of stimulation (e.g., Friesen et *al.* [13], Dorman and Spahr [14], Garnham *et al.* [15], Wilson *et al.* [16]). To put it in perspective, the number of independent auditory filters in normal hearing is about 39, and is about 28 for the frequency range important for speech perception [17, 18]. Clinical sound processing strategies either stimulate all electrodes or select a subset of electrodes per stimulation cycle (in a time-interleaved fashion). This subset of electrodes is selected based on speech features. If CI users are only able to utilize 8 – 10 channels of stimulation, a channel-selection strategy that optimizes which electrodes are picked for activation based on speech features, environment, and spatial location of electrode contacts in the cochlea may help implant users to make better use of putative stimulation sites.

1.2 Dissertation Goals and Contributions

Up to this point in time, the lack of knowledge with respect to the spatial relationship between the electrodes and the neural stimulation sites has resulted in a generic default frequency mapping paradigm with the expectation that all CI users will learn to adapt to the clinically-assigned frequency locations of stimulation. In this dissertation, our primary goal has been to optimize some aspects of the presentation issues in present-day cochlear implantation technology by devising patient-specific, customized mapping strategies that are unique for each implant recipient and are based on the individual's electrode placements. This is achieved by leveraging image-guided procedures to determine the true location of individual electrodes with respect to the nerve fibers and designing patient-specific frequency allocation schemes that help to minimize sub-optimal frequency-place mapping distortions in CIs. In addition, sound processing strategies that aim to optimize the channel selection process on an individual basis are explored. The effectiveness of the proposed techniques are evaluated with groups of both normal hearing individuals using acoustic simulations of cochlear implants as well as with implant recipients. The proposed patient-specific sound coding and fitting schemes, together with findings from the studies reported in this dissertation may help with our understanding of adaptation to electric hearing with distorted and sparse representation of sound cues. The data may help with advances in CI technology.

Specifically, this dissertation aims to explore the following four research directions:

- **Image-guided customization of frequency-place mapping in cochlear implants:** This is the primary goal of the research conducted in this dissertation. A patient-specific, image-guided frequency allocation strategy will be proposed that aims to tailor-fit sound processor frequency assignments based on the electro-neural characteristics of each individual. The proposed frequency fitting strategy will be shown to minimize the frequency-place distortions by achieving a balance between frequency matching and frequency compression to provide a better representation of the sound signal.

- **Assessment of speech recognition with different configurations of spectral mismatch with groups of both normal hearing and cochlear implant users.** The acute and semi-chronic effects of various frequency-maps on the speech recognition performance are assessed systematically using acoustic simulation of cochlear implants with normal hearing individuals. In addition, the effectiveness of the proposed customized frequency mapping strategy will be evaluated in a semi-chronic study of three months with cochlear implants users.

- **Image-guided customization and optimization of channel selection.** An image-guided strategy to optimize channel selection in *n-of-m* sound coding strategies is proposed. The strategy leverages the spatial location of electrode contacts to adaptively control the activation of electrodes that are most likely to cause channel interaction. Simulation data is provided to demonstrate the potential effectiveness of the proposed scheme.

- **Optimization of channel selection based on signal-to-noise characteristics of the acoustic environment.** A method to optimize the channel selection process of *n-of-m* sound processing strategies in adverse listening conditions is proposed. The proposed scheme selects the most-information rich channels in each stimulation cycle based on the location of formant frequencies and the instantaneous signal-to-noise ratio. The data from acute experiments with cochlear implant users will indicate the potential benefit of the proposed scheme in adverse listening conditions.

1.3 Dissertation Outline

While the area of hearing science, hearing impairment, and cochlear prostheses is too vast to cover in a single dissertation, we have made every effort to address the relevant aspects in the following chapters. Fortunately, there have been a number of excellent quality of prior research and publications in this domain. References are provided throughout the text to facilitate enthusiastic readers to explore further. The dissertation is outlined as follows:

- **Chapter 2** provides an overview of the background literature review and gives further details on the state of the art in cochlear implant technology; factors that lead to variability in implant outcomes with special emphasis on some existing challenges including patient specific audiological and neuro-physiological reasons that may be responsible for reduced performance with the implant are presented.
- **Chapter 3** gives a detailed literature review on frequency-place mismatch in cochlear implants, its impact on the listening performance, and the role of perceptual learning.
- **Chapter 4** is the core of the dissertation and discusses an image-guided patient-centric approach towards optimization of frequency place function. This is accompanied by listening studies and experimental results from normal hearing individuals and cochlear implant users.
- **Chapter 5** provides details on strategies to optimize channel selection criteria in the signal processing chain that can help minimize unwanted channel interactions caused by the

current spread and attempts to optimize the listening experience in challenging listening environments.

- **Chapter 6** concludes the dissertation and summarizes the thesis contributions. It also provides limitations of the study and guidelines for advancing future research directions in this domain.

Appendices are provided at the end of this dissertation that contain relevant supplementary details.

CHAPTER 2

SOUND PROCESSING ADVANCEMENTS IN COCHLEAR IMPLANTS

At the time of this writing (Dec. 2015), there are three manufacturers that have FDA approved cochlear implant devices in the United States namely, Cochlear Corporation, Advanced Bionics, and Medical Electronics (MED-EL). In addition to the above three, Oticon Medical (formerly Neurelec) and Neurotron Biotechnology are emerging as new players but the lack of FDA clearance has limited their market shares primarily to France and China, respectively. Despite differences in sound processing strategies, number of electrodes, stimulation patterns, and value-added features, the basic operational principle as well as implant outcomes across vendors largely remain the same [8, 19-23]. On average, cochlear implant users are able to obtain 80% open-set sentence recognition in quiet and about 50% for word recognition with 10% - 30% standard deviation regardless of the device brand [4, 11]. However, there remains a huge variability in performance outcomes across implant recipients of each manufacturer, with the highest performers gaining speech recognition ability close to 100% while others are at 0% in complex listening environments. Wide variability across users has baffled scientists and clinicians since the early days of cochlear implantation and continues to present a challenge, primarily because there is no fixed formula to predict the outcomes of cochlear implantation success prior to surgery. Factors that work in one individual do not necessarily hold equal significance in others, and at times it is not possible to trace the cause and effect relationship for reasons that are not clearly understood today. This argument not only holds true for recipients who do not receive any speech recognition ability even years after surgery but equally valid for the "super-star" performers that continue to amaze those in the field on the remarkable ability of implants (given sparse mimicking of the peripheral auditory system). However, scientific studies and clinical data collected over the years has provided the community a better understanding of likely factors that may contribute to varying performance with implants.

We can broadly classify the factors that influence variability in performance outcomes into four (overlapping) categories, namely technological, neuro-physiological, audiological, and cognitive factors. The technological aspects relate to the device technology, mainly sound processing strategy, algorithms for speech enhancement in realistic environments, electrode design, and electrical stimulation patterns that are generated by the implant technology. Neuro-physiological factors include neuronal survival, health of auditory pathways, electrode-nerve interface, cochlear morphology, and others. Audiological factors refer to the etiology and duration of hearing loss, length of deprivation, and experience with the implant. Cognitive factors relate to central processing, i.e., the brain's ability in deciphering the artificial stimulation, neural plasticity, cross-modal cortex encroachment, learning, memory, attention, and language processing. It was initially thought that poor fitting or loss of auditory neurons was the likely cause of poor performance with implants. While this may be true to an extent, accumulating and compelling evidence points to audiological and cognitive factors as significant contributors to variability in outcomes [24-31]. This is not to say that neurophysiology or technological advances are not likely to favor the outcomes. It is the combination of these factors that ultimately has a synergistic effect on perception with cochlear implants. In the following sections, we will focus on how sound processing advancements have contributed to implant outcomes in light of available data and our experience with cochlear implants.

Although the basic design of a cochlear implant has not changed since the mid-1990s, cochlear implant devices have continued to improve over the years due to technological advances in microelectronics, computing, digital signal processing, battery technology, electrode designs, stimulation strategies, and better surgical techniques. These improvements at the system level have not only enhanced the functionality, usability, and safety of the device but has also led to remarkable improvements in performance. Percutaneous to transcutaneous link, body worn to behind the ear processors, single to dual microphones – all these system-level upgrades have enhanced the usability of the device.

Single channel to multi-channel cochlear implant was the first paradigm shift that enabled speech understanding ability for some recipients. For the larger part, improvements in sound processing have been, (and continue to), be responsible for the constant and steady improvements in performance outcomes [8]. In a retrospective study, Lenarz and colleagues at the Medical University of Hannover, Germany assessed the effect of technological advancements on speech perception performance in a cohort of 1005 cochlear implantees that received an implant over 5 virtual generations from 1984 to 2008 [32]. They found a positive influence of a maturity of the technology, and attributed improvements in electrode designs and speech processing strategies, largely responsible for improved speech perception performance. The following subsections provide further details with respect to cochlear implant technology

Sound processing advancements have manifested in two distinct ways: (i) improvement in speech processing strategies and sound coding algorithms, and (ii) improvements in digital signal processor hardware. While the first deals with how to effectively and efficiently encode the sound information to generate a perceptually meaningful neural stimulation, the latter enables implementation of sound coding algorithms to run on an embedded DSP in real-time. For the latter part of the discussion, we will focus on sound processing strategies as they relate to variability in implant outcomes. Let us start with a brief introduction of the peripheral auditory system.

In the normal auditory system, acoustic waves enter the pinna (outer ear), travel through the ear canal and vibrate the tympanic membrane (ear drum). These mechanical vibrations are translated by a series of transformations by the three middle ear bones - malleus, incus, and stapes. The foot of the stapes vibrates the oval window which creates traveling waves in the fluid inside the cochlea and causes the basilar membrane to vibrate. The physiology of the basilar membrane (BM) (width, stiffness) makes it akin to a spectrum analyzer (a frequency decoding system) with location and magnitude of the displacement along the BM corresponding to the frequency and intensity of the input sound respectively. Motion of the BM is sensed by the hair cells (located within the organ of corti) which generate electrical signals by the process of ion exchange. These electrical signals are sensed by the auditory nerve fibers (spiral ganglion) that branch out of the modiolus and group together to form the auditory nerve. The auditory nerve finally relays these electrical signals through a network of auditory pathways to the brain for subsequent central processing.

The goal of cochlear implants is to mimic, to the extent possible, the electrical excitation patterns at the spiral ganglion. All contemporary multiband sound processing strategies exploit the tonotopic map of the cochlea and filter the acoustic signal into different frequency bands in

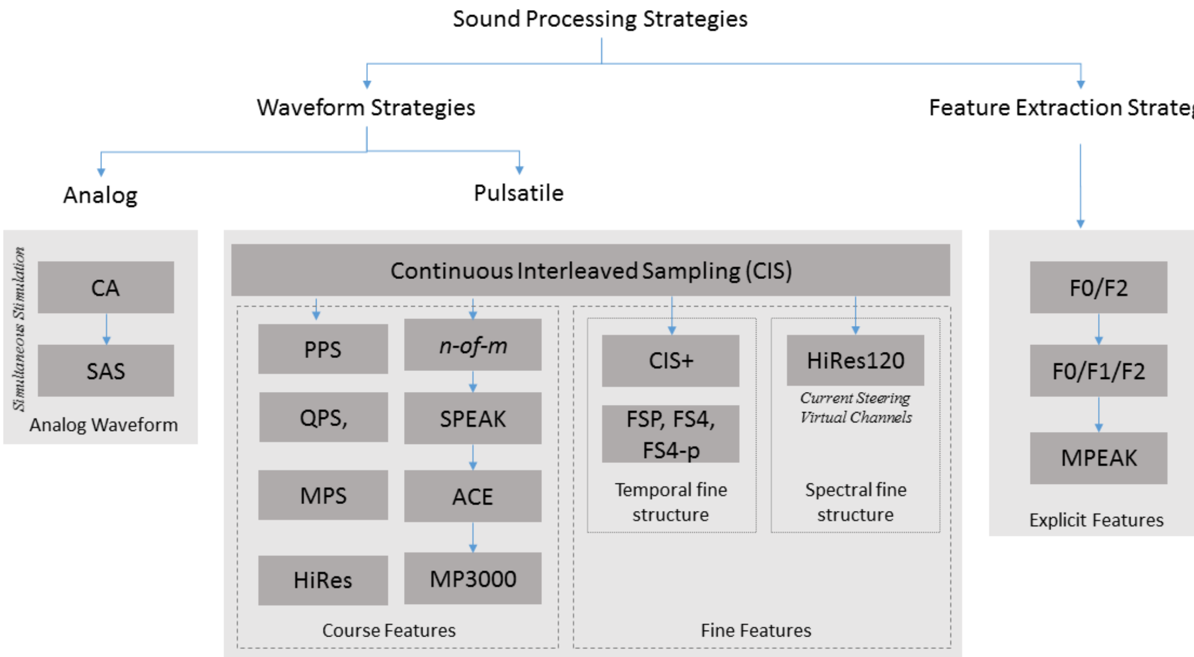


Figure 2.1. Classification of sound processing strategies.

an attempt to mirror the functionality of the cochlea. Figure 2.1classifies popular signal processing strategies that have been devised for cochlear implant systems over the years. Sound coding strategies can be categorized as waveform (temporal) or feature-based strategies. A relatively newer generation of fine-feature strategies that claim to deliver the fine temporal and spatial features of sound have also emerged. Explicit feature-based strategies, now obsolete, were used in early generations of cochlear implants. F0/F2 was the first strategy developed in the early 1980s that explicitly provided formant features (fundamental frequency and the second formant) to the appropriate electrodes [33, 34]. This was later refined to include the first formant frequency (F1) and was called the F0/F1/F2 sound coding strategy and became available in the Nucleus Wearable Speech Processor (WSP) in 1985 [35, 36]. Further improvements to the F0/F1/F2 strategy came with the MULTIPEAK (MPEAK) strategy that provided high-frequency information content in addition to formant frequencies [37]. These strategies provided a very coarse representation of the formant peaks over time and allowed implant users to obtain open-set speech recognition for the first time [38-40]. Around the same time, a fundamentally different approach was being used in the Ineraid device manufactured in Utah [41]. It used a filtered waveform together with automatic compression to deliver simultaneous stimulation to four intra-cochlear electrodes in an analog form. The approach called the Compressed-Analog (CA) provided open set speech recognition to its users and yielded superior performance in comparison to single-channel implants [41, 42]. With the advent of pulsatile waveform strategies in the early 1990s, analog and feature-based strategies were almost entirely replaced. The theoretical basis of waveform or envelope-based strategies is based on Dudley's VOice CODER (VOCODER) [43] (i.e., sound processing is carried out on temporal envelopes of band-passed acoustic signal). Continuous Interleaved Sampling (CIS) algorithm formulated by Wilson *et al.* [44] can be considered as the de-facto standard for all multi-channel sequential stimulation sound coding strategies that have emerged to date.

In the CIS strategy, the acoustic signal is first pre-emphasized and passed through a set of band-pass filters. The number of band-pass filters *'m'* is equal to the number of stimulating electrodes in the electrode array. The envelope of each band is extracted either by a Hilbert transform, or more commonly by rectification followed by low-pass filtering. The extracted envelope is then compressed by a logarithmic mapping function to account for relatively low dynamic range of electric hearing (5 – 15 dB) as compared to the acoustic dynamic range which is on the order of 100 dB. The mapping is patient-specific and accounts for the lowest and most comfortable levels of perceivable sound at each stimulation site. The implant receiver modulates biphasic electric pulses with the compressed envelopes and delivers modulated pulses to the corresponding electrodes in a time-interleaved fashion. The time-interleaved approach allows accurate and minimally interfering representation and delivery of band-limited envelope cues to the auditory nerve. The signal flow in the CIS algorithm is depicted in Figure 2.2. This elegant and yet effective approach addresses a broad range of issues from front-end processing (e.g., spectral weighting, filtering, and envelope extraction) to back-end processing (e.g., amplitude compression, mapping, and constant high-rate, interleaved, sequential pulsatile stimulation), to engineering implementation and device fitting. All these factors eventually led to a high level of speech understanding ability with cochlear implants and are the reasons that CIS continues to be used in almost all major implant systems.

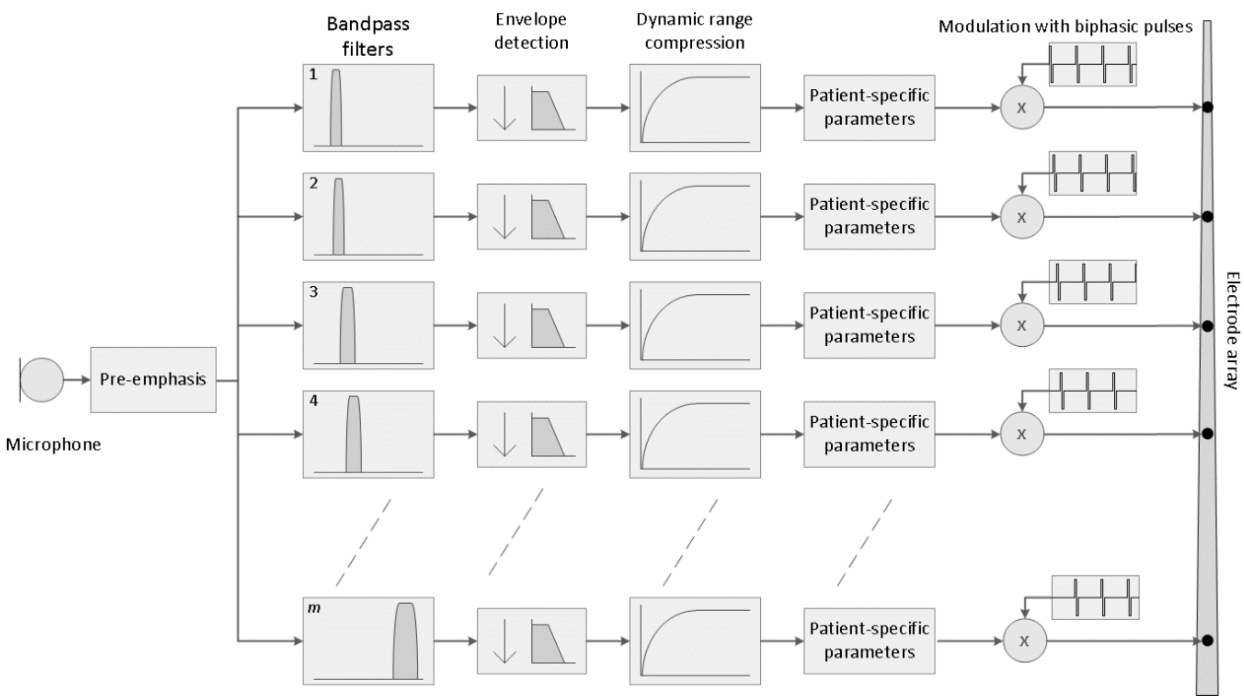


Figure 2.2. Signal flow in Continuous Interleaved Sampling (CIS) algorithm.

The *n-of-m* strategies [45] are a variant of the CIS strategy and are typically used in implant systems that have a larger number of electrodes (though CIS is not limited to number of electrodes). The basic signal pipeline remains the same – band-pass filtering followed by envelope extraction, however there are two major differences. The first difference is that processing is carried out on temporal frames that are typically 2 – 8 ms in duration. The second difference is in channel selection, (i.e., in each processing frame, only *n* out of possible *m* electrodes are selected for stimulation). Channel selection is based on the bands with the highest energy (corresponding to spectral peaks in that stimulation cycle). A typical value of *n* is between 8 – 10, and for Nucleus devices $m = 22$ corresponding to a 22-electrode scala tympani electrode array. One of the earlier flavors of *n-of-m* strategies was the SPEAK strategy (from Spectral Peak) [46] which has evolved into Advanced Combinations Encoder (ACE) strategy [47, 48] and is used in most devices by Cochlear Corporation. The major difference between the two is the fixed stimulation rate in the former (250 Hz), while the later can be programmed to run at higher or equal stimulation rates (990 Hz is seen as a typical stimulation rate in most implant systems using the ACE strategy – but this can be changed by the audiologist according to user preference). A more recent variant of ACE strategy is called MP3000 (also PACE) strategy that uses psychoacoustic masking models to optimize the channel selection process and select perceptually relevant channels for stimulation [49]. With its conceptual roots in MP3 compression, the MP3000 algorithm aims to reduce the spread of excitation by reducing the number of clusters (neighboring channels) that are selected in a typical *n-of-m* approach. It is argued that masking functions used in the algorithm are tuned to normal hearing and therefore may not equally apply to electrically evoked hearing. However, the performance of MP3000 is demonstrated to be on a par with some of the previous methods [50]. The major advantage of MP3000 is in the increased battery performance relative to ACE due to reduced stimuli. The MP3000 approach is not currently used in commercial devices.

Since conventional envelope-based strategies provide limited fine temporal and spectral features, there has been a strong push towards better representation of fine features in sound coding strategies. Med-El uses a fine-structure processing (FSP) scheme to provide additional information about the temporal fine structure of the signal [21, 51-53]. This is achieved by delivering bursts of high rate stimulus pulses on the apical-most electrodes that are triggered by positive zero-crossings of the band-pass filtered waveform. A significantly different approach employed in Advanced Bionics processors aims to enhance the spatial spectral resolution by using current steering to invoke virtual channels [54]. The strategy called HiRes120, utilizes eight different ratios of currents to create eight virtual channels per adjacent pair of physical electrodes which leads to 120 sites in total from an electrode array consisting of 16 electrodes. These additional (virtual) pitches claim to improve temporal and spatial resolution of the stimulation patterns; however, there is no evidence for improvement in speech understanding ability or music perception over other strategies [55, 56].

The performance of all of the above sound processing strategies has been evaluated extensively and reported in the literature. Figure 2.3 summarizes the sentence recognition scores in quiet as a function of time for different speech processing strategies and device models [57]. Clearly, with advancements in sound processing, speech recognition performance of implant users has consistently improved over the years. However, despite the differences in device types, electrodes, and speech processing strategy, there appears to be no significant difference in performance by users of different devices. More importantly, users of the same device and similar speech processing strategy show a wide range of outcomes.

Another aspect that is worth exploring is the role of various speech enhancement algorithms in improving speech recognition performance particularly in noise. There have been many speech enhancement algorithms reported in the literature over the years [58, 59]. While most have been limited to experimental and/or laboratory trials, a few techniques have made their way into commercial devices and are used in conjunction with sound processing strategies. Notable among these techniques are Adaptive Dynamic Range Optimization (ADRO) [60-62], automatic gain adjustment, and ClearVoice [63] that are being used in Cochlear, Med-El, and Advanced Bionics devices. All these approaches in conjunction with a sound processing strategy claim to provide better speech understanding ability in everyday listening conditions on average.

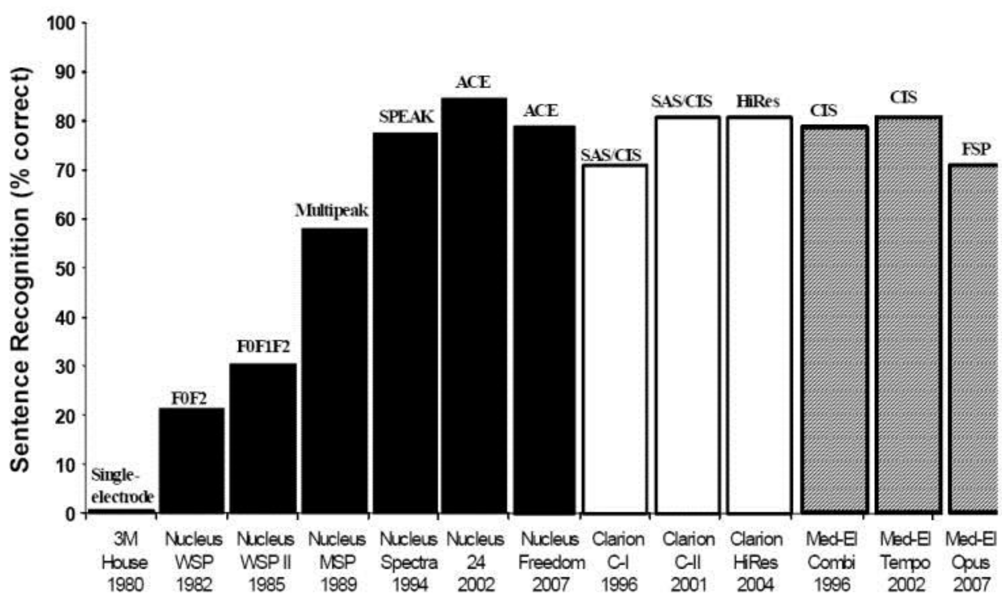


Figure 2.3. Sentence recognition scores in quiet as a function of sound processing strategies and cochlear implant devices over the years.Reprinted with permission from [57]. Copyright 2008, IEEE.

A word of caution when conducting experiments with cochlear implant users is to be careful about any possible interactions of these add-on processing schemes, with the experiment in hand since these schemes do not necessarily produce equal outcomes for all users. For example, an experiment that aims to evaluate the effectiveness of a speech enhancement algorithm may use pre-processed audio files that are played in free field during the experiment. The add-on processing strategies could have a combined effect that may result in higher individual variations. In [64], we investigated the effect of ADRO on speech understanding ability of cochlear implant users in adverse listening conditions. We tested speech recognition in anechoic quiet, noisy, reverberant, noisy reverberant, and reverberant noisy conditions. Although no significant effect of ADRO processing on speech intelligibility was observed, a huge variability across subjects in all listening conditions was seen. Seven out of ten subjects had equal or better scores without ADRO for speech in the presence of reverberation and noise together. Our results indicated that while ADRO is turned on by default in all Nucleus devices, it may have detrimental effects for some listeners in some listening conditions, and therefore it may be left as an optional setting for the user. Further details about the study are given in Appendix A.

Contemporary multiband signal processing strategies try to mimic the normal peripheral auditory system in an approximate way. The deterministic electrical stimulation patterns that are provided by the present scala tympani electrodes do not have the fine temporal resolution; nor the spatial resolution needed to provide sharp speech nuances needed to represent speech and music with full detail. In the normal auditory system, highly nonlinear filtering, feedback loops, level dependent compression mechanisms, and random spontaneous activity in auditory neurons occur. The lack of these factors in CIs may well limit the perceptual abilities of implant recipients.

CHAPTER 3

FREQUENCY-PLACE MISMATCH

Auditory nerve fibers have intrinsic “characteristic frequencies” (CFs) and are tonotopically organized in the cochlea, (e.g., nerve fibers located at deeper sites along the length of the cochlea have lower CFs and thus when they are simulated, lower pitched sounds are perceived). CIs exploit this phenomenon by providing electrical stimulation along the length of the cochlea via an electrode array, which is threaded into the cochlear bony labyrinth during surgery. Insertion depth of the electrode array, number of electrodes, electrode configuration, degree of neuronal survival, positioning and proximity of electrodes to the auditory nerve fibers, largely determine which tonotopically mapped groups of nerve fibers are stimulated by each electrode. Deeper electrode insertions generally favor improved speech recognition in CIs [4-9]. This is due to accessibility of the apical regions of the cochlea which correspond to lower frequencies and hence theoretically more low-frequency speech information (e.g., location of F0 and the first formant frequency) which can be provided without spectral distortion. However, a deeper electrode insertion has challenges of its own, including, insertion trauma and optimal electrode placement (electrode array migrating to neighboring cochlear canals during the surgery). Variations in electrode insertion depth result in differences in the accessible tonotopic range among implant recipients, (i.e., range of CF stimulated at the corresponding electrode locations). For example, as noted in Chapter 2, with insertion depth of 30 mm from the round window, the most apical electrode would correspond to CF of approximately 185 Hz, while a shallower insertion of 20 mm will correspond to 1170 Hz. This offset/mismatch requires user adaptation for effective human decoding of speech through the CI. Despite the variations in electrode insertion depth and the fact that the final positions of the electrodes in relation to nerve fibers are generally unknown and unique for each patient, contemporary CI sound processors use a standard mapping strategy. Each electrode is programmed by the CI processor to stimulate nerve fibers corresponding to a predefined frequency-bandwidth. The standard mapping assigns default frequency allocation to all implantees and maps the full acoustic range of speech (approximately 100 - 200 up to 8500 Hz) to the tonotopic location of the electrode array with the expectation that CI users will learn to adapt to the modified map over time. Such a mismatch between frequency analysis bands of the CI sound processor and the CF of the nerve fibers that are stimulated can result in frequency-place shifting (frequency offset), frequency compression, expansion, warping, or a combination of the above. These factors not only deteriorate spectral characteristics of the perceived sound but may degrade human decoding performance.

This chapter starts with the differences in the tonotopic map of the organ of corti and spiral ganglion followed by discussion on the effects of frequency-place mismatch on speech recognition performance of cochlear implant users. A detailed literature review is provided and findings from prominent studies conducted with cochlear implant recipients and normal hearing individuals in this regard are discussed.

3.1 Cochlear morphology and frequency map of the human cochlear spiral ganglion

The frequency map of the cochlea is calculated based on the trajectories of the peripheral processes (dendrites) of the spiral ganglion cells. In normal hearing, the site of electric spike initiation is governed by the hair cells that depolarize at the organ of corti and the innervating nerve fibers. Thus, for normal hearing, spatial locations along the organ of corti could serve as possible frequency coordinates, given that the true length of the organ of corti is known. Until very recently, the Greenwood frequency-position function [65, 66] served as the benchmark standard to estimate the tonotopic frequencies represented by neurons along the organ of corti. The Greenwood function is an empirically derived logarithmic relationship between frequency and its place of representation along the basilar membrane that has been shown to be consistent for many mammalian species (when scaled to appropriate cochlear length and audible frequency range). While the Greenwood function may provide a good estimate of the frequency map for normal hearing (at the level of organ of corti), it may not be accurate at the level of spiral ganglion, particularly for a compromised peripheral auditory system. The anatomy and physiology of the cochlea limits applicability of Greenwood function for electrical stimulation provided by scala tympani implants, primarily because the site of the electric spike initiation is not determined by the organ of corti, but by the neural elements that are stimulated by the electrodes. In addition, orientation of the hair cells, and the trajectories of the nerve fibers play a significant role. In the basal cochlear coil, the nerve bundles form radial projections from the modiolus to the organ of corti. In the middle and apical turns, the trajectory of the fibers deviates significantly because Rosenthal’s canal does not extend all the way to the apical turn [67], but terminates in a bulge in the middle turn in which the spiral ganglion cells are densely packed wall to wall [68]. Cell projections beyond the middle turn follow vertical trajectories. In a compromised cochlea, dendrites soon start to degenerate with the onset of impairment at the periphery [69]. Although, spiral ganglion cell somata may survive in larger populations long after deafness and CI use [70], the effective tonotopic map at the spiral ganglion may be significantly different from the organ of corti map. Furthermore, the estimates computed by the Greenwood function are only as accurate as the exact length of the organ of corti, which cannot be determined in most temporal bone and imaging studies. Estimates of frequency based upon the average OC length could be inaccurate due to significant individual variations [71, 72]. These factors have serious implications for scala tympani cochlear implants that rely on a valid/accurate cochlear tonotopic map to assign bandlimited sound signals to their corresponding electrodes. More importantly, different electrode array types may establish varying frequency-to-electrode relationships with similar insertion depth within the same implantee if positioning of the electrode array differs (i.e., if the distance of the electrodes to the modiolor wall is significantly different).

Kawano *et al.* [71] utilized computer-aided three-dimensional reconstruction of eight adult cochleae to measure the lengths of organ of corti (mean ± SD, 35.58 ± 1.41 mm), Rosenthal's canal (15.98 ± 1.33 mm), outer (40.81 ± 1.97) and inner (18.29 ± 1.47 mm) walls of scala tympani. They found that Rosenthal's canal did not appear to be linearly related to the length of the organ of corti, and ranged from 1.75 to 2 turns in all specimens. As a comparison, the number of cochlear turns range from 2.5 – 3 turns (with a majority having between 2.5 to 2.75 turns [73]). These results indicate that the frequency map of the spiral ganglion may well be different than that at the organ of corti.

Stakhovskaya and colleagues at the University of California at San Francisco studied cadaver cocheae with precise imaging techniques to devise the frequency map of the dendrites of the SG neurons as they enter the Rosenthal's canal [74]. They also formulated an empirical relationship between the tonotopic maps of the organ of corti and the spiral ganglion. Their data indicated length of spiral ganglion to be significantly shorter than that of organ of corti, with an average OC length equal to 33.13 mm and an average SG length equal to 13.69 mm in their specimen. They found strong correlation between the diameter of the basal coil and the length of organ of corti, and they suggested using the basal coil diameter to better predict the length of OC in preoperative imaging to estimate the insertion angles needed to cover a specific cochlear extent. An apparent example of clinical implication of these findings directly relates to the design of electrode arrays. The length of perimodiolar (modiolus hugging) electrodes, for example, would need to be shorter than that of the standard electrode arrays (which are positioned along the lateral wall of the scala tympani) to cover the same cochlear extent (frequency range). For implant recipients with residual hearing, a pre-calculated insertion angle based on the frequency map of spiral ganglion would result in a more accurate frequency range needed to be covered with electrical stimulation.

Figure 3.1 illustrates the mean characteristic frequencies along the cochlear spiral at the organ of corti and spiral ganglion measured by Kawano *et al.* [71] and Stakhovskaya *et al.* [74]. In comparison to the Greenwood frequency function, the model provided by Stakhovskaya *et al.* [74], seems to provide a relatively better approximation of the frequency map of the human cochlear spiral ganglion which is why it is gaining popularity with the cochlear implant research community. In the present work, we use the same model – the frequency map of the spiral ganglion to estimate the “place” of electrical stimulation.

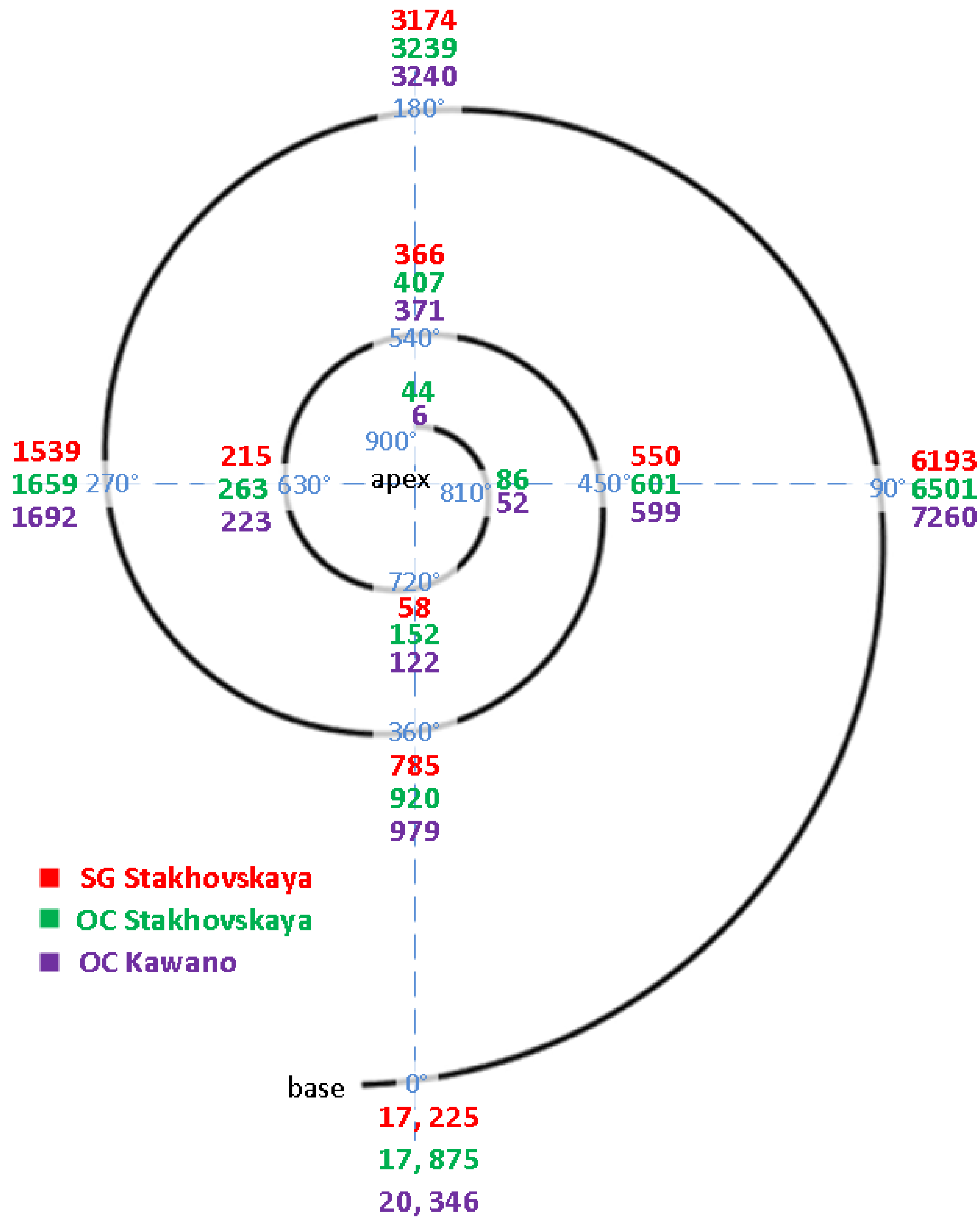


Figure 3.1. Mean frequencies along the organ of corti and spiral ganglion at different angles of rotation in cochlea. Numbers printed in purple (lower-most) represent mean tonotopic frequencies along the organ of corti measured by Kawano *et al.* [71]. Numbers printed in green (middle) and red (upper-most) represent the tonotopic frequencies at the organ of corti and the spiral ganglion level, respectively, computed by Stakhovskaya *et al.* [74]. Rotational angles are marked (over the spiral) in blue and represent angles from the base of cochlea.

3.2 Speech recognition as a function of altered spectral distribution

Speech recognition is a robust process that can be accomplished under conditions of severe distortions in the original signal, possibly because speech itself is a robust signal. Under ideal listening conditions, the central pattern recognition and linguistic access mechanisms have a rich and redundant set of peripheral cues. As listening conditions deteriorate, or under conditions of peripheral pathology, the central pattern recognition must work with a reduced or distorted set of cues from the periphery, which may introduce ambiguity or additional cognitive load for human speech recognition.

Mismatch between the frequency bands of the CI sound processor and the characteristic frequencies of the nerve fibers that are stimulated by the scala tympani implants can manifest into several kinds of spectral distortions in the perceived sound. As previously noted in Chapter 2, these include frequency-place shifting (frequency offset), frequency compression, frequency expansion, frequency warping, or a combination of the above. These distortions occur primarily due to the variability observed in the insertion depth that may result in stimulating groups of nerve fibers that are not tonotopically matched to the frequency bands assigned by the sound processor. In addition, dead regions in the neural population can create spectral holes. On the other hand, there is evidence that electric fields inside the cochlea may stimulate spiral ganglion located in the adjacent turns as well. Such a phenomenon may exacerbate the spectral warping challenge. Figure 3.2 - Figure 3.7 illustrate the above spectral distortions as they may appear in a cochlear implant system. Figure 3.2 demonstrates an ideal scenario, in which analysis filters of the sound processor are matched exactly with the tonotopic place of stimulation. Figure 3.2 shows frequency matching with shallow electrode array insertion. Figure 3.2 illustrates a linear shift in frequency-to-place mapping. The frequency compression and frequency expansion phenomena are shown in Figure 3.2 and Figure 3.2 respectively. Finally, the combination of above spectral distortions, as they may appear in an actual CI system is illustrated in Figure 3.2

Perfect frequency-place alignment

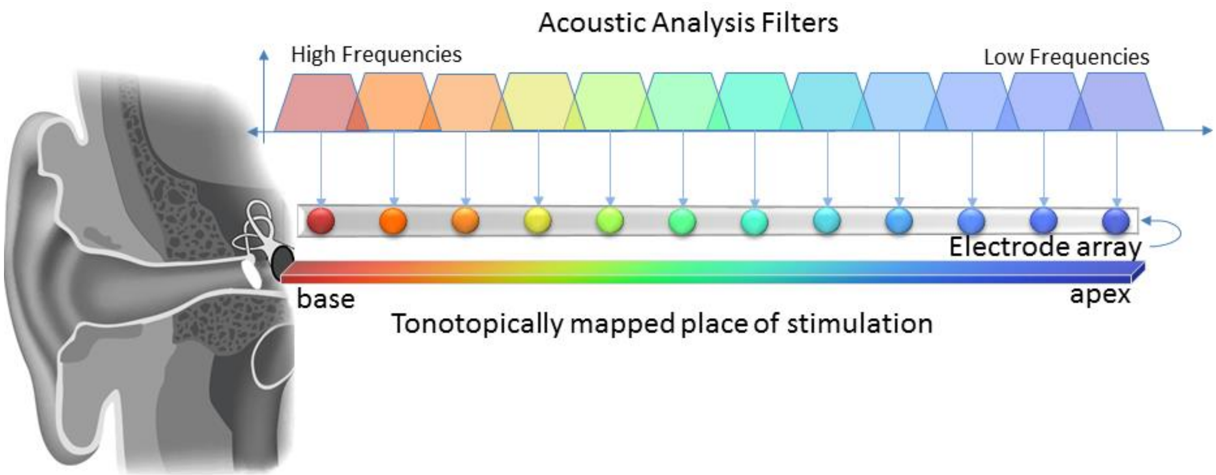


Figure 3.2. Perfect alignment (matching) of analysis filters with the tonotopic place covering full extent (frequency range) of cochlea. This and the following figures provide a conceptual representation of different frequency-place scenarios. Acoustic analysis filters represent the frequency bands of a sound processor. Cochlea is stretched open, and the tonotopic place of stimulation represents the characteristic frequencies of the auditory nerve fibers.

Shallow insertion with frequency-place alignment

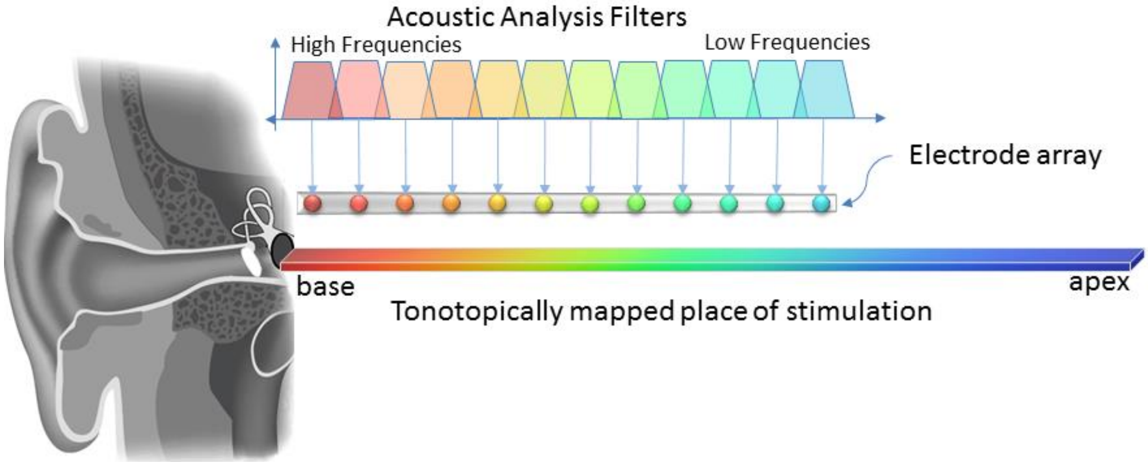


Figure 3.3. Shallow insertion in which analysis filters are matched with the tonotopic place. The lower-most frequencies are truncated.

Linear frequency shift

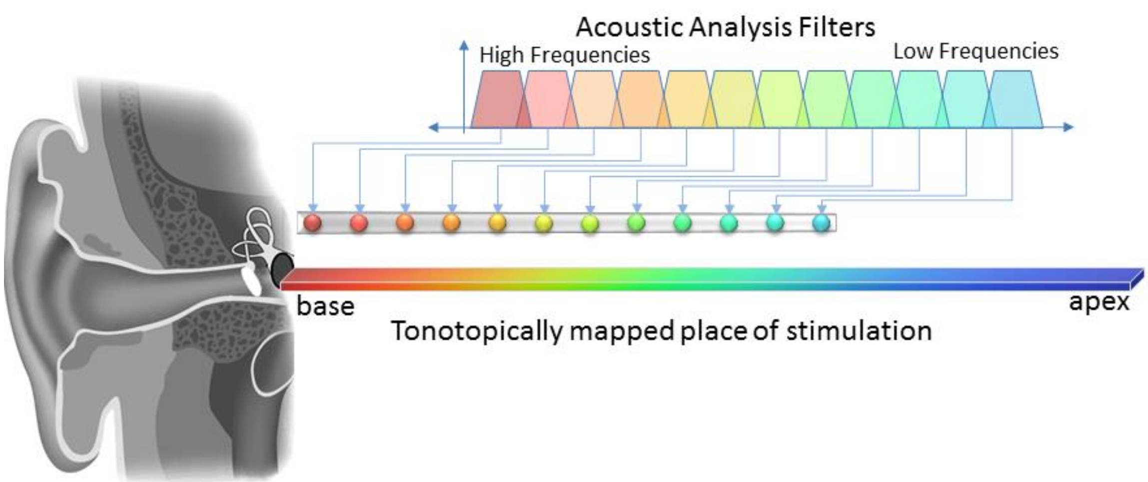


Figure 3.4. Linear shift/Frequency offset.

Frequency compression

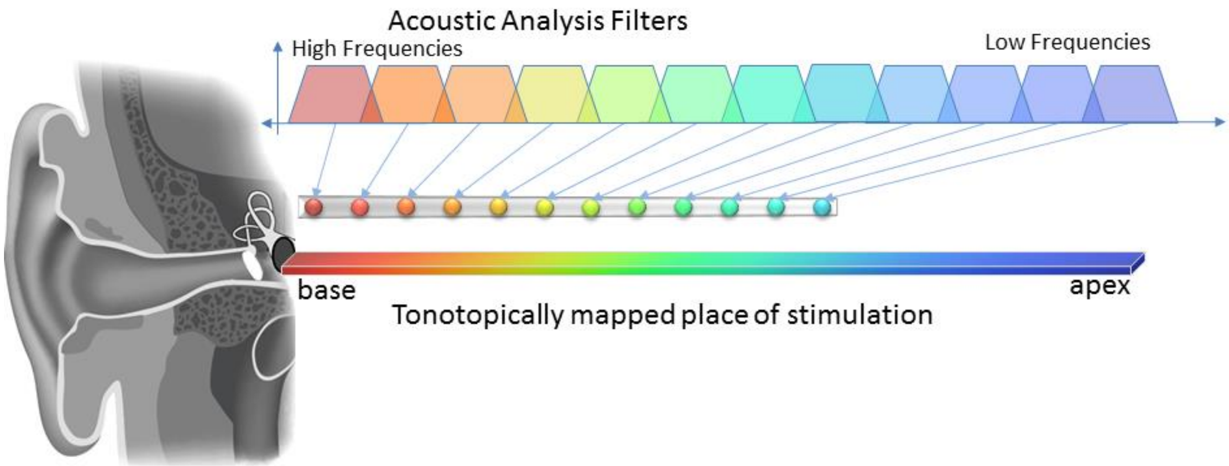


Figure 3.5. Frequency Compression: Full acoustic range is compressed and mapped to a relatively smaller cochlear place.

Frequency expansion

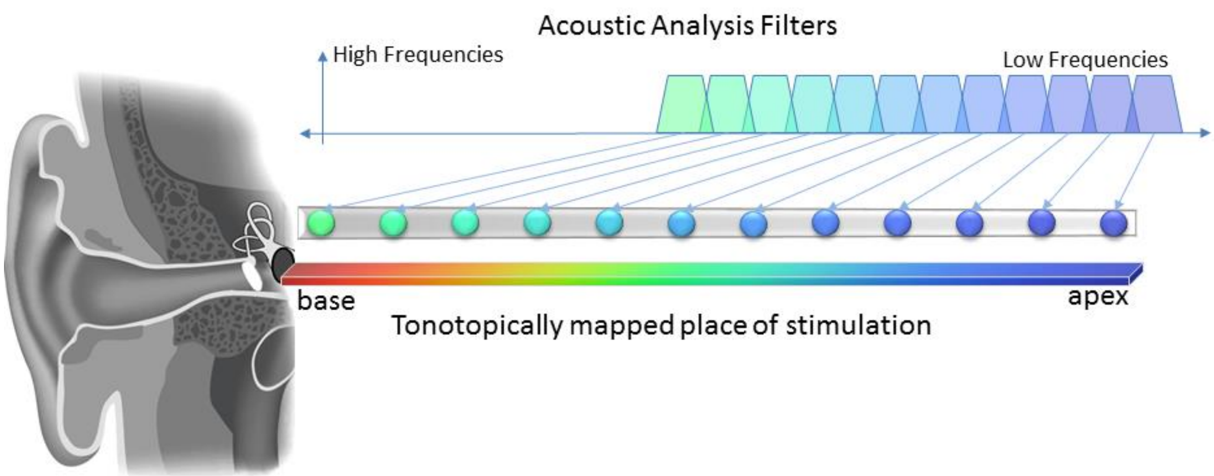


Figure 3.6. Frequency Expansion: Narrow acoustic range is spread out to a larger cochlear place.

Combination of frequency-place mismatch

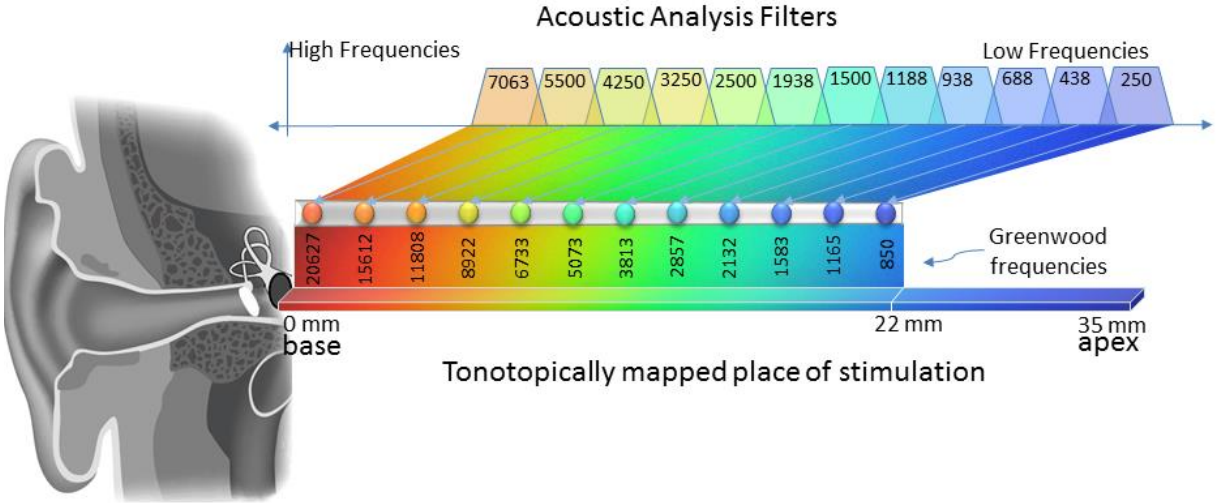


Figure 3.7. Illustration of an example scenario depicting the frequency place mismatch observed in an actual cochlear implant with insertion depth of 22 mm. The spectral distortions include frequency shift, frequency compression, and frequency warping. Frequencies are based on the Greenwood function.

Speech recognition as a function of altered spectral distribution has been widely studied in both CI recipients, as well as with acoustic simulations using normal hearing listeners [49-77]. The findings suggest that peak performance is achieved when the full acoustic range is mapped to the tonotopic map in a matched scenario (i.e., analysis bands correspond to the tonotopic map of the cochlea in a 1-to-1 manner, see Figure 3.2); however, minor mismatch does not account for the significant reduction in human listener performance. There are two opinions regarding frequency-place mismatch in cochlear implants. Some research groups are of the opinion that frequency mismatch could potentially account for differences in performance among the recipients, and that such mismatch can only be accommodated to a limited degree; therefore, these distortions should be minimized. On the other end of the spectrum, some groups argue that these mismatches can be compensated for over time and that neural plasticity of the cortex will eventually create a modified frequency map in the auditory cortex, therefore it is important to deliver full information via the sound processor and let the brain work its reorganizational magic. In the following sections, findings from the literature in this context will be discussed.

Some classical studies that investigated the effect of spectral and temporal distortions on the intelligibility of speech signal include those conducted by Tiffany and Bennet (1961) [75], Daniloff *et al.* (1968) [76], Nagafucci *et al.* (1976) [77], and Reed *et al.* (1983) [78], amongst others. Daniloff *et al.* [76], for example, studied various degrees of time and frequency compression on vowel recognition. They found that spectral degradation impacted the intelligibility of vowels more than the temporal distortions, and that the second formant frequency (F2) played a more critical role in vowel perception than the first formant frequency (F1). Also, relative or proportionate shifts of up to 40% in formant structure could be tolerated without significantly impacting phonemic quality; however, frequency compression ratio beyond fifty percent could severely impact vowel recognition. Nagafucci *et al.* [77] demonstrated that vowel recognition decreased sharply with frequency compression and expansion in comparison to temporal changes. The performance reduced to half for frequency distortions caused by 50% frequency compression or 200% frequency expansion (frequency compression had a relatively more deleterious effect than frequency expansion). While vowel discrimination was relatively unaffected by time expansion, performance decreased significantly beyond 50% time compression (a similar trend was demonstrated by Tiffany and Bennet [75]). Beasley *et al.* [79] demonstrated that with 15 days of training with 35% frequency-shifted/time-compressed speech, hearing impaired children accommodated to the altered pattern of speech cues and showed improvements relative to the baseline performance measured on the first day. Braida *et al.* (1979) [80], in a review of past research on linear amplification, amplitude compression, and frequency lowering schemes for hearing aids, concluded that frequency lowering, (for example, to match the hearing range of a hearing impaired individual) did not improve speech recognition, and often decreased performance compared to simple amplification schemes. Later, Reed *et al.* (1983) [78] systematically studied the effects of frequency lowering with linear and nonlinear spectral warping on consonant recognition ability by normal hearing listeners. Their results indicated that the best performance achieved on any of the frequency-lowering schemes was roughly equivalent to the performance achieved by low-pass filtering to an equivalent bandwidth. These results confirmed that frequency lowering did not improve consonant recognition and that linear frequency compression (compressing the whole frequency range) was more detrimental than frequency warping-compression (i.e., only higher frequencies are compressed and lowered).

The findings from these above-mentioned classical studies serve as a foundation for research on speech degradation and its impact on cochlear implants. Some of the earlier studies that investigated the effects of spectral distortions in cochlear implants were conducted by Drs. Dorman, Loizou, Shannon, Zeng, and Fu in the late 90s [81-83]. In one of these preliminary studies, Dorman *et al.* studied the effect of cochlear implant insertion depth with normal hearing (NH) individuals using five channel acoustic simulations of cochlear implants [81]. They found that simulated insertion depth had a significant effect on the identification accuracy of consonants, vowels, and sentences presented in quite (without learning). Vowels, in particular, were more susceptible to frequency shifts. Although performance was significantly poorer at 22-

and 23-mm simulated insertion depths, performance at 25-mm simulated insertion depth was, comparable to normal. Their results indicated that insertion depth of 25mm should be sufficient to allow for an effective level of speech understanding which may be achieved with deeper insertions. (see Figure 3.8)

Shannon *et al.* [83] reported similar findings with vocoder-based simulations in a series of experiments using a 4-channel vocoder that preserved temporal cues and systematically distorted spectral cues by shifting and non-linear warping. By manipulating the frequency extent of the individual analysis and synthesis bands to create a warping in the spectral distribution of

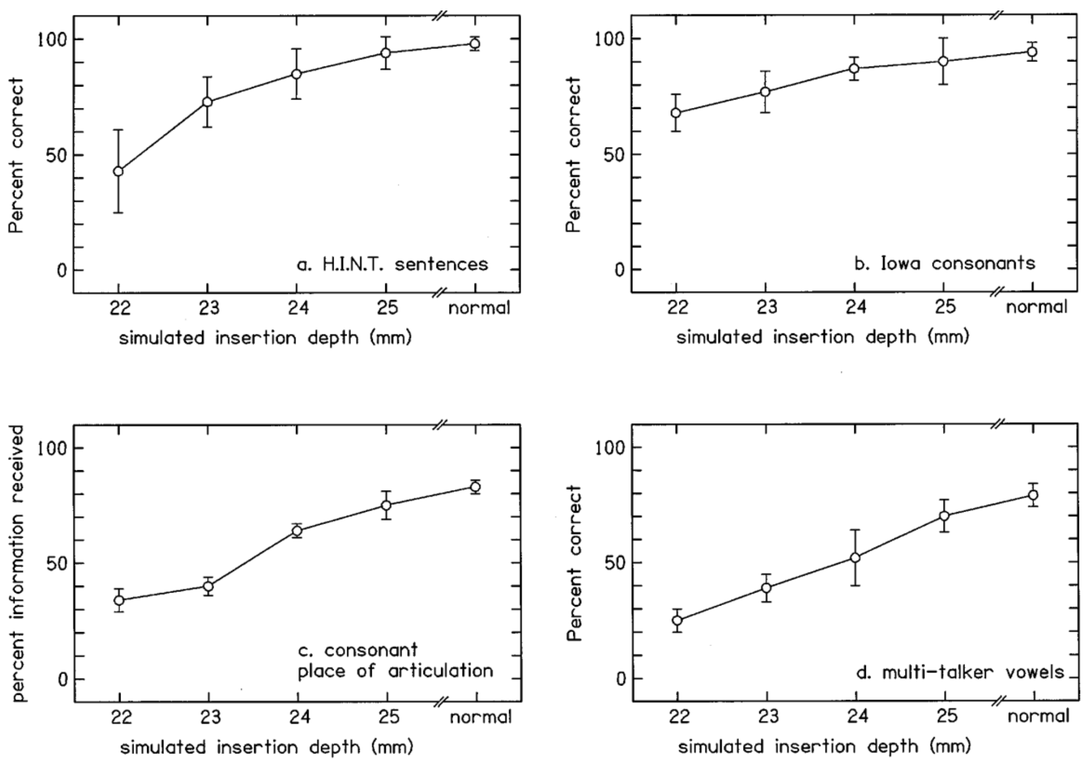


Figure 3.8. Average test scores as a function of simulated insertion depth from Dorman *et al.* [81]. Reprinted with permission from [81]. Copyright 1997, Acoustical Society of America.

envelope cues, they demonstrated a dramatic decrease in performance to the point that speech was completely unintelligible. They also observed large deficits in speech recognition performance when the synthesis bands were linearly shifted in frequency, mimicking the tonotopic shifts that may occur due to the basal position of electrodes in a cochlear implant. In line with previous studies, their results demonstrated that vowel recognition and consonant place of articulation required more spectral detail for high levels of recognition, indicating that vowel recognition[1] depended more on spectral cues than on temporal cues. Their results indicated that the central pattern recognition mechanisms are not robust to linear translations of the pattern along the neural array, at least acutely.

In another study conducted at the House Ear Institute with normal hearing and cochlear implant listeners, Fu and Shannon [82] assessed the recognition of spectrally degraded and frequency-shifted vowels. Spectral degradation was achieved by simulating 4-band, 8-band, and 16-band vocoders, and frequency shifting was achieved by systematically varying the frequency extent of the analysis and synthesis filters. Their results showed that vowel recognition was sensitive to both spectral resolution and frequency mismatch and more interestingly, both factors were orthogonal to each other in terms of intelligibility (i.e., each had a separate and independent effect on speech recognition). One consequence of this result is that greater spectral resolution (e.g., by increasing the number of independently perceptual bands/electrodes) would not compensate for or provide additional robustness against the deleterious effect of tonotopic shift and mismatch. In their acute study (i.e., without any learning), the best performance was observed with frequency matched maps, but a relative shift of up to 3 mm (basal or apical) could

[1] Vowels contribute little to overall speech intelligibility.

be tolerated without any significant loss in performance. However, performance dropped significantly beyond a 3 mm shift in either direction. A 3 mm basal shift corresponds to the insertion depth of 25 mm from the base (or lowest corner frequency of 513 Hz) based on the Greenwood function (assuming 28 mm as full insertion from the round window).

Başkent and Shannon studied the frequency place mismatch and their effects on speech recognition performance of cochlear implant recipients in a series of experiments conducted at University of Southern California and House Ear Institute [84-87]. In [84], they explored the effects of frequency compression and frequency expansion on speech intelligibility in a vocoder-based simulation study with 6 normal hearing listeners. They simulated two commonly observed electrode insertion depths, 25 mm and 20 mm from the round window. The frequency ranges of the carrier filters were kept constant and those of the analysis filters were varied to simulate the frequency-place mismatch condition, which varied from -5 mm to +5 mm relative shifts (extreme expansion to extreme compression). All testing was done acutely (i.e., without any training). The best speech recognition (for consonants, vowels, and TIMIT sentences) was always observed in the matched condition and both frequency compressed or expanded maps produced poorer performance. Generally, frequency expanded maps produced poorer performance than frequency compressed maps, indicating that spreading of the critical speech spectral region to a larger range in the cochlea (like an acoustic fovea) has detrimental effects on human speech recognition performance. The frequency compression phenomenon is similar to the typical mapping approach used in cochlear implant processors, in which an acoustic frequency range of 150 – 10,000 Hz is generally mapped on to tonotopic locations that range from 500 – 6000 Hz (for 5 mm basal shift with16 mm electrode array length, resulting in a compression of approximately 2 octaves (based on the Greenwood function)). This simulated condition resulted in at least 20 percent deficit in sentence recognition as compared to the matched condition. The data from their acute experiments indicated that without learning, speech patterns in the central nervous system can tolerate only a relatively small amount of distortion (2 – 3 mm) in tonotopic space. Beyond a 2 - 3mm shift, speech understanding was severely impacted. This study was later repeated with the 6 Med-El Combi 40+[2] cochlear implant recipients in [85]. Despite the uncertainties in true frequencies of nerve fibers at the stimulation sites, there results showed a similar trend to the above-mentioned simulations study conducted with normal hearing listeners [84]. The best performance was obtained with maps that were matched and had the least amount of spectral distortion. Both frequency compression and frequency expansion schemes reduced performance. This effect was more pronounced for vowels, which are more susceptible to spectral distortions. Again, the study measured only the acute effects of using different frequency maps and the results may not hold equally valid for any accommodation that may occur after long term use of the maps. However, their data indicated that the choice of the frequency-place map had a significant effect on speech recognition, and if a recipient was initially fitted with an optimum map that had a minimum overall frequency-place distortion, any further accommodation could start from this highest level of performance.

In a later study [86], Başkent and Shannon investigated the effect of electrode insertion depth on speech recognition abilities of 4 Med-El Combi 40+ cochlear implant users. They simulated 10 different insertion depths that varied from shallow insertion of 7.2 mm (with single

---

[2] Combi 40+ is a relatively long electrode array from Med-El that may achieve insertion depth of up to 31 mm.

electrode) to deeper insertion of 28.8 mm (10 active electrodes) by systematically switching off the apical-most electrodes within the same subject. For each simulated insertion depth, a frequency-matched map and a frequency-compressed map was tested for speech recognition of vowels, consonants, and sentences. Their results indicated that for insertions deeper than 20 mm from the round window, the frequency matched map produced better results than the clinically assigned frequency compressed map. Here, 20 mm corresponds to 1168 Hz on the Greenwood scale. This implies that providing more spectral information (by performing compression) was actually detrimental for speech understanding in these subjects. For insertion depths less than 20 mm, they found that a mild amount of frequency-place compression was better than truncating the lower frequencies. For example, for insertion depths of 19.2 mm and 16.8 mm, the peak performance was observed with compressions of +2 to +3 mm and +3 to +4.5 mm respectively (see Figure 3.9). These compressions equate to cut-off frequencies of 643 Hz and 715 Hz (center frequencies = 821 Hz and 948 Hz) for the lowest most frequency band respectively. However, that study only considered the acute effects of frequency-place maps and completely ignored the role of long-term adaptation. While CI recipients may be able to accommodate to distorted pattern of tonotopic activity, they probably cannot overcome the loss of information caused by truncating the frequency range. It is for these reasons that despite individual variations in electrode insertion depths, clinical implant systems continue to provide the full acoustic range through their sound processors. However, the data from these studies indicate that severe frequency-place mismatch (due to extremely compressed frequency maps) can be a disadvantage. It should be noted that the exact amount of time for full adaptation to various degrees of frequency-place mismatch in most subjects is not known, and further research is

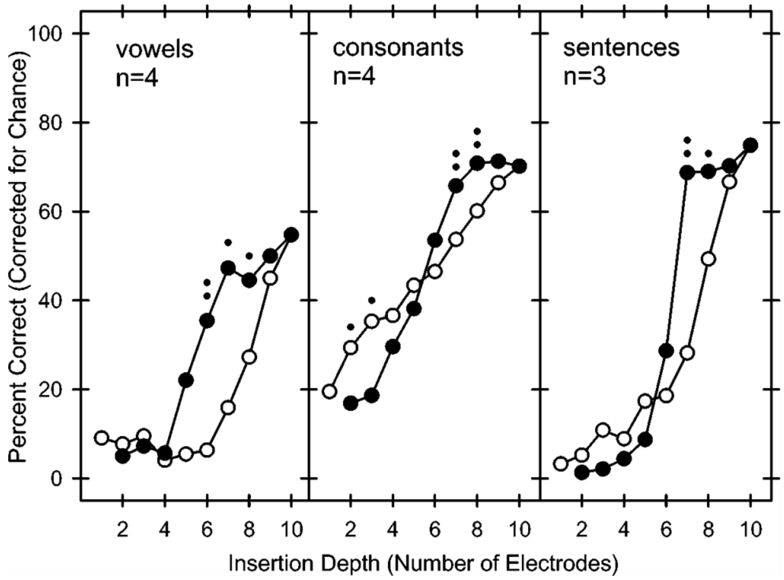


Figure 3.9. Average test scores as a function of simulated insertion depth in Combi 40+ implant users from Başkent and Shannon [86]. The open circles indicate a frequency compressed map with full acoustic range. Filled circles show the scores from the matched map (lower frequencies were truncated). Reprinted with permission from [86]. Copyright 2005, Acoustical Society of America.

needed in this regard.

The above-mentioned studies investigated the effects of frequency shifting, compression, and expansion separately; however, in cochlear implants these distortions may occur in combination. In their later work, Başkent and Shannon [87] explored the combined acute effects of frequency shift and compression-expansion on speech recognition using acoustic simulations of cochlear implants with five normal-hearing subjects. Similar to the findings of previous studies, they found that the matched map generally produced the best results, and applying either of the three spectral distortions independently resulted in reduced performance. However, they noticed a compensatory effect for some conditions when the two degradations were applied simultaneously. For all apical shifts of -5 mm (and some conditions of -3 mm), the shifted and compressed map produced better performance than the frequency-shifted map alone. Also, the performance was higher for maps, where low frequencies were better matched as compared to the maps where higher frequencies were matched. Their analysis showed that this compensatory effect was more commonly observed for maps where the combination of the shift and compression resulted in alignment of low-frequencies, particularly within the 1000 – 2000 Hz range, even when the remaining frequencies did not necessarily match. These findings are consistent with studies on speech perception. The Speech Intelligibility Index (SII) which predicts speech recognition performance from the amount of information available, weights frequency information form 1 – 3 kHz as the most critical [88]. Başkent and Shannon concluded that for shallow insertions (19 mm or less), a mild amount of compression was better than matching the complete frequency range to avoid over-truncation, and that matching the frequency range between 1000 – 2000 Hz, which contains the most critical speech information, can be more beneficial for speech recognition than matching the entire frequency range in certain conditions. In the next chapter of this dissertation, we will explore how we utilize these experimental data to devise user-customized frequency assignment strategies.

In contrast to the aforementioned studies that varied the lower frequency boundaries to create frequency mismatch conditions, Goupell *et al.* [89] investigated the effects of upper-frequency boundary and warping on (German) sentence recognition of seven Med-El CI users and the effects of adaptation with six normal hearing individuals. They tested eighteen different variations of processing schemes that varied in frequency- place mapping, stimulation range, and upper boundary of the frequency. They found that spectrally unwarped conditions often produced the best performance, but variations of up to 0.77 octave for the most basal electrode could be tolerated without impacting human listener performance. For the unwarped conditions, only eight to ten frequency matched channels were sufficient for equivalent performance to the normal 12-electrode sound processing. In their second experiment, they found that frequency matched and frequency expanded maps showed improvement with feedback training in acoustic simulations of the cochlear implant with NH listeners; however, no improvement was observed for the frequency compressed maps.

3.2.1 The role of adaptation

To this point, we have explored the acute effects of spectral distortion on speech recognition. Contrary to the aforementioned studies, some research groups argue that results from acute studies underestimate the effect and/or benefit of learning/adaptation, and that neural plasticity can help facilitate and thereby create an acceptable “adapted electric map” over time, (i.e. the listener gradually learns to live with the current configuration and adapts to the altered pattern of stimulation)

Rosen *et al.* investigated the adaptation to upward spectral shifts of speech in a simulation study with four normal hearing listeners [90]. They simulated a 6.5 mm basalward shift (1.3 – 2.9 octaves, depending on the frequency) using a 4 band noise vocoder and tested consonant, vowel, and sentence recognition acutely and with connected-discourse tracking (CDT) [91]. Consistent with other studies, they found that without any auditory training, matched conditions led to the best performance, and the spectrally shifted condition resulted in negligible speech recognition. However, after about 3 hours of experience (nine 20-minute CDT sessions), performance increased from near zero to about one-half the performance in the matched conditions (30% of words). Their results suggested that listeners could learn to adapt to spectral shifts that arise due to frequency-place mismatch in cochlear implants (e.g., due to shallow insertions). Although, the final level of performance with spectrally shifted maps was significantly lower than the matched condition even after the extremely short training period, it was clear that acute-experiments with acoustic simulations of cochlear implants seriously underestimate the effect of learning. This study concluded with many unanswered questions, (e.g., what degree of spectral mismatch could be accommodated to? Duration of time required for adaptation? Is full accommodation to spectrally degraded maps even possible?) Rosen *et al.* [90] suggested that speech perception difficulties which implant users experience as a result of frequency-place mismatch may be a short-term limitation that could be overcome with experience.

Harnsberger *et al.* [92] demonstrated adaptation to an upward spectral shift in experienced cochlear implant listeners. In this study, the implant users were asked to indicate their perception of synthetic vowels generated by varying F1 and F2 characteristics, in an attempt to create "perceptual vowel" spaces in hopes to find their true frequency space. Their idea was that if recipients had not adapted to upward spectral shifts, the perceived vowels will most likely have lower first and second formants than natural vowels. However, there was no evidence of such effects, instead, the vowel spaces differed from one another in the relative size of their vowel categories. Their findings suggest that perpetual (frequency) map may be altered with experience and that differences in format frequency discrimination may account of individual differences in vowel perception in cochlear implant users.

Fu *et al.* [93] investigated the perceptual learning patterns following changes in the frequency-to-electrode assignments with 3 Nucleus CI users. In their study, the participants used an experimental map that was shifted apically by 2 – 4 mm (1 – 0.68 octave) relative to their clinical (normal) map over a period of three months. Their results showed that although performance with the frequency shifted maps dropped drastically immediately upon activation, their performance continued to improve over the course of three months, indicating adaptation to the altered speech patterns. However, none of the subjects could exhibit full accommodation in the three month time frame. Consonant and HINT sentence recognition scores gradually approached a performance level that was comparable to, but still below, baseline performance (with the clinical map). On the other hand, vowel and TIMIT sentence recognition remained significantly lower even at the end of the study. This indicates that long-time exposure may not compensate for the deficit in performance caused by a 2 – 4 mm shift in tonotopic location of stimulation, at least not within a 3 month time period. Interestingly, the clinical map resulted in the same performance both at the start and the end of the study, indicating that the central representation of frequency space (spectral patterns) had not been reshaped after 3-months of experience with the frequency shifted map (i.e., the users retained both frequency representations). This result is quite surprising, as it indicates that at some level multiple maps may co-exist, perhaps with differing saliencies. It should be noted that Fu *et al.* considered an apical shift and had only a vague knowledge of electrode array insertion depth (from the rings on the array that resided outside the cochlea, as indicated by the surgical report). The exact tonotopic-place map (that was being stimulated) was unknown. If most apical nerve fibers (representative of low frequencies) were not being stimulated (which is quite likely for normal insertions), a basal-ward frequency shift would be observed. In such a case, creating an apical shift in the map would further increase the frequency-place mismatch, which might reduce overall performance (and adaptation trends as well).

Faulkner *et al.* [94] simulated tonotopically mapped frequency-maps that varied in insertion depth in a group of eight normal hearing individuals using noise-excited vocoders. They simulated five insertion depths ranging from 25 mm to 17 mm, a range that is commonly observed in recipients depicting normal to relatively shallow insertion. In all five simulated conditions, analysis filters were matched to the synthesis filters to create matched frequency-place conditions. The center-frequency of the analysis filter ranged from 502 Hz (25 mm) to 1851 Hz (17 mm). All participants were given 20 minutes of CDT training with each frequency map prior to testing. The results were as follows. Speech recognition was found to be proportional to the electrode array insertion depth. Identification of consonants and sentences remained at an asymptotic level for insertion depths of 25 mm, 23 mm and 21 mm, below which (21 mm or less), the performance dropped significantly. Vowel recognition showed a progressive decline with a decrease in insertion depth. Mean results from their study are shown in Figure 3.10. The trend in results seems to follow predictions from the Articulation Index (AI) weightings [88]. AI studies show that the loss of information below 1 kHz will significantly reduce intelligibility of unprocessed speech, and that additional higher frequency information gained from matching will be of slight benefit. Overall, their results indicate that if frequency assignments of sound processors were to be matched with the tonotopic place of stimulation, users with electrode insertion depths of 19 mm or lower will obtain significantly poorer speech recognition performance.

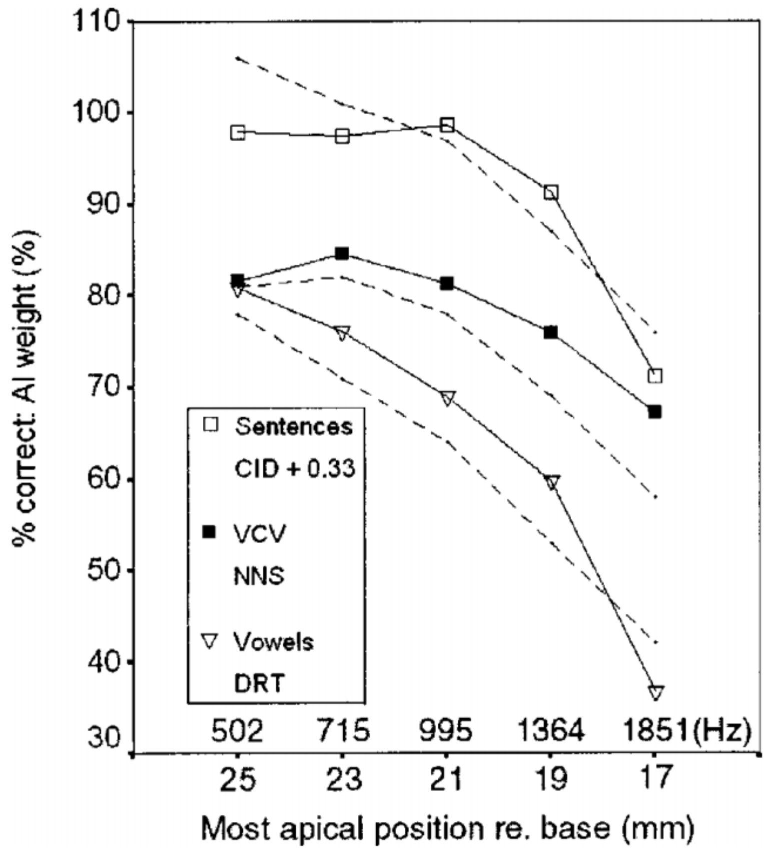


Figure 3.10. Summary of test scores as a function of simulated insertion depth over speech materials with normal hearing listeners from Faulkner *et al*.[94]. Three types of speech materials were tested, namely BKB sentences, consonants in vowel-consonant-vowel (VCV) context, and vowels. The dotted lines represent predictions from the AI weightings for comparable material. Reprinted with permission from [94]. Copyright 2003, Acoustical Society of America.

Faulkner, Rosen, and colleagues, in their follow-up research [95], compared frequency matched maps with upward-shifted frequency maps to simulate shallow insertion depth of 16.9 mm (insertion depth of 16.9 mm from the round window corresponds to the most apical electrode residing at 1851 Hz on tonotopic place according to the Greenwood function). The study was conducted using vocoder-based simulations with eight normal hearing individuals who were tested with frequency matched and frequency-shifted maps before and after a three hour training session with each map. Their results showed that training improved performance with both maps. Speech recognition of male talker was found to be better with the frequency shifted map than with the frequency matched map. On the contrary, for the female talker sentence recognition was found to be better with the matched map, and vowel recognition had an equal level of performance for both maps. In general, higher adaptation to frequency-shifted maps was observed. They concluded that full acoustic information may matter more than frequency-place alignment for shallow insertion depths (at least at 16.9 mm simulated condition) in the long run, assuming that individuals will continue to adapt to the frequency shifted maps. The results from this study, however, should be interpreted with caution. First, the authors simulated insertion depth of 16.9 mm from the round window and tested a frequency shifted map that would mimic a 6 mm basalward shift. The center frequency of the most apical filter for 6 mm equates to 715 Hz. All frequencies below the cutoff (601 Hz) were truncated. Assuming a full insertion of 28 mm from the round window, the most apical electrode for the frequency matched map is located (28-16.9) ~11 mm from the apex (i.e., 11 mm shift from the optimum place). On the other hand, the frequency shifted map has a (28 – 23) 5 mm relative shift for the most apical band from the optimum place. Generally, a CI sound processor will provide a full acoustic range to the electrodes regardless of the insertion depth. For such a case, a clinical processor will allocate 250 Hz to the most apical electrode (Cochlear Corp. frequency table) that would result in a relative shift of 11 mm. The frequency allocation scheme used by Faulkner *et al* in [94] underestimates the frequency mismatch that would be caused by an actual CI processor and therefore these results may not be applied to patients who have shallow insertions. Furthermore, the results from the study are reported independently for the male and female talker.

Table 3.1. Center frequencies (CFs) of analysis filters used in the study by Faulkner *et al.* in [95].

CFs of freq-matched filters (Hz)	1851	2492	3338	4453	5923	7861	10416	13783
CFs of freq-shifted filters (Hz)	715	995	1364	1851	2492	3338	4453	5923

In a relatively recent work, Faulkner and colleagues assessed the effect of unilateral spectral shift in binaural hearing in a study lead by Siciliano [96]. Their hypothesis was that given sufficient training, listeners would receive a binaural advantage for speech information that had a unilateral shift but was matched contralaterally. The speech recognition performance with the binaurally mismatched processor was always significantly lower than the processor that had just three unshifted bands. Despite 10 hours of training, the subjects did not show a binaural advantage for the binaurally mismatched frequency-place map. This resistance to learning, they concluded, could indicate a constraint on speech perceptual neural plasticity to instances where the relative frequency order is preserved. These findings raise thought-provoking clinical implications for bilateral cochlear implants. Their data suggest that it may be important to minimize frequency place mismatches not only in the same ear, but also minimize mismatches across both ears. Optimization of bilateral cochlear implants, therefore, may be achieved by minimizing distortions across ears. This is not surprising as projections of the auditory pathways from each ear are bilateral. The pathways from both ears cross paths at the cochlear nuclei, superior olives, and inferior colliculus before reaching the primary cortex. See Figure 3.11. Since, signals from each ear are combined so early in the process, and sent to both sides of the brain, the complexity of these pathways makes frequency-place mismatch across the ears intricate and difficult to study.

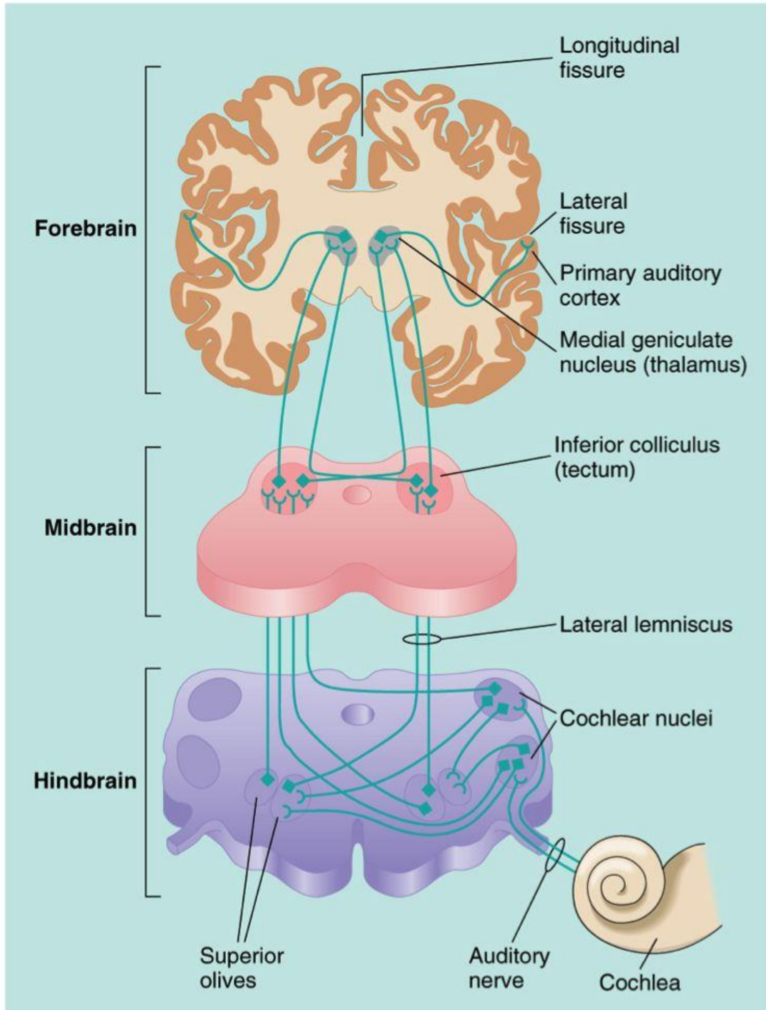
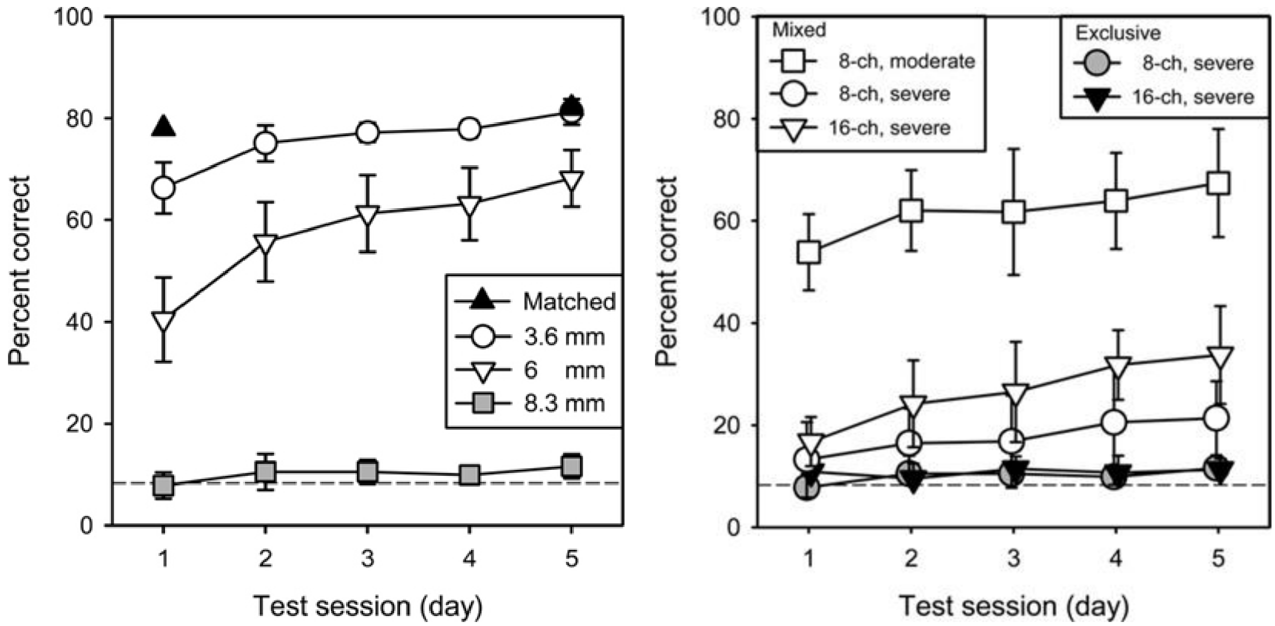


Figure 3.11. Projections of the auditory pathways from ear to the cortex. Reprinted with permission from Pinel, J. P. J., Biopsychology, 9/E [97]. Copyright 2014, Pearson.

Li, Galvin and Fu [98] studied the interactions between unsupervised learning and the degree of spectral mismatch on short term perceptual adaptation in normal hearing listeners to spectrally shifted vowels with acoustic simulations of cochlear implants. In their first experiment, they simulated three insertion depths with shifts of 3.6 mm, 6 mm, and 8.3 mm towards the base, indicating slight, moderate, and severe mismatch respectively. The center frequencies of the analysis filters and synthesis filters are given in Table 3.2 (synthesis filters reflect tonotopic frequencies at the cochlea, Greenwood assumed). Listeners were repeatedly tested for vowel recognition over a 5-day study period without any explicit training or feedback. Subjects showed complete adaptation to the 3.6 mm shift, achieving scores equal to the matched condition and at least partial adaptation to 6 mm shift. No improvement was seen for the severely shifted condition, indicating 8.3 mm spectral shift could not be accommodated for over a period of 5 days. In the second experiment, two groups (4 participants per group) were tested with a severe shift (8.3 mm) over a period of five days. The first group was tested exclusively with the 8.3 mm shifted map, while the second group followed a mixed exposure protocol that involved alternating exposure to the moderate and severe shifts with 8 and 16 channels of processing. By the end of the 5th day, they found no adaptation to the severe mismatched map in the first group which was given exclusive exposure to the 8.3 mm severe shift. On the contrary, the second group (which was given mixed exposure), showed significant adaptation to the severely shifted speech with 8-channels and even greater adaptation with 16-channels. The asymptotic level of performance with severe mismatch was still significantly lower than that with moderately shifted

Table 3.2. Center frequencies (CFs) of synthesis filters for different simulated insertion depths used in the study by Li *et al.* in [98].

Matched filters – 0 mm shift (Hz)	272	464	744	1155	1754	2629	3905	5769
Slight shift – 3 mm shift (Hz)	544	763	1052	1433	1936	2598	3471	4622
Moderate shift – 6 mm shift (Hz)	805	1108	1507	2033	2726	3640	4845	6433
Severe shift – 8 mm shift (Hz)	1168	1584	2136	2863	3821	5084	6748	8942



(a) (b)

Figure 3.12. Mean vowel recognition scores as a function of test session reported by Li *et al.* [98] (a) Experiment 1: Different simulated insertion depths. (b) Experiment 2: Severe spectral shift (8.3 mm) with different training protocols. Error bars indicate one standard deviation. Reprinted with permission from [98]. Copyright 2009, Ear and Hearing.

maps, indicating only a partial adaptation. The results from both experiments are shown in Figure 3.12 and indicate that listeners are able to adapt completely to shifts of up to 3 mm and exhibit at least partial adaptation to spectral shifts equivalent to 6 mm with passive learning. For shifts beyond 6 mm, some passive adaptation could be observed given gradual or mixed exposures to a smaller spectral shift. These findings are in line with studies conducted by Fu and Galvin [99], and Svirisky *et al.* [100] that suggest gradual or mixed exposure to shifted maps may be more beneficial that an abrupt exposure to a severely shifted map. Although the findings from this study are quite interesting, it is unclear if these adaptation trends would equally hold for other speech materials, listening conditions, and for actual cochlear implant users. For example, Faulkner *et al.* demonstrated that results in the patterns of phonetic training did not necessarily generalize to speech recognition. Also, Parikh and Loizou [101] showed that speech perception in noisy listening conditions may well be quite different than that in quiet. While F2 cues are more important in vowel recognition in quiet, F1 cues together with the F2 envelope play a more dominant role in noise. Furthermore, the number of perceptually discriminable channels in the cochlear implant recipients is generally very limited (at most about 8). As discussed in the previous chapter, there is evidence from many studies that vowel recognition plateaus at 4 - 6 frequency bands [13-15]. Therefore, the availability of 16 perceptually independent stimulation sites for vowel recognition may well be beyond the capacity of current cochlear implants. Although Fu *et al.* [82] in their previous study demonstrated that the number of channels and frequency shifts are completely independent processes, the results from this study indicate that this might not hold equally true for accommodation to 16-channels. Further research on adaptation to severely distorted speech cues with different speech materials with varying complexities and noise conditions is warranted to make any reasonable arguments on "real-world" implications with cochlear implant users.

Studies conducted by Reiss *et al.* provide evidence to changes in the perceived pitch with long-term use of hybrid cochlear implants [102-104]. In [102], they measured electric pitch sensations in hybrid cochlear implant users at various stages of implant use to examine the effect of experience to perceived pitch. They found that electric pitch perception often shifted in frequency, sometimes by as much as two octaves, during the first few years of implant use. They observed that this change in perception could be attributed to two adaptation trends, namely, short-term changes, and slow systematic changes. Short-term or early pitch sensations may more closely reflect peripheral innervation patterns, while later pitch sensations may be due to higher-level, experience dependent changes. In a study conducted on 20 hybrid CI users over a period of 24 months, they found that pitch perception at individual electrode locations more closely resembled the frequency map assigned by the processor than place-frequency predicted from the cochlear place [103]. A very interesting case study about electric pitch perception of a bilateral cochlear implant user who used a standard electrode (24 mm) in one ear and a short (hybrid) electrode array (10 mm) in the contralateral ear was reported by Reiss *et al.* in [104]. The standard electrode array had 20 active electrodes, while the short array had only 6 electrodes. Despite the significant variation in array types and putative insertion depths, the recipient preferred the full acoustic range (188 – 7938 Hz) in both ears and thus both processors were programmed to allocate same (full) frequency range. After 2 years of continuous use, the researchers used psychophysical methods to pitch match the electrode pairs in both ears and surprisingly found that pitch-matched electrode pairs between the CIs were aligned closer to the processor-provided frequencies rather than the actual cochlear place. This result suggests that pitch perception may have adapted over 2 years to reduce the perceived spectral discrepancies between the two ears, despite the 2 – 3 octave difference in tonotopic mapping. The evidence from these above-mentioned studies suggests that the brain may adapt to spectral mismatches by remapping pitch over long-term use of cochlear implants.

3.3 Summary and Conclusion

The findings from the afore mentioned studies suggest that spectral distortions that occur due to frequency-place mismatch in cochlear implants may have a detrimental effect on speech reception depending upon the degree of mismatch and distortion. Shifts of up to 3 mm can be tolerated without having a significant impact on speech perception. Although, mismatches from

3 – 6 mm show worse performance acutely, listeners may show at least some partial adaptation over time. Continued exposure and auditory training may lead to better results. The results from most studies indicate that shifts beyond 7 mm may not be accommodated to, at least within a short span of time. The degree of accommodation and time to adapt to severe mismatch may be variable across listeners, and it is not yet clear as to what extent neural plasticity can continue to facilitate reception of severely distorted speech. Also, there is no evidence to explain if adaptation trends observed with moderate to severe frequency mismatch will extrapolate to complex listening situations, such as music perception, speech recognition in the presence of noise, and reverberation.

The results from these studies at least hold some implications for cochlear implant fitting. For minor mismatches (up to 25 mm insertion depth) compressing the full acoustic range to the electrode array will generally result in equal performance to the matched conditions over time. For moderate to severe mismatches, clinical frequency maps (that compress the full acoustic range) may be detrimental for speech reception at least acutely. For such cases, a balanced approach that compresses the lowest acoustic frequencies and matches intermediate frequencies, may result in a better performance, at least acutely. Listeners may be able to adapt to moderate to severe mismatches, at least partially; however, intermediate balanced maps during the accommodation period may facilitate accommodation [100]. Also, it is fair to acknowledge that the neural plasticity may be unique for each recipient, and therefore, the final level of adaptation and the time to adapt to the frequency-mismatch may vary from individual to individual. A frequency map that minimizes the mismatch, but at the same time is able to represent sound frequencies without a significant loss of speech cues that are important for understanding, may lead to quicker adaptation or improve outcomes or both.

CHAPTER 4

USER-SPECIFIC FREQUENCY-PLACE MAPPING

In the previous chapter, effects of frequency-place mapping on speech recognition were explored. Scientific studies conducted with normal hearing and cochlear implant users suggest that if the spatial relationship between electrodes and the auditory nerve fibers (that they stimulate) could be known, a more effective frequency-place mapping could be devised. The goal of these mapping schemes would be to maximize delivery of the full acoustic information, but at the same time keep frequency-place mismatch to a minimum. A mapping scheme that could also better mimic the normal peripheral frequency map would be desirable. Unfortunately, these two goals tend to be incompatible with each other for electric hearing via current cochlear implants. The exact placement and insertion depth of scala tympani electrode arrays relative to the tonotopic place in the cochlea is generally unknown. For moderate to shallow insertions, for example, matching the sound processor’s analysis frequency bands with the tonotopic place of stimulation, may result in a considerable loss of the low-frequency information (that could be vital for speech recognition). Similarly, extremely compressive maps may be less than optimum and reduce perception. There is evidence that a compromise between the two may result in an overall optimum configuration (i.e., a map that is not drastically different from the frequency map of the auditory cortex). Unfortunately, until very recently, positioning of the electrodes and their relationship with the spiral ganglion cells could not be known. Although radiograms (e.g., x-rays) can capture the electrode array image in the cochlea, they lack the ability to visualize the fine nerve fibers. Similarly computer tomography (CT) imaging could be used, but nerve fibers here are too fine to be visible on CT images and the lack of contrast for the SG cells makes it extremely hard to delineate directly.

Noble *et al.* at Vanderbilt University have devised a new image-processing technique based on weighted active shape modeling which determines the location of electrode contacts and their spatial relationship to the peripheral neural processes, from both pre and post implantation CT scans of recipients' cochlea [105-107]. In the present work, we have collaborated with their group to utilize image-guided procedures to determine electrode-nerve spatial-relationship in each CI recipient's cochlea. We use these image maps to customize frequency-place functions on an individual basis. In the following section, details about the image-processing technique are provided.

4.1 Image processing technique

External cochlear boundaries are registered relatively clearer in CT images and are easier to segment, but due to a lack of contrast of the SG, delineation of the fine nerve fibers is not possible by human vision or conventional image processing segmentation algorithms. Noble *et al.* [105] overcame this problem by first identifying external cochlear features that are sharply registered in CT images and can be segmented easily (i.e., identifying landmarks which are reliable). Subsequently, a binary point distribution model (PDM) was created that represents the visibility of cochlear points (landmarks) on the CT image. PDM was used to create a reference template or statistical-shaped model (SSM) of the cochlea. SSM was derived from μCT images of six cadaveric cochleae specimen in vitro, and includes the spiral ganglion (SG). The tonotopic mapping of the spiral ganglion or "active region" was computed using SG models from

Stakhoskaya *et al* [74]. The basis of the approach is to iteratively fit the model on the target CT image that has visible external cochlear features. Thus, by utilizing a segmented external cochlear image in conjunction with SSM, approximations to the intracochlear anatomy and the position of spiral ganglion are made. Upon segmentation, the tonotopic map from the model is finally transferred to the target image (see Figure 4.1 for an example of the processing sequence). The overall approach works by taking two CT scans. The first scan is taken prior to the surgery for an accurate SG segmentation without having to address issues concerning metallic artifacts in the image. Second scan is taken after surgery to determine exact electrode locations. The electrode array is generally well contrasted in the CT. The electrode contacts with spacing

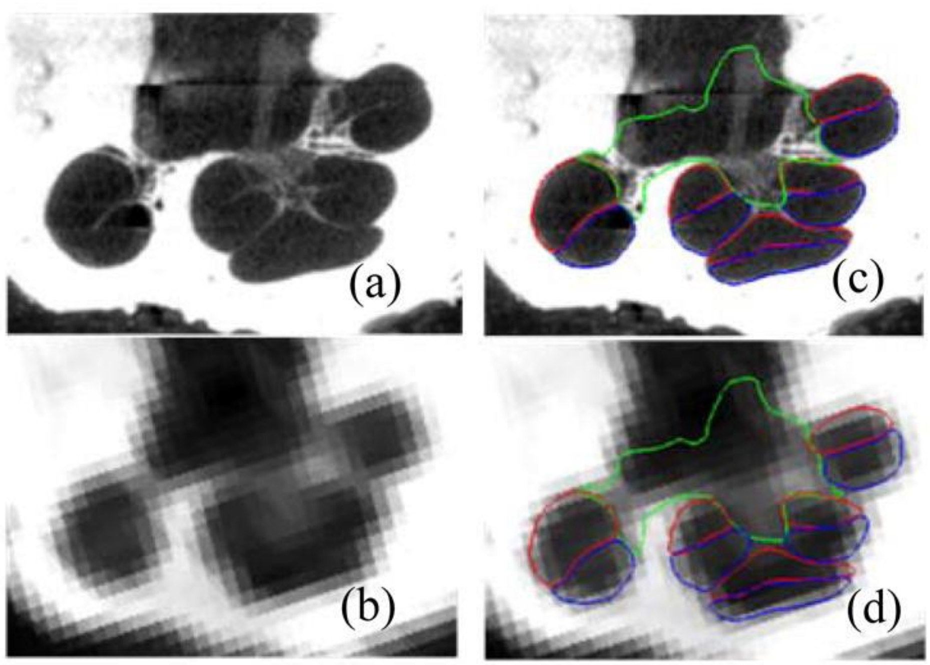


Figure 4.1. Slice of (a) µCT and (b) CT images of human cochlea. Same slices are segmented to delineate scala vestibule (blue), scala tympani (red), and spiral ganglion (green). Reprinted with permission from [106]. Copyright 2013, IEEE.

greater than or equal to 1 mm (for example Advanced Bioncis, and Med-El arrrays) are easy to identify. Closely spaced electrode contacts (for example, Cochlear Ltd. arrays) can be identified by superimposing electrode models on the target image. Finally, image reconstruction is

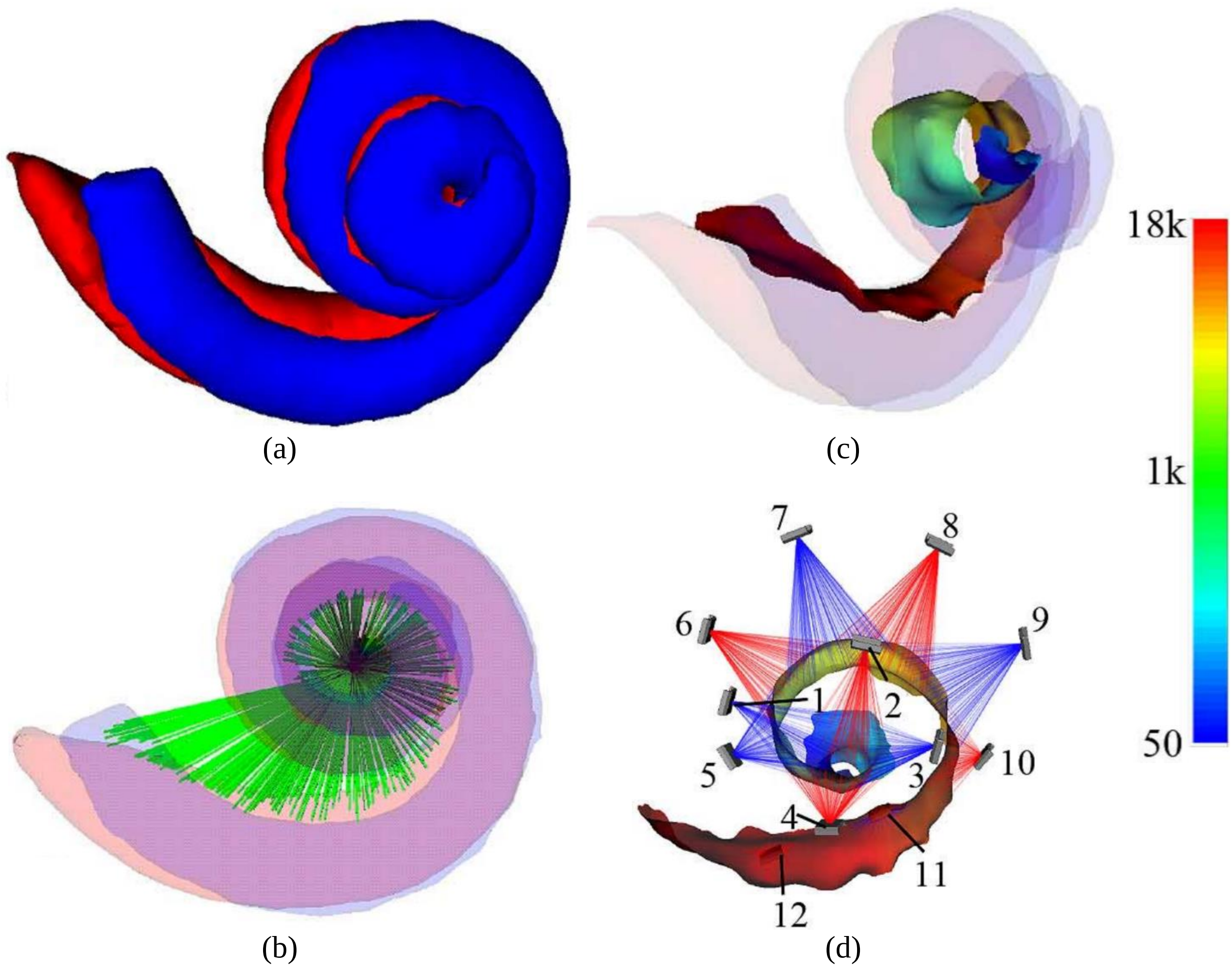


Figure 4.2. Spatial analysis of a cochlear implant recipient's cochlea. Shown in (a) – (c) are scala tympani in red and scala vesitbuli in blue. Rendering of the auditory nerves is shown in (b) in green. Tonotopic place of stimulation is color coded to represent the characteristic frequencies of the SG in (c) and (d). Rendering of implanted electrodes and the illustration of current spread is shown in (d). Blue and red colors are used to represent alternative electrode pairs. Reprinted with permission from [106]. Copyright 2013, IEEE.

completed by a transformation that correctly merges both pre- and post-operative scans to create a spatial representation of the electrode contacts with respect to the putative locations of the spiral ganglion. An example sequence is shown in Figure 4.2.

In order to facilitate a visualization of the programming-relevant information, these images are transformed to a 2-dimensional plot of distance-vs-frequency curves, an example of which is shown in Figure 4.3. Each blue or red DVF curve in the plot corresponds to an electrode in the array. A DVF curve defines the Euclidean distance from the corresponding electrode to the closest tonotopically mapped neural stimulation site. Distance is shown on the y-axis and the tonotopic frequency of the neural sites is varied on the x-axis. Thus, a DVF curve defines the

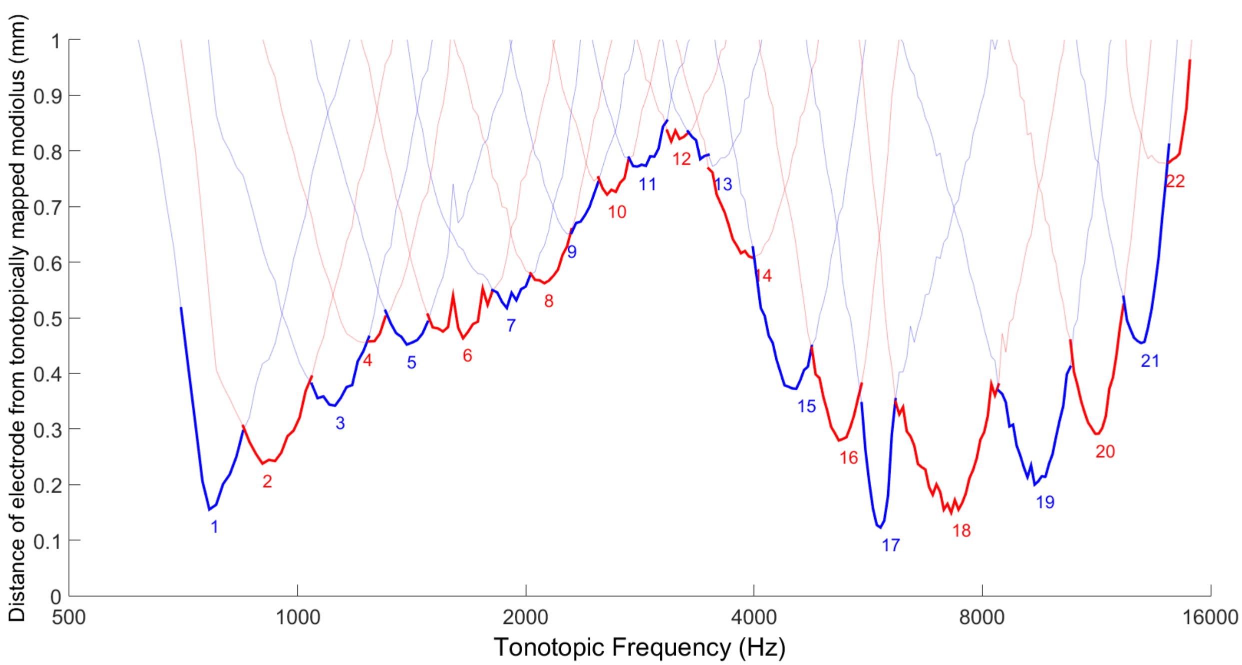


Figure 4.3. An example of electrode distance-vs.-frequency curves of a random cochlear implantee shown as a sequence of blue and red segments. The recipient has a 22-channel electrode array from Cochlear Ltd.

distance from various neural sites to the corresponding electrode. These curves provide a unique insight not only into the CF of each stimulation site, but also the degree of spectral overlap potentially caused by the neighboring electrodes (inference on potential level of current spread). DVF curves shown in Figure 4.3 indicate that most apical electrode is residing roughly 1 turn from the round window and is aligned with cochlear tonotopic frequency of 765 Hz. Conversely, the default frequency allocation strategy employed in clinical systems would map 765 Hz to 250 Hz (for Cochlear Nucleus device). That is at least a difference of 1.5 octaves. Furthermore, a comparison of the DVF curve minimas and center frequencies employed in clinical processors, as shown in Table 4.1, reveals a high degree of frequency-place mismatch. These frequency-place mismatch artifacts manifest into frequency compression, expansion, and warping across

Table 4.1. Default frequency allocation table for analysis bands in 22-channel ACE sound processing strategy. All frequencies are in Hertz (Hz).

Channel Number	1	2	3	4	5	6	7	8	9	10	11
Lower cut-off frequency	188	313	438	563	688	813	938	1063	1188	1313	1563
Center frequency	250	375	500	625	750	875	1000	1125	1250	1438	1688
Higher cut-off frequency	313	438	563	688	813	938	1063	1188	1313	1563	1813
Bandwidth	125	125	125	125	125	125	125	125	125	250	250

Channel Number	12	13	14	15	16	17	18	19	20	21	22
Lower cut-off frequency	1813	2063	1313	2688	3063	3563	4063	4688	5313	6063	6938
Center frequency	1938	2188	2500	2875	3313	3813	4375	5000	5688	6500	7438
Higher cut-off frequency	2063	1313	2688	3063	3563	4063	4688	5313	6063	6938	7938
Bandwidth	250	250	375	375	500	500	625	625	750	875	1000

the frequency range, and as discussed in the previous chapter, may decrease speech recognition performance. Here, we propose a frequency allocation scheme that utilizes the recipients' imaging data, more specifically the DVF curves, and attempts to reduce the frequency-place mismatch artifacts in hopes of improving implant performance outcomes. The details concerning the proposed custom frequency-allocation scheme are given in the next section.

4.2 Custom frequency allocation – Matching versus loss of low frequencies

Before describing the algorithm, we would first take a look at set of principles we utilize to create the frequency-allocation scheme. From a signal processing standpoint, an ideal scenario would be one in which the electrode contacts were distributed along the entire length of the spiral ganglion and were assigned acoustic frequencies in a matched one-to-one scenario, providing the full acoustic range to the periphery. Unfortunately, the electrode neural interface is typically less than ideal. Results from imaging studies indicate a wide range of insertion depths with Nucleus devices, which range from 2 to 26 mm from the round window [108-110]. Results from studies described in Chapter 3 suggest that if insertion depth is deep enough, matching acoustic frequency bands with the tonotopic place would be better for speech recognition. However, if insertion depth is moderate to shallow, both overly-compressed maps and frequency matched maps may result in a less than optimum performance. The speech intelligibility index [88] and experimental evidence make it quite clear that low frequencies are important for speech recognition, particularly for the intelligibility of vowels (although vowels are relatively less important for overall speech understanding). In general, better representation of formant frequencies, will result in better speech recognition performance. Therefore, over-truncating the low frequencies in an attempt to match the remaining frequency range (for example, for shallow insertions) may be detrimental. Although listeners may be able to adapt to the distorted spectral representation of speech, they probably cannot overcome the loss of information resulting due to truncation of a frequency range. This raises three inter-related key questions: First, to what extent could lower-frequencies be truncated without significantly impacting speech recognition performance (e.g., in an attempt to match the remaining frequencies)? Secondly, what is the range of frequencies that should be matched to the tonotopic place if a compromise between truncation and matching is to be achieved? Finally, to what could extent frequency compression be tolerable (or adapted to) without degrading performance?

Fu and Shannon [82] found that vowel recognition was only mildly affected by frequency allocation as long as analysis and carrier bands were matched. For simulated insertion depth of 21.25 mm, that places the corner frequency of most apical frequency band at 960 Hz at the Greenwood scale, the loss of intelligibility was only 20% from the best performance (i.e., for truncating all frequencies below 960 Hz). Their analysis further revealed that by matching analysis and carrier bands up to 25 mm insertion depth (i.e., cut-off frequency of apical most band = 513 Hz), would result in a performance level similar to that achieved with full insertion. In their experiments, with normal hearing and cochlear implant listeners, 25 mm insertion depth (or 3 mm shift in either direction) (i.e., 513 cutoff frequency) was the cross-over point beyond which signs of decline in performance started to appear. The findings from Başkent and Shannon [86] suggest that for insertions deeper than 26.4 mm, matched and compressed maps produced similar results, and for insertion ranges from 24 mm to 19.2 mm, performance was better with the matched map than with the frequency compressed map (26.4 mm and 19.2 mm from base correspond to cutoff frequencies of the most apical electrode equal to 393 Hz and 1332 Hz, respectively, see Figure 3.9). Faulkner *et al.* [94] demonstrated that matching frequencies to the tonotopic place up to 21 mm (995 Hz) did not result in any noticeable difference in the consonant and sentence recognition performance, and frequency matching for insertions lower than 19 mm was detrimental to overall speech recognition. Similar trends were demonstrated by Dorman *et al.* [81] and Li *et al.* [98], among others. Başkent and Shannon [87] demonstrated that a combination of frequency shift and compression that aligned the frequency range between 1 – 2 kHz resulted in the best performance, even when remaining frequencies did not match. The trend in results observed from the aforementioned studies follow predictions from the Articulation Index (AI) weightings very closely [88]. AI studies show that the loss of information below 1 kHz will significantly reduce intelligibility. The Speech Intelligibility Index (SII) weights the frequency range between 1 – 3 kHz as the most critical for overall speech understanding.

Please note that the above mentioned studies do not model the anatomy of the electrically stimulated cochlea and assume the Greenwood frequency map for computing frequency coordinates along the length of the cochlea. As noted in Section 3.1, the spiral ganglion map is different from the organ of corti frequency map, and is more relevant for cochlear implants. Overall, the general results from these studies indicate that optimal performance could be achieved with a compromise between the two extremes, (i.e., frequency compression and matching). In the next section, we would describe a frequency-allocation scheme that attempts to reach a compromise for an optimal representation of the speech information by minimizing frequency-place artifacts based on results from perceptual studies with normal hearing individuals and CI users.

4.3 Algorithm

The proposed frequency allocation scheme derives frequency bandwidths of the analysis filters from the DVF curves of each implant recipient uniquely. Each curve in Figure 4.3 corresponds to the spatial proximity of an electrode to the individual nerve fibers, where the minimum points on the curve represent the center CF stimulated by that electrode. We use these CFs of the stimulation sites as a reference to design the analysis filter-banks. The frequency space is first divided into the following four broad sub-bands: B_0, B_1, B_2, and B_3, with frequency ranges of ω_0= [0.2–0.5] kHz, ω_1= [0.5–1.0] kHz, ω_2= [1.0–3.0] kHz, and ω_3= [3.0–8.0] kHz respectively. From the DVF curves, we first determine the number of electrodes, n_i, whose CFs lie in each of the $i = 0, 1, 2, 3$ sub-bands, and perform the following set of procedures:

Step#1: If $n_0>0$ (i.e., deep insertion which enables access to the place frequencies lower than 500 Hz), n_0 filter(s) with center-frequencies is set equal to the CFs of the n_0 electrode(s) are assigned to create the perfect matching conditions.

If $n_1 \geq 2$, n_1 filters with center-frequencies is set equal to the CFs of the electrodes are assigned to create the perfect matching conditions.

Similarly, n_2 filters in B_2 space ([1.0–3.0] kHz) are allocated by perfectly matching the center frequencies of the filters with the corresponding CFs of the curves.

Step#2: If $n_1<2$, borrow (2- n_1) filters from the B_2 frequency space and map them on to the B_1 space. Introduce a mild frequency compression in the lower-most bands of the B_2 space (to compensate for the filters allocated to B_1), while maximizing the frequency matching of the remaining filters in the B_2 space with the CFs of the curves.

Step#3: Design n_3 filters in the B_3 frequency space by using a logarithmic filter spacing.

The aim of this simple 3 step rule-set is to maximize frequency matching at lower frequencies (less than 3 kHz) while ensuring that the lowest frequencies are not truncated (see Figure 4.4). For shallow insertion depths, instead of matching frequencies and thus truncating the lower frequencies, a mild frequency compression is used, while maximizing the frequency match between 1 - 3 kHz. In order to avoid loss of the low frequencies in case of shallow insertions, a minimum of 2 filters are always allotted in the B_1 space (500 Hz – 1 kHz). Although frequencies below 500 Hz could be useful, this limit was selected as a compromise to achieve matching in the B_2 frequency space (1 kHz – 3 kHz). This decision was based on experimental results from perceptual studies reported in [82, 86, 94] (see Section 4.2 for details). For deeper insertions, which provide tonotopically accurate access to frequencies lower than 500 Hz, filter-banks are

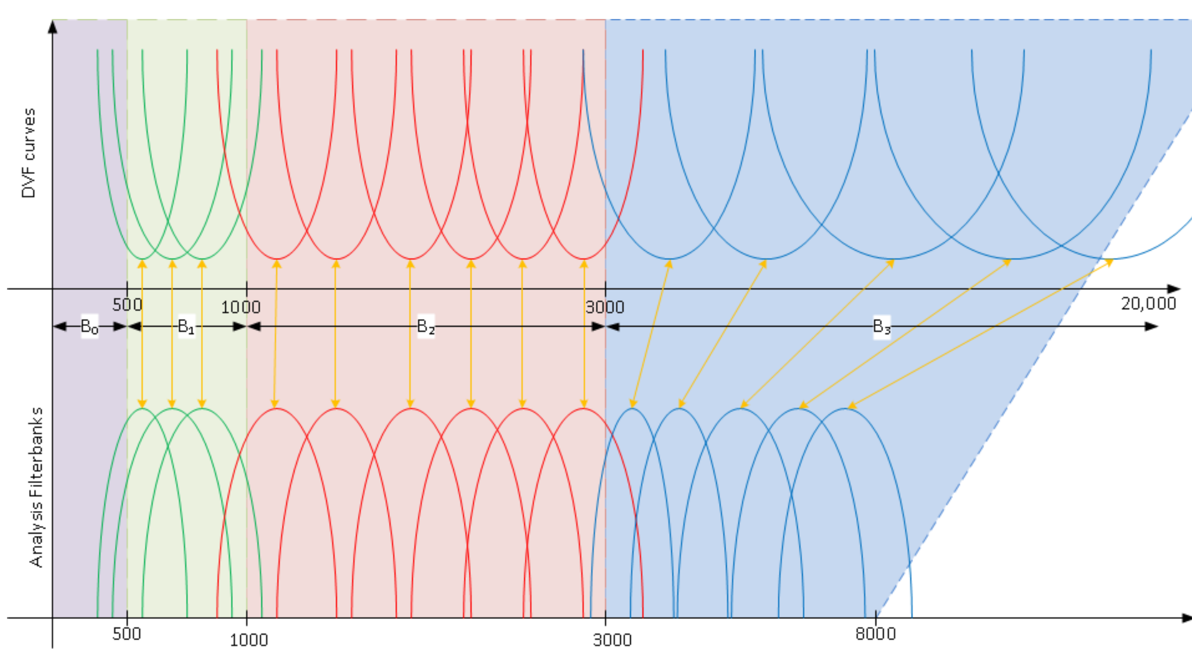


Figure 4.4. Custom Frequency allocation algorithm.The figure shows how DVF curves are mapped on four frequency spaces described in the algorithm to maximize matching at low frequencies, without excessively truncating the low frequencies.

matched according to the DVF curves. These rules are based on evidence from previous research studies as discussed earlier in this section, and aim to achieve a balanced frequency map with minimum frequency distortions and adequate representation of lower frequencies. In addition, bandwidth continuity constraints are imposed to avoid abnormally broad bandwidth filters. The use of an implant recipient's own imaging data, results in a unique frequency table that may provide a better match to the electrode neural frequency characteristics of that recipient. Thus, each recipient would have a custom, and typically different, frequency allocation table.

Figure 4.6 shows an example of relationship between the electrode locations in the cochlea, their tonotopic frequencies, and frequency-to-place mapping in (b) the standard/default fitting technique and (c) the user-customized mapping technique proposed here. The tonotopic map is derived from the DVF curves of an implant user and varies across CI recipients. Figure 4.6c shows a reduction in spectral shift and frequency-compression in the customized map as compared to the standard map; however, it is achieved at a cost of both decreasing the number of analysis bands in low frequencies, as well as some truncation of the lowest-most frequencies.

The proposed algorithm could optionally be combined with the electrode deactivation scheme as discussed in [106]. DVF curves that highly overlap with each other, indicate that neighboring electrodes are likely to cause channel interaction. Absence of curve minimas, high degree of overlap from the adjacent electrodes, and frequency bandwidths less than 125 Hz[3] on the DVF curves could be considered as likely indicators to deactivate electrodes to reduce spectral overlap. However, care must be taken not to switch off excessive number of electrodes, particularly the apical electrodes - the ones delivering the low frequency information, to avoid

[3] 125 Hz is the frequency bandwidth of each frequency bin in the Nucleus processors.

significantly reducing the spectral resolution. In an ideal CI sound coding scheme, the number of perceptually discriminable frequency channels should be as high as possible. This is typically achieved by designing narrow-band filters for lower frequencies to deliver better spectral resolution. Excessively switching off electrodes at lower frequencies may potentially be detrimental than the benefits achieved from reducing spectral overlap. Further research is needed to systematically evaluate the relationship between the number of electrodes at lower frequencies and the associated current spread.

4.4 Evaluation

The proposed custom frequency assignment scheme was evaluated in the following three listening studies:

i) Study 1: Acoustic simulations of cochlear implants with normal hearing listeners,
ii) Study 2: Simulating the effect of adaptation to frequency-place functions in normal hearing listeners, and
iii) Study 3: Evaluation with cochlear implant recipients.

In the following subsections, details about each of the experiment are provided.

4.4.1 Study 1: Acute simulations with normal hearing listeners

Acoustic simulations of cochlear implants (using noise-band vocoder) were conducted with normal hearing listeners to assess the efficacy of the proposed frequency assignment scheme and compare it against the clinical mapping strategy acutely.

Subjects

Forty-two normal-hearing (NH) listeners between the ages of 18 to 24 years participated in this study. All participants were native speakers of American English language, and had pure tone audiometric thresholds equal or better than 20 dB HL at octave frequencies from 250 to 8000 Hz. All subjects were paid for their participation and the study protocol was approved by the Institutional Review Board (IRB) of the University of Texas at Dallas. Each subject was tested with a single unique frequency-place map which was determined from the imaging data (DVF curves) of 14 unique CI users (1 map/subject). Each image map was tested with three different subjects.

Stimuli and Procedure

Speech recognition was assessed with four sets of test materials, namely vowels, consonants, sentences in quiet, and sentences in noise presented at +10 dB signal-to-noise ratio (SNR). Recorded IEEE sentences [111] were used as the stimuli for testing speech understanding in quiet and noise. Each listener was presented with 20 sentences per test condition. The root mean square level of all sentences was equalized and the acoustic stimuli were presented at 65 dB sound pressure level (SPL) in the free field from a single loudspeaker in a double-wall sound booth. Speech-shaped noise (SSN) with the same average long-term spectrum as the IEEE corpus was used to generate the noisy signals at 10 dB SNR. Vowel stimuli consisted of 12 medial vowels presented in /h/-vowel-/d/ context (had, hod, hawed, head, hayed, heard, hid, heed, hoed, hood, hud, who'd) [112]. Consonant stimuli consisted of 20 medial consonants presented in /a/-consonant-/a/ context (aba, acha, ada, afa, aga, aja, aka, ala, ama, ana, apa, ara, asa, asha, ata, atha, ava, awa, aya, aza) [113]. Both vowel and consonant stimuli were acquired from [113]. Each test material was presented in both male and female voices, with the test material order for all the test conditions randomized across subjects.

The performance was measured acutely without any training; however, participants were given glimpses of generic vocoder-processed stimuli for five minutes before the start of the experiment. In order to avoid any learning effects, no repetitions were allowed in any test condition. Subjects were asked to repeat back sentences presented in quiet and noise. All words were marked for correctness, and percent correct scores were computed by dividing the number of correct responses by the total number of words in that list. Consonant and vowels tests were conducted using a custom-developed MATLAB GUI[4] application, which had a grid of buttons (20 for consonants and 12 for vowels) each marked with one of the possible responses. Stimuli tokens were presented in a random order, and subjects were asked to respond by pressing the appropriate button. Subjects were instructed to make an educated guess or choose not to respond, if they were not sure. For each processing condition, 40 consonant tokens (20 male and 20 female) were presented, separately for both speakers. Similarly for vowels, 12 tokens were presented twice, once in a male and once in a female voice. The order of speakers was randomized across the trials. Percent correct scores were computed by dividing the number of correct responses by the total number of tokens presented. Subjects were given regular breaks and each test took approximately 3 hours per subject.

Signal Processing

In order to simulate the CI sound processing, a noise-band vocoder was implemented as shown in Figure 4.5. The input signal was first pre-emphasized using a second order Butterworth filter

---

[4] GUI stands for Graphical User Interface.

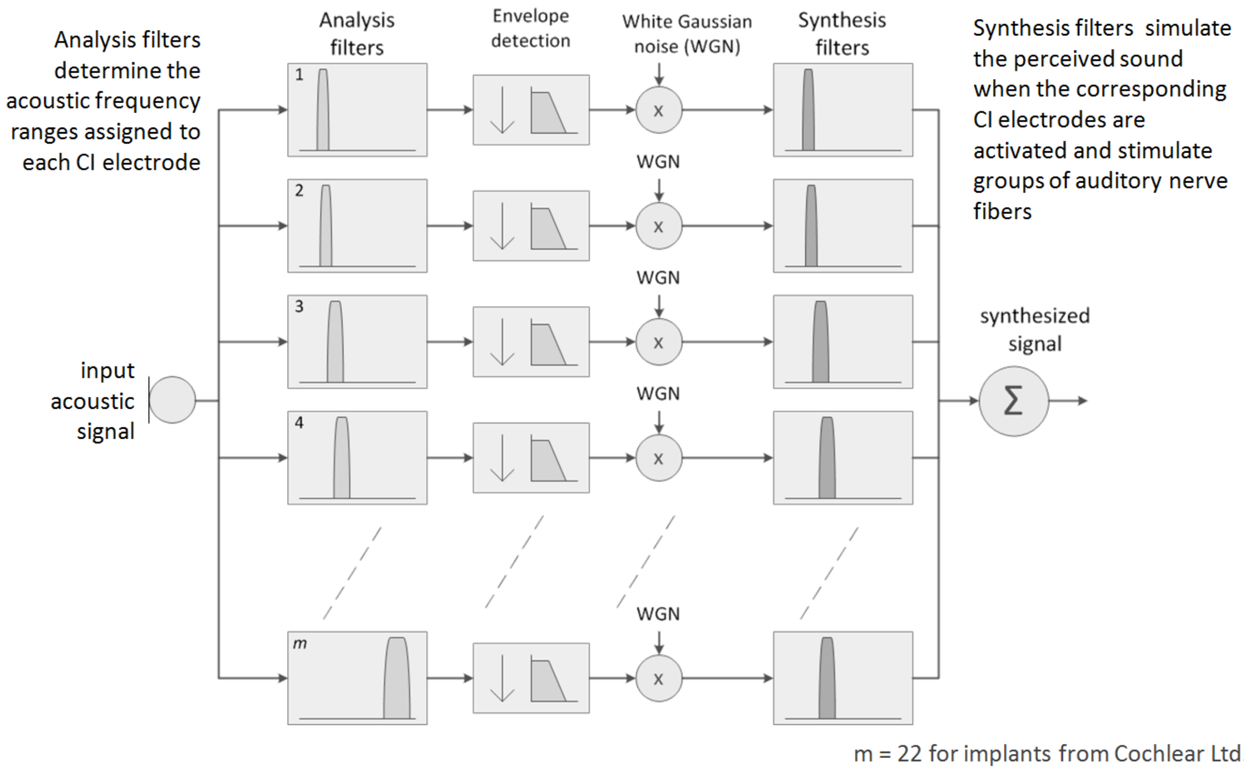


Figure 4.5. Vocoder to simulate cochlear implant sound processing. Analysis filters determine the acoustic frequency ranges assigned to each CI electrode, whereas synthesis filters simulate the perceived sound when the corresponding CI electrodes are activated and stimulate groups of auditory nerve fibers.

with a cut-off frequency of 1 kHz. The pre-emphasized signal was then passed through a set of bandpass analysis filters (3rd order Butterworth). The number of filters were typically 22, corresponding to the number of electrodes in implant systems manufactured by Cochlear Corp[5]. Next, envelopes from each frequency channel were extracted via rectification and low-pass filtering (2nd order Butterworth with a cutoff frequency equal to 400 Hz). The envelope of each band modulated white noise and the resulting multiband signal was passed through a set of

[5] For some maps, fewer than 22 channels were used depending on the DVF curves. Some electrodes were simulated as de-activated to decrease the cross-channel interference as in [106, 107]. For each map and condition, the number of active channels are provided in the next section.

bandpass synthesis (carrier) filters (3^{rd}-order Butterworth). The signals were finally summed across all the bands to produce a single vocoder-processed acoustic signal. The RMS value of the vocoded (vocoder-processed) signal was equalized to match the original signal.

The frequency characteristics (cutoffs, bandwidths, center frequencies) of analysis filters determine the acoustic frequency ranges assigned to the corresponding CI electrodes. Each synthesis filter simulates the frequency space on the cochlear place that is stimulated with the activation of the corresponding CI electrode. The frequency characteristics of analysis and synthesis filters were manipulated to simulate the following four mapping conditions:

Condition#1: Default frequency allocation with ideal electrode position: In this condition, same set of analysis and synthesis filters were used to simulate an ideal scenario in which the acoustic frequencies were matched to the cochlear place in a matched one-to-one scenario. This condition used the default frequency allocation table which is used in ACE coding strategy. The frequency characteristics of the filters used in this condition are given in Table 4.1.

Condition#2: Default frequency allocation with true electrode position: In this condition, we tried mimicking the actual listening perception experienced by CI users. This is achieved by using the default ACE filterbanks at the analysis stage, and filterbanks derived from the DVF curves at the synthesis stage. The resulting signal typically has frequency-place mismatches.

Condition#3: Custom frequency allocation with true electrode position: In this condition, custom frequency allocation was used at the analysis stage and filterbanks derived from the DVF curves were used as synthesis filters to simulate the perceived sound. Custom filter-banks were designed according to each individual's DVF curve data using the methods described in Section 4.3.

Condition#4: Frequency allocation matched with true electrode position: The analysis and synthesis filter-banks were chosen identically from the DVF curves. This condition corresponds to perfect matching of the acoustic filterbanks with the cochlear tonotopic locations of the electrode contacts.

The four conditions are summarized in the Table 4.2 and are depicted graphically in Figure 4.6. Details about each of the frequency allocation map are given in Appendix B. The characteristic frequencies of nerve bundles stimulated by the most apical electrode in each image map are given in Table 4.3. Other than condition 3, all other conditions were always simulated with 22 frequency bands. In condition 3, some electrodes were simulated to be switched-off to decrease cross-channel interference. The number of frequency bands ranged from 14 to 22.

Table 4.2. Summary of frequency mapping conditions for experiment 1.

Condition	Analysis Filters	Synthesis Filters	Number of bands
1. Default Frequency allocation, with ideal electrode positioning	Default ACE	Default ACE	22
2. Default Frequency allocation, with true electrode positioning	Default ACE	Image-based	22
3. Custom frequency allocation, with true electrode positioning	Custom	Image-based	Variable (14 -22)
4.. Frequency allocation matched with true electrode positioning	Image-based	Image-based	22

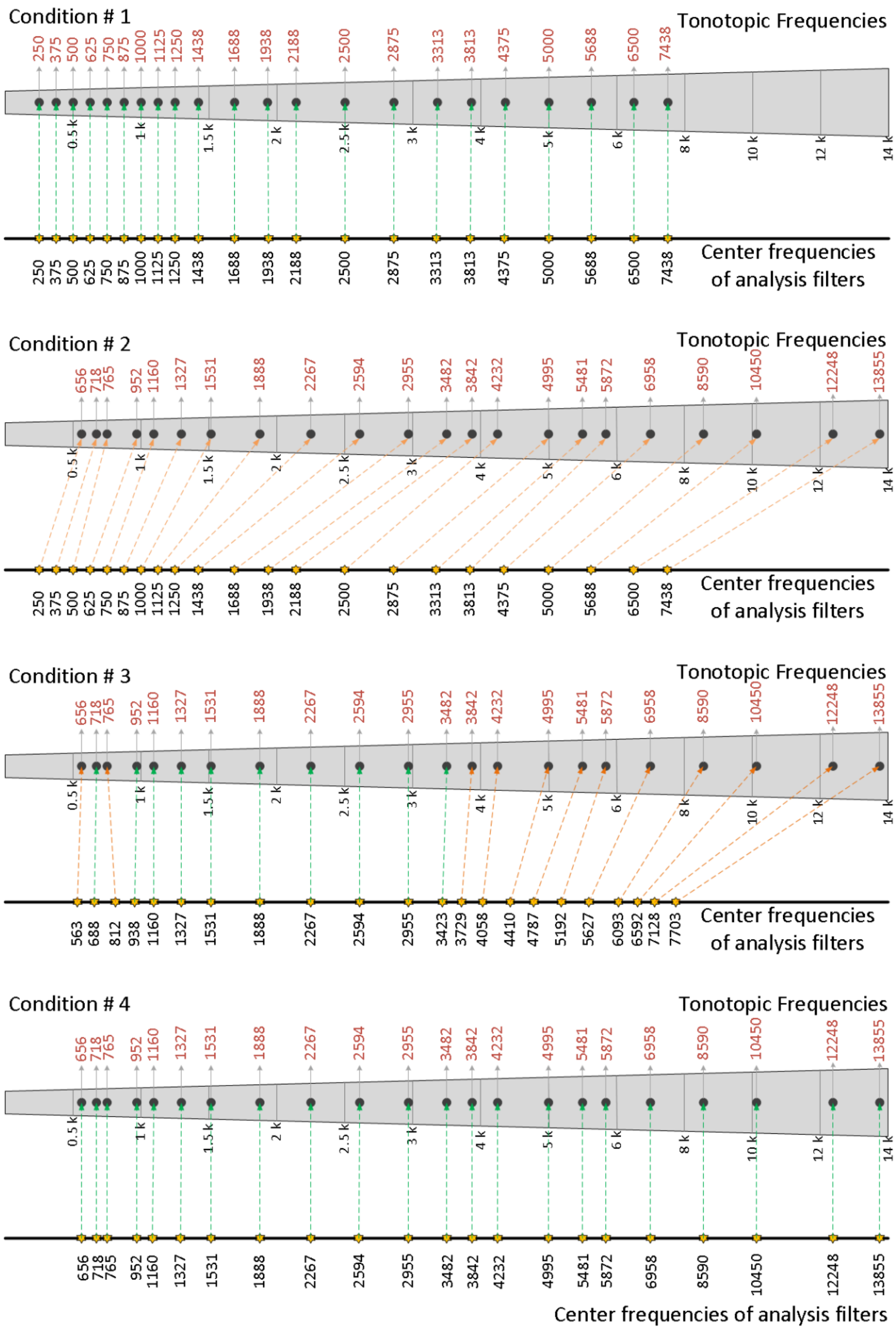


Figure 4.6. An example of graphical depiction of frequency mapping conditions for experiment 1. Conditions 2 – 4 use a randomly selected image map (DVF) of a CI user. Figure not to scale.

Table 4.3. Characteristic frequencies of the nerve bundles stimulated by the most apical electrode for each image map based on SG map.

MAP ID	M1	M2	M3	M4	M5	M6	M7
CF of most apical electrode (Hz)	291	337	656	702	733	765	797
MAP ID	M8	M9	M10	M11	M12	M13	M14
CF of most apical electrode (Hz)	847	847	864	864	934	1101	1201

Subjects were tested with all four frequency mapping conditions. The order of conditions was randomized across the subjects. Each image map was tested with three individual subjects and the scores were averaged.

Results

Figure 4.7 shows mean speech understanding scores for each of the four mapping conditions as a function of test material. Consistent with findings from the previous studies, the results here indicate peak performance with ideally matched condition (Cond#1) (i.e., full range of acoustic information is matched exactly across analysis and synthesis filter banks). However, since Cond#1 is generally not achievable in real life, the aim of the study was to compare the performance of Conds #3 and #4 against Cond#2, which simulates the mapping conditions that could be achievable with frequency-place configurations of the clinical systems. Results indicate Cond#3 generally performed equal or better with all test material as compared to Conds #2 and #4, with largest improvement observed for speech (sentences) presented in quiet (+15% improvement), followed by speech in noise (+12% improvement), and vowel identification (+8% improvement).

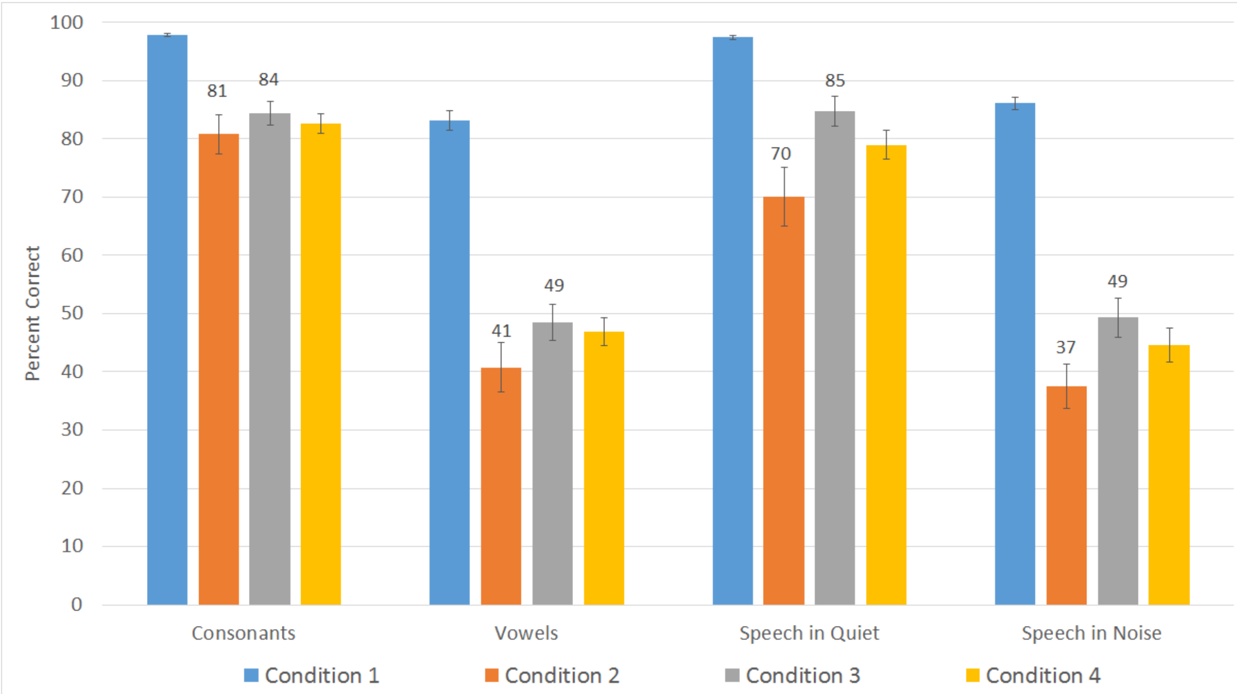


Figure 4.7. Average speech recognition scores of 42 participants from Study 1. Percentage correct scores for recognition of consonants, vowels, speech in quiet, and speech in noise (SNR = 10 dB) with respect to four frequency mapping conditions. Condition 1: Default Frequency allocation, with ideal electrode positioning. Condition 2: Default Frequency allocation, with true electrode positioning. Condition 3: Custom frequency allocation, with true electrode positioning. Condition 4. Frequency allocation matched with true electrode positioning. Error bars represent standard errors of means.

Repeated-measures analysis of variance (ANOVA) was performed to assess the effects of mapping conditions and speech material on the speech understanding scores with an α factor set to 0.05. Subjects were considered a random factor, while mapping conditions and speech material were used as the main analysis factors. ANOVA revealed a significant main effect of mapping condition ($F[3,123]=168.59$, $p<0.001$) and test material ($F[3,123]=310.92$, $p<0.001$) on speech understanding scores. The interactions between mapping conditions and test material were statistically significant ($F[9,369]=43.57$, $p<0.001$). A Post-hoc Bonferroni test for pairwise comparisons between the four mapping conditions found all conditions to be statistically different from each other ($p<0.001$). The comparisons between the speech material indicated that all test material significantly differed from each other, except for vowels and sentences presented in noise ($p = 1.000$). Post-hoc Bonferroni-corrected comparisons for each test material indicated that with the exception of following two, 1) consonant understanding in conditions 2 and 4 ($p = 0.276$), and 2) vowel identification in conditions 3 and 4 ($p = 0.414$), all test speech material/conditions were significantly different from each other. All other test conditions were found to be significantly different from each other for the four tested speech materials.

Figure 4.8 shows mean sentence recognition scores presented in quiet for 14 different image maps with conditions 2 and 3. The figure shows one-to-one comparison of performance levels achieved with the clinical and custom frequency mapping. Individual maps are depicted on the horizontal axis and are ordered with decreasing electrode insertion depth. The characteristic frequencies of nerve bundles stimulated by the most apical electrode in each image map are given in Table 4.3. For all the image maps tested, the proposed frequency allocation produced better performance than the clinically assigned frequency allocation table for sentence recognition in quiet. On average, the proposed solution resulted in 15% improvement as compared to the clinical map. A 1-way repeated measures ANOVA was conducted to compare the effect of two mapping conditions on overall sentence recognition scores (in quiet). A statistically significant effect of frequency allocation scheme was found ($F[1,41]=42.05$, $p<0.001$). A 1-way repeated-measures ANOVA was recomputed with 14 image maps as between-subject factors to investigate if frequency allocation scheme had a consistent effect across all image maps. The ANOVA revealed a significant main effect of the mapping condition

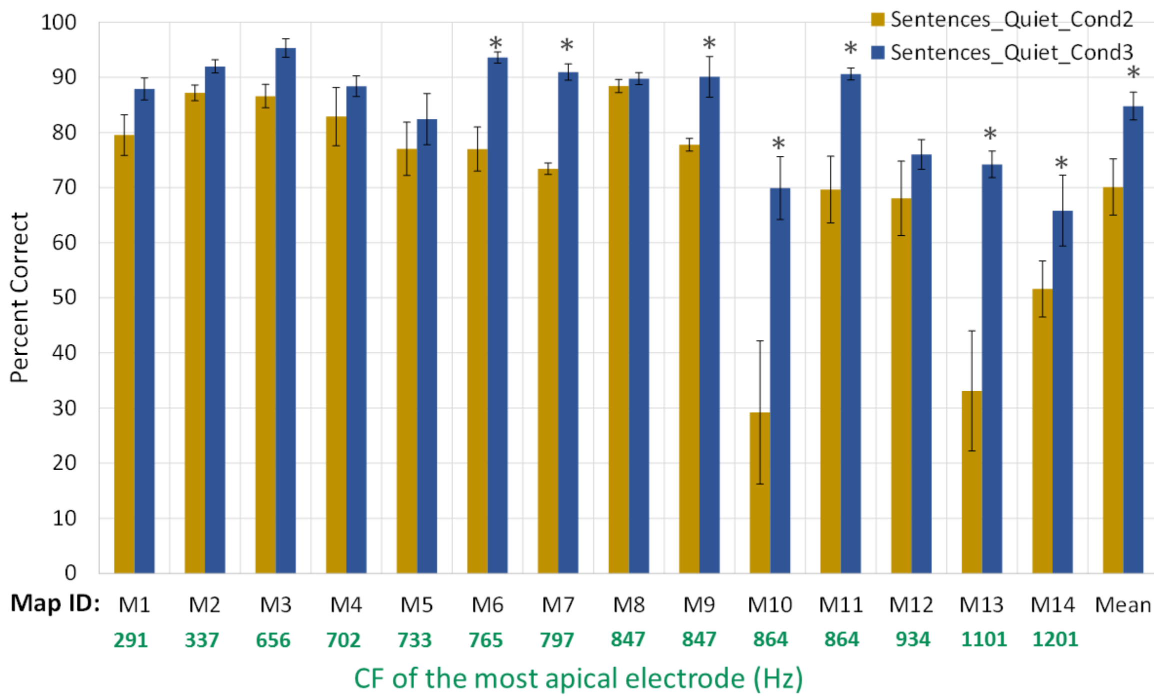


Figure 4.8. Average sentence recognition scores in quiet of 42 participants tested with 14 unique maps (3 subjects/map) from Study 1.Scores from condition 2 and condition 3 are shown. Each bar represent average scores from 3 participants. Error bars represent standard errors of means. Significantly different scores (*p* values less than 0.05) are marked with asterisk.

(F[1,28]=89.52, *p*<0.001), as before. The interactions between frequency allocation and image maps were also statistically significant (F[13, 28]=4.56, *p*<0.001). This was followed by pairwise comparisons for each image map. The performance levels for the maps which were found to be significantly different (*p* values < 0.05) for the two frequency allocation schemes are marked by asterisk on Figure 4.8.

Figure 4.9 shows average sentence recognition scores in the presence of noise (at 10 dB SNR) for the two frequency allocation schemes with 14 image maps. On average, sentence recognition scores were 12 percentage points better with the custom frequency maps than the clinically assigned frequency allocation table. In order to test the statistical significance of the

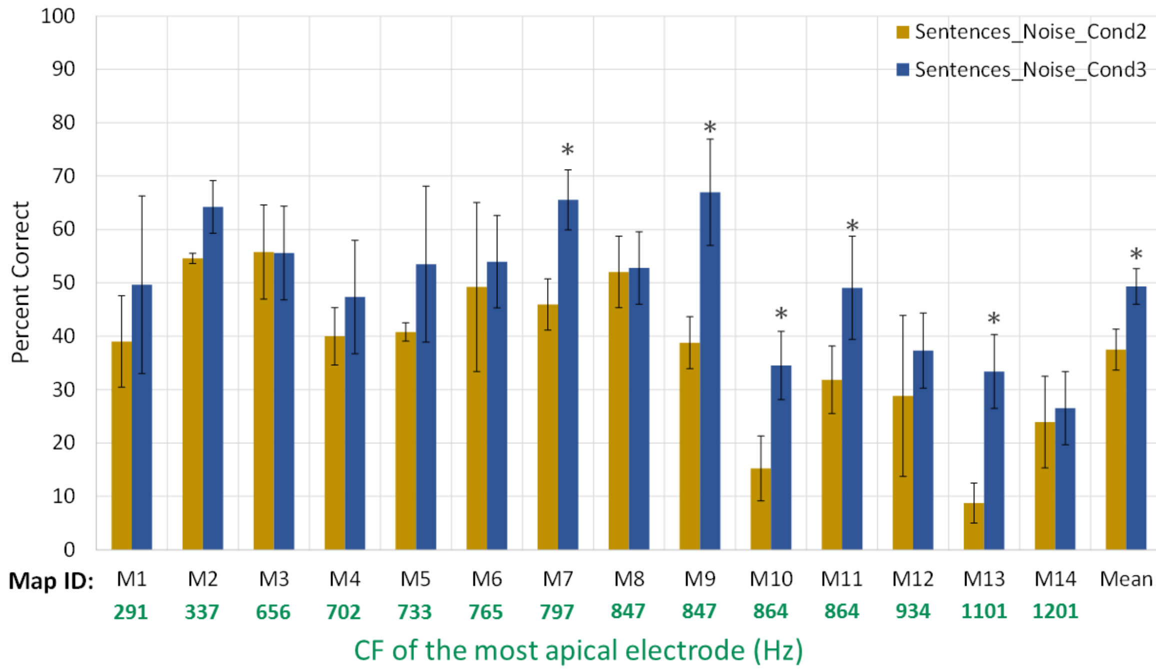


Figure 4.9. Average sentence recognition scores in noise (10 dB SNR) SSN of 42 participants tested with 14 unique maps (3 subjects/map) from Study 1. Each bar represent average scores from 3 participants. Error bars represent standard errors of means.

scores with two frequency allocation schemes, a 1-way repeated measures ANOVA was conducted. A statistically significant effect of frequency allocation scheme was found ($F[1,41]=31.10$, $p<0.001$), indicating that the proposed frequency allocation scheme produced statistically better results for sentences presented in noise as compared to the default frequency mapping. A 1-way repeated-measures ANOVA was re-performed with 14 image maps as between-subject factors to investigate if the frequency allocation scheme had a consistent effect across all image maps. A statistically significant effect of frequency mapping was found ($F[1,28]=35.01$, $p<0.001$). The interactions between frequency allocation and image maps were not statistically significant ($F[13, 28]=1.38$, $p=0.22$). Differences in the two condition that were found to be statistically significant in the individual F-tests for each image map are marked by asterisk in Figure 4.9.

Consonant identification scores with the two frequency mapping schemes are shown in Figure 4.10 for each individual image map. As compared to sentence recognition scores, consonant identification showed roughly equal performance level for all image maps. Despite little numerical differences in the identification scores between the two mapping conditions, a 1 way repeated measures ANOVA revealed an overall statistically significant difference between the two mapping conditions ($F[1,41]=5.49$, $p=0.024$). A 1-way repeated-measures ANOVA conducted using image maps as between-subject factors showed a significant effect of the frequency allocation scheme ($F[1,28]=7.48$, $p=0.011$), as well as significant interactions between image maps and the two frequency allocation schemes ($F[13, 28]=77.78$, $p=0.044$). Pairwise comparisons between the two frequency allocation schemes for each image map revealed that only two image maps, namely M10 ($p<0.001$) and M12 ($p=0.025$), produced statistically significant differences in performance.

Vowel identification scores with each image map are shown in Figure 4.11. On average, vowel recognition was better by 8 % points with the proposed custom frequency assignment strategy as compared to condition 2 (default frequency mapping). In order to assess the statistical significance of the differences in scores with the two frequency mapping conditions, a 1-way repeated measures ANOVA was performed. Results indicated an overall significant effect of frequency-mapping condition ($F[1,41]=16.387$, $p<0.001$). The analysis was repeated to assess the contribution of each map in the observed differences, by keeping image map as between-subjects factor. Significant effect of frequency-mapping condition was observed ($F[1,28]=19.14$,

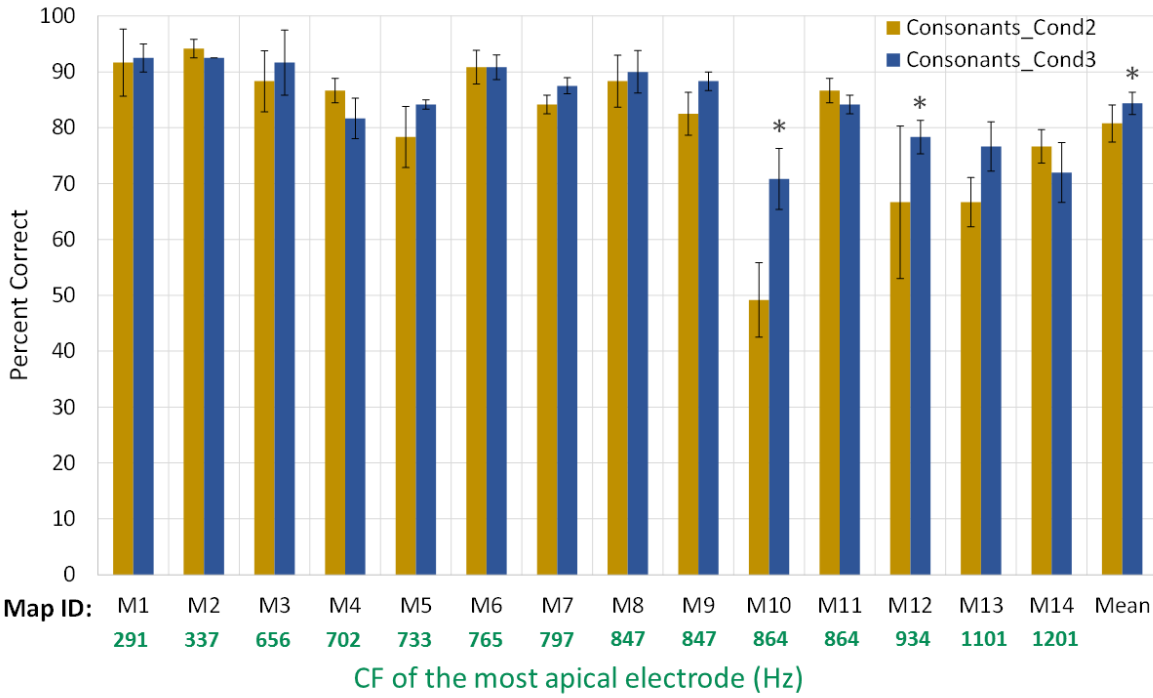


Figure 4.10. Average consonant recognition scores in condition 2 and condition 3 of 42 participants tested with 14 unique maps (3 subjects/map) from Study 1.Each bar represent average scores from 3 participants. Error bars represent standard errors of means.

p<0.001). However, the interactions between the image maps and the frequency allocation schemes were not statistically significant. (F[13, 28]=1.53, *p*=0.168). Differences in the condition that were found to be statistically significant in individual F-tests for each image map are marked by asterisk in Figure 4.11.

The effect of insertion depth on speech recognition performance was evaluated using a 2-way repeated measures ANOVA by keeping the recognition scores as blocked factor and CF of the most apical electrode as between-subjects test factor. Three frequency allocation schemes were evaluated, namely condition 2 (default frequency allocation), condition 3 (custom frequency allocation), and condition 4 (matched, i.e., analysis and synthesis filters matched with

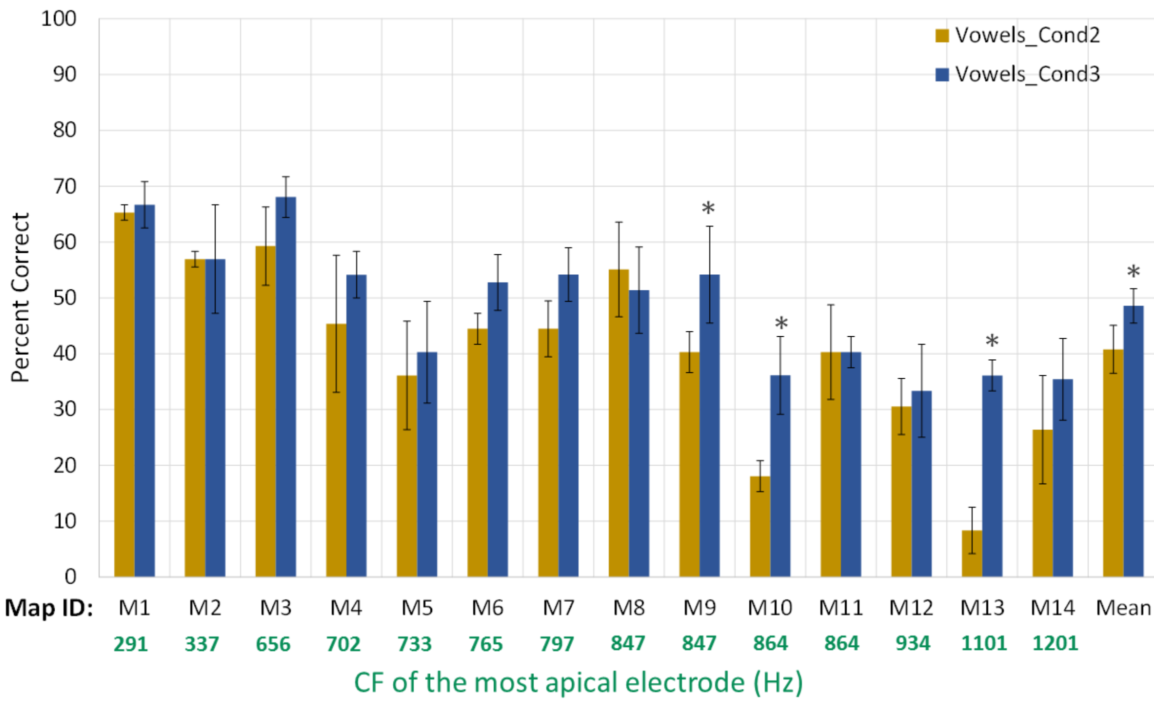


Figure 4.11. Average vowel recognition scores in condition 2 and condition 3 of 42 participants tested with 14 unique maps (3 subjects/map) from Study 1.Each bar represent average scores from 3 participants. Error bars represent standard errors of means.

Table 4.4. Analysis from 2-way repeated measures ANOVA. Analysis is presented for 3 frequency allocations conditions, with insertion depth as with in subjects' factor. Interactions refer to interactions between the frequency mapping condition and insertion depth (CF of most apical electrode).

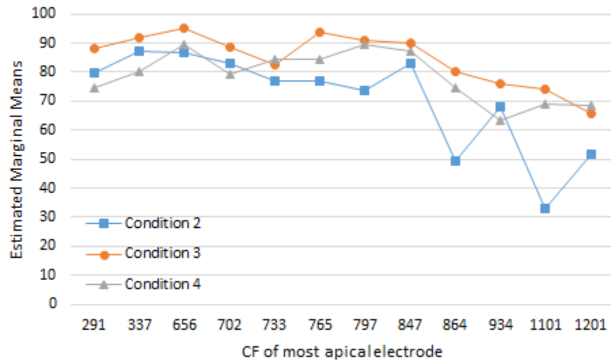
	Sentences in Quiet	Sentences in Noise	Consonants	Vowels
Condition	$F[2, 60] = 39.765$ $p<0.001$	$F[2, 60] = 16.054$ $p<0.001$	$F[2, 60]= 2.338$ $p=0.105$	$F[2, 60] = 7.532$ $p<0.001$
Interactions	$F[22, 60] = 4.059$ $p<0.001$	$F[22, 60] = 1.702$ $p=0.054$	$F[22, 60]= 1.492$ $p<0.112$	$F[22, 60] = 1.196$ $p=0.286$

tonotopic place). The ANOVA results are tabulated in Table 4.4. The results indicated that there was a significant effect of insertion depth on the identification of vowels and sentence recognition, both in quiet and in noise. For the simulated insertion depths, consonant identification performance was not impacted by the simulated insertion depths with either of the three frequency allocation schemes. This may be explained by the fact that consonant recognition is primarily known to be dependent on temporal details than the spectral details. For sentences presented in quiet, a significant interaction was found between the frequency allocation condition and the simulated insertion depth. Pairwise comparisons between frequency allocation scheme and insertion depth were conducted. The results of the analysis are visually represented in Table 4.5. The entries marked with asterisk in the table were statistically different. No statistically significant interactions were observed for consonants, vowels, and sentences presented in noise. Estimated marginal means for all four types of speech stimuli with three frequency allocation schemes are plotted in Figure 4.12.

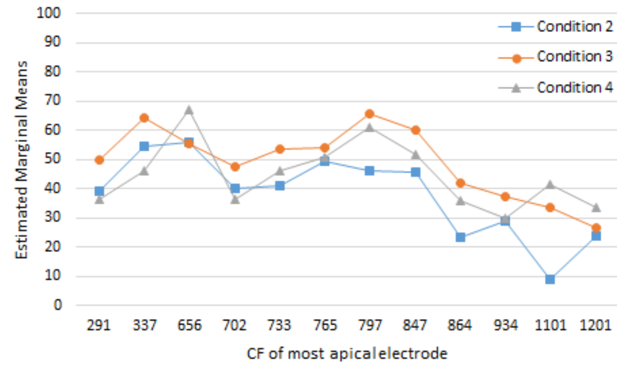
It is noteworthy to analyze at least two trends in Figure 4.12. First, for all speech material, performance decreased with decreasing insertion depth. However, sentence recognition and vowel identification was significantly impacted by the loss/mismatch of frequency due to the

Table 4.5. Visual representation of pairwise comparisons between frequency allocation conditions and insertion depths for sentences presented in quiet. Boxes shaded in dark represent statistically significant interactions (p<0.001)

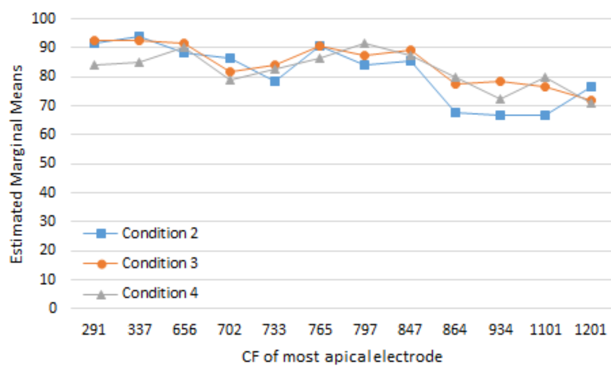
Apical Frequency (Hz)	291	337	656	702	733	765	797	847	864	934	1101	1201
Cond2 - Cond3						*	*		*		*	*
Cond2 - Cond4							*		*		*	*
Cond3 - Cond4	*	*		*		*			*	*		



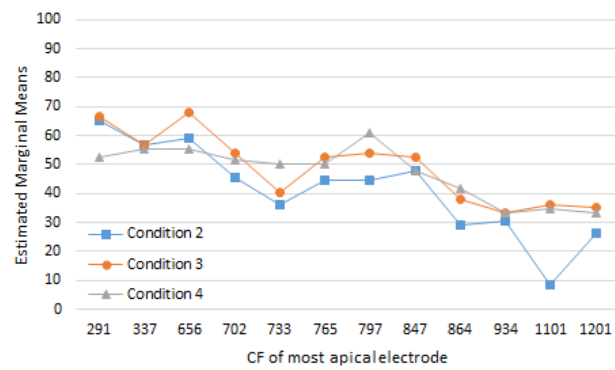
(a) Sentences in Quiet



(b) Sentences in Noise



(c) Consonants



(c) Vowels

Figure 4.12. Estimated marginal means for three frequency allocation conditions as a function of insertion depth (CF of nerve fibers stimulated by most apical electrode). Average scores from 42 subjects who participated in Study 1.

decreasing insertion depth. Consonants were relatively not much impacted by the simulated insertion depths. Secondly, the loss in performance was more profound for condition 2 (clinical/default frequency allocation scheme) as compared to other schemes. The proposed/custom frequency allocation scheme (condition 3) generally outperformed condition 2. These trends are in line with the statistical analysis presented earlier in this section.

Discussion

Experiment 1 aimed at investigating the acute effects of different combinations of frequency place maps on speech recognition. Specifically, four frequency allocation schemes were tested, i) ideal frequency-place mapping, which is not usually achieved in a clinical setting, ii) default frequency allocation and its acute effects with true electrode positions, iii) custom (proposed) frequency allocation scheme with reference to the true electrode positions, and iv) frequency allocation matched with the true electrode positions. Electrode locations were derived from image maps of 14 implant recipients, and each map was tested with 3 participants. Please note, that all the maps had unique place frequencies, and locations of the most apical electrodes across the maps ranged from tonotopic frequencies[6] of 291 to 1201 Hz (Table 4.3). Condition 1 (ideal), as expected, produced best results with all the test materials. In addition to mapping full acoustic range (178 – 8000 Hz), which was aligned with the tonotopic place of stimulation, it is noteworthy to see that the condition 1 also provided better spectral resolution at lower frequencies than any other mapping condition. At least 5 independent frequency bands were assigned for frequency range from 178 – 1000 Hz. While the number of perceptually discriminable (utilizable) channels in cochlear implants may be less than the total number of electrode contacts on the array (perhaps due to current summation), normal hearing listeners are able to make use of higher number of frequency bands in complex listening environments.

Condition 3 generally performed equal or better than condition 2, at least acutely with the tested speech material. One of the potential reasons for better performance may be attributed to the lower frequency-place mismatch with the proposed solution, and hence lower spectral

---

[6] Spiral ganglion frequencies according to Stakhovskoya *et al.* [74].

distortion in the perceived sound. Within test materials, the largest difference was observed for sentences in presented in quiet (+15%); however the level of performance was significantly lower than the ideal condition, indicating that spectral distortions due to the loss of low frequencies, loss of spectral resolution (due to wider filter banks), residual frequency-place mismatch, and absence of learning effects could account for the remaining gap in the loss of performance. The performance may improve with extended exposure to the frequency allocation schemes, as can be seen with studies conducted by Rosen *et al.* [90] and Faulkner *et al*. [94, 95]. Furthermore, adaptation trends and learning patterns (final performance levels and adaptation times) may differ for different frequency maps. It would not be surprising to observe easier and perhaps complete adaptation to the maps with lower mismatch as compared to the extremely distorted maps; the later may only result in partial adaptation. In the next experiment, we will explore the effects of training and adaptation to different frequency allocation schemes.

4.4.2 Study 2: Simulating the effect of adaptation to frequency-place mismatch with normal hearing listeners

The effects of auditory training with acoustic simulations of cochlear implants on the speech recognition of normal hearing listeners were investigated in this study. The aim of the experiments was to study the interactions between the training time and frequency allocation schemes, which were presented in the previous sections.

Subjects

Ten normal-hearing (NH) listeners between the ages of 18-24 participated in this experiment. All participants were native speakers of American English language, and had pure tone audiometric thresholds equal or better than 20 dB HL at the octave frequencies from 250 to 8000 Hz. Five image maps, which had shallow insertion and showed relatively low performance in study 1, were selected for these experiments. These image maps correspond to maps M5, M10, M12, M13 and M14 of study 1. Table 4.6 provides the characteristic frequencies of the nerve fibers stimulated by the most apical electrode for each image map. For details on the frequency characteristics of each image map, please refer to Appendix B. Half of the participants (N = 5) were tested with condition 2 of the previous study (default frequency allocation, with true electrode positions), and the other half (N = 5) were exclusively tested with custom frequency allocation strategy (condition 3).

Table 4.6. Characteristic frequencies of the nerve bundles stimulated by the most apical electrode for each image map used in Study 2.

MAP ID	M5	M10	M12	M13	M14
CF of most apical electrode (Hz)	733	864	934	1101	1201

Stimuli and Procedure

Speech recognition was assessed using the same set of speech stimuli as used in the experiment 1, namely, 12 medial vowels, 20 medial consonants, IEEE sentences presented in quiet, and in noise at +10 dB signal-to-noise ratio (SNR). The order of the test material was randomized within subjects and across trials. The testing and grading procedure essentially remained the same as that of the first study. However, in this experiment, performance was measured at four different time intervals of training as illustrated in Figure 4.13. The first testing was conducted

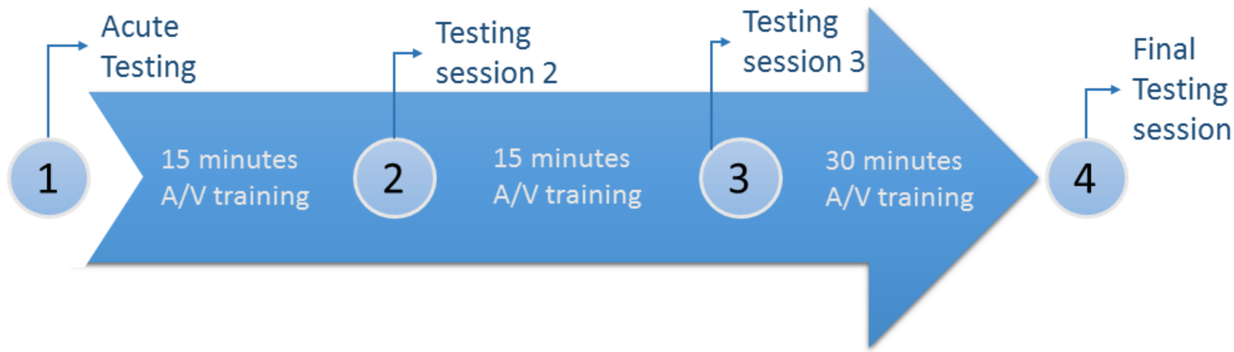


Figure 4.13. Test and training procedure for study 2. Subjects were tested acutely (with minimal training) at the start of the test, and then progressively given audio/video (A/V) training sessions and tested at intervals 2, 3, and finally at 4.

acutely, i.e., with minimum training – about 5 minutes to acquaint listeners with vocoder processed sounds and speech material. The participants were then given three audio-video (A/V) training sessions, and the performance was measured after each training session. First two training sessions were 15 minutes each, and the last session was 30 minutes in duration. A/V training consisted of watching and listening to a video embedded with processed audio. The material comprised of inspirational (most popular) TED videos to encourage participants to listen. The audio was extracted and processed with the test condition (vocoder implemented with the frequency allocation scheme) and merged back to the original video to create an A/V training material. In total, four videos were used, two from male and two from female talkers.

In order to be consistent, no repetitions were allowed. Subjects were asked to repeat back the words they perceived for sentences presented in quiet and noise. All the words were marked for correctness. 40 tokens of consonants (20 in male and 20 in female voice) were presented per condition. Similarly, 24 vowel tokens (12 in male and 12 in female voice) were presented per condition. Consonants and vowels identification tests were conducted by responding on the computer screen. After each testing session, participants were given a break of 5 – 10 minutes. However, no breaks were given within the training and testing session. Tests took approximately 4 hours per subject.

Signal Processing and Conditions

Noise-band vocoder, similar to the one used in Study 1, was used for these experiments. Two frequency allocation schemes were tested:

Condition#2: Default frequency allocation with true electrode position to mimic the actual frequency-place map of the CI recipients. Default ACE filterbanks were used at the analysis stage, and filterbanks derived from the DVF curves were used at the synthesis stage.

Condition#3: Custom frequency allocation with true electrode position to evaluate the proposed strategy. A custom frequency allocation table was used at the analysis stage and filterbanks with CFs derived from the DVF curves were used as synthesis filters.

Results

Figure 4.14 shows mean speech understanding scores at four time intervals for the two mapping conditions with different test materials. In general, both conditions showed improvement in performance with auditory training, suggesting that listening performance may improve with long-term exposure of distorted speech, at least for the range of maps tested in this experiment. The greatest improvement was observed for sentences presented in quiet for both conditions, 17% and 11% points from the baseline acute performance for conditions 2 and 3, respectively. On all measures of speech, condition 3 (custom frequency allocation) generally performed numerically better than the condition 2 (default frequency allocation scheme). Largest difference between the two conditions at the final testing session was observed for sentences presented in quiet (23 %) and in noise (30%), followed by vowels (20%). Consonant recognition remained relatively unchanged for both the maps, but there was an overall numeric improvement of 11% with condition 3. It is difficult to say if the performance reached at an asymptotic level by the last test session; however, trends in Figure 4.14 indicate that most significant improvement was observed just after 15 minutes of training, after which listener performance continued to improve

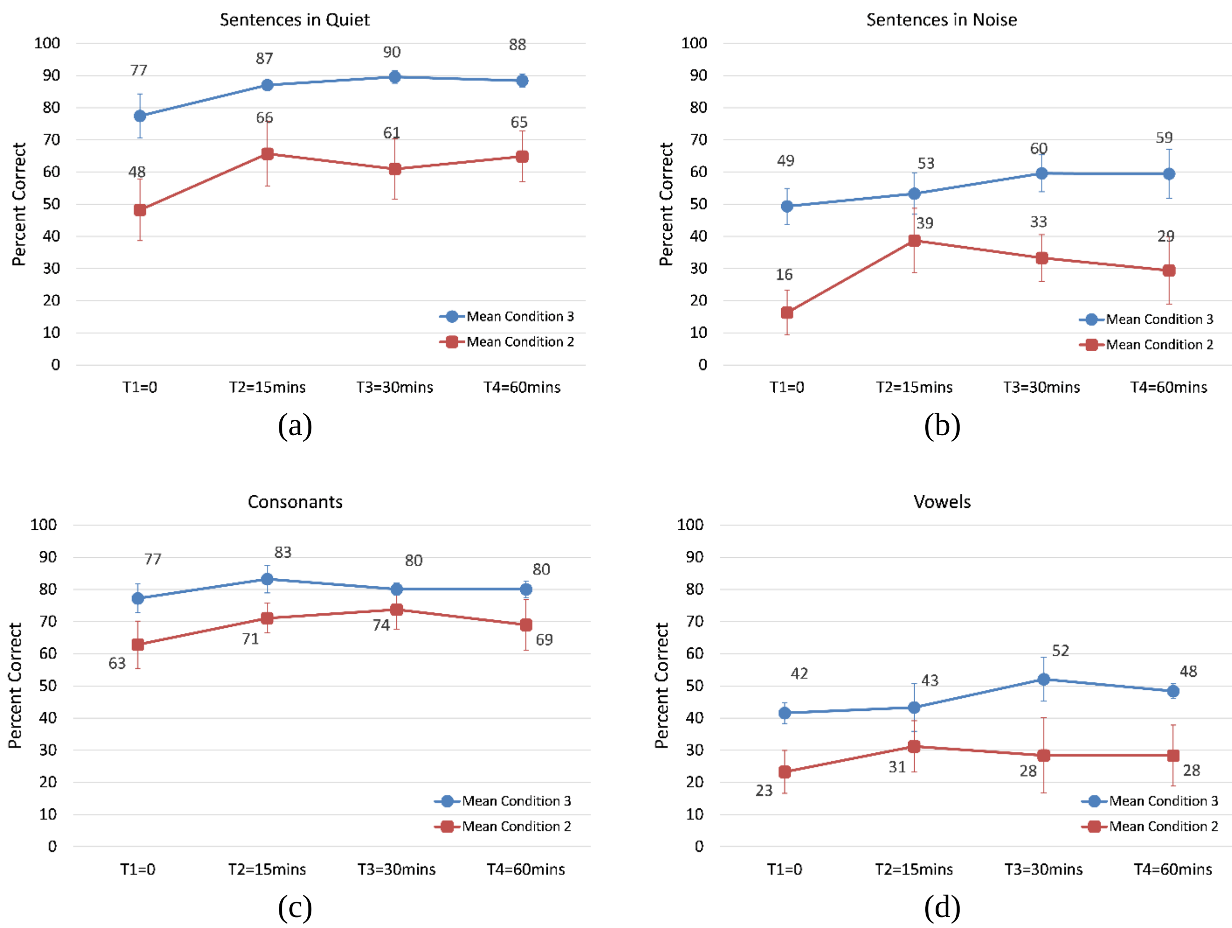


Figure 4.14. Average test scores measured at four time intervals in study 2. Subjects were tested acutely (with minimal training) at the start of the test, and then progressively given audio/video (A/V) training sessions and tested at intervals 2, 3, and finally at 4. Condition 2 mimics default frequency allocation with true tonotopic place, Condition 3 depicts proposed custom frequency allocation scheme with true tonotopic place.

at incremental levels (this impression was confirmed by statistical analysis, details are given below).

A mixed-design (split-plot) ANOVA was performed to assess the effects of learning (test session) and mapping condition on speech recognition performance. Speech material (4 levels: consonants, vowels, sentences in quiet, and sentences in noise) and test sessions (4 levels: T1 (acute), T2, T3, and T4 (final test session)) were considered as within-subject factors and mapping condition was considered as between-subject factor. A significant effect of test session ($F[3,24]=12.810$, $p<0.001$) and speech material ($F[3,24]=66.110$, $p<0.001$) was observed. All interactions were insignificant: test session × condition ($F[3,24]=1.882$, $p=0.160$), speech material × condition ($F[3,24]=2.136$, $p=0.122$), test session × speech material ($F[9, 72]=0.793$, $p=0.624$), test session × speech material x condition ($F[9,72]=0.821$, $p=0.599$). Pairwise comparisons between the test sessions indicated that acute scores (first test session, T1) were significantly different from all subsequent test sessions ($p<0.015$). However, there were no statistically significant differences among the remaining test sessions, indicating that any learning occurred just after the first training session. Furthermore, pairwise comparisons between each speech material revealed that performance levels measured with consonants and sentences in quiet, as well as performance levels measured with vowels and sentences in noise were statistically similar ($p \leq 0.001$). Other pairs were significantly different. A significant effect of the frequency mapping condition was observed ($F[1, 8]=7.478$, $p=0.026$). Mean performance level on all speech material was 46.57% and 66.82% with conditions 2 and 3, respectively. Overall, there was a mean improvement of 20.31 percent points with the custom frequency maps. This improvement was statistically significant.

A series of separate mixed design ANOVAs were conducted for each test material to assess the effect of frequency mapping condition and training on speech recognition. Test sessions were considered as within-subject factors and mapping conditions were considered as between-subject factor. The results are summarized in Table 4.7. With the exception of vowel identification, a significant effect of test session was observed for the other three test materials, indicating that training improved the understanding ability of consonants, sentences in quiet, and sentences in noise; however, vowel identification was relatively unchanged, at least for the training period administered. This trend is also observed in Figure 4.14 (d) which shows that acute and post-test vowel identification performance did not significantly change over time. The frequency mapping condition had a significant effect only on the sentence recognition performance, both in quiet and in noise. The interactions between test session and mapping condition were not significant for any tested speech material, indicating that irrespective of frequency mapping condition, the training improved listeners' performance. The post-training mean speech recognition scores as a function of test material for the both mapping conditions are shown in Figure 4.14. A summary of the pairwise Bonferroni comparisons between conditions 2 and 3 at every test session with the four tested types of speech material is given in Table 4.8. Results that were significantly different are marked with asterisk. The scores, as noted above, indicate that by the end of the adaptation period, the improvements with the custom frequency maps were only statistically significant for sentence recognition in quiet and in noise.

Figure 4.16 and Figure 4.17 provide another perspective towards data visualization. They demonstrate average speech recognition performance as a function of electrode insertion depth. It is interesting to see that for insertion angle of ~370° (i.e., location of most apical electrode at

Table 4.7. Summary of mixed design ANOVA for study 2. Analysis is presented for performance difference over 4 test sessions with two test conditions.

	Sentences in Quiet	Sentences in Noise	Consonants	Vowels
Test Session	F[3, 24] = 11.197 p<0.001	F[3, 24] = 6.866 p=0.002	F[3, 24]= 3.156 p=0.043	F[3, 24] = 0.901 p=0.455
Condition	F[1, 8] = 7.672 p=0.024	F[1, 8] = 7.312 p=0.027	F[1, 8] = 2.823 p=0.131	F[1, 8] = 4.910 p=0.058
Interactions	F[3, 24] = 0.947 p=0.434	F[3, 24] = 2.711 p=0.067	F[3, 24]= 0.841 p=0.485	F[3, 24] = 0.471 p=0.705

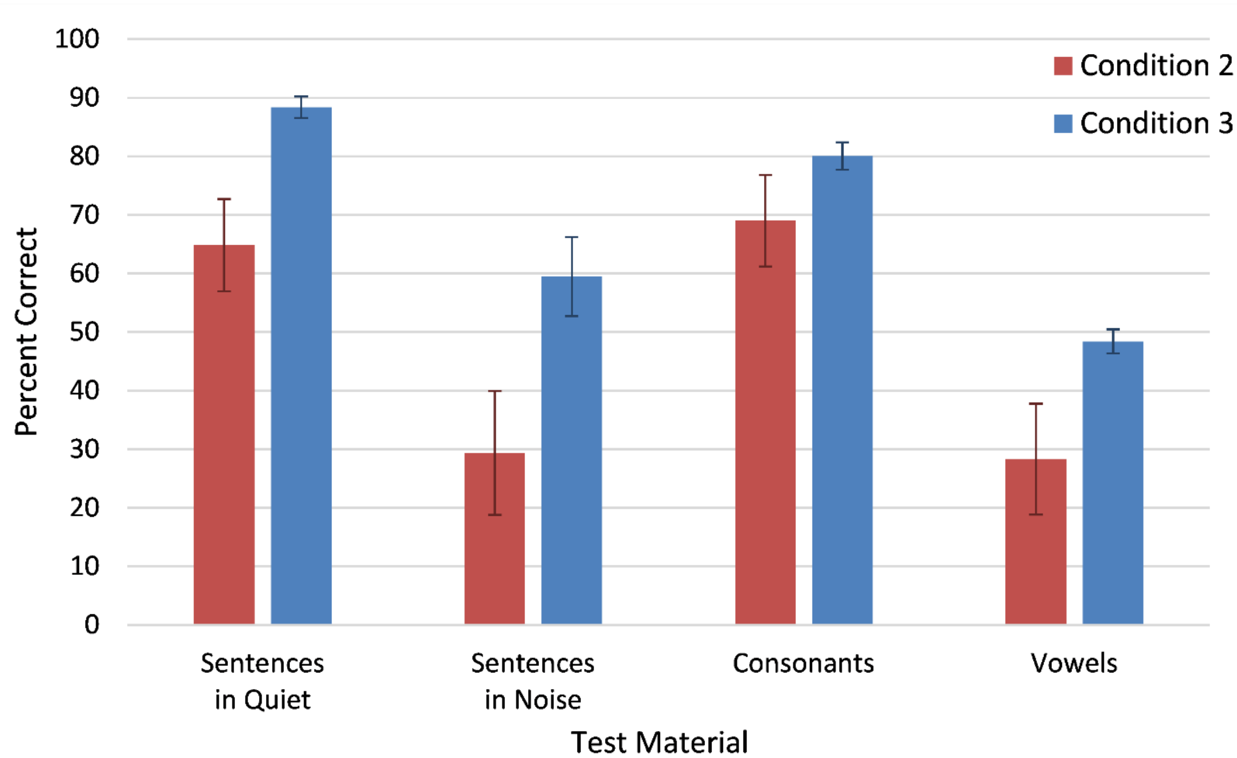


Figure 4.15. Post-training test scores for conditions 2 and 3 for study 2.

Table 4.8. Bonferroni pairwise comparisons between condition 2 and condition 3 at every test session with different test material. Statistical significant results are marked with an asterisk.

Cond2-Cond3 pairwise Bonferroni comparisons	Sentences in quiet	Sentences in noise	Consonants	Vowels
Test Session 1	**p*=0.033	**p*=0.005	*p*=0.124	**p*=0.036
Test Session 2	*p*=0.071	*p*=0.244	*p*=0.077	*p*=0.277
Test Session 3	**p*=0.017	**p*=0.018	*p*=0.346	*p*=0.110
Test Session 4	**p*=0.020	**p*=0.043	*p*=0.213	*p*=0.072

733 Hz), although the speech recognition with the clinical frequency map was 10 percentage points lower than the custom maps prior to the training, the scores converged to the same level of performance after the training. For shallower insertions, although there was an improvement with both maps with training, scores with the custom maps were significantly better than the clinical maps.

Discussion

Acute studies with vocoder processed speech may grossly underestimate the effect of learning and perceptual accommodation to the distorted spectral cues. This study aimed at investigating the effects of learning on speech recognition performance and assess if trends observed in acute experiments (Study 1) would mirror if listeners were provided with an auditory training to the processed sound. In this study, normal hearing listeners were tested at four test sessions, acutely and after progressive auditory training. In general, the performance improved for both frequency

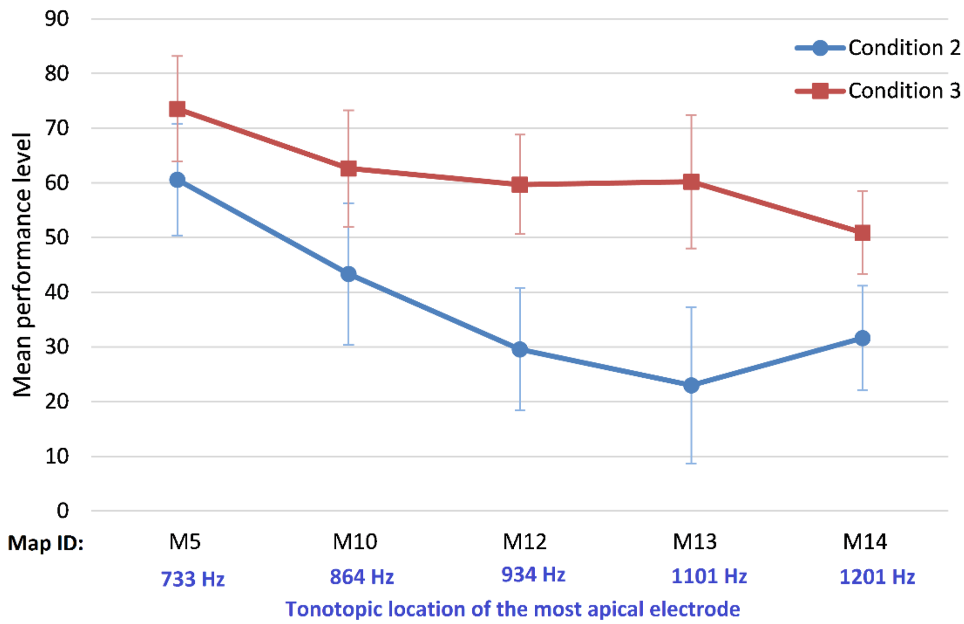


Figure 4.16. Pre-training test scores for conditions 2 and 3 as a function of electrode insertion depth.

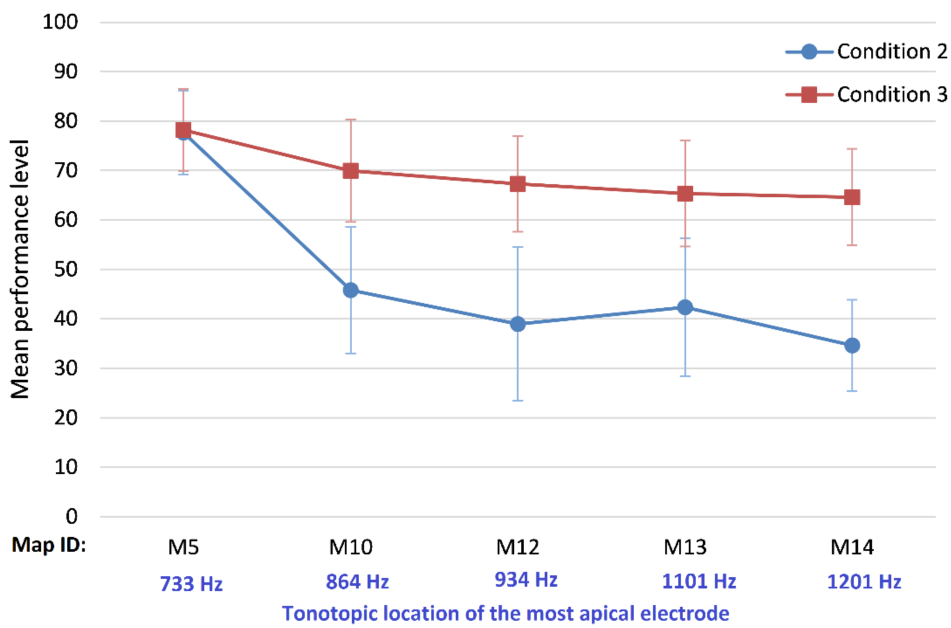


Figure 4.17. Post-training test scores for conditions 2 and 3 as a function of electrode insertion depth.

mapping schemes, suggesting that listeners show at least partial accommodation to vocoder processed speech, irrespective of the mapping condition. On all measures of speech, the custom frequency allocation scheme performed better than the default clinical frequency mapping scheme. Statistically significant improvements were observed for sentence recognition in quiet and in noise. Consonant and vowel identification remained roughly at the same level pre and post training for both the maps.

In this study, we were particularly interested in exploring the differences in asymptotic level of performances and the time it would take to accommodate to clinical and custom frequency maps, at least over a period of 3 -4 hours. The 4 hour test window is most likely too short time to observe full accommodation. It is unclear if listeners will continue to improve with long term training and if complete adaptation with either of the frequency mapping schemes would ever be possible with these maps. Speech recognition performance with either of the frequency allocation schemes was still significantly lower than the ideal frequency/electrode placement (Study 1, Condition 1). Although sentence recognition in quiet with condition 3 is close to the ideal scores, performance with consonant, vowel and sentence recognition in noise was far from ideal. It is difficult to conclude, to what extent, if any, listeners will continue to show improvement on difficult speech recognition tasks with training.

Clinical Implications: The data from this study suggest that shallow insertions that cause extreme frequency-place mismatch result in spectral distortions in the perceived sound that are difficult to overcome with training, at least within a 4-hour tested window using simulations with normal hearing listeners. However, a balanced frequency map, as compared to an extremely compressed/distorted map, may help with speech recognition, at least acutely. However, extreme care must be taken to extrapolate these results to cochlear implant listeners. Pre-lingual CI recipients, who receive the implant later in their lives (i.e., post critical language acquisition period) may not have an equivalent normal tonotopic representation in the auditory cortex (or may not have any map to start with). For such a group of recipients (who are starting from a blank slate), compressed maps may be a better choice as they may create a higher-level tonotopic organization that is matched with the processor assigned frequencies, and provide a full acoustic range. Furthermore, some experienced cochlear implant listeners may find the compressed frequency maps more intelligible due to re-organization of the higher-level tonotopic representation with experience. If complete adaptation to the clinical (compressed) frequency maps is observed in an implant user, it may be better to stay with the clinically assigned map. On the other hand, implant users, particularly new recipients, with moderate to extreme frequency-place mismatch, who demonstrate difficulty in adapting to the compressed frequency maps, the proposed solution may be a better mapping strategy, at least for the short-run. Evidence from a recent study by Svirsky *et al.* [100] suggests that gradual adaptation with intermediate frequency maps may help improve implant outcomes as compared to fitting with an extremely mismatched maps.

4.4.3 Study 3: Evaluation with cochlear implant users

The simulation experiments reported in Section 4.4.1 and Section 4.4.2 investigated the effects of spectral distortions caused by different frequency allocation schemes on speech perception. It was generally observed that the custom frequency allocation scheme, which aims to reach a balance between frequency matching, frequency compression, and truncation, could potentially yield a better speech recognition performance relative to the clinically-used frequency maps, at least acutely with normal hearing listeners. In the following study, we evaluated the custom frequency assignment strategy with users of cochlear implants.

Subjects

Nine adult post-lingually deafened cochlear implant recipients were recruited for this study. Three out of nine subjects were not motivated to try the experimental map for extended use due to listening difficulties with the experimental map, and thus were dropped out of the study on the first day. Of the remaining six, one subject dropped-out after one month. In total five subjects continued their participation for the entire length of the study, three months or beyond. At the three months testing session, subject V4 was only available for a few conditions. All subjects were native speakers of American English and used devices from Cochlear Corporation or Med-El. All listeners used their devices routinely and had a minimum of 1 year of experience with their implants. Further biographical data for the subjects is presented in Table 4.9.

Stimuli

Speech recognition was assessed using eight sets of speech materials, including six open-set recognition tasks and two closed set identification tasks. Open-set tasks included administration of consonant-nucleus-consonant (CNC) monosyllabic words [114], CNC phonemes, AzBio [115] sentences in quiet, AzBio sentences in 10-talker babble noise at 10 dB and 5 dB SNR levels, and IEEE sentences in quiet [111]. The closed-set tasks included multi-talker vowel and consonant identification.

In the AzBio corpus, sentences are spoken by multiple male and female talkers using a conversational style than a deliberate speaking style, and they have fewer contextual cues than other tests of sentence recognition. There are 15 lists of AzBio sentences, with 20 sentences per

Table 4.9. Details of the study participants.

ID	Age, years	CI use, years	Implant type	Processor	Number of electrodes	SG frequency of the most apical electrode (Hz)	Sound processing strategy	Stimulation Rate, Hz	Contralateral ear
V1	51	2.5	Med-El Concert Standard	Opus2, RONDO	9	257	FSP	2000	Normal (SSD)
V2	72	1.8	Med-El Concert Flex28	Opus 2, RONDO	11	434	FSP	1186	Hearing-aid
V3	88	1.8	Cochlear Nucleus Freedom CI24RE	N5	22	733	ACE	500	Hearing-aid
V4	54	4.7	Cochlear CI512	N5	22	1100	ACE	900	Hearing-aid
V5	51	1.3	Cochlear Nucleus CI24RE (CA)	N6	22	1461	ACE	900	No hearing

list and 4 talkers (2 male and 2 female). Each sentence ranges from 4 – 12 words and all lists are equated (i.e., equal intelligibility across lists) [116]. One list (20 sentences) per test condition was administered. AzBio sentences are recommended by the new minimum speech test battery (MSTB) as an assessment material for adult CI recipients [117]. IEEE sentences, on the other hand, are phonetically balanced and are relatively more difficult to understand than the AzBio sentences.

Recorded IEEE sentences were presented in both male and female voice, 2 lists per condition (1 list in male, and 1 in female voice).

CNC word test consists of lists of monosyllabic words with equal (and same as English language) phonemic distribution across the lists [114]. There are ten lists, with 50 words per list, and each word comprises of three phonemes. One list (50 words) per condition was administered, and listeners were assessed for both phoneme and word recognition.

Vowel recognition was assessed using a 12-alternative identification paradigm for 12 phonemes, including 10 monophthongs and 2 diphthongs, presented in /h/-vowel-/d/ context (had, hod, hawed, head, hayed, heard, hid, heed, hoed, hood, hud, who'd). Consonant stimuli comprised of 20 medial consonants presented in /a/-consonant-/a/ context. Both vowels and consonants comprised of productions from 5 male and 5 female talkers and were acquired from [113]. Closed-set tests were conducted independently for male and female speakers, and all tokens were randomly drawn from different talkers. In addition to the speech recognition tasks, spectral resolution was assessed via spectral modulation detection (SMD) task. SMD threshold for an implanted ear is reported to be highly correlated with speech understanding, and is thus a non-linguistic, psychoacoustic index for speech recognition performance. The quick SMD task was drawn from Gifford *et al.* [118] and included a 3-interval, forced-choice procedure to identify spectrally modulated noise from flat-spectrum noises. Spectral modulation was achieved using logarithmically spaced, sinusoidally-modulated (125 Hz – 5600 Hz) broadband carrier stimulus. Five modulation depths (10, 11, 13, 14, and 16 dB) were tested at frequencies of 0.5 and 1.0 cycles/octave and with (6 trials for each modulation depth). A total of 60 trials were administered and each trial was scored correct or incorrect based on the user response. The chance level of performance was 1/3.

Hearing quality was assessed using the Speech, Spatial, and Qualities of Hearing (SSQ) questionnaire [119]. SSQ comprises of three parts, namely speech hearing (14 questions), spatial hearing (17 questions), and qualities of hearing (18 questions). It employs a visual analog scale (ranging from 0 to 10) to register responses and aims at measuring self-reported auditory disability across a wide variety of domains that reflect the reality of hearing in the everyday listening environments. These domains include speech in variety of competing contexts, spatial hearing associated with direction, distance and movement components, and the overall quality of speech such as clarity, fullness, naturalness, and ease of listening. Higher scores represent better speech understanding, spatial hearing, and sound quality.

Signal processing programming strategy

Frequency allocation scheme for each subject was derived from his/her imaging data. DVF curves for each subject were fed as an input to the frequency-place optimization algorithm (discussed in section 4.3). Some subjects (V1 and V2) had previously participated in electrode deactivation study [106, 107]. For the above two and all other subjects, only those electrodes which were active in the clinical map were used to create the custom frequency maps. Therefore, no electrodes were switched on or off, even if it contradicted with the programming principles of the algorithm[7].This was done to study the effect of only one factor (i.e., changing frequency allocation table) in this experiment.

Procedure/Method

Experimental (custom) frequency maps were created in advance of the testing day. Subjects were tested with a battery of hearing and speech recognition tests, in up to two listening modes: electric-only (implant only), and electric + acoustic (implant and hearing aid). An ABA (test/re-test: Clinical/Proposed/Clinical) study model, as described in the following section, was employed. First, baseline measurements were taken with the default clinical map, the map which participants walked in with (and were likely accommodated to). For bimodal subjects who used

---

[7] Electrode activation/de-activation schemes could potentially interfere with the algorithm design. Ideally, better spectral resolution can be delivered by increasing the number of perceptually discriminable stimulation sites (electrodes). (Although, scientific evidence suggests that CI users are not able to utilize greater than 10 electrodes.) Electrode deactivation schemes, e.g., by Noble *et al.* [106, 107], aim to reduce the channel interference caused by the electrodes that reside very close to each other spatially, and are likely to stimulate the same group of nerve fibers. With such electrode deactivation scheme, they demonstrated an improvement in performance by many CI users, despite reducing the number of active electrodes. The present frequency-allocation scheme derives its sound analysis filter banks from the number of electrodes that are available. Electrode-deactivation strategy, thus, could interfere with the algorithm principles in the following ways. In general, low frequencies are represented by multiple narrow-band analysis filters, while wide band filters are used to represent higher frequencies. This aids in better representation of lower frequencies that are generally more critical for speech understanding and also follows logarithmically spaced tonotopic map of the spiral ganglion. For example, frequency maps used in the devices by Cochlear Corporation use at least 12 out 22 electrodes to represent frequency range of 288 – 1600 Hz. The remaining filters are used to represent frequencies up to 7938 Hz. Such a scheme provides better representation of lower frequencies and produces better speech recognition results than linear spacing. This frequency allocation trend is also consistent with other implant manufacturers. Now imagine, a scheme that switches of electrodes on the basis of spatial location of electrodes. If a handful of electrodes are deactivated (pruned) in the lower frequencies, this may result in a significant reduction in the possibility to represent low frequencies with narrow filter-banks with the proposed custom frequency allocation strategy. In such cases, representation of low frequencies with broad filters can actually be detrimental. Care must be taken when investigating similar frequency-allocation schemes, especially for the confounding factors that are known to be inversely related to each other.

hearing-aid or had good hearing in the contralateral ear, the measurements were repeated in the best-aided condition. Following baseline testing, participants' CI processor was programmed with the custom frequency map. Other sound processing parameters, such as timing parameters, current levels and environmental settings, were not changed and were kept similar to the original clinical map. After fitting with custom maps, the subjects were tested with the same battery of speech material in CI-only and under best-aided conditions, when possible. The only auditory training provided to the listeners before the acute testing was reading the rainbow passage [120] that took about 2 – 5 minutes. The experimental session on the first day took about 4 hours in total (baseline + acute testing). After the testing session, the subjects were given the experimental map to take home and were encouraged to use it in their normal daily lives, as they would normally use their clinical processor. Their processors were configured with only one map, with volume and sensitivity settings equivalent to their original clinical maps. This was done to ensure that they only used the experimental map throughout the length of the study. Five subjects returned back for a follow-up session a week after the first programming session, during which they were retested with the experimental map, which they had been using in the past one week. Finally, five participants returned back after three months for a semi-chronic performance evaluation with the experimental map. During the final testing session, participants were first tested with the custom frequency map in electric-alone and best-aided conditions, after which the participants' processors were reprogrammed with their original clinical map (which they have been using before the start of the study). The performance was quantified on several measures, including adaptation trend observed with the experimental map acutely, 1-week, and 3-months post-activation, as well as quantitative comparison with the clinical map. Finally, to measure the performance differences qualitatively, participants completed SSQ questionnaire before the start of the study and at the end of the 3-month period.

Results

Figure 4.18 - Figure 4.22 show mean scores for different test materials with clinical and custom frequency maps as a function of the test session. The first and last data points in each chart represent scores from the clinical map pre- and post- fitting of the custom frequency map. Scores with custom frequency maps were collected acutely on day 1 (without experience), at 1 week, and after 3 months of use. Acute performance from the custom maps was considerably lower than the clinical map on all measures, except spectral modulation detection (SMD). This is expected, because when changes to a map are made, quantitative and qualitative hearing scores generally tend to favor the original map [121]. However, participants displayed at least partial

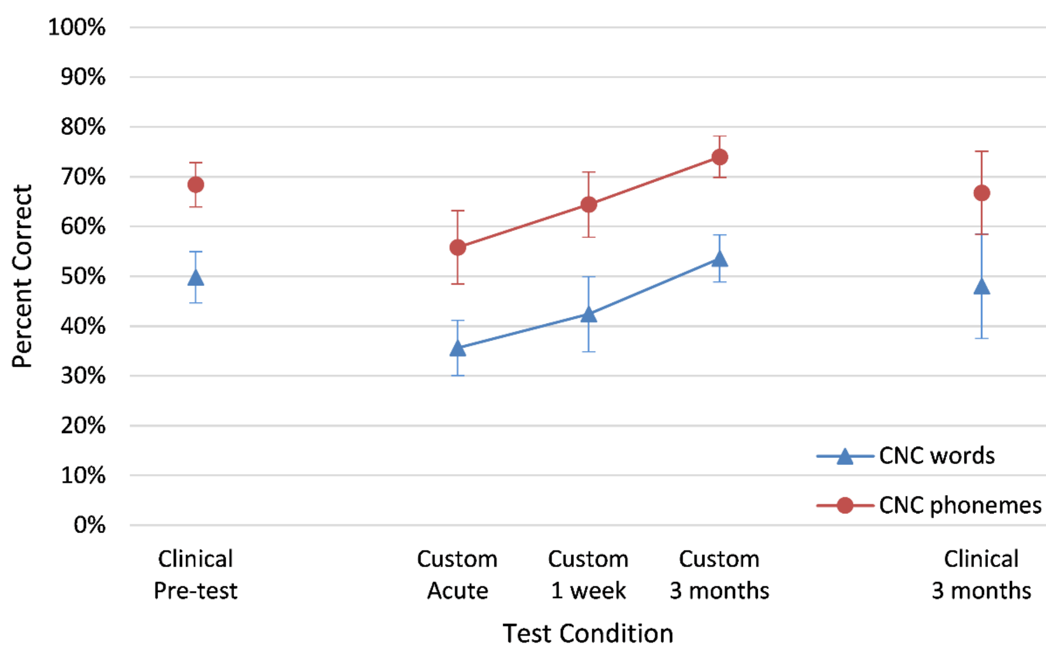


Figure 4.18. Mean CNC words and phonemes recognition scores of 5 CI users from Study 3. Error bars represent standard errors of means.

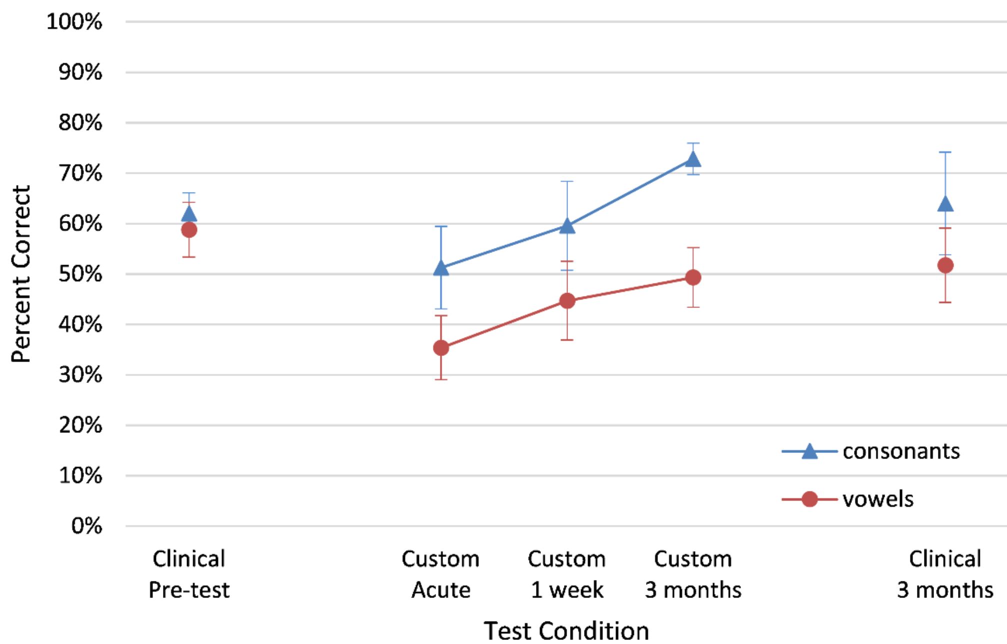


Figure 4.19. Mean consonants and vowels identification scores of 5 CI users from Study 3. Error bars represent standard errors of means.

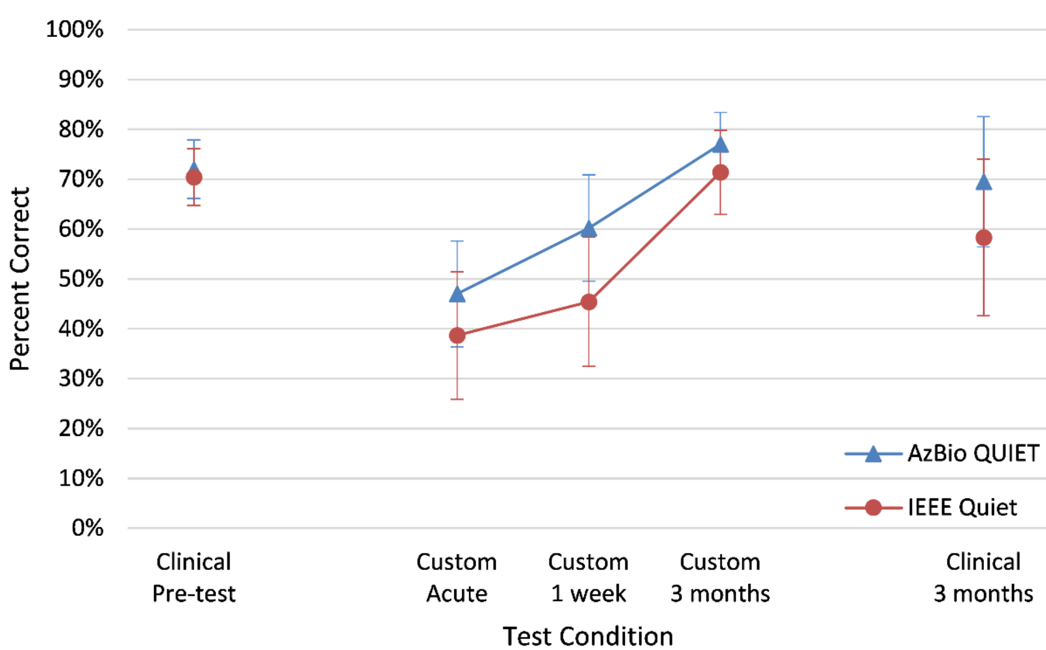


Figure 4.20. Mean sentence recognition scores in quiet of 5 CI users from Study 3. Error bars represent standard errors of means.

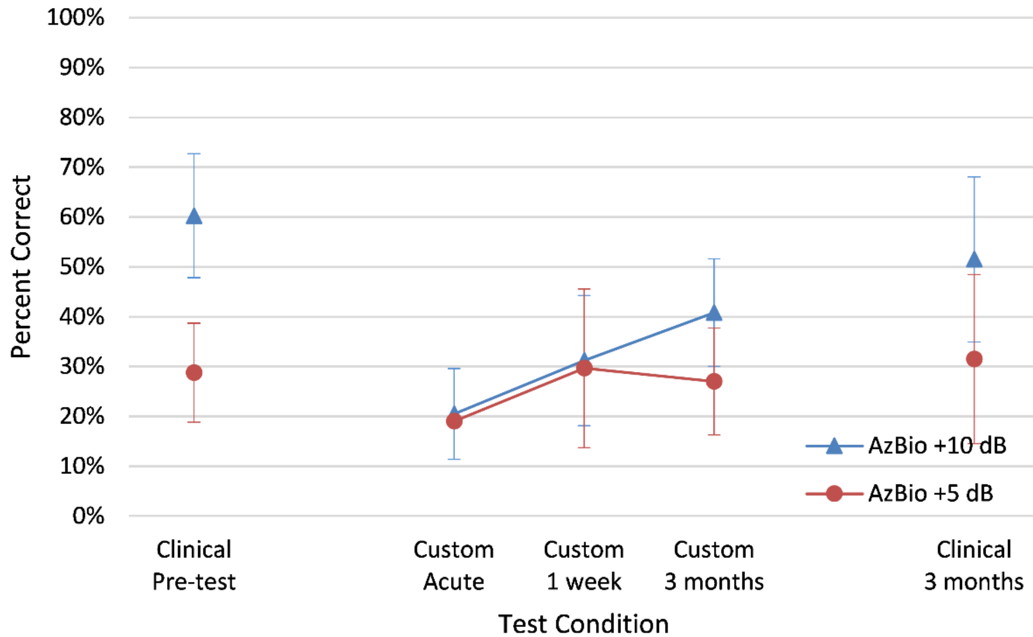


Figure 4.21. Mean sentence recognition scores in noise of 5 CI users from Study 3. Error bars represent standard errors of means.

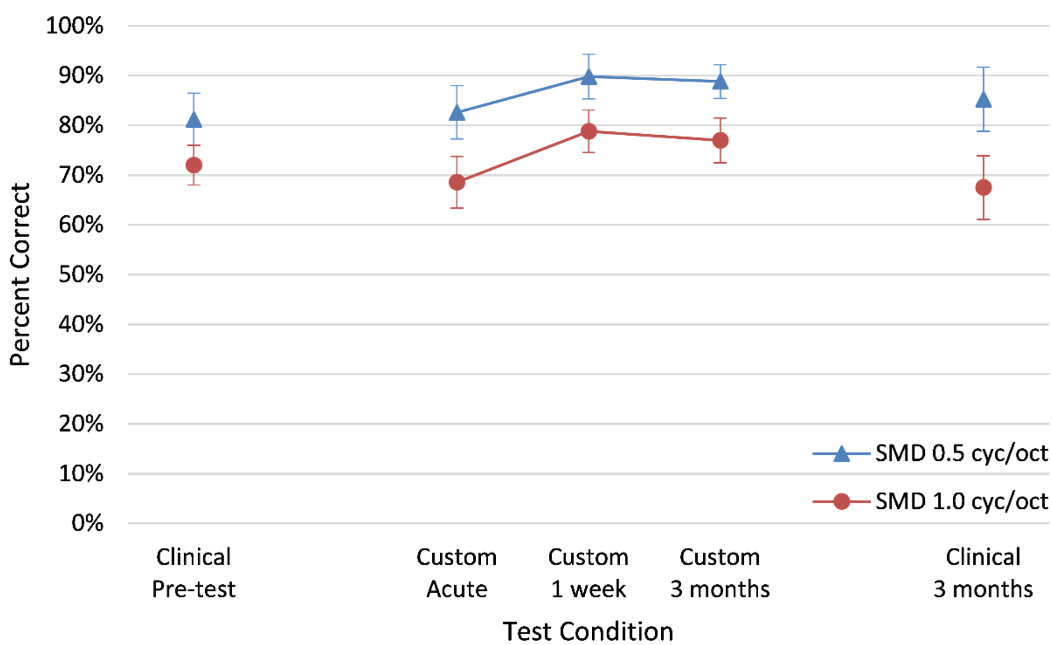


Figure 4.22. Mean spectral modulation detection (SMD) scores of 5 CI users from Study 3. Error bars represent standard errors of means.

accommodation to the custom maps during the three month-time window. Statistical analysis of the data is provided as follows.

Repeated measures multi-variate ANOVA (MANOVA) was performed with frequency mapping conditions (clinical vs custom), speech materials (10 test measures), and test session (2 levels: pre and post) as the main analysis factors. No overall statistically significant effect of mapping condition (p = 0.152) or test session (p = 0.061) was observed. There was a significant effect of speech material (p <0.001). Two-way interaction between speech material and test session, as well as 3-way interactions among all three factors were statistically significant. The other two-way interactions were non-significant. The statistics are summarized in Table 4.10. Please note that for the above analysis only two test sessions were considered. For clinical map, this equates to tests conducted with the clinical map prior to the start of study and after 3 months. For custom map, acute and 3-months post-activation data points were used. The data from subject V4 were excluded from analysis due to missing data points.

Post-hoc Bonferroni tests showed that phoneme (p=0.020) and consonant recognition (p=0.007) scores changed with the test session. In addition, pairwise comparisons (t-tests) revealed that performance with the clinical maps at day 1 and at 3 months (re-test condition) was statistically equivalent (p = 0.868)[8]. However, test session had a significant effect on overall performance with custom mapping (p = 0.003). This could be explained by the effect of learning/adaptation with the custom maps over three months of extended use of the experimental program. The analysis on the effects of learning and test session is reported later in this section.

---

[8] This was also confirmed with a separate 2-way repeated measures ANOVA, with test session and speech material as the main analysis factors. Analysis revealed no statistically significant difference in scores between the test sessions (F[1, 3] = 0.033, p = 0.868]. See Table 4.9

Table 4.10. MANOVA statistics for study 3 with cochlear implant users. Test Sessions (pre vs post), frequency mapping conditions (clinical vs custom), and speech materials (10 levels) were considered main analysis factors.

Factor	F statistic	Significance
Test Sessions (pre and post)	$F[1, 3] = 8.597$	$p = 0.061$
Mapping Conditions (clinical and custom)	$F[1, 3] = 3.656$	$p = 0.152$
Speech Materials (10 levels)	$F[9,27] = 12.472$	$p < 0.001$
Mapping Condition × Speech Material	$F[9, 27] = 1.985$	$p = 0.082$
Mapping Condition × Test Session	$F[1, 3] = 4.178$	$p = 0.134$
Speech Material × Test Session	$F[9, 27] = 2.362$	$p = 0.041$
Mapping Condition × Speech Material × Test Session	$F[9, 27]= 2.337$	$p = 0.043$

A series of separate 2-way repeated measures ANOVAs were conducted to compare the effects of mapping conditions across test sessions. Mapping conditions and speech material were considered as the main analysis factors. First, acute performance with the custom frequency maps was compared against the clinical maps. Analysis revealed a significant effect of the mapping condition ($F[1,4] = 15.743$, $p = 0.017$) as well as speech material ($F[9,36] = 17.013$, $p < 0.001$). The interaction between the mapping conditions and speech material was statistically significant ($F[9,36] = 4.020$, $p = 0.001$). This significant ANOVA was followed with Fisher's Least Significance Difference (LSD) paired *t*-test, to investigate the effects of mapping conditions on each tested speech material. Pairwise comparisons revealed that half of the speech material (5 out of 10) was relatively unaffected by the mapping condition. The performance levels with speech materials that were statistically different in either of the mapping condition were vowels (*p*=0.008), AzBio sentences in quiet (*p*=0.016), AzBio sentence in 10dB SNR (*p*=0.016), AzBio sentence in 5 dB SNR (*p*=0.034), and IEEE sentences in quiet (*p*=0.046).

Next, the effect of experience with the experimental program was analyzed. Performance levels across test sessions (acute, 1 week, and 3 months) on all speech measures were evaluated using a 2-way repeated measures ANOVA. Results showed a significant effect of test session on performance ($F[2,8] = 18.950$, $p = 0.001$), as well as a significant effect of speech material ($F[9,36] = 19.064$, $p < 0.001$). The interactions between test session and speech material were also significant ($F[18,72] = 2.397$, $p = 0.005$). A post-hoc Bonferroni test showed that performance difference with learning was only significant after 3 months of experience with the experimental map ($p = 0.006$). Although performance improved progressively with each test session (Session 1, acute: 43.97±5.48), (Session 2, 1 week: 53.43±7.41), (Session 3, 3-months: 62.50±4.24); sessions 1 and 2 (p =0.086), as well as sessions 2 and 3 (p=0.182) were not statistically different from each other. Interestingly, performance levels at 1 week with the custom maps were not statistically different from their clinical baseline scores ($F[1,4] = 2.952$, $p = 0.161$). However, there were statistically significant interactions between mapping conditions and speech material ($F[9, 36] = 2.945$, $p = 0.010$). A paired *t*-test showed that performance on neither of the speech recognition tasks to be significantly different with the custom or the clinical maps. The interactions were most likely of statistical significance due to the individual differences in the test materials.

The final levels of performances at three months with the experimental (custom frequency) maps were compared against the baseline performances measured with the clinical maps on day 1 (pre-test). Estimated marginal means (averaged across all speech measures) for the two test conditions along with confidence intervals are given in Table 4.12. The two mapping conditions produced similar performances. A 2-way repeated measures ANOVA revealed no statistically significant difference between the two mapping conditions (F[1,4] = 0.298, *p* = 0.614). A significant effect of speech material was observed (F[9,36] = 20.106, *p* <0.001). Interactions between test session and speech material were not significant (F[9,36] = 0.122, *p* = 0.122).

Please note that, excluding subject V4 from analysis, for whom scores for consonants, vowels, IEEE Quiet and AzBio 5dB were extrapolated for the custom map at three months, the ANOVA revealed an overall better performance with the custom maps as compared to the clinical map (F[1,3] = 98.065, *p* = 0.002). Although mean scores with the custom maps were only slightly higher (63.58±5.29) than the clinical map (60.64±5.29), the improvement was statistically significant.

Table 4.11. Estimated marginal means for clinical maps on day 1 and after 3 months. Baseline scores from the clinical map (measured at Day 1) were compared against acute scores obtained from clinical map at 3 months, during which subjects used an experimental program. Data excludes scores from V4 who was not available for at 3 months for complete testing.

Mapping Condition	Mean	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
Clinical Baseline	60.64	5.29	43.81	77.47
Clinical acute 3 months	59.40	9.93	27.79	91.08

Table 4.12. Estimated marginal means for baseline clinical and custom maps. Baseline scores from the clinical map (measured at day 1) were compared against 3-months scores with the custom frequency allocation scheme. (Scores from subject V4 were extrapolated for four test materials with the custom maps.)

Mapping Condition	Mean	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
Clinical Baseline	61.46	4.18	49.86	73.06
Custom at 3 months	62.50	4.24	50.72	74.28

Table 4.13. Estimated marginal means for custom and clinical maps at 3-months. Scores from 3-months post-activation of the experimental maps were compared against acute scores obtained from clinical map at 3 months. (V4 excluded from analysis.)

Mapping Condition	Mean	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
Custom at 3 months	63.58	5.30	46.72	80.43
Clinical, acute 3 months	59.40	9.93	27.79	91.02

The performance levels measured at 3 months with the experimental (custom frequency) maps were also compared against the measurements taken with the clinical map acutely after 3 months. Estimated marginal means (averaged across all speech measures) for the two test conditions along with respective confidence intervals are given in Table 4.13. Although mean scores from the custom frequency maps were numerically slightly better that the clinical map (mean difference of 4.17%), a 2-way repeated measures ANOVA did not indicate mapping conditions to have any statistically significant effect on the performance levels ($F[1,3] = 0.397$, p

= 0.573). However, a significant effect of the speech material was observed ($F[9,27] = 8.695$, p <0.001). The interactions between the test session and speech material were not significant as well ($F[9,27] = 1.472$, $p = 0.208$).

Figure 4.23 shows mean speech recognition performance of five CI subjects with clinical and custom frequency maps. Average across all ten speech measures was considered as a mean score for each mapping/testing condition. Scores from the clinical map were acquired at the start of the study (bar 1), and at 3 months (3^{rd} bar). In order to investigate if individual subjects performed significantly different with any of the two mapping conditions, a repeated measures ANOVA was performed with subjects and mapping conditions as the main analysis factors. Since data for subject V4 were not complete, only 4 subjects were considered for analysis. The performance levels with the clinical map at the start of the study, and with the custom map at the 3-months post-activation session were used as the main analysis factors. Results indicated that individual differences between the subjects were statistically significant ($F[3,27]=11.273$, $p<0.001$). Frequency mapping conditions did not have any statistically significant effect ($F[1,8]=1.984$, $p=0.193$) on performance. The interactions between subjects and maps were also not significant ($F[3,27]=0.043$, $p=0.988$).

Average qualitative scores obtained from the SSQ questionnaires are shown in Figure 4.24. SSQ questionnaires were filled at the start of the study with the clinical map and at the end of the study (3-months) with the custom maps. Overall, there was no significant perceptual differences between the two maps. Average scores from four subjects changed from 4.905 ± 2.14 (clinical) to 4.503 ± 2.26 with the custom map.

To this point, we have presented analysis of mean speech recognition performance across all subjects and test material. In the following section, performance levels of each subject on individual test material are provided separately to highlight the differences between the subjects and the test materials.

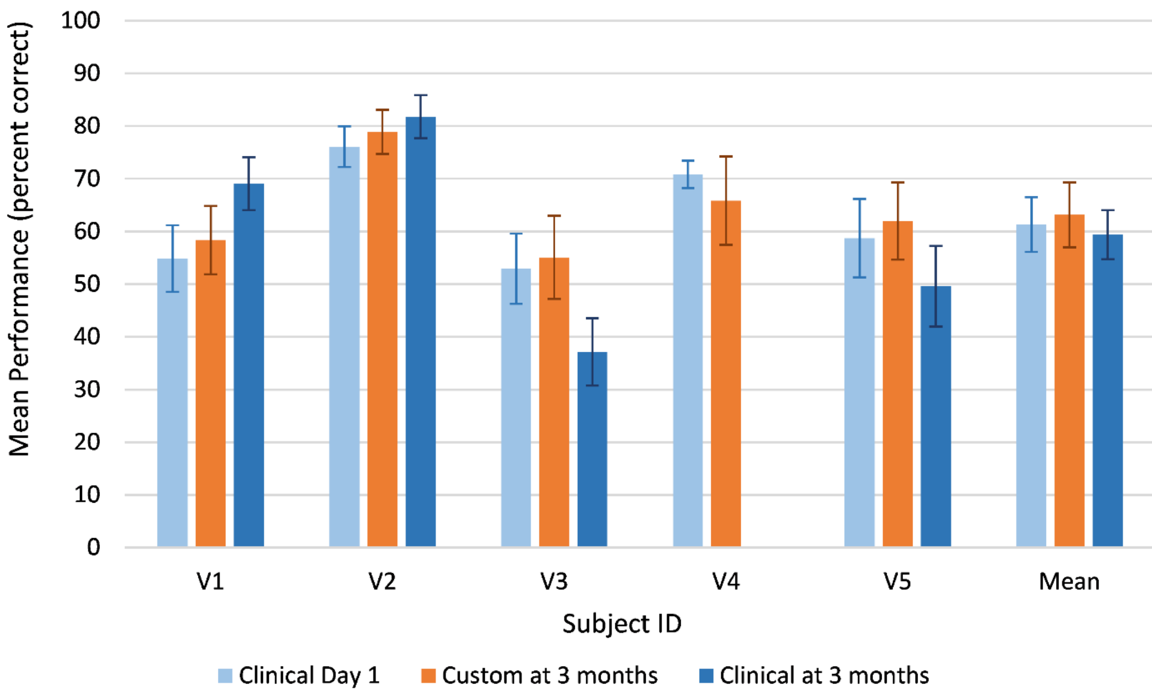


Figure 4.23. Mean performance of 5 CI subjects on 10 speech recognition tasks with clinical and custom frequency maps. Error bars represent standard errors of means.

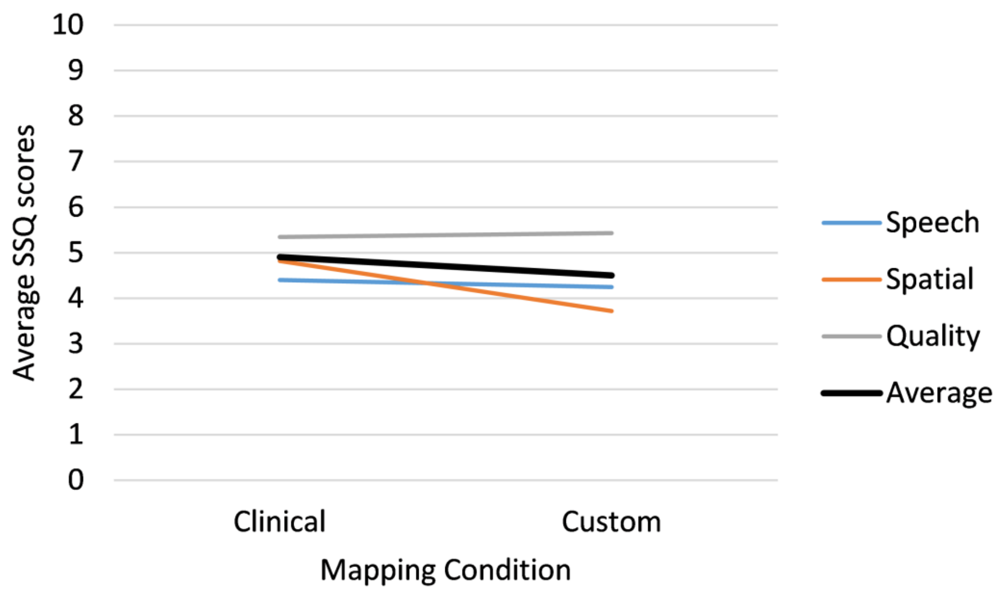


Figure 4.24. Mean SSQ scores of 4 CI subjects with clinical and custom frequency maps.

Subject V1:

The frequency characteristics of the clinical and custom frequency maps along with center frequencies of the image filters (curve minimas from the DVF curves) for subject V1 are shown in Table 4.14. Speech recognition scores with various test material are presented in Figure 4.25.

A 2-way repeated measures ANOVA was performed using the clinical and custom frequency maps at two testing sessions (pre and post), as the main analysis factors. The results indicated a significant effect of mapping condition (F[1, 9] = 5.166, p = 0.049), and test session (F[1, 9] = 10.250, p = 0.011). The interactions between mapping conditions and the test sessions

Table 4.14. Frequency allocation tables for subject V1. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were not active.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	100	168	237	257	125	250	375
2	-	-	-	410	-	-	-
3	237	334	431	627	375	625	875
4	-	-	-	797	-	-	-
5	431	570	710	1082	875	1082	1290
6	710	912	1115	1653	1290	1653	2016
7	1115	1411	1707	2394	2016	2394	2773
8	1707	2140	2574	3614	2773	3387	4000
9	2574	3211	3849	5418	4000	4737	5474
10	3849	4788	5728	6146	5474	6063	6651
11	5728	7114	8500	12542	6651	7295	7939
12	-	-	-	15182	-	-	-

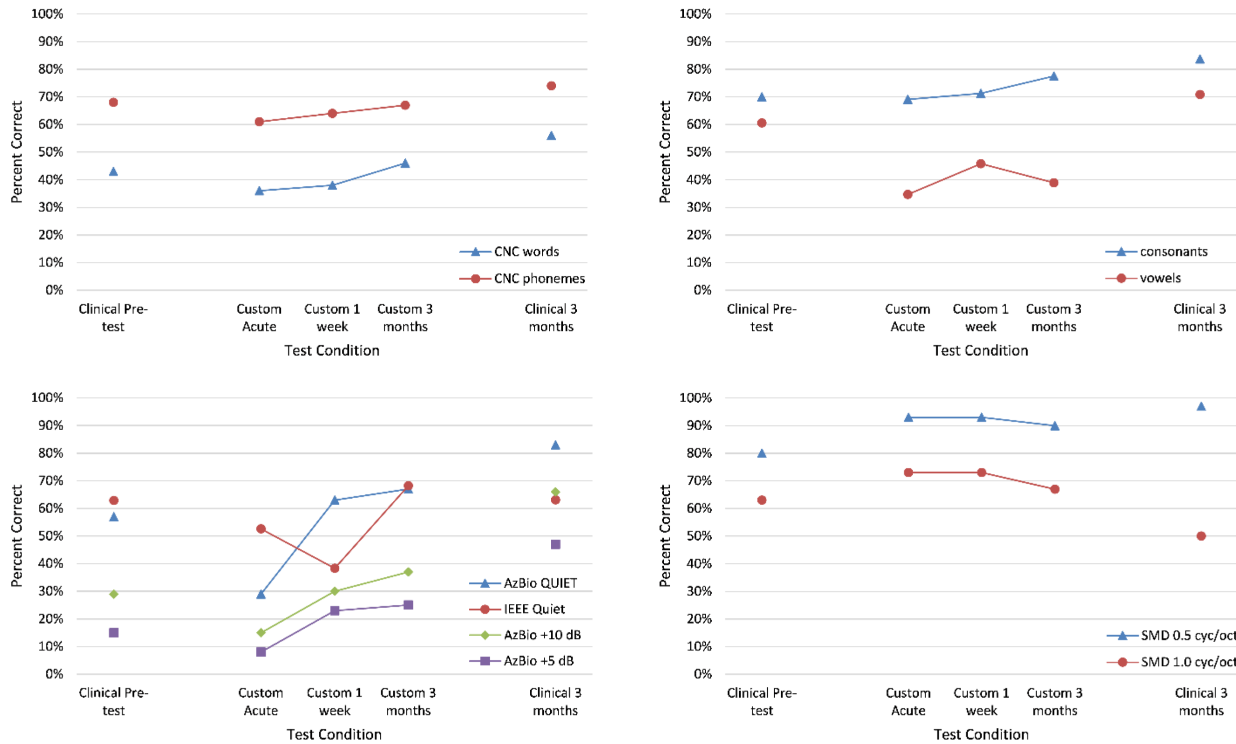


Figure 4.25. CI-only score profile of subject V1 on 10 speech recognition tasks with clinical and custom frequency maps.

were not significant (F[1, 9] = 0.636, p = 0.446). Pairwise t-tests were conducted between different pairs of test sessions and mapping conditions. The results are summarized in Table 4.15.

Remarks/Discussion: The trends in the above results are very interesting. The p-values of first three pairs indicate that there was no statistically significant difference between the two maps. By the end of the three months study period, the custom map was on average 3.5 percentage points better than the clinical map. However, when the participant was shifted back to his old clinical map, his scores with the clinical maps were significantly improved (14 percentage points, on average). Pairwise comparison of scores with the custom map and the clinical map at 3 months indicated borderline results (p=0.05).

Possibly, the custom map acted as an intermediate map, which may have boosted the performance with the clinical map. By the end of the study, the subject kept both custom and clinical frequency maps.

The subject was a single-side deafened (SSD) adult. His scores in best aided condition were at ceiling levels for all test materials with both the maps at all test sessions and with all speech material.

Table 4.15. Paired samples *t*-test for electric-only condition with subject V1.

Pair	Mean Difference	Std. Error Mean	*t*	*p*-value
Clinical_Session1 Custom_Session1	7.708	13.235	1.842	0.099
Clinical_Session1 Custom_Session2	0.901	11.658	0.244	0.812
Clinical_Session1 Custom_Session3	-3.518	9.549	-1.165	0.274
Clinical_Session1 Clinical_Session3	-14.219	14.925	-3.013	0.015
Clinical_Session3 Custom_Session3	10.701	4.725	2.265	0.050

Subject V2:

The frequency characteristics of the clinical and custom frequency maps along with center frequencies of the image filters (curve minimas from the DVF curves) for subject V2 are given in Table 4.16. Speech recognition scores with various test material are given in Figure 4.26. A 2-way repeated measures ANOVA was performed, with performance from the clinical and custom frequency maps at two testing sessions (pre and post) as the main analysis factors. Results indicated a significant effect of mapping condition ($F[1, 9] = 10.196$, $p = 0.011$) and test session ($F[1, 9] = 12.798$, $p = 0.006$). The interactions between mapping conditions and test sessions were statistically significant ($F[1, 9] = 9.715$, $p = 0.012$).

Table 4.16. Frequency allocation tables for subject V2. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were not active.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	100	168	208	410	250	375	500
2	-	-	-	569	-	-	-
3	208	280	352	749	500	625	750
4	352	448	545	899	750	875	1000
5	545	675	806	1180	1000	1187	1375
6	806	983	1160	1484	1375	1550	1725
7	1160	1402	1643	2205	1725	1975	2225
8	1643	1973	2303	3071	2225	2538	2850
9	2303	2756	3208	4546	2850	3275	3700
10	3208	3829	4450	5805	3700	4250	4800
11	4450	5302	6155	9173	4800	5500	6200
12	6155	7328	8500	13370	6200	7100	8000

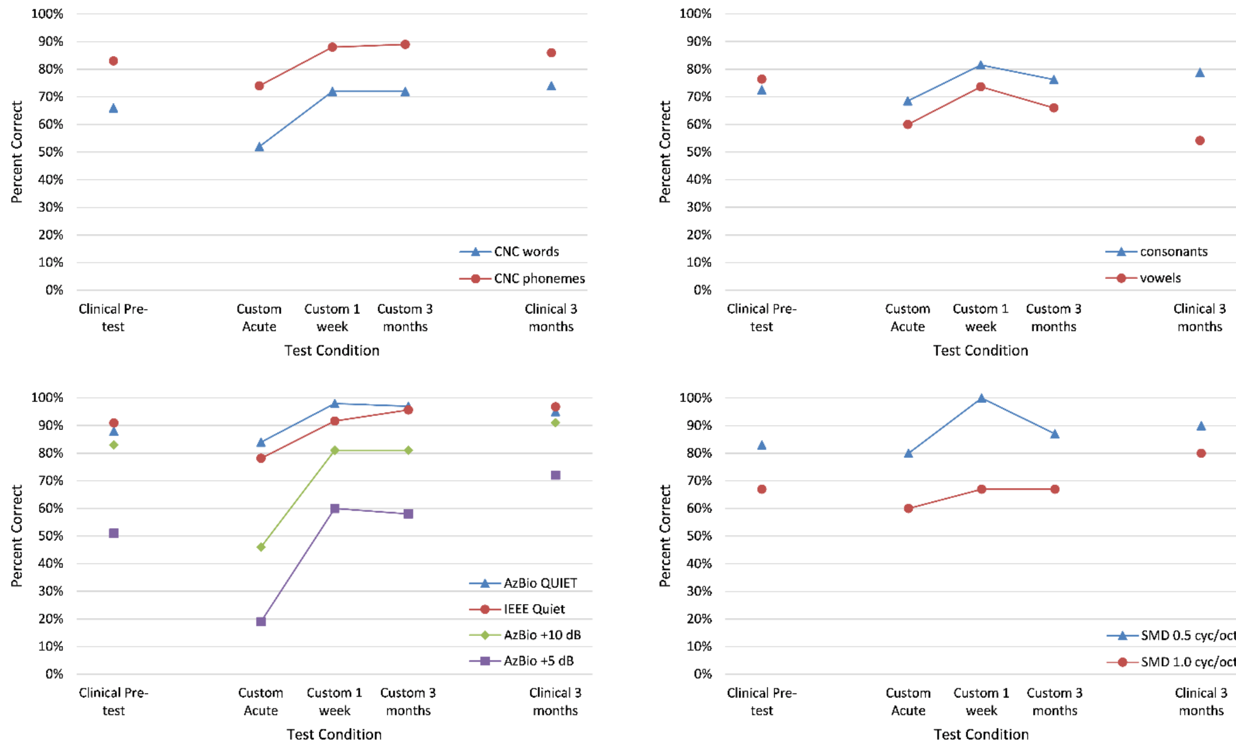


Figure 4.26. CI-only score profile of subject V2 on 10 speech recognition tasks with clinical and custom frequency maps.

Pairwise *t*-tests were conducted between different pairs of test sessions and mapping conditions. The results are summarized in Table 4.17. The results indicated that acute scores with the custom map were significantly lower than the clinical map (day 1) (p=0.005), but after 1 week of extended use, performance levels with the custom maps was significantly better than the clinical map (p=0.028). Interestingly, the performance at 3-months with both maps was not significantly different (p=0.151). Comparisons of performance levels observed with the clinical map before and after 3-months indicated that there was no significant difference between the two (p=0.136) maps.

Table 4.17. Paired samples *t*-test for electric-only condition with subject V1.

Pair	Mean Difference	Std. Error Mean	t	*p*-value
Clinical_Session1 Custom_Session1	13.919	3.734	3.728	0.005
Clinical_Session1 Custom_Session2	-5.197	1.985	-2.618	0.028
Clinical_Session1 Custom_Session3	-2.811	1.789	-1.571	0.151
Clinical_Session1 Clinical_Session3	-5.693	3.478	-1.637	0.136
Clinical_Session3 Custom_Session3	2.882	2.483	1.161	0.276

Bimodal (Electric+Acoustic) scores: Figure 4.27 shows the speech recognition performance in best aided conditions with an implant in one ear and a hearing-aid in the other of subject V2. Due to the lack of time, some conditions were not tested. The results indicated that subject's performance with custom maps was generally comparable (perhaps equivalent) to the clinical maps in bimodal condition. A 2-way repeated measures ANOVA revealed no statistically significant difference between the two maps by the end of the study (Clinical_Session3 and Custom_Session3) ($F[1, 7] = 0.048$, $p = 0.833$).

Remarks/Discussion: The above analysis for the electric-only condition indicates that the subject's performance dropped significantly with the custom map at the first testing session, but after 1-week of use, subject accommodated to the custom map and his average scores were better than his clinical map. Scores after 3 months,

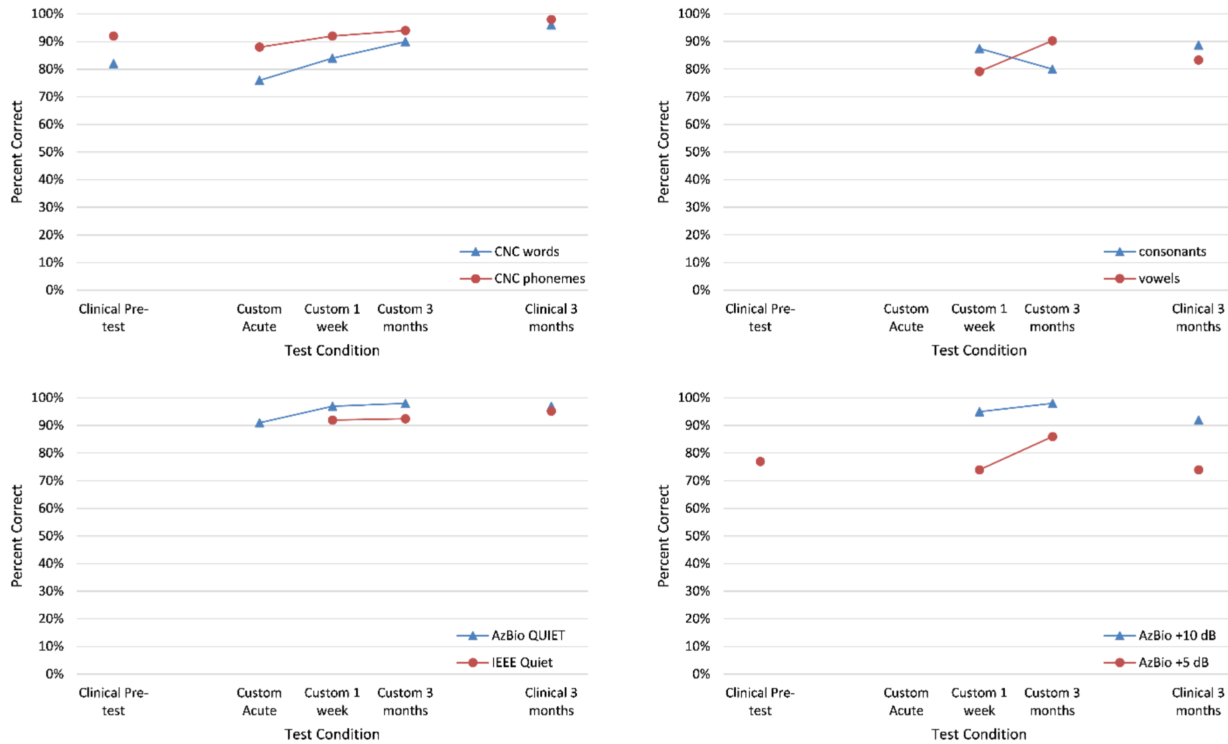


Figure 4.27. EAS score profile of subject V2 on 8 speech recognition tasks with clinical and custom frequency maps.

were not significantly better than the clinical map. It appears that the performance had reached an asymptotic level after one week. Interestingly, as with subject V1, the performance with the clinical map after 1 month was numerically better than that at the start of the study. For the bimodal conditions, the subject's performance was equivalent with the two maps at the start and at end of the three month period.

By the end of the study, the subject chose to keep both clinical and custom frequency programs for everyday use.

Subject V3:

The frequency characteristics of the clinical and custom frequency maps along with center frequencies of the image filters (curve minimas from the DVF curves) for subject V3 are given in Table 4.18. Speech recognition scores with various test materials are presented in Figure 4.28. A 2-way repeated measures ANOVA was performed by considering performance with the two frequency maps at two test sessions (pre and post) as the main analysis factors. The results indicated no significant effect of mapping condition ($F[1, 9] = 0.06$, $p = 0.940$), or test session ($F[1, 9] = 1.507$, $p = 0.251$). However, there was a significant interaction between mapping conditions and test session ($F[1, 9] = 20.680$, $p = 0.001$). Pairwise *t*-tests were conducted between the different pairs of test sessions and mapping conditions. The results are summarized in Table 4.19. The results indicated that acute scores with the custom map were significantly lower than the clinical map (day 1) (p=0.024), but after 1 week of daily use, performance with the custom maps was equivalent to the clinical map (p=0.067). The performance continued to improve numerically over 3 months. At three months, scores with custom maps were equivalent to the clinical map (p=0.594); however, the performance with custom maps at this stage was significantly better than the performance re-evaluation with the clinical map at 3 months (p = 0.001). This was due to the decreased performance observed with the clinical map after 3 months relative to the measures obtained at day 1 (p=0.003).

Table 4.18. Frequency allocation tables for subject V3. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were not active.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	733	500	583	667
2	313	375	438	831	667	750	833
3	438	500	563	970	833	917	1019
4	563	625	688	1121	1019	1121	1151
5	688	750	813	1180	1151	1180	1265
6	813	875	938	1349	1265	1349	1429
7	938	1000	1063	1508	1429	1508	1658
8	1063	1125	1188	1807	1658	1807	2006
9	1188	1250	1313	2205	2006	2205	2332
10	1313	1438	1563	2459	2332	2459	2579
11	1563	1688	1813	2698	2579	2698	3049
12	1813	1938	2063	3110	3049	3401	3313
13	2063	2188	2313	3658	3313	3694	3563
14	2313	2500	2688	4182	3563	4011	3938
15	2688	2875	3063	4766	3938	4354	4313
16	3063	3313	3563	5418	4313	4726	4688
17	3563	3813	4063	6007	4688	5127	5063
18	4063	4375	4688	8130	5063	5562	5563
19	4688	5000	5313	10678	5563	6033	6063
20	5313	5688	6063	12542	6063	6542	6688
21	6063	6500	6938	13950	6688	7094	7313
22	6938	7438	7938	14865	7313	7690	7938

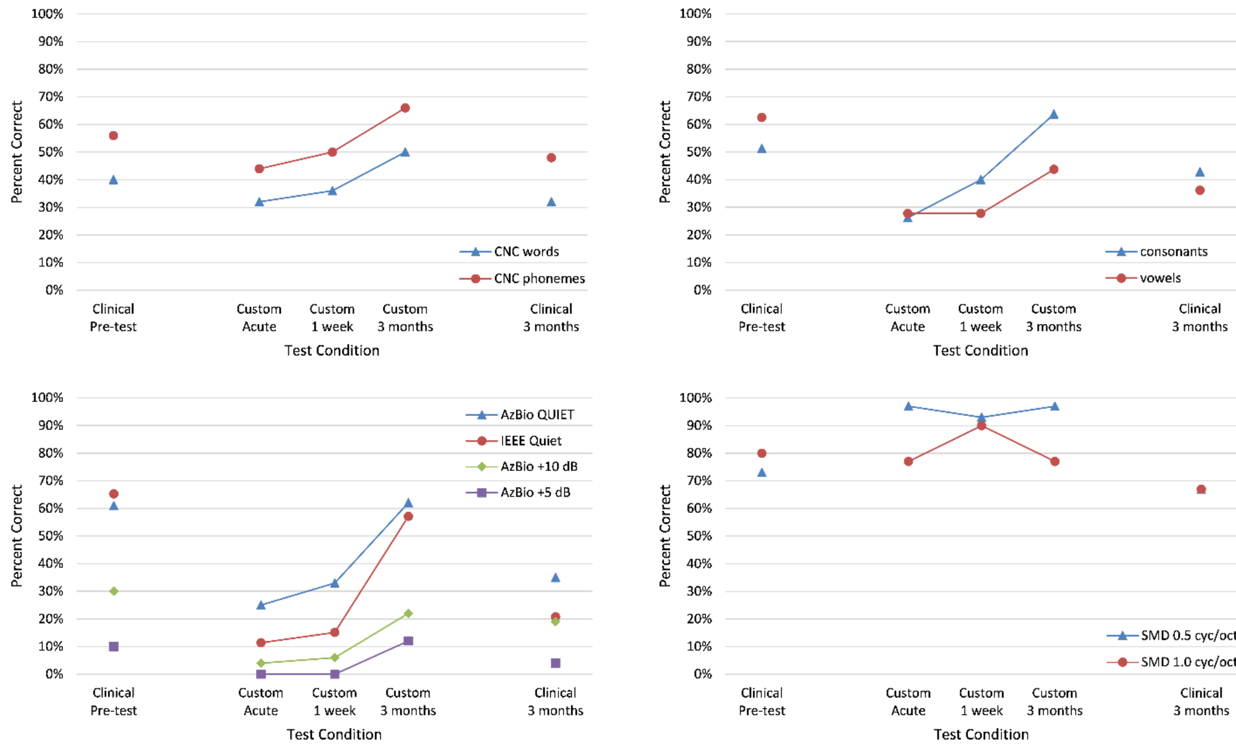


Figure 4.28. CI-only score profile of subject V3 on 10 speech recognition tasks with clinical and custom frequency maps.

Bimodal (Electric+Acoustic) scores: Figure 4.29 shows speech recognition performance in best aided conditions, with an implant in one ear and a hearing-aid in the other. A 2-way repeated measures ANOVA revealed no statistically significant difference between the two maps ($F[1, 7] = 2.846$, $p = 0.135$), or performance differences between the test sessions (pre vs post) ($F[1, 7] = 3.301$, $p = 0.112$). The interactions between the mapping conditions and the test sessions were statistically significant. Pairwise *t*-tests (Table 4.20) indicated that performance with the custom map was significantly lower than the clinical map acutely and at 1-week test session; however, performance was equivalent to the baseline clinical map by 3 months. No statistical difference between pre- and post- evaluation with the clinical maps was observed.

Table 4.19. Paired samples *t*-test for electric-only condition with subject V3.

Pair	Mean Difference	Std. Error Mean	*t*	*p*-value
Clinical_Session1 Custom_Session1	18.459	6.823	2.705	0.024
Clinical_Session1 Custom_Session2	13.815	6.6278	2.084	0.067
Clinical_Session1 Custom_Session3	-2.162	3.917	-0.552	0.594
Clinical_Session1 Clinical_Session3	15.738	4.003	3.931	0.003
Clinical_Session3 Custom_Session3	-17.90	3.447	-5.193	0.001

Table 4.20. Paired samples *t*-test for electric+acoustic condition with subject V3.

Pair	Mean Difference	Std. Error Mean	*t*	*p*-value
Clinical_Session1 Custom_Session1	13.750	4.395	3.129	0.017
Clinical_Session1 Custom_Session2	7.774	3.156	2.463	0.043
Clinical_Session1 Custom_Session3	1.511	2.968	0.509	0.626
Clinical_Session1 Clinical_Session3	7.540	4.084	1.846	0.107
Clinical_Session3 Custom_Session3	-6.029	3.105	-1.941	0.093

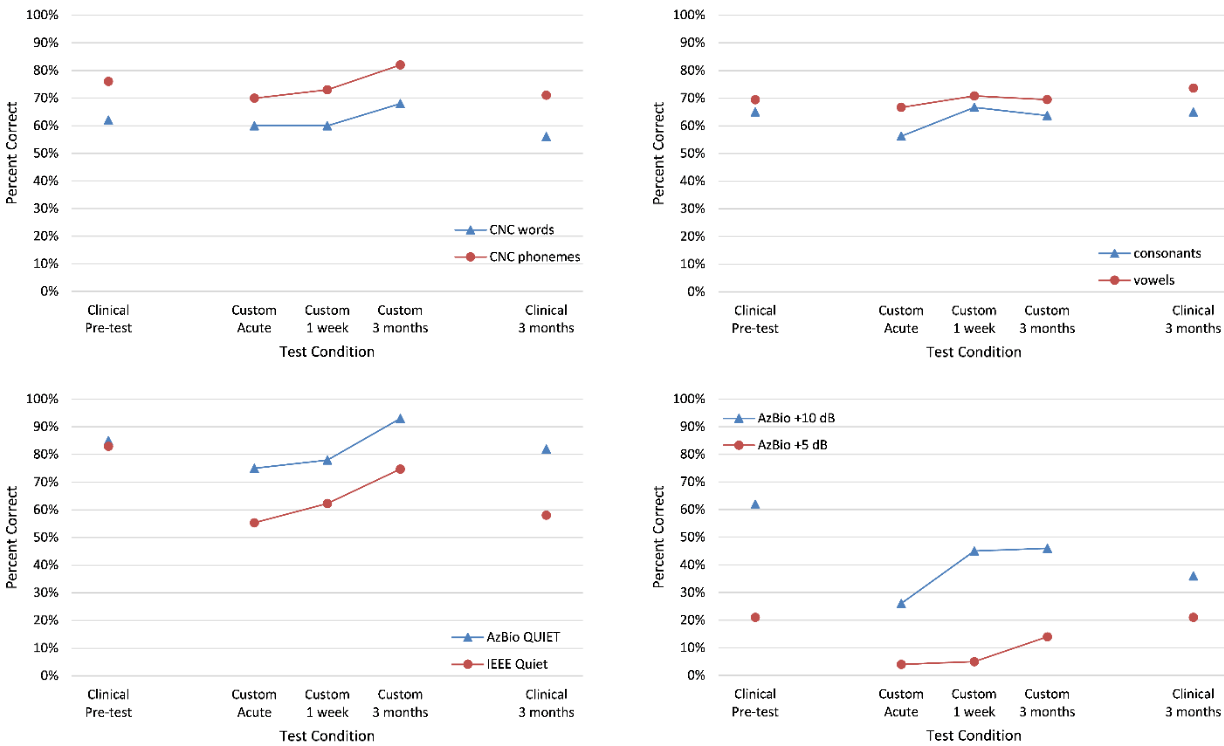


Figure 4.29. EAS score profile of subject V3 on 8 speech recognition tasks with clinical and custom frequency maps.

Remarks/Discussion: The above analyses for electric-only and bimodal conditions indicate that the subject's performance dropped significantly with the custom map acutely, but after 1-week of daily use, the subject accommodated to the custom map and his performance levels were comparable or equivalent to the clinical map. Scores continued to improve during the three months period. At the three months test session, scores from both maps were equivalent for both electric-only and electric+acoustic stimulation.

By the end of the study, the subject chose to keep the custom frequency map exclusively for extended use.

Subject V5:

Subject V5 was a unilateral implant user. He had no hearing in his contralateral ear. Frequency characteristics of the clinical and custom frequency maps along with the center frequencies obtained from the image filters (curve minimas from DVF curves) for subject V5 are given in Table 4.21. Speech recognition scores with various test material for subject V5 are shown in Figure 4.30. A 2-way repeated measures ANOVA was performed to assess performance with the clinical and custom frequency maps at two test sessions (pre and post). Mapping conditions and test sessions were the main analysis factors. The analysis of scores indicated no significant effect of mapping condition ($F[1, 9] = 0.039$, $p = 0.847$). A significant effect of test session was observed ($F[1, 9] = 5.362$, $p = 0.046$), and there was a significant interaction between the mapping condition and the test session ($F[1, 9] = 81.187$, $p < 0.001$). Pairwise *t*-tests were conducted between the different pairs of test sessions and mapping conditions. The results are summarized in Table 4.22. The results indicated that performance levels with the custom frequency maps on day 1 and at 1-week were significantly lower than the clinical map, but after three months the performance with custom maps was equivalent to the clinical map (p=0.414). For this subject, the performance with his old clinical program at 3-months was significantly worse than the baseline performance with the same map at day 1 (p<0.001). Pairwise comparison of speech performance with the custom and clinical maps at three months indicated significantly better performance with the custom maps ($p = 0.004$).

Table 4.21. Frequency allocation tables for subject V5. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were not active.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	1438	500	625	750
2	313	375	438	1604	750	875	1000
3	438	500	563	1861	1000	1188	1375
4	563	625	688	2057	1375	1563	1750
5	688	750	813	2298	1750	1938	2125
6	813	875	938	2459	2125	2313	2500
7	938	1000	1063	2698	2500	2688	2875
8	1063	1125	1188	3190	2875	3063	3250
9	1188	1250	1313	3354	3250	3375	3500
10	1313	1438	1563	3704	3500	3625	3750
11	1563	1688	1813	4082	3750	3875	4000
12	1813	1938	2063	4710	4000	4125	4250
13	2063	2188	2313	5294	4250	4438	4625
14	2313	2500	2688	5872	4625	4813	5000
15	2688	2875	3063	7116	5000	5188	5375
16	3063	3313	3563	9376	5375	5563	5750
17	3563	3813	4063	11029	5750	5938	6125
18	4063	4375	4688	12542	6125	6313	6500
19	4688	5000	5313	13803	6500	6688	6875
20	5313	5688	6063	14708	6875	7063	7250
21	6063	6500	6938	15670	7250	7438	7625
22	6938	7438	7938	16173	7625	7813	8000

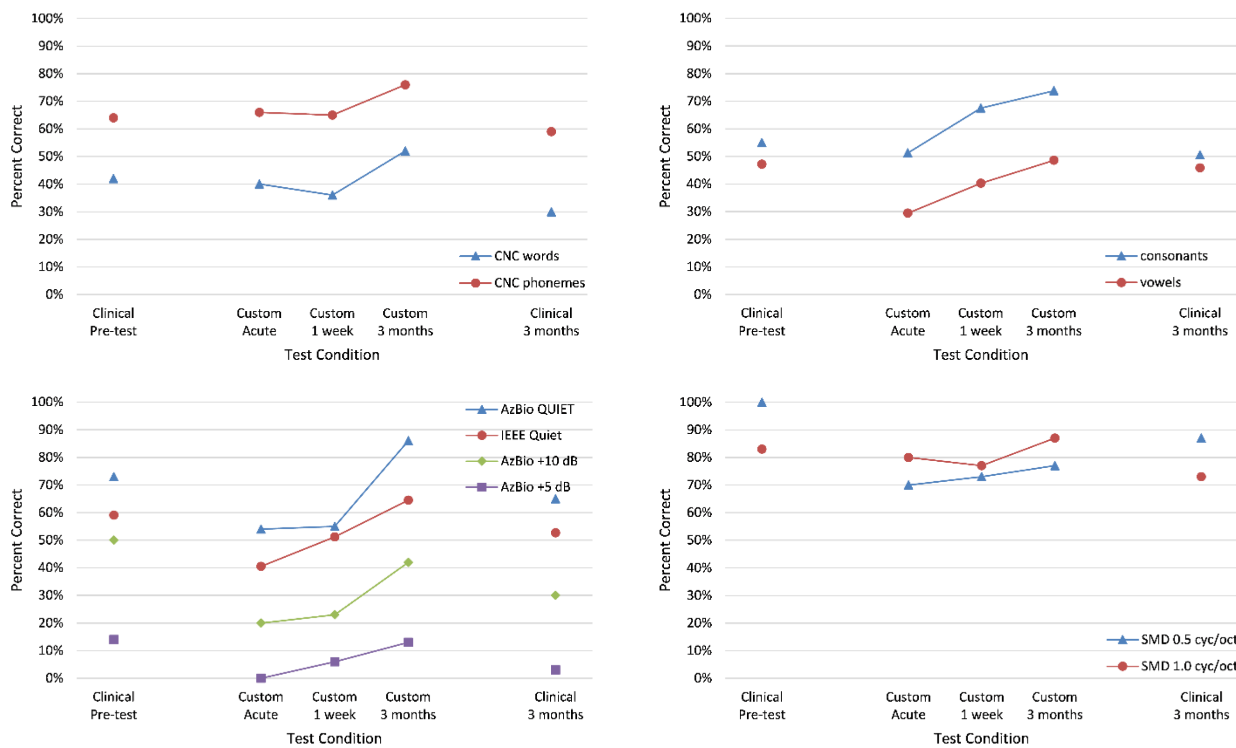


Figure 4.30. CI-only score profile of subject V5 on 10 speech recognition tasks with clinical and custom frequency maps.

Remarks/Discussion: Subject V5's clinical map was 2.5 octaves shifted in frequency for the most apical electrode, indicating severe frequency-place mismatch (assuming peripheral processes stimulated by the electrodes had CFs equal to the ones computed by the DVF curves). The custom frequency map created for this subject was drastically different than the clinical map (see Table 4.21). However, despite this difference between the clinical and custom maps, the subject showed progressive improvement over three month study period with the custom map. At the end of the three month period, his average scores were equivalent to his baseline performance with the clinical map. By the end of the study, the subject chose to keep the custom frequency map exclusively for extended use.

Table 4.22. Paired samples *t*-test for electric-only condition with subject V5.

Pair	Mean Difference	Std. Error Mean	t	*p*-value
Clinical_Session1 Custom_Session1	13.606	3.646	3.732	0.005
Clinical_Session1 Custom_Session2	9.336	3.822	2.443	0.037
Clinical_Session1 Custom_Session3	-3.255	3.800	-0.857	0.414
Clinical_Session1 Clinical_Session3	9.122	1.678	5.436	0.000
Clinical_Session3 Custom_Session3	-12.377	3.175	-3.899	0.004

General Discussion

This study aimed at investigating customized frequency mapping schemes for cochlear implant recipients. Five experienced, post-lingually deafened adult users of CIs participated in this semi-chronic study. Imaging data of the recipients' cochleae provided relationships between spatial locations of the electrode contacts and the characteristic frequencies of the nerve fibers. These imaging data were used to optimize and "tailor-fit" frequency-place functions for each participant. At the start of the study, participants' performance was measured on ten speech recognition tasks with their clinically assigned frequency map, after which their processors were configured with the custom frequency allocation tables determined by the proposed algorithm. Upon fitting, speech recognition tests were re-administered in this acute testing condition with the custom maps. Subjects continued to use the experimental program exclusively for 3 months in their daily lives. A follow-up session was conducted at 1-week post-activation of the experimental programs, and a final assessment was carried out after 3 months. At both of these sessions, speech recognition performances were re-measured.

Consistent with typical clinical observations, performance levels with the experimental program dropped significantly lower than the clinical processor for all CI users. However, all subjects showed progressive improvement with extended use of the experimental programs on all measures of speech recognition. Although the performance levels improved after 1-week post-activation, scores were not significantly different from acute scores with the custom maps measured at day 1. Interestingly, the levels of performance at 1 week were not significantly different than the baseline scores with the clinical maps. At three months, performance levels with the experimental programs were significantly better than the acute scores, and also not statistically different than the clinical program. Overall the progressive improvement with the experimental programs indicate effects of learning, and at least partial adaptation to the custom frequency maps. All subjects displayed the effects of learning with varying degrees. It is impossible to say if performance levels reached an asymptotic level at 3 months. Perceptual studies with CIs suggest that performance levels continue to improve at least up to 24 months post-activation of new maps. However, most significant improvement occurs in the first few months.

In this study, we made significant changes to the frequency allocation tables of CI processors. In some cases, the modifications were extreme relative to their clinical, default frequency allocation schemes. It is surprising that despite these extreme modifications, performance levels were similar to the clinical maps just after 1 week of daily use. Fu *et al.* [93]

demonstrated with three CI users that the deficit in the performance caused by a spectral shifts of 2 – 4 mm in the tonotopic location of stimulation cannot be compensated with at least 3 months of exposure to the new patterns of stimulation. The data from the current study suggest that, if information is delivered with minimum distortion relative to the normal acoustic frequency map, listeners may not only accommodate, but may show improved performance over time.

Comparison with Simulations

The image maps of CI subjects recruited in this study were not the same as that of study 1 (section 4.4.1) or study 2 (section 4.4.2). In order to make equivalent comparisons with simulation data, the same image maps were retested using acoustic simulations of cochlear implants. Similar to study 1 and 2, a noise band vocoder was implemented. Frequency characteristics of the analysis and synthesis filters were varied to reflect the following mapping conditions, i) ideal matched, ii) clinically assigned frequency map in relation to the true tonotopic map, iii) custom frequency map in relation to the true tonotopic place, and iv) analysis filters exactly matched with the tonotopic frequencies of electrode locations. A total of ten subjects participated in this experiment. Five unique frequency maps (of CI subjects who participated in the 3-month clinical study) were used, and each map was tested with 2 subjects. The following set of speech stimuli was used for testing: i) AzBio sentences in quiet, ii) AzBio sentences in 10-talker babble noise at 10 dB SNR, iii) 20 medial-consonants, and iv) 12-medial vowels presented in both male and female voices at 65 dB SPL in a double-walled sound booth.

Figure 4.31 shows mean speech understanding scores of the ten subjects who were tested acutely with the four frequency mapping conditions. Similar to the findings of study 1, condition

1 (ideal) performed significantly better than the other three conditions. A 2-way repeated measures ANOVA was performed to assess the effects of mapping conditions and speech material on speech recognition performance. Mapping conditions (4 levels) and speech material (4 levels) were considered as the main analysis factors in the design. The results showed a significant effect of mapping condition ($F[3, 21] = 24.432$, $p<0.001$), and speech material

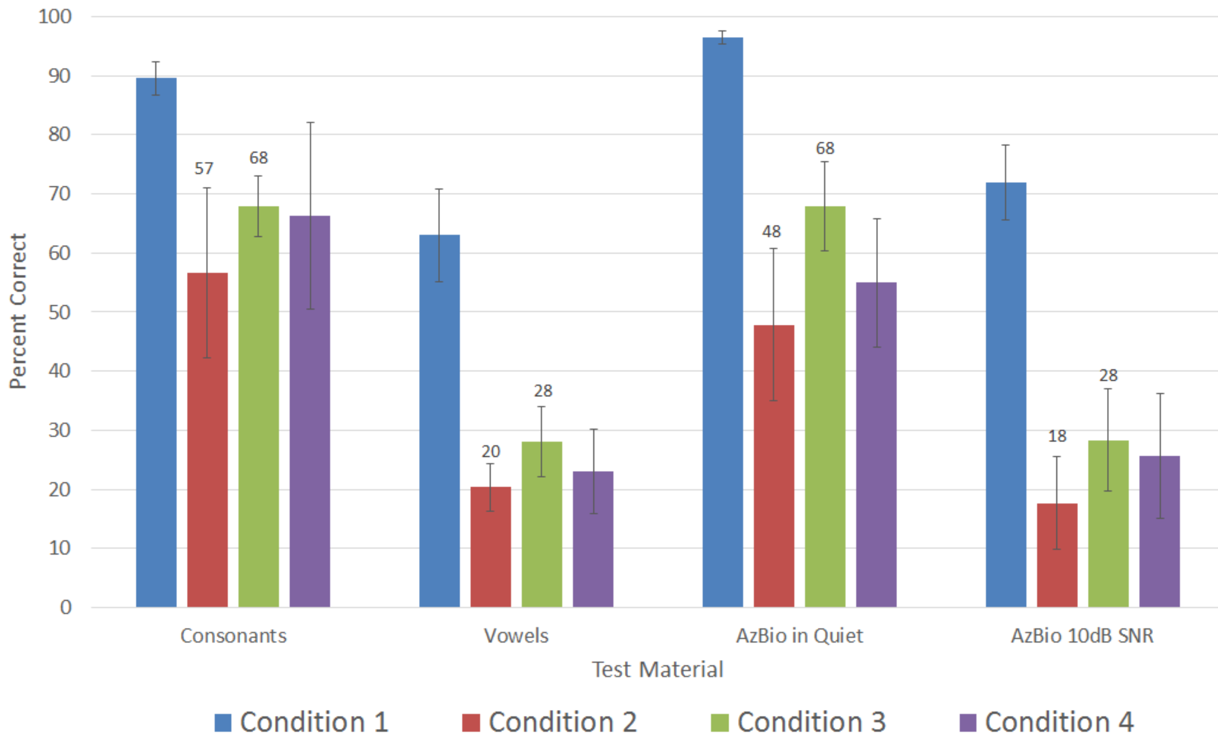


Figure 4.31. Average speech recognition scores of 10 participants from simulation study conducted with image maps of study 3 (CI participants). Percentage correct scores for recognition of consonants, vowels, AzBio sentences in quiet, and AzBio sentences in noise (SNR = 10 dB) with respect to four frequency mapping conditions. Condition 1: Default Frequency allocation, with ideal electrode positioning. Condition 2: Default Frequency allocation, with true electrode positioning. Condition 3: Custom frequency allocation, with true electrode positioning. Condition 4. Frequency allocation matched with true electrode positioning. Error bars represent standard errors of means.

(F[3, 21] =73.996, $p<0.001$). The interactions between mapping condition and speech material were statistically significant (F[9, 63] = 5.633, $p<0.001$). This was followed by a post-hoc Bonferroni test. Results indicated that on all measures of speech tests, condition 1 (ideal) was significantly better than all other mapping conditions ($p<0.05$). Other comparisons had mixed outcomes. A chart showing all pairwise comparisons is presented in Table 4.23. To summarize, the proposed custom frequency allocation scheme (condition 3) was significantly better than the clinical mapping scheme (condition 2) on all measures of speech material, except vowel identification. However, on neither speech recognition task, the proposed solution could reach the level of performance obtained with the ideal mapping. Condition 4 (the matched condition) was not significantly different than conditions 2 and 3 on all speech tests.

Following the acute testing, half of the participants were given 20 minutes of audio-visual auditory training with clinically assigned frequency maps (condition 2), while the other half was provided training with the custom frequency maps (condition 3). A/V training followed the structure similar to that administered in study 2. Participants listened to an audio/video session from 2 talkers, male and female, 10 minutes each talker. Twenty minutes training session was chosen on the basis of data obtained from study 2, which indicated that in comparison to four hours of auditory training, greatest difference occurs just after first 15 minutes of auditory training. Although the performance may not reach the maximum asymptotic level, the performance after 15 minutes is significantly improved in comparison to the acute testing. Such a scheme allowed us to perform quick evaluation with some auditory training, and hence complete the experiments within the same test session. In this way, each image map was tested with both frequency maps (clinical and custom), and performance after 20 minutes of auditory training

Table 4.23. Pairwise Bonferroni comparisons of mapping conditions for each test material.

Speech Material	(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
Consonants	1	2	32.999*	7.174	.015	6.914	59.083
		3	21.606*	5.278	.028	2.417	40.796
		4	23.281*	6.059	.038	1.251	45.312
	2	1	-32.999*	7.174	.015	-59.083	-6.914
		3	-11.392*	2.465	.015	-20.353	-2.432
		4	-9.718	4.053	.286	-24.455	5.020
	3	1	-21.606*	5.278	.028	-40.796	-2.417
		2	11.392*	2.465	.015	2.432	20.353
		4	1.675	2.368	1.000	-6.936	10.286
	4	1	-23.281*	6.059	.038	-45.312	-1.251
		2	9.718	4.053	.286	-5.020	24.455
		3	-1.675	2.368	1.000	-10.286	6.936
Vowels	1	2	42.665*	7.905	.006	13.923	71.407
		3	34.960*	8.356	.025	4.580	65.340
		4	39.966*	8.516	.013	9.004	70.928
	2	1	-42.665*	7.905	.006	-71.407	-13.923
		3	-7.705	3.119	.257	-19.043	3.633
		4	-2.699	3.178	1.000	-14.252	8.855
	3	1	-34.960*	8.356	.025	-65.340	-4.580
		2	7.705	3.119	.257	-3.633	19.043
		4	5.006	2.677	.622	-4.727	14.739
	4	1	-39.966*	8.516	.013	-70.928	-9.004
		2	2.699	3.178	1.000	-8.855	14.252
		3	-5.006	2.677	.622	-14.739	4.727
AzBio Quiet	1	2	48.583*	9.549	.009	13.864	83.301
		3	28.634*	5.766	.010	7.668	49.599
		4	41.505*	7.936	.007	12.652	70.358
	2	1	-48.583*	9.549	.009	-83.301	-13.864
		3	-19.949*	4.661	.022	-36.894	-3.004
		4	-7.078	3.229	.387	-18.817	4.662
	3	1	-28.634*	5.766	.010	-49.599	-7.668
		2	19.949*	4.661	.022	3.004	36.894
		4	12.871	3.742	.065	-.733	26.476
	4	1	-41.505*	7.936	.007	-70.358	-12.652
		2	7.078	3.229	.387	-4.662	18.817
		3	-12.871	3.742	.065	-26.476	.733
AzBio 10 dB SNR	1	2	54.301*	9.197	.004	20.862	87.740
		3	43.643*	9.466	.015	9.227	78.058
		4	46.314*	11.139	.026	5.816	86.811
	2	1	-54.301*	9.197	.004	-87.740	-20.862
		3	-10.659*	1.644	.002	-16.636	-4.682
		4	-7.988	4.249	.613	-23.434	7.459
	3	1	-43.643*	9.466	.015	-78.058	-9.227
		2	10.659*	1.644	.002	4.682	16.636
		4	2.671	4.082	1.000	-12.168	17.511
	4	1	-46.314*	11.139	.026	-86.811	-5.816
		2	7.988	4.249	.613	-7.459	23.434
		3	-2.671	4.082	1.000	-17.511	12.168

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

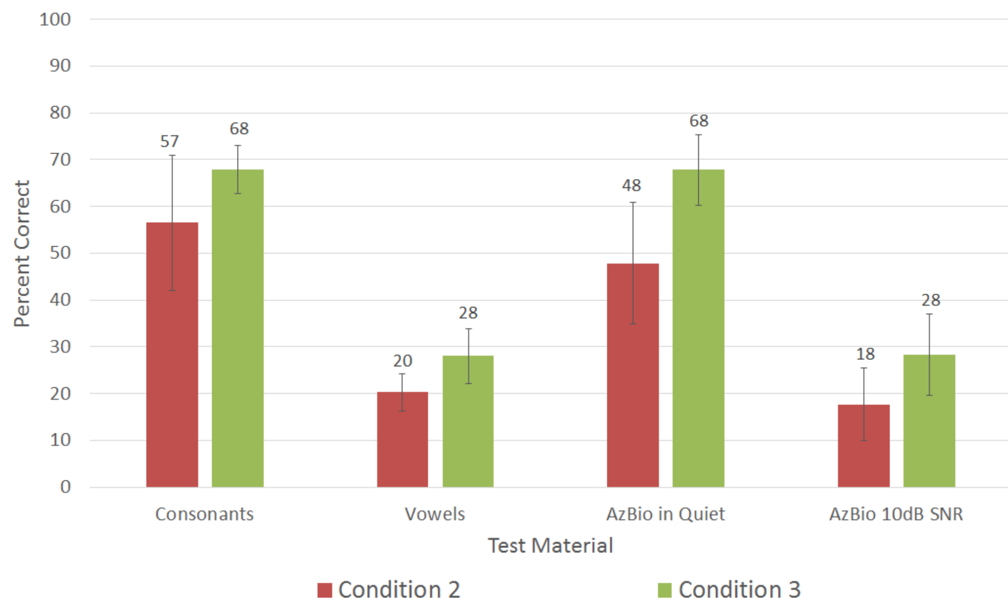
b. Adjustment for multiple comparisons: Bonferroni.

were measured and compared against each other. Figure 4.32 shows mean speech understanding results for condition 2 and condition 3 before and after the training session. The level of improvement with training varied by mapping condition and test material. The largest improvement was observed for sentences presented in quiet.

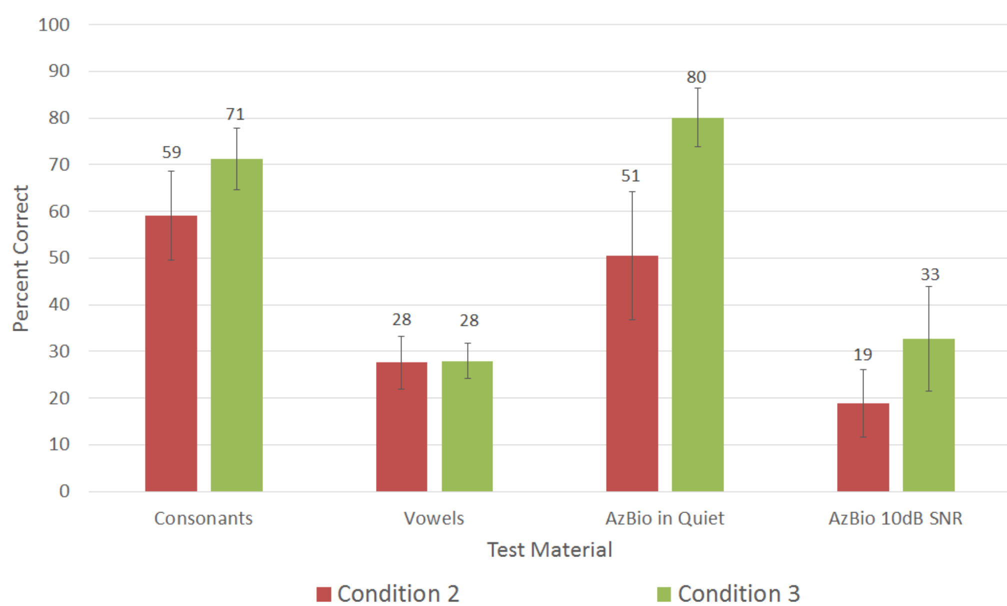
A mixed design multi-variate repeated-measures ANOVA (MANOVA) was conducted with speech material and test session (pre and post training) as within subject factors, and mapping condition as between-subject factors. The analysis revealed a significant effect of speech material ($F[3,18]=97.732$, $p<0.001$). The effect of test session was on borderline ($F[1,6]=5.937$, $p = 0.051$), and was thus considered statistically not significant. The following interactions between the test session, speech material and mapping condition were not statistically significant: (test session × mapping condition), (speech material × test session), (speech material × test session × mapping condition). A significant interaction was observed between speech material and mapping conditions ($F[3,18]=5.039$, $p=0.010$). Post-hoc Bonferroni pairwise comparisons of mapping conditions were conducted for each speech material. The results are summarized in Table 4.24. No statistically significant differences between the two mapping conditions were found after auditory training.

Discussion

In this study, with 10 normal hearing individuals, we simulated the same set of maps that were used by CI users who participated in study 3. The aim of the experiments was to investigate if simulation scores could be compared with actual speech performance observed with CI users. In reality, these comparisons are often difficult to make. It is generally not possible to draw exact conclusions, because the mechanics of normal hearing are significantly different from



(a)



(b)

Figure 4.32. Average speech recognition scores for default (condition 2) and custom (condition 3) frequency maps. (a) Acute results (b) with auditory training from simulation study conducted with image maps of study 3 (CI participants).

Table 4.24. Summary of statistics from mixed design MANOVA

Speech Material	Mapping Condition	Mean	Std. Error	*p* value
Consonants	Condition 2	59.30	7.39	0.441
	Condition 3	67.93		
Vowels	Condition 2	23.44	4.61	0.592
	Condition 3	27.13		
AzBio Quiet	Condition 2	47.98	9.24	0.087
	Condition 3	74.67		
AzBio 10 dB SNR	Condition 2	18.40	8.53	0.383
	Condition 3	29.76		

electrically evoked hearing by cochlear implants. A number of factors that are responsible for speech perception with CIs, for example, audiological, cognitive, and neuro-physiological aspects, cannot be simulated with vocoder-based studies. The main objective behind this study was to evaluate if proposed frequency assignment strategy would provide equal or better speech recognition in comparison to the clinical frequency maps. In line with data from Study 1, the results from this experiment indicated that speech recognition performance with the custom frequency assignment scheme was generally better than the clinical mapping strategy, at least acutely. With 15 minutes of auditory training, the speech understanding increased for both maps. However, after training the difference was no longer statistically significance. One possible reason for this could be the small sample size. This could be a likely factor, because standard deviations in the scores were significantly higher than the previous vocoder-based experiments.

As opposed to study 1, where we tested each map with 3 listeners, in this study each map was tested with only two listeners acutely. The training data were collected from one listener per map. Increasing the sample size may help reduce standard deviations.

4.5 Summary and Discussion

The electrode-neural interface for cochlear implant (CI) recipients is generally far less than ideal. The placement of electrodes relative to the spiral ganglion not only determines the spatial specificity of neural excitation, but also the characteristic frequencies of neural clusters. Large variation in electrode insertion depths across recipients generally results in a unique frequency-place relationship for each CI user. Despite this mismatch, contemporary CI sound processors are usually programmed to assign a generic, pre-defined frequency allocation to the electrode contacts across all users, with the expectation that CI users will accommodate to the frequency-place distortion with experience. The degree of spectral mismatch and an individual's ability to accommodate to the distorted spectral representation of the perceived sound may at least be partially responsible for degraded performance or slower accommodation to electrically-evoked hearing with CIs.

In this chapter, we proposed a user-specific frequency assignment strategy that aims to minimize sub-optimal frequency-place mapping distortions in CIs. The algorithm leverages image-guided procedures to determine the true location of individual electrode contacts with respect to the nerve fibers that may survive, and tailor-fits a frequency place function based on an individuals' electrode-neural interface. The proposed custom frequency mapping strategy was evaluated in three studies. i) acute simulations with normal hearing individuals, ii) simulation of the effect of learning and adaptation to the clinical and proposed frequency maps with normal hearing listeners, and iii) evaluation with cochlear implant listeners in a semi-chronic experiment.

The first study conducted with 42 normal hearing listeners using acute acoustic simulations of cochlear implants, indicated significantly better speech recognition scores than the default clinical mapping scheme on all measures of speech material. Since acute simulation scores may underestimate the potential effects of learning and a user's adaptation to spectrally shifted speech, we investigated the proposed technique in a semi-chronic paradigm. Ten normal-hearing listeners were provided with approximately four hours of auditory training and tested at different intervals with both the default frequency mapping scheme and the proposed custom mapping strategy. For all measures of speech reception, all participants showed significantly better performance with the proposed custom mapping strategy, at least within the four hour test period. The data indicated that listeners adapt to both schemes, but the overall final level of performance with the proposed scheme was still significantly better than the default clinical strategy.

The above two simulation studies with normal hearing subjects served as a viable proof-of-concept for the follow-on investigation with cochlear implant users. Five adult post-lingually deafened CI users participated in a semi-chronic study that lasted three months. Electrode placements were derived from patients' cochlear CT scans, which were used to create custom frequency-analysis tables for each individual. Patients were fitted with these experimental programs and their listening performance on various open-set and closed-set speech recognition tasks was evaluated i) acutely, ii) after 1 week, and iii) after 3 months. Results indicated that acute performance with the custom frequency allocation scheme was significantly lower than their clinical programs, but over the course of three months all subjects displayed improvement with the experimental program. By the end of the three months, the average speech recognition with the custom maps reached to the same level of performance as of their original clinical map, indicating adaptation to the customized frequency maps.

By the end of the study, all participants chose to keep the experimental program either exclusively, or with their old clinical programs. Perceptual studies with CIs suggest that performance levels continue to improve at least up to two years post-activation of new maps. The data from this study indicate that patient-centric optimization of frequency fitting may hold potential for improving implant outcomes, particularly for recipients with moderate to high degree of frequency-place mismatch. Pitch percepts elicited by cochlear implants that are aligned or not drastically different from normal cortical acoustic map could improve the bottom-up presentation of acoustic cues that may potentially lead to overall better speech perception.

4.6 Limitations

The limitations of the proposed custom frequency allocation strategy are discussed as follows: First, it was assumed that electrode locations computed by the image processing algorithm provided an accurate spatial relationship of the electrode contacts to the spiral ganglion cells. The frequency-place relationship could be compromised if there were any inaccuracies in the image processing technique (Noble *et al.* [106]) or the frequency map of the spiral ganglion (Stakhovskaya *et al.* [74]) used in this study. The image processing technique assumed pristine survival and radial projections of the peripheral processes (nerve fibers) throughout the length of the cochlea. In reality, the dendrites in a compromised cochlea, soon start to degenerate with the onset of impairment at the periphery [69]. Dead regions (due to the absence of neurites) may create spectral gaps (spectral holes) that could introduce spectral deterioration. Furthermore, the projections of the auditory nerve fibers beyond the middle turn follow vertical trajectories, which create additional challenges in the electrical stimulation of the apical regions of the cochlea, which is assumed vital for the delivery of low frequency information in this dissertation.

The experiments with various configurations of frequency-mapping assume that peripheral processes stimulated by the electrode contacts have the same intrinsic characteristic frequencies that are computed by the DVF curves. The spread of excitation, distance of electrodes to the putative stimulation sites, and the trajectories of nerve fibers may not result in a one-to-one relationship between spatial location of electrodes and the tonotopic place of stimulation. The physiology of electric hearing was not completely simulated in either of the simulation studies presented in Sections 4.4.1 and 4.4.2. Study 2 only considered four hours of training, which is a very short time duration to observe full adaptation to spectral distortion. Extensive training and longer experience times should be administered to observe learning effects and the extent of accommodation to frequency mismatched speech. Study 3 with cochlear implant users employed an ABA (test/re-test: Clinical/Proposed/Clinical) study model. The retest session with the clinical map after 3 months was an acute measure. For a true one-to-one comparison, subjects should be given experience with the clinical map for a three-month period, after which they should be retested. Finally, the three-month time frame considered in this study may be too short to see the full effect of learning and accommodation. The study period should ideally be extended to 12 – 24 months for a reliable analysis.

CHAPTER 5

CHANNEL SELECTION OPTIMIZATION

A multi-channel cochlear implant system provides electrical stimulation at multiple sites in cochlea via an intracochlear electrode array. These electrodes activate (i.e., provide electrical stimulation to) auditory nerve fibers that are tonotopically organized. Nerve fibers at the base of the cochlea have higher characteristic frequencies, and those at the apex are associated with low pitch percepts. Cochlear implants exploit this tonotopic arrangement of nerve fibers to deliver sound information encoded in multiple acoustic frequency bands to the nerve fibers via an electrode array. The symphony of electrical stimulation patterns elicits sound sensation at the periphery. In principle, cochlear implant sound processor essentially acts as a discrete-time spectrum analyzer which decomposes sounds in multiple frequency bands, performs sound coding operations on each band and relays the processed frequency bands to the corresponding electrodes in a sequential pulse by pulse basis. Theoretically, any sound frequency range can be mapped to any electrode contact. Also, timing of pulses and stimulation levels can be controlled quite precisely. These complex temporal and spectral stimulation patterns are largely responsible for sound sensation and speech recognition.

5.1 Overview on the significance of spectral and temporal cues for speech understanding

The speech signal can be represented by its time and frequency components, and speech intelligibility is largely determined by the availability of these spectral and temporal cues to the listener. Cochlear implant sound processing strategies, thus, aim at optimizing the temporal and spectral encoding of sound characteristics. However, the design of scala tympani implants very much limits the delivery of “high-resolution” temporal-spectral acoustic components. First, the limited number of electrode contacts (12 – 22 electrodes in commercial systems) is virtually incomparable to approximately 20,000 hair cells tuned to different pitches. This is further aggravated by the current spread in cochlea that further limits the number of perceptually independent stimulation sites to no more than ten, even in the best performers. The result is severe degradation of spectral resolution with scala tympani implants, the effects of which are easy to observe (e.g., CI users’ ability to listen in noisy environments). Although cochlear implant listeners, in general, obtain high level of speech recognition in quiet, their performance is significantly worse even in moderate amounts of noise. Poor music perception with CIs is also another aspect that is primarily related to poor spectral encoding. Research indicates that good level of speech recognition in quiet can be achieved by as low as 3 – 4 channels of spectral information. However, listening in complex environments requires significantly higher spectral resolution. For example, Shannon *et al.* [12] demonstrated that recognition of complex speech materials can require up to 30 or more independent spectral channels for an equivalent level of performance, a feat, at least, not possible with current systems.

On the other hand, the temporal encoding in CIs is also far less than ideal. Temporal information is commonly classified as 1) speech envelope (low frequency modulations/fluctuations), 2) low frequency harmonics related to the fundamental frequency of speech (F0), and 3) temporal fine structure (TFS) [122]. While speech envelope can be conveyed more robustly than the latter two, it also has its limitations, and may not always be correctly coded by the sound processing strategies. In theory, a high stimulation rate can provide better timing resolution to deliver fluctuations in the temporal envelope more accurately; however, in practice, high stimulation rates can have both positive and negative effects. Research indicates that at high stimulation rates, similar to natural response patterns, the neuronal discharge patterns are more stochastic rather than deterministic [16, 123, 124]. On the contrary, the refractory period of the auditory nerve fibers puts a limit on the maximum stimulation rate that can convey high resolution temporal envelope without affecting other confounding factors, such as spectral smearing [125]. Research studies are divided on this argument, some groups reporting benefit with higher stimulation rate, while others not a significant difference. On average, perceptual studies do not indicate statistically significant benefit with high stimulation rates [19, 126]. Temporal fine structure, on the other hand, relates to the variations (fast fluctuations) in acoustic waveform within the same period of periodic sounds. TFS helps with pitch perception, sound localization, and binaural segregation of competing sound sources. Pitch perception with CIs is generally very poor, and this is in part due to minimal or complete absence of TFS representation in sound coding schemes, or due to poor availability of periodicity cues, and to some extent the presence of frequency-place mismatch (incorrect place of stimulation). The limited availability of temporal cues in turn connects with the dilemma of poor spectral resolution with the implants

– limited number of perceptually discriminable stimulation sites hinder the ability to provide precise accurate spectral location of pitch and harmonics to the right place.

To summarize, cochlear implants provide only a very sparse representation of acoustic cues. Robustness of the speech signal against spectral and temporal distortions has actually resulted in impressive outcomes with CIs. However, factors such as complex listening situations, tonal/foreign languages, and accents, for example, severely impact the listening performance with CIs. With current implant technology, the solution to this conundrum seems to rely heavily on effectiveness of the sound coding strategies. Throughout the history of CIs, sound processing strategies have been largely responsible for most significant advances in CI performance outcomes. CI processing schemes aim to maximize the encoding of temporal and spectral cues for delivery with the present-day implants. A review of the existing CI processing strategies was presented in Chapter 2. The strengths of clinical and experimental sound processing strategies in improving various aspects of information coding were explored. FSP strategy employed in devices from Med-EL, for example, uses variable rate encoding across electrodes to provide better temporal fine structure. Advanced Bionics processors use HiRes120 strategy that aims to improve fine spectral features by employing current steering techniques to create virtual channels. Spectral maxima sound coding algorithms, for example *n-of-m* strategies [45], used in numerous commercial CI systems, utilize a channel picking strategy to activate a subset of information-rich channels per stimulation cycle to avoid unnecessary stimulation on channels that are likely to hold extra or ‘unwanted’ details. Advanced Combination Encoder (ACE) is a prime example of *n-of-m* sound coding scheme and is the most widely used clinical strategy in commercial systems [47]. In the remainder of this chapter, we will expand our discussion on *n*-

of-m strategies, their benefits, drawbacks, and techniques for further enhancements that may result in improved performance.

5.2 Channel selection process in *n-of-m* strategies

The *n-of-m* strategies are a variant of the CIS strategy, and are typically used in implant systems that have a larger number of electrodes (though CIS is not limited to the number of electrodes). The basic signal pipeline remains the same – band-pass filtering followed by envelope extraction, however there are two major differences. The first difference is that the processing is carried out on temporal frames that are typically 2 – 8 ms in duration. The second difference is in channel selection, (i.e., in each processing frame, only n out of possible m electrodes are selected for stimulation). Typically, channel selection is based on the bands with highest energy (corresponding to spectral peaks in that stimulation cycle). A typical value of n ranges from 8 to 10, and for Nucleus devices m = 22 corresponding to a 22-electrode scala tympani electrode array. One of the earlier flavors of *n-of-m* strategies was the SPEAK strategy (from Spectral Peak) [46] which has evolved into Advanced Combinations Encoder (ACE) strategy [47] and is used in most devices by Cochlear Corporation.

Channel selection is the most critical aspect of *n-of-m* strategies. The efficacy of this technique relies on how efficiently meaningful channels are picked for stimulation. In one aspect, the channel selection strategies work to our advantage, as they could ideally pick only those channels for stimulation that contain meaningful information, and discard the frequency bands that are dominated by noise. This can be a very useful noise reduction strategy, only if it correctly discards the noise-dominant channels. On the contrary, if channel picking strategy is not robust against noise, for example, it could be picking up noise dominant channels, that could result in a performance that could be far worse than stimulating all channels. In the later sections, we would see how spectral maxima channel selection works in different environment types. Furthermore, channel selection can have a significant effect on overall current fields inside the cochlea, and thus, could mitigate or exacerbate channel interference issues. Electrode contacts that are distributed evenly within cochlea generally benefit from interleaved stimulation patterns from the CIS-like *n-of-m* strategies. On the contrary, consider a case of two electrode contacts that are physically very close to each other (for example, due to array curling up), and are stimulating the same group of nerve fibers. If channel selection strategy picks such electrode pair for stimulation, the same group would be stimulated twice, thus causing excessive 'unwanted' stimulation, which may have negative effects on performance. Traditionally, if such an electrode pair could be identified, one of the electrodes would be de-activated to reduce the channel interference with the neighboring electrodes. Scientific evidence suggests that such a scheme could improve outcomes [106, 107, 127-129]. Unfortunately, knowledge of such electrode pairs (or groups) is generally not known clinically. Psychophysical assessment may provide some indication about the problematic electrodes, true estimate is not possible without imaging techniques. Also, de-activating a handful of electrodes, particularly at the low-frequencies may be detrimental than helpful as it may increase the frequency-place distortions.

In the next sections, we present two schemes to optimize the channel selection process. In the first scheme, we have aimed to enhance channel selection process by adaptively assigning weights to each time-frequency unit based on the formant locations of the speech signal and instantaneous signal to noise ratio. The performance of the proposed technique was evaluated acutely with three cochlear implant users in different noise scenarios. In the second scheme, we utilize image-guided procedures to customize channel selection processes based on individual electro-neural characteristics. The details of the proposed techniques are given next.

5.3 Channel selection optimization based on spectral features and signal to noise ratio

The goal of *n-of-m* strategies is to represent the meaningful sound features in a limited number of channels, and this is typically achieved by selecting frequency bands with highest energy in each stimulation cycle. This technique works well in quiet, but is inherently problematic in noisy conditions when noise dominates the target, and noise-dominant channels may mistakenly be selected for stimulation. This could potentially be one of the reasons that CI users are unable to tease apart meaningful features of the target speech from noise because the target-dominant channels may not get activated. Therefore, an intelligent channel-selection strategy, which is able to classify and select channels with the highest amount of target-dominant speech, (and not necessarily just energy) could be useful in adverse listening situations.

A number of noise reduction algorithms for cochlear implants have been proposed over the years which are either based on signal pre-processing (e.g., [130-133]) or signal conditioning integrated with sound coding (e.g., [134, 135]). While the former approach can work well in hearing aids, it is potentially susceptible to unwanted signal distortion which may easily be enhanced by CI processing (e.g., compression function emphasizes low energy sounds logarithmically, musical noise is a prime example), or it may be computationally intensive, thus making it unfavorable for CI processors. The later approach generally relies on spectral modification or modifying channel-selection based on the signal to noise ratio (SNR). Hu and Loizou [134], for example, used a sigmoidal-shaped function that applies attenuation to the noisy envelopes (computed by CIS strategy) inversely proportional to the estimated SNR in each channel. In their later work [135], they reported restoration of speech intelligibility to the level attained in quiet by discarding channels with SNR levels less than 0 dB (binary masking), and varying the number of active channels when the channel SNR was known (ideal condition). While this approach can work well in ideal conditions, one clear disadvantage is that binary masking would completely discard channels containing speech components that are essential for speech intelligibility, but are either unfortunately dominated by noise or wrongly classified by the noise estimation algorithm. The technique here takes its inspiration from the former two approaches, and shapes the weighting functions used in the ACE processing based on the instantaneous SNR of each time frequency (TF) unit. In addition, priority is assigned to channels containing the three speech formants, F1, F2, and F3.

5.3.1 Method

In the clinical/standard ACE (STD_ACE) strategy (Figure 5.1, inside the dotted block), the acoustic signal is sampled at 16 kHz, pre-emphasized, and buffered using a Blackman window into 8 ms (128 samples) analysis frames. Frame overlapping (or analysis rate) typically depends on the channel stimulation rate. For each analysis frame, 128-point FFT and magnitude squared spectrum is computed; thereby, giving 64 frequency bins, with each bin having a frequency resolution of 125 Hz. These bins are passed through 22 weighting filters (Figure 5.2), which essentially compute the envelopes of each channel. Next, 8 - 12 channels with the highest amplitudes are selected and compressed to the current levels using a loudness growth function (LGF) and the patient's clinical MAP, which maps the acoustic amplitudes to the patient's electrical dynamic range.

Figure 5.1 shows the block diagram of the proposed technique (in conjunction with the STD_ACE routine). The proposed technique operates based on two principles, 1) by assigning priority to the formant bands and 2) by assigning weights to each TF unit based on the instantaneous SNR.

Assigning priority to the formant bands

The frequencies of the first three speech formant (F1 – F3) peaks as well as their trajectory over time provide valuable cues to listeners for vowels, glides and stop-consonant perception [133]. This is the reason that feature-extraction strategies [33, 34, 37, 136] have been popular in earlier generation CI processors of 1980s and 1990s. F0/F2 [33, 34] and F0/F1/F2 [136] strategies extract formant locations (F1 and F2) and stimulate the corresponding electrodes at a rate of F0 pulses/sec (pps) for voiced segments and an average rate of 100 pps for unvoiced segments. The MULTIPEAK (MPEAK) strategy [37] stimulates four electrodes at a time and always activates electrode numbers 4 and 7 for F1 and F2 respectively, and then selects the remaining two based on the spectral content of the speech signal.

Spectral maxima-based sound strategies were later adapted to encode the entire spectrum of the speech signal, of whom ACE is the prime example. The shortcoming in spectral maxima algorithms, as noted earlier, comes from the fact that channel selection is based on the largest

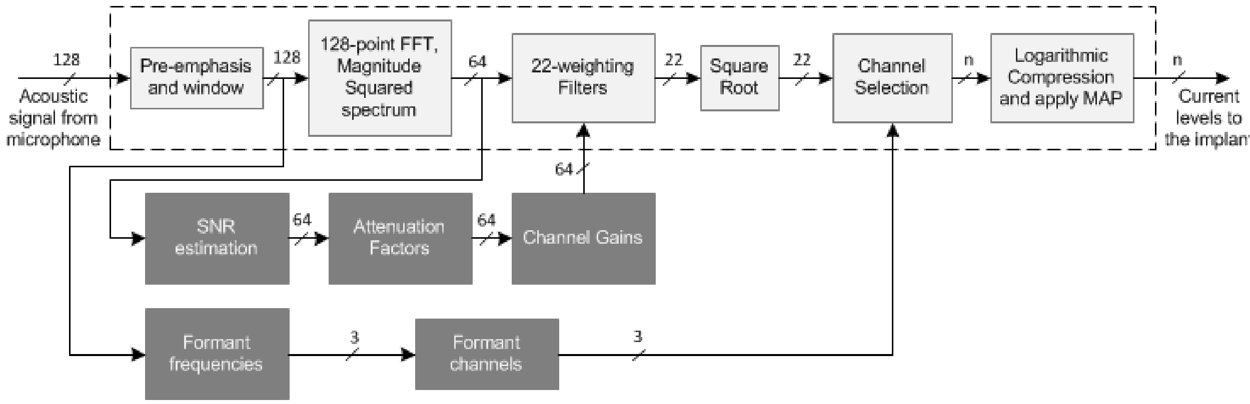


Figure 5.1. Signal flow in the standard ACE strategy (shown inside the dotted block). Processing blocks for the proposed technique are shown in the darker tone. Numbers on connecting arrows represent the frame size in number of samples at each step.

filter amplitudes which may not necessarily be the spectral (formant) peaks, and hence may not encode the major formant frequencies. Several maxima may come from a single spectral peak [137]. This can be problematic in noise, which tends to reduce the dynamic range of the spectrum as well as the spectral contrast (peak-to-valley ratio on LPC spectrum). Thus, preference may be given to noise dominant channels irrespective of the presence/absence of spectral peaks. In fairness, the F1 spectral peak is preserved to a certain degree in noise which works to the advantage of ACE. Although, the location of peaks of higher formants may not be affected as much in noise, spectral smearing and reduced spectral contrast could give preference to the noise dominant channels.

The proposed technique continuously computes the first three formant peaks (F1 – F3) in each analysis cycle and assigns priority to the channels corresponding to the formant frequencies during the channel selection process. Formant frequencies are computed by solving the roots of the linear prediction coefficients (LPC). Formant continuity constraints are imposed to avoid unwanted distortion.

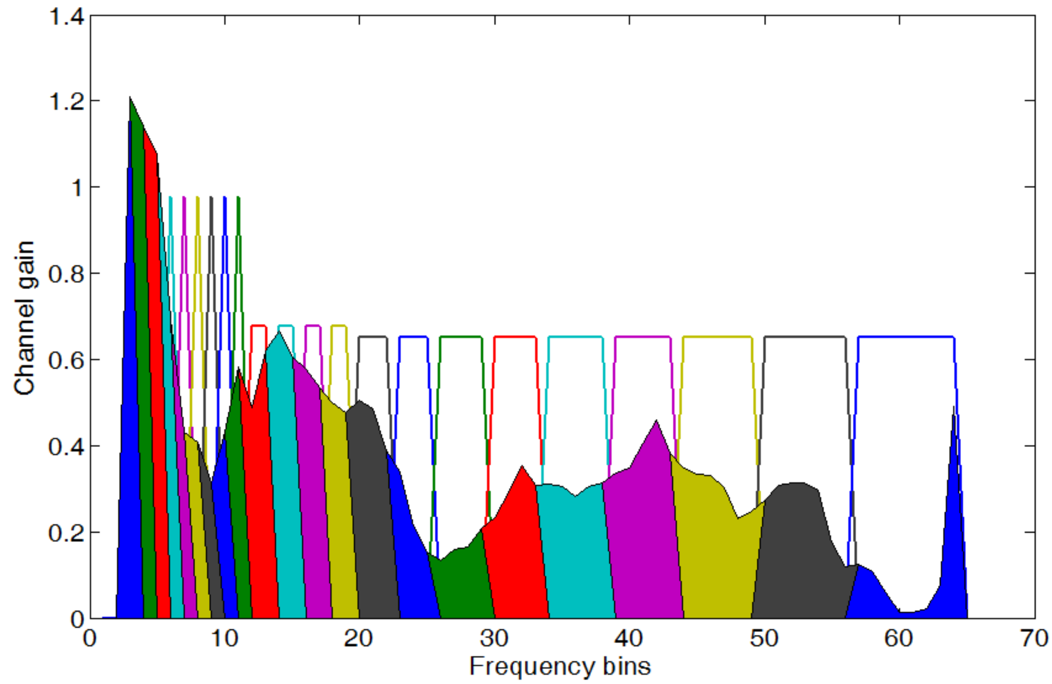


Figure 5.2. Weighting filters used in the ACE strategy plotted in solid line. Spectrum shaping by soft-masking technique shapes the spectrum based on SNR of each TF unit as shown by the color-filled plots.

Assigning attenuation factor based on the SNR

Hu and Loizou [134, 135] applied binary and soft masking techniques to channels with low SNR. Given that a channel can comprise of as many as 8 frequency bins, (filters with broader bandwidth would have higher number of bins), channel classification would be compromised for bin-widths of 2 or more. (If the number of bands are less than 20, the number of bins per frequency band will increase). The proposed technique estimates the SNR for each TF unit, $X(i, j)$, where X is the magnitude squared spectrum of the i^{th} analysis frame and jth frequency bin. This yields a total of 64 SNR values for each analysis frame (stimulation cycle). Based on the computed SNR, an attenuation factor is generated. In the present study, we analyzed both binary and soft masking techniques. In binary masking, a weight of 0 was assigned to the

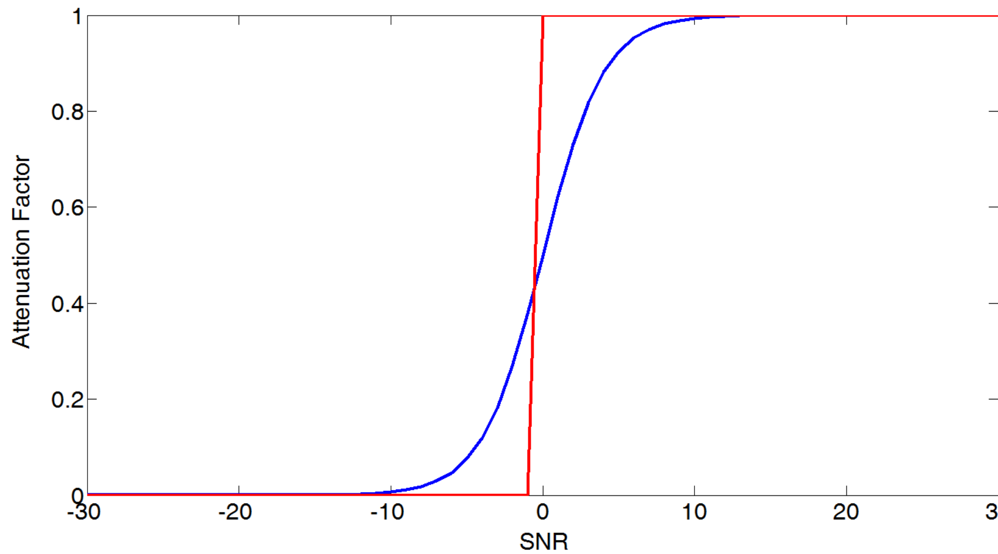


Figure 5.3. Binary weighting (red) and soft-masking (blue) attenuation functions.

$SNR(i,j) < 0\,dB$, and for the rest, (i.e., $SNR(i,j) \geq 0\,dB$), a binary value of 1 was assigned. In the soft masking approach, a sigmoidal-shaped function was considered which plateaus for $SNRs > 15\,dB$ and floors to 0 for $SNRs < -15dB$. Both weighting functions are shown in Figure 5.3. The 64 weighting values generated for each TF unit are then used to shape the gain of weighting functions. This is illustrated in Figure 5.2 for the soft masking technique.

5.3.2 Evaluation

In order to evaluate the effectiveness of the proposed technique, evaluation was first carried out with the SNR of each TF unit known a priori. The results from this experiment would validate if the proposed technique is effective. In the second phase, the instantaneous SNR of each TF unit was estimated using improved minimum controlled recursive average (IMCRA) algorithm [138]. While any SNR estimation algorithm could be used, IMCRA utilizes the spectrum components of each TF unit, $X(i,j)$, which are already computed by ACE, and therefore, no separate processing was needed. Furthermore, the advantages of the IMCRA method are particularly notable in adverse environments involving non-stationary noise, weak speech components, and low SNR conditions.

Subjects

A total of 3 CI users participated in this acute experiment. All participants were native speakers of American English and were fitted with Nucleus 24 multichannel device manufactured by Cochlear Corp. All participants used ACE as their speech processing strategy.

Test Material and Procedure

IEEE sentences [111] were used as the speech stimuli for testing. 10 sentences for each test condition were used. The algorithms were implemented offline in MATLAB and stimuli were presented via UT Dallas's PDA-based research platform [139].Two sets of experiments were conducted. In the first set, the effectiveness of assigning priority to formant channels (F1, F2, and F3) was evaluated both in terms of speech intelligibility and perception quality. All words were marked for correctness. A total of 6 conditions were tested in experiment 1, namely speech in quiet, speech in 10 dB SNR speech shaped noise (SSN), speech in 5 dB SNR SSN, speech in 10 dB SNR white Gaussian noise (WGN), speech in reverberation with reverberation time T_{60} = 600 ms, and finally, speech in reverberation (T_{60} = 600 ms) and 10 dB noise.

In the second experiment, four techniques were assessed separately, namely ideal-binary (IdBinary), ideal-soft (IdSoft), estimated-binary (EsBinary), and estimated-soft (EsSoft). Ideal conditions represent when the SNR was known a priori, while the estimated conditions represent when the SNR was estimated using IMCRA. Binary conditions correspond to the output of the attenuation factor to binary values, while the soft conditions represent output from the sigmoidal attenuation function, as described earlier. Speech intelligibility and quality measures were assessed for each technique with speech in 10 dB SNR SSN, and speech in 5 dB SNR SSN. The results from STD_ACE were used as baseline scores for comparison.

For the speech quality tests, the same sentences processed with STD_ACE and FRMNTS_ACE strategies were streamed back to back and CI users were asked to rate the quality of the second sentence as compared to the first in terms of being pleasant, clear and free of sound distortions on a scale of -3 to 3, with 0 being ‘about the same’, -3 being much worse, and +3 indicating much better.

Results

Figure 5.4 shows the mean intelligibility scores for STD_ACE and formant-based ACE (FRMNTS_ACE) technique. While there is very little to no significant improvement in intelligibility at high SNRs, there was an improvement of 17% for speech at 5 dB SNR SSN, and 20% when noise was added to the reverberant signal (Rev600+n). Modest improvement was also observed for speech in WGN at 10 dB SNR. For speech quality, the user response was between +1 and 0, on average, indicating slightly better to no difference in all test conditions. The subjects reported that words “popped out” more in the FRMNTS_ACE strategy.

The mean intelligibility scores for experiment 2 are presented in Figure 5.5. The results showed that both IdBinary and IdSoft techniques were able to restore speech intelligibility to the level equivalent to speech in quiet. This establishes the effectiveness of the proposed technique in masking noise if cues to the SNR are available. Figure 5.5 also presents the intelligibility scores when the noise was estimated. There was no significant improvement at 10 dB SNR level. However, at 5 dB SNR, improvement was observed with both EsBinary and EsSoft, but the later

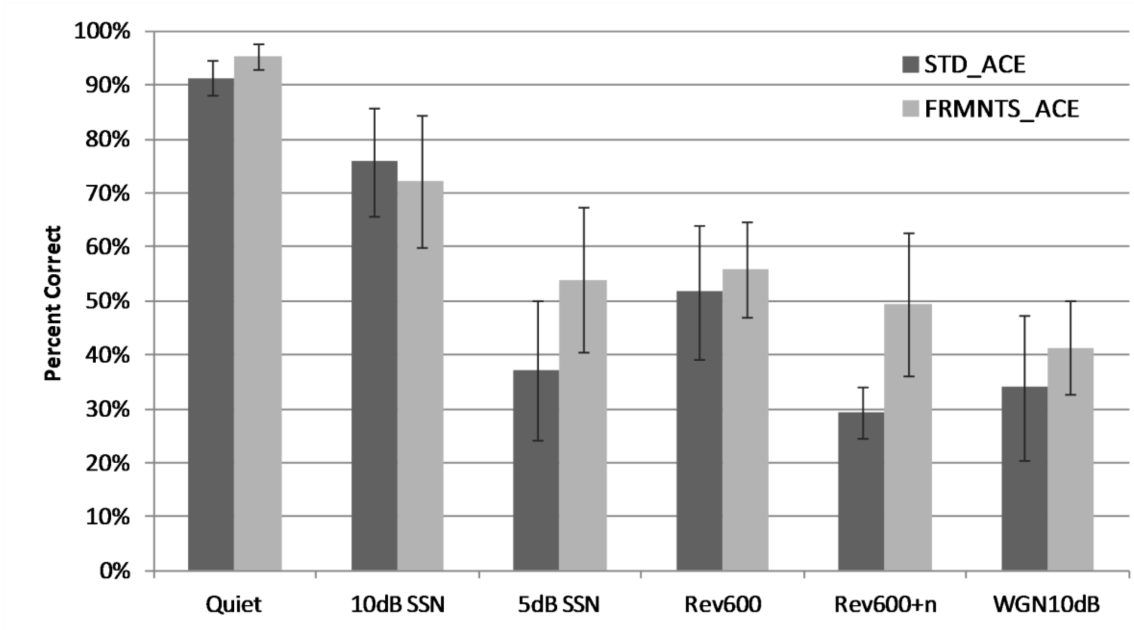


Figure 5.4. Mean speech intelligibility scores of 3 CI users in Experiment 1. Error bars represent standard errors of the means.

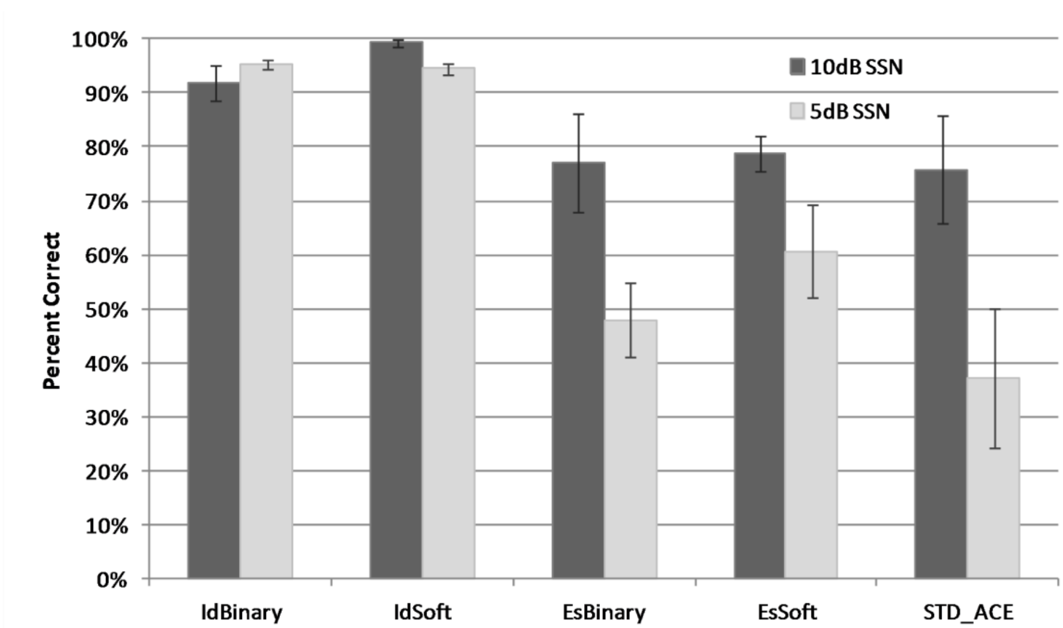


Figure 5.5. Mean speech intelligibility scores of 3 CI users in Experiment 2. Comparison of the proposed technique using IdBinary, IdSoftm, EsBinary, and EsSoft approaches with STD_ACE strategy. Error bars represent standard errors of the means.

resulted in relatively more significant gains in intelligibility (>20 percent). The results indicate that the proposed technique can potentially improve speech intelligibility in low-SNR conditions.

Quality tests for Experiment 2 indicated an average score of “Much Better” as compared to the STD_ACE (unprocessed) for both ideal conditions. For the estimated SNR at 10 dB and 5 dB SNR levels, the average score was (+2) corresponding to “Better” as compared to the STD_ACE. The quality scores between binary and soft masking approaches indicated slight preference for the soft masking approach, with an average score of +0.5.

5.3.3 Summary and Discussion

This study considered two approaches for improving the channel selection process of spectral maxima sound coding algorithms for cochlear implant systems. The first approach assigned priority to the channels containing the formant frequencies, while the second approach adaptively assigned weights to each time-frequency (TF) unit based on the estimated SNR in each stimulation cycle. Two types of maskers, binary and soft, were used to assign attenuation factor to each TF unit based on SNR. Both approaches were evaluated independently and synergistically on 3 CI users. The intelligibility scores indicated significant improvement at low SNR levels, with speech at 5 dB SNR and speech masked by reverberation and noise. The quality scores revealed very little preference to the formants-based approach over the standard ACE, and soft-masking over the binary masking. However, the noise masking approach was greatly preferred over standard ACE strategy. The speech recognition scores with the two approaches revealed some benefit with the formants-based technique over the standard ACE, and soft masking over binary masking at low SNR levels.

The proposed technique is inherently limited by the accuracy of formant locations and SNR estimation. Speech intelligibility reached to the level obtained in quiet when SNR was known *a priori*. The proposed technique could potentially be used to improve speech understanding performance of CI users in adverse listening environments.

5.4 Image-guided customization of channel selection

A Distance-Vs-Frequency (DVF) curve profile derived from an implant recipient's CT scan, an example of which was shown in Figure 4.3, provide insight not only into the characteristic frequency of each stimulation site, but also the degree of spectral overlap caused by the neighboring electrodes (inference on potential level of current spread and spatial specificity of the neural excitation). The spatial location of electrode contacts could be used to determine which electrodes are likely to cause channel interaction. Consider a case of two electrode contacts that are physically very close to each other (for example, due to array curling up), and are stimulating the same group of nerve fibers. If channel selection strategy picks such electrode pair for stimulation, the electrode group would stimulate the same neural population, thus causing excessive current summation, which may have negative effects on performance. As noted previously, electrode deactivation strategies for such problematic electrodes could potentially improve outcomes [106, 107, 127-129]. In the proposed approach, we use patients' imaging data to identify the problematic electrodes and rather than switching off these electrodes, we activate electrodes in such a fashion that electrodes that are likely to cause channel interaction are not selected in the same stimulation cycle. This approach aims to preserve the fine spectral structure to a greater degree than electrode de-activation strategies.

Figure 5.6 illustrates two different scenarios of spatial locations of electrodes. Column 1 depicts ideal electrode placement (i.e., neighboring electrodes are well spaced, and thus not causing channel interaction issues), whereas column 2 depicts electrode 3 (E_3) spatially close to the electrodes 2 (E_2) and 4 (E_4) and thus, it may result in possibly higher channel interaction. A depiction of their DVF curves is also shown in the same figure. In the conventional electrode de-activation strategy, electrode 3 could be turned off to overcome the channel interaction problem. While this is a decent approach; however, it results in completely switching-off electrodes. Some of the consequences of electrode deactivation are modifications to the filter frequency assignments (frequency-place distortion) and broadening of the filter-bandwidths, both of which may alter the spectral cues. The ability to provide fine spectral structure, is thus, further compromised due to the decrease in attainable frequency resolution. This may become more challenging if a handful of electrodes are switched off (see example 2 below).

One way to make use of the electrode(s) that are likely to cause channel interaction, is by optimizing the channel selection process. Figure 5.7 shows the electrode selection process for three different scenarios – i) conventional channel selection when all 4 electrodes are ON; ii) channel selection in standard electrode-deactivation approach (electrode 3 is switched off); and iii) proposed channel selection (for this particular example). Each column in Figure 5.7 represents one stimulation cycle, and each box represents one time-frequency unit. A filled (black) box represents an active electrode in that particular stimulation cycle. Figure 5.7 (c) shows the proposed approach, in which we alternate activation between electrodes 2 and 3, (i.e., E_2 and E_3 cannot be active in the same stimulation cycle). By time-interleaving the activation of channels (much like the CIS approach), we can make use of the frequency spaces of E_2 and E_3

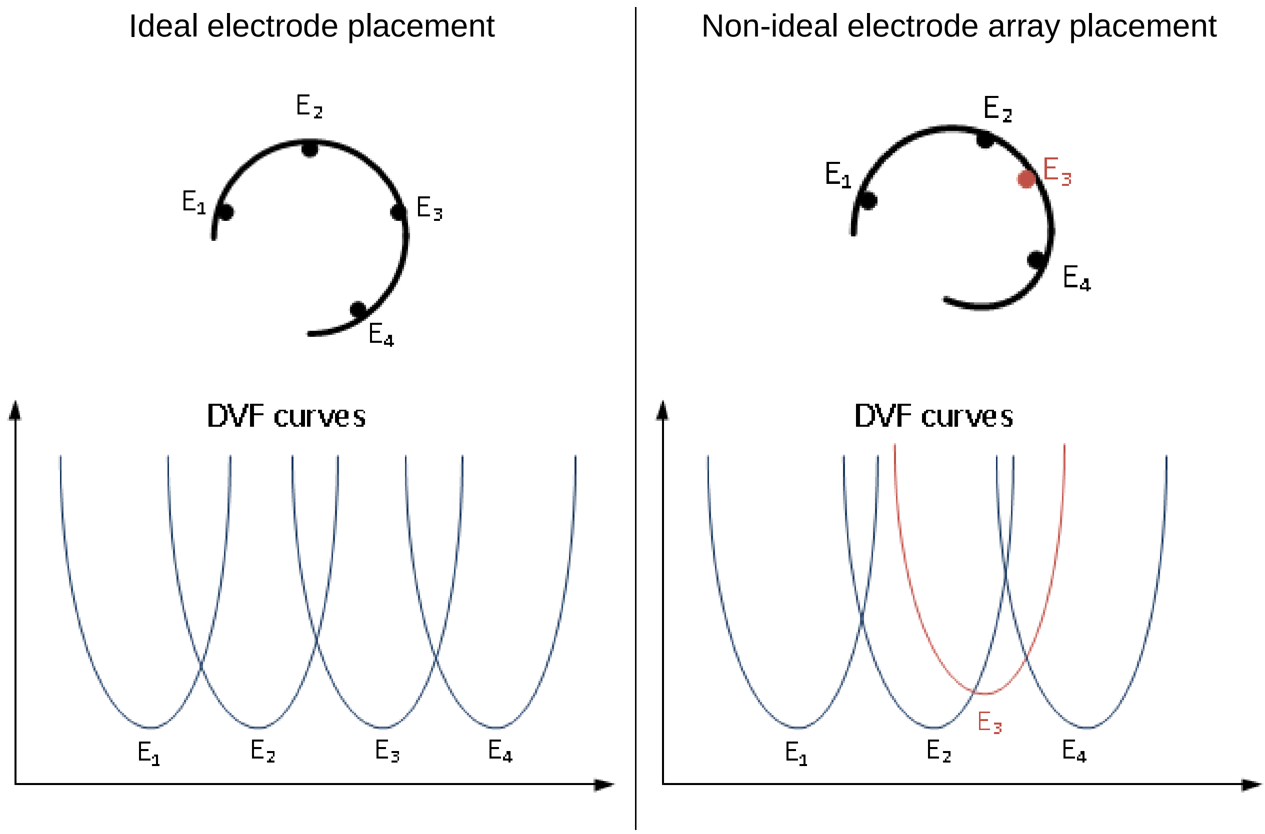
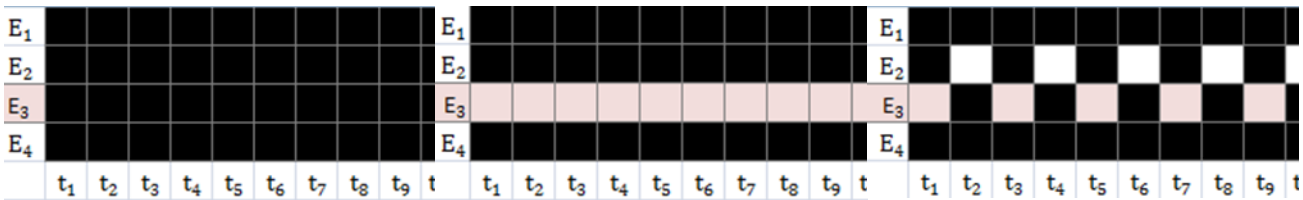


Figure 5.6. Depiction of ideal and non-ideal electrode placement scenarios, along with the corresponding DVF curves.

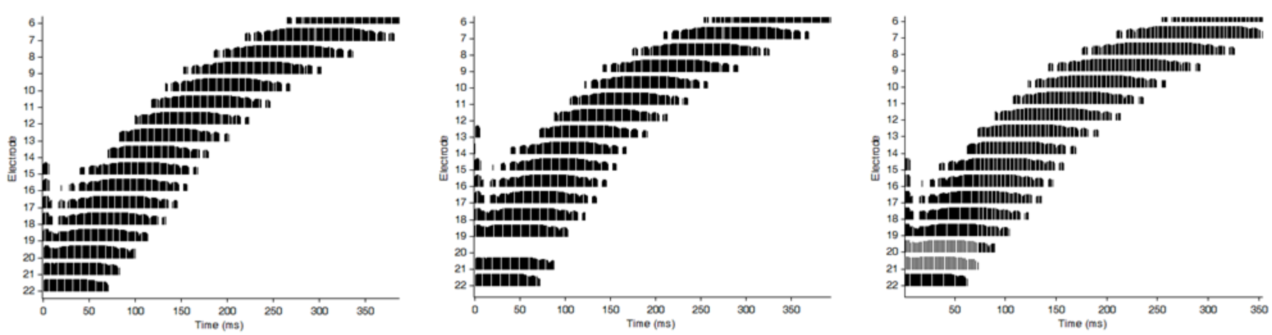
tonotopic place without turning off any electrode completely. We call this approach image-guided Time-Interleaved Channel Selection (TICS) strategy.

Figure 5.8 shows electrodograms of a chirp signal obtained from ACE processing strategy with three approaches. The proposed approach can be extended to more complicated scenarios when one or more electrodes are interacting with each other. The following section gives details of the proposed algorithm.



(a) Standard (b) Electrode-deactivation (c) Proposed TICS

Figure 5.7. Channel Selection process depicted for 3 scenarios for the example shown in Figure 5.6. Each shaded box represents an active time-frequency unit. (a) Standard approach which keeps all electrodes on. (b) Electrode-deactivation strategy. Electrode likely to cause channel interaction is switched off. (c) the proposed Time-Interleaved Channel Selection (TICS) approach.



(a) Standard (b) Electrode-deactivation (c) Proposed TICS

Figure 5.8. Electrodograms of a chirp signal for the three scenarios for the example shown in Figure 5.6. In (c), E20 and E21 (representative of bands # 3 and 3) are stimulated on alternative cycles, (thus stimulation rates of E20 and E21 are half of the other electrodes)

5.4.1 Image-guided Time-Interleaved Channel Selection (TICS) - Algorithm

In the ACE processing strategy, *n-of-m* electrodes (channels/bands) are selected. Typical value of m=22 for Cochlear Corp. electrode arrays, and n=8. Each of the m bands are sorted in descending order at the channel selection process.

Say, the sorted channels are represented by: $c_1, c_2, c_3, \ldots, c_m$, $m = 22$

and λ_{c_i,c_j}represents the proximity index of channels c_i and c_j with each other. For simplicity, let's assume λ_{c_i,c_j} is binary[9] and can either be 1 or 0. A value of 1 indicating that channel c_i is likely to cause channel interaction with electrode c_j, and a value of 0 indicating that electrodes are not likely to cause channel interaction. Thus for each channel c_i, proximity indices can be written as:

$$\Lambda_{c_1} = \left[\lambda_{c_i,c_1}, \lambda_{c_i,c_2}, \lambda_{c_i,c_3}, \cdots, \lambda_{c_i,c_m}\right]$$

The proximity indices for all m electrodes can be written as $m \times m$ matrix:

$$\mathbf{\Lambda} = \begin{bmatrix} \Lambda_{c_1} \\ \Lambda_{c_2} \\ \vdots \\ \Lambda_{c_m} \end{bmatrix} = \begin{bmatrix} \lambda_{c_1,c_1} & \lambda_{c_1,c_2} & \cdots & \lambda_{c_1,c_m} \\ \lambda_{c_2,c_1} & \lambda_{c_2,c_2} & \cdots & \lambda_{c_2,c_m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{c_m,c_1} & \lambda_{c_m,c_2} & \cdots & \lambda_{c_m,c_m} \end{bmatrix}$$

where $\mathbf{\Lambda}$ is a binary symmetric matrix, i.e., $\lambda_{c_i,c_j} = \lambda_{c_j,c_i}$ and $sum(\Lambda_{c_i})$ gives number of interactions = n_{c_i} of each channel. If $n_{c_i} = 0$, it implies that channel i (c_i) has no known channel interactions with other electrodes. If $n_{c_i} > 0$, it implies that channel c_i has channel interactions with other electrodes. In order to alternate activation/selection of channels with $n_{c_i} > 0$, we keep channel selection history $\mathbf{\Psi}$, which is given by:

$$\mathbf{\Psi} = \begin{bmatrix} \Psi_{c_1,t_\tau} \\ \Psi_{c_2,t_\tau} \\ \vdots \\ \Psi_{c_3,t_\tau} \end{bmatrix} = \begin{bmatrix} \Psi_{c_1,t_1} & \Psi_{c_1,t_2} & \cdots & \Psi_{c_1,t_m} \\ \Psi_{c_2,t_1} & \Psi_{c_2,t_2} & \cdots & \Psi_{c_2,t_m} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{c_m,t_1} & \Psi_{c_m,t_2} & \cdots & \Psi_{c_m,t_m} \end{bmatrix}$$

---

[9] Instead of the binary weights, a soft-masking strategy may also be considered. A soft-masking strategy would assign a value between 0 and 1 based on the proximity index.

Here, Ψ_{c_i,t_τ} represents history of channel i at time τ (in the previous τ cycle of stimulation) and holds a binary value – 1 corresponding to active and 0 corresponding to not-active condition. For each Ψ_{c_i,t_τ}, history of at least n_i cycles is required for that channel.

Let $\emptyset_{c_i}$ represent decision to activate the channel c_i at the current stimulation cycle. A value of 1 represents that channel will be selected for stimulation, whereas a value of 0 represents that channel will not be selected for stimulation.

$$\emptyset_{c_i} = \begin{cases} 1, & if\ (n_{c_i} = 0)\ or\ (\varphi_{c_i} = 0) \\ 0, & otherwise \end{cases}$$

The following rule is used for channel selection. If $n_{c_i} > 0$, φ_{c_i}will be '0' if and only if $\Psi_{c_i,t_{n_{c_i}}} = 0$ (i.e., there is no history of activation of channel c_i in the last n_{c_i} cycles) AND the channels that are likely to interact with c_i are not active in the current stimulation cycle.

In the following section, we will explore two example DVF maps of actual CI users, and the resulting electrodogram profiles from different channel-selection schemes using the ACE processing strategy.

5.4.2 Example 1: Low level of channel interaction

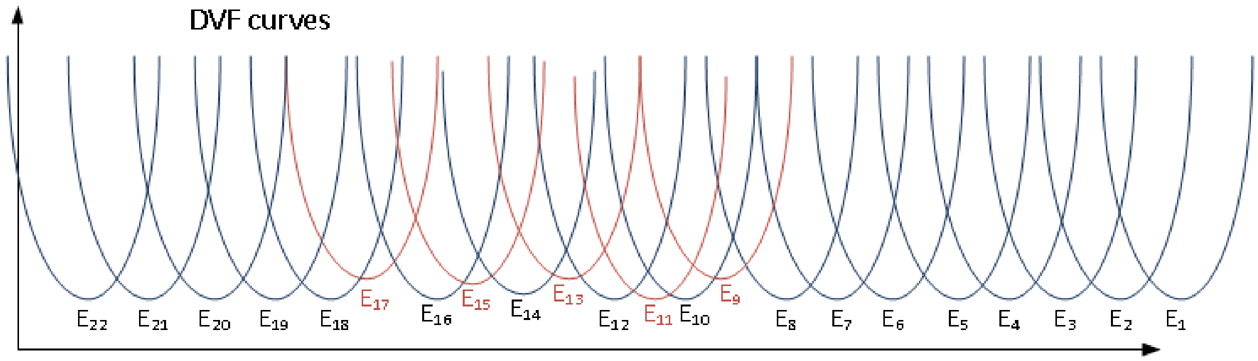


Figure 5.9. Example of a DVF curves profile of a CI. Curves in the red are likely to cause channel interaction.

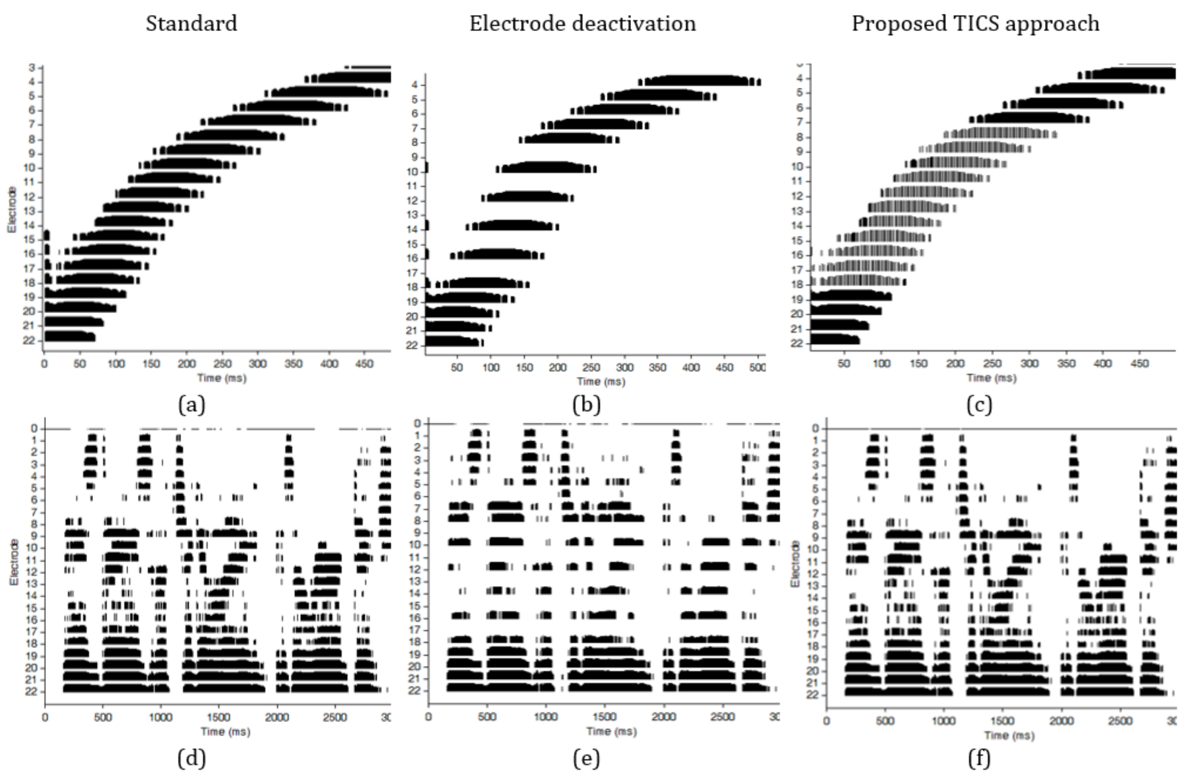


Figure 5.10. Electrodograms of (a) – (c) chirp signal; and (d) – (f) IEEE speech sentence obtained from three channel selection schemes. Column 1: standard – all electrodes are on; Column 2: electrode de-activation strategy – switch off electrodes that are likely to cause channel interaction; and Column 3: Proposed Time-Interleaved Channel Selection (TICS) scheme for the DVF curves shown in Figure 5.9, Example 1.

5.4.3 Example 2: High-level of channel interaction

The following example demonstrates a scenario with higher level of spatial channel interaction from neighboring electrodes. Comparisons of the electrodograms generated from the proposed TICS strategy with the electrode de-activation scheme shows better preservation of spectral detail.

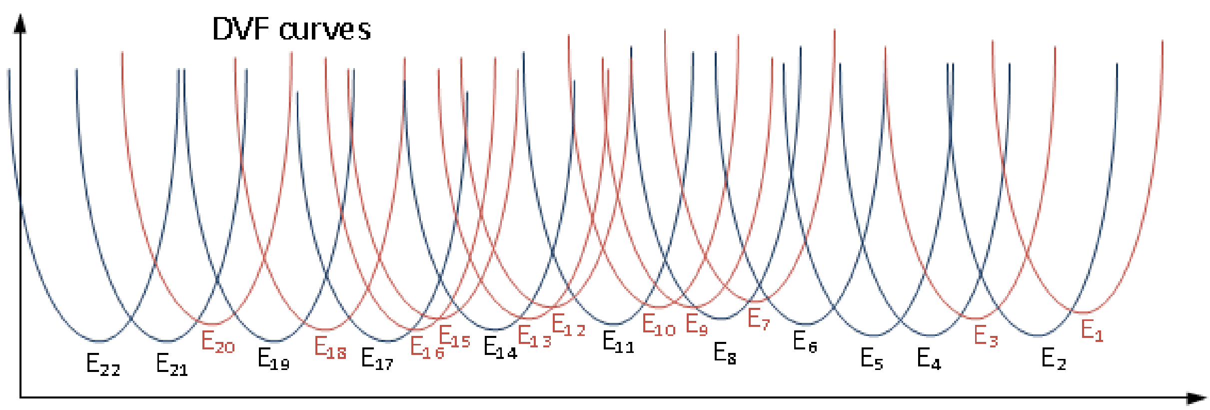


Figure 5.11. Example of a DVF curves profile of a CI with high level of spatial electrode interactions. Curves in the red are likely to cause channel interaction.

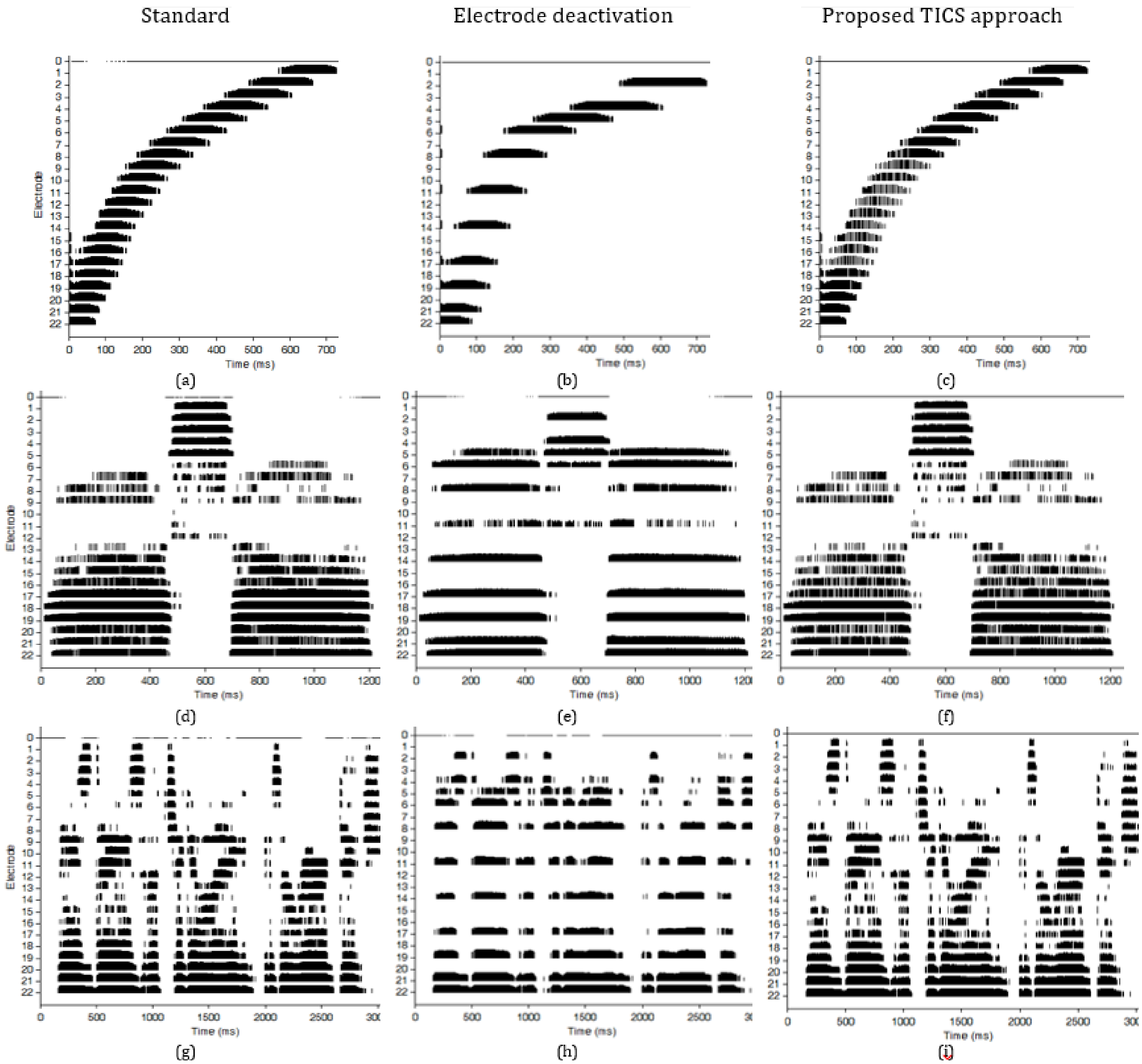


Figure 5.12. Electrodograms of (a) – (c) chirp signal; (d) – (f) consonant “ASA”; and (g) – (i) IEEE speech sentence obtained from three channel selection schemes. Column 1: standard – all electrodes are on; Column 2: electrode de-activation strategy – switch off electrodes that are likely to cause channel interaction; and Column 3: Proposed Time-Interleaved Channel Selection (TICS) scheme for the DVF curves shown in Figure 5.11, Example 2.

5.4.4 Summary/Discussion

Image-guided procedures were employed in a new customized channel selection processes based on the individual electro-neural interface on a subject-by-subject basis. An image-guided time-interleaved channel selection algorithm was proposed. This technique identified electrode pairs that were likely to cause channel interaction with neighboring electrodes from patients' CT scans, and channel selection scheme was customized in a way to ensure that the pair (or set of adjacent electrodes) were not activated in the same stimulation cycle. The channel selection scheme was successfully modified to accomplish time-interleaved activation of problematic electrodes. In theory, the algorithm works in a similar way to the standard CIS approach [44] (sequential stimulation), but here we utilized image-guided procedures to further optimize the channel selection process to stimulate only those electrodes that are presenting independent and non-overlapping information. It is also quite similar to the MP3000 strategy [49], which uses psychoacoustic masking models to optimize the channel selection process, and selects perceptually relevant channels for stimulation. While MP3000 algorithm aims to reduce the spread of excitation by reducing the number of clusters (neighboring channels) that are selected in a typical *n-of-m* approach, it does not account for spatial location of electrodes and their interaction in the physical space on individual to individual basis. The techniques proposed in this chapter could potentially be used as a user-specific channel selection strategy that could potentially work in conjunction with customized frequency mapping scheme discussed in the Chapter 4 to overall improve sound processing for implant users.

CHAPTER 6

SUMMARY AND CONCLUSIONS

The scope of cochlear implant research has expanded significantly over recent years, which is revealing newer insights and better understanding of hearing mechanics and the auditory system, as a whole. Although research is advancing at a high pace, performance levels with implants have only made incremental advancements in the last decade. Large variability in outcomes, limitations inherent in the design of current generation of multi-channel scala tympani implants, and exceeding evidence on the dominant role of higher level cortical functions in implant outcomes, raises a critical thoughtful question: “Have we done enough at the periphery?” This dissertation has been aimed at exploring customized processing and fitting paradigms for cochlear implants which may serve as a better interface between bottom-up and top-down processing. By devising image-guided patient-specific frequency-place functions, and channel activation strategies, we have attempted to minimize the mismatch between stimulation patterns generated by implants and those that occur naturally at the spiral ganglion, with the hope that the central pattern recognition will have “less of a cognitive effort” in the decoding of artificial stimuli. Therefore, this dissertation could potentially serve as a first step in this direction.

6.1 Key Contributions

Specifically, the thesis contributions stemming from this study may be divided into two main areas. 1) image-guided patient specific frequency mapping for cochlear implants, and 2) optimization of channel selection schemes for cochlear implant sound processing. A summary from both aspects are presented in the next sections, followed by suggestions for future work.

6.1.1 Image-guided customization of frequency-place functions

A lack of knowledge on the spatial relationship between electrode locations and the corresponding stimulation sites has resulted in a generic *one-size-fits-all* frequency mapping paradigm with the hope that CI users will learn to adapt to the incorrect frequency locations of stimulation. Suboptimal electrode array placement, variations in insertion depth, and exact positioning and proximity of electrodes to nerve fibers can all, and do, result in mismatch between intended and actual pitch perception. This frequency mismatch holds potential for reducing the efficacy of coded speech information to the auditory cortex and, consequently, limits speech recognition. In this dissertation, we have proposed a patient-specific frequency assignment strategy which helps to minimize sub-optimal frequency-place mapping distortions in CIs. The algorithm leverages image-guided procedures to determine the true location of individual electrodes with respect to the nerve fibers and tailor-fits a frequency place function based on an individuals' electrode-neuro interface.

Algorithm

The proposed strategy developed in Chapter 4, was shown to utilize pre and post implantation CT scans of the recipients' cochleae to determine the precise spatial location of electrode contacts and their corresponding neural stimulation sites and thus generate an optimal user-customized frequency-place function which is used to derive frequency characteristics of the filterbanks. This is achieved by maximizing the frequency match at lower frequencies (frequency range of first three formants), and introduced a mild compression as needed to avoid truncation (e.g., due to shallow insertion). Mid and high frequency bands were assigned along a conventional logarithmic filter spacing.

Simulation Study with normal hearing listeners

Performance of the proposed strategy was evaluated with 42 normal hearing (NH) listeners using acute acoustic simulations of a cochlear implant with actual electro-neuro image maps of CI users. The simulation data indicated significantly better speech recognitions scores as compared to the default clinical mapping scheme on all measures of speech. Although the improvements were observed for all image maps that had various degrees of frequency mismatch, the proposed strategy produced significant improvements in moderate-to-extreme frequency-place mismatch conditions.

Perceptual Adaptation – Simulation study

Since acute simulation scores may underestimate the true potential effects of learning and user-adaptation to speech with a reduced and degraded set of spectral cues, we also investigated the proposed technique in a semi-chronic paradigm. Ten normal-hearing listeners were provided with approximately four hours of auditory training and tested at different intervals with both the default frequency mapping scheme and the proposed custom mapping strategy. On all measures of speech, all participants showed significantly better performance with the proposed custom mapping strategy, at least within the four hour test period. The data indicated that listeners adapt to both schemes, but the level of accommodation and final level of performance with the proposed scheme was still significantly better than the default clinical strategy.

The above two simulation studies with normal hearing subjects served as a viable proof-of-concept for the follow-on investigation with cochlear implant users.

Evaluation with cochlear implant listeners.

Five experienced, post-lingually deafened adult CI users participated in this semi-chronic study. Imaging data of the recipients' cochleae provided the needed relationships between spatial location of electrode contacts and the characteristic frequencies of the nerve fibers. These physiological blueprints were used to optimize and "tailor-fit" frequency-place functions for each participant.

Each participant's performance was measured on ten speech recognition tasks with their clinically assigned frequency map, as well as with the proposed custom frequency assignment scheme. Participants used the experimental program for three months and their performance levels were measured at three stages during the study, i) acutely, ii) after 1 week, and iii) after 3 months. Consistent with typical clinical observations, performance levels with the experimental program dropped significantly lower than the clinical processor for all CI users. However, all subjects showed progressive improvement with extended use of the experimental programs on all measures of speech recognition. The performance levels improved after 1-week post-activation, and were not significantly different than their baseline scores with the clinical maps. At three months, performance levels with the experimental programs were significantly better than the acute scores and were also not statistically different than their original clinical strategy. Overall the progressive improvement with the experimental programs indicated the effects of learning and at least partial adaptation to the custom frequency maps. By the end of the study, all participants chose to keep the experimental program either exclusively, or with their old clinical programs.

Perceptual studies with CIs suggest that performance levels continue to improve at least up to two years post-activation of new maps. The data from this study indicates that patient-centric optimization of frequency fitting may hold potential for improving implant outcomes, particularly for recipients with moderate to high degree of frequency-place mismatch. Pitch percepts elicited by cochlear implants that are aligned or not drastically different from normal cortical acoustic map could improve the bottom-up presentation of acoustic cues that may potentially lead to overall better speech perception.

6.1.2 Optimization of channel selection in sound coding strategies

The next major area for thesis contribution was based on channels selection process in the sound processing strategies for cochlear implants. The channel selection process in *n-of-m* sound coding strategies, abundantly used in commercial cochlear implants, plays a critical role in overall listening performance. Two schemes were proposed to optimize the channel selection process. The first scheme considered instantaneous signal to noise levels to give preference to channels that were dominant in speech information rather than noise. In the second scheme, image guided procedures were utilized to ensure activation of electrode subsets that could potentially minimize channel interaction.

Strategy 1: Formant and SNR-based channel selection optimization

This study considered two approaches for improving the channel selection process of existing spectral maxima sound coding algorithms for cochlear implant systems. The first approach assigns priority to channels containing formant frequencies, while the second approach adaptively assigns weights to each time-frequency (TF) unit based on the estimated SNR in each stimulation cycle. Two types of maskers, binary and soft, were explored to assign effective attenuation factors to each TF unit based on SNR. Both approaches were evaluated independently and synergistically with 3 CI users. Intelligibility scores indicated significant improvement at low SNR levels, with speech at 5 dB SNR and speech masked by reverberation and noise. Quality scores revealed that the noise masking approach was greatly preferred over standard program. The results from this study indicated the potential of the proposed technique to improve speech perception of CI users in adverse listening environments.

Strategy 2: Image-guided customization of electrode activation schemes

Image-guided procedures were employed in a new customized channel selection processes based on the individual electro-neural interface on a subject-by-subject basis. An image-guided time-interleaved channel selection algorithm was proposed. This technique identified electrode pairs that were likely to cause channel interaction with neighboring electrodes from patients' CT scans, and channel selection scheme was customized in a way to ensure that the pair (or set of adjacent electrodes) were not activated in the same stimulation cycle. The channel selection scheme was successfully modified to accomplish time-interleaved activation of problematic electrodes.

6.2 Future Work

The work presented in this dissertation is a first towards customizing sound processing and fitting for implant users. There is a huge potential to expand/improve the current work and to further explore research avenues which can help create customized solutions for implant recipients. The following directions can be considered for future research activities:

Chronic evaluation of customized frequency maps

The customized frequency fitting strategy discussed in Chapter 4 was evaluated with groups of normal hearing and cochlear implant recipients in acute and semi-chronic studies. In order to assess the true potential of algorithms and strategies which rely on user-adaptation and reorganization of the higher level cortical functions, long-term, chronic evaluations are a necessity.

The simulation study conducted with normal hearing listeners only considered a four hour training period. It is very likely that listeners will continue to show further adaptation with extensive training. In order to find the asymptotic levels of performances with different frequency-place functions, a long-term training protocol should be administered. Both clinical and custom frequency strategies should be evaluated with a larger sample size, both to improve the statistical significance, as well as make one-to-one comparisons.

Clinical evaluation of the proposed custom frequency maps was conducted with five CI recipients in a three-month semi-chronic study. Perceptual studies with CIs suggest that performance levels continue to improve at least up to two years post-activation of new maps. Therefore, it is important to assess the long-term adaptation trends with custom frequency mapping approach in order evaluate the full potential of the proposed strategy. A minimum study period of six to twelve months, and more preferably twenty-four months, should be considered for chronic evaluation. A longer-time period is essential because higher level cortical re-organization requires time to settle and learn to make use of the sparse patterns of neural activity provided by the CIs. It may be argued that the proposed approach should be relatively easier to accommodate to since it is more close to the normal cortical acoustic map; however, it should be understood that cortical representation of frequency space in the experienced cochlear implant listeners recruited in our study most likely had been already reorganized to match the distorted representation presented by the clinical maps. For such a case, learning to reinterpret a better (undistorted) information is akin to taking a step back and relearning. This re-learning will require to time reach full benefit with experienced implant users. Also, the cortical plasticity may vary from individual to individual. Some listeners may learn to adapt quicker than others. All these factors must be considered when conducting studies with post-lingually deafened and experienced implant users.

Start from a blank slate

The frequency-fitting strategy proposed in this dissertation was evaluated with experienced implant users who had been using the clinical map for at least over a year. It is very likely that implant users would have adapted to the distorted spectral representation over time. In order to make fair (apples-to-apples) comparisons, a clinical evaluation of the strategy should be conducted with new implant recipients, who have no experience with electrical stimulation. Two groups of implant users should be considered. The first group should be assigned clinical frequency maps, whereas sound processors of the second group should be configured with custom frequency map. It would be important to have equal distribution of electrode insertion depths, as well as audiological parameters (length of deprivation, age, etiology of hearing loss, etc.) in both groups for fair comparisons. Speech recognition performance should be assessed in a long-term clinical trial of twelve to twenty-four months.

Increase the sample size

The logistics of running chronic clinical studies with implant recipients are understandably demanding. However, a larger sample size is needed not only for increasing the statistical significance, also to better understand the trends and factors that may influence outcomes. Huge variability observed with implant outcomes makes this even more important. An interesting study would be the one which would consider groups of implant recipients with low, moderate and extreme tonotopic mismatch, and evaluate the proposed frequency-mapping strategy with these groups.

Systematic evaluation of frequency-place mapping in simulations

The simulations conducted with frequency-place map using normal hearing listeners considered actual image-maps of cochlear implant recipients. While numeric improvements were observed for all tested maps (with normal, moderate, and extreme mismatch) with the customized frequency maps as opposed to the clinical maps, largest improvements were observed in conditions of extreme tonotopic mismatch. It would be interesting to conduct a systematic evaluation of the proposed scheme (in simulations, both acutely and semi-chronically) that would investigate trends in performance outcomes as a function of insertion depth and make comparison with the clinical maps. This data would further elucidate the potential benefit from the proposed approach with various degrees of frequency-place mismatch that arise due to variations in insertion depth alone.

Intermediate maps for cochlear implantation

A recent study conducted by Svirsky *et al.* [100] found that gradual accommodation to auditory mismatch decreased the accommodation time to spectrally-degraded, frequency shifted input.

They tested two groups of implant recipients. The first group was exposed to the clinical map which had extreme tonotopic mismatch, while the second group was given an intermediate map with lower frequency-place mismatch, prior to shifting to the clinical map. By the end of the study, they found that the learning process was faster for listeners who were exposed to gradual approach. These results suggest that gradual rather than sudden exposure may facilitate perceptual learning in the face of a spectrally degraded, frequency-shift input. The evidence from this study could reciprocate with customized frequency-mapping paradigm presented in this dissertation. The proposed frequency fitting strategy aimed at minimizing the frequency-place distortions by achieving a balance between frequency matching and frequency compression to provide a better representation of the sound signal. The maps constructed with the proposed strategy could be considered as intermediate frequency maps before shifting to clinical maps that provide a complete acoustic range compressively. Future research could investigate if customized maps could decrease the accommodation time and improve the final level of performance with the implants.

Improvements in the imaging processing techniques

The image-guided strategies presented in this dissertation rely on image processing techniques developed by Noble *et al.* [106]. This image analysis procedures involve using pre and post implantation CT scans to determine the spatial location of electrode array and its proximity to the tonotopically mapped modulus. One of the short-comings of this approach is that it assumes a healthy distribution of neural population. In reality, the spiral ganglion cells may not be distributed evenly along the Rosenthal’s canal, especially in a compromised auditory system. Neural atrophy and dead zones may create additional complicacies, such as spectral warping or spectral holes in the perceived sound. Future cochlear imaging research could potentially focus on delineating the fine spiral ganglion cells from the imaging data to find the true distribution of healthy neurons in an implant recipient. The frequency mapping and channel selection procedures mentioned in this dissertation should be modified to address accordingly.

Furthermore, the imaging technique used in this study requires pre and post implantation CT scans. Given that most existing CI users may not have their CT imaging data prior to surgery, these techniques may not be extended to the large existing population. An image analysis technique that could find electrode-nerve relationships of existing CI users would be extremely helpful to expand the benefit of the proposed research to a wider population.

Evaluation of image-guided channel selection strategy with cochlear implant users

The image-guide time-interleaved channel selection (TICS) strategy presented in Chapter 5 of this dissertation was only evaluated in simulations due to logistical problems. The simulation data indicated that TICS approach could potentially be beneficial as opposed to both standard clinical, as well as electrode de-activation strategies. A clinical evaluation of the strategy with cochlear implant users would help establish if the proposed approach could bring meaningful improvements in speech recognition, particularly in adverse listening conditions. The future research could be directed towards the evaluation of the proposed approach in a clinical study with CI users.

Evaluation of SNR-based channel selection optimization and formant strategies with cochlear implant users

The strategies presented in Section 5.3 make use of spectral features (formants) and environment SNR to improve channel selection process in sound processing for cochlear implants. Due to logistics, the strategy was only evaluated with three CI users. Although the sample size was small, the proposed strategy resulted in significant gains in speech recognition performance in adverse listening environments. Future research could focus on evaluating the proposed strategies with a larger sample size, and in various environment types. Furthermore, the strategies could be implemented on a portable real-time sound processor (e.g., UT Dallas Cochlear Implant Research Interface [139]) to evaluate the efficacy of the proposed strategy in field trials and everyday listening environments.

As can be seen, there are numerous directions and possibilities to expand the existing work. The research work presented in this dissertation offers a foundation and presents guidelines for future research directions and advancements in customizing cochlear implant sound processors for implant recipients.

6.3 Concluding Remarks

With current trends in modern medicine leading to personalized medicine, customization of prosthetic devices represents a reasonable and auspicious direction for future advancements to help each recipient with customized, user-centric treatments and rehabilitation strategies. Huge variability observed in performance outcomes with cochlear implantation has long been a challenge for the research community. Patient-specific fitting and sound processing schemes may help poor performers to gain better speech understanding abilities in everyday life. In this dissertation, we have presented two image-guided patient-specific optimization techniques. The first strategy minimized the frequency-place mismatch artifacts that are commonly observed with suboptimal electrode array placement. This was achieved by devising unique frequency-to-place functions for each recipient that better reflect the electrode-neural interface of the implants. The second approach aimed at customizing and optimizing channel selection in sound coding strategies based on environments and spatial electrode interactions. Experimental evidence suggests that customized approaches discussed in this dissertation may improve overall performance levels and potentially lower adaptation times to electric hearing.

The overall aim of the dissertation was to improve the quality of life of cochlear implant users by providing them customized solutions, and hence the ability to gain better sound perception from their devices. The goal has been to improve CI devices for implant users, rather than simply perform a CI procedure and expect implantees to adapt to electrically evoked hearing with hope that they will learn to make use of sparsely coded information. Pitch percepts elicited by cochlear implants that are aligned or not drastically different from normal cortical acoustic map could improve the bottom-up presentation of acoustic cues. The image-guided sound processing and fitting strategies developed in this study may help bridge this gap and provide possibilities for optimizing CI processing for each individual.

APPENDIX A

SPEECH INTELLIGIBILITY WITH ADAPTIVE DYNAMIC RANGE OPTIMIZATION IN ADVERSE LISTENING CONDITIONS

The aim of this study was to investigate the effect of Adaptive Dynamic Range Optimization (ADRO)[10] on speech identification for cochlear implant (CI) users in adverse listening conditions. In this study, anechoic quiet, noisy, reverberant, noisy reverberant, and reverberant noisy conditions were evaluated. Two scenarios were considered when modeling the combined effects of reverberation and noise: (a) noise is added to the reverberant speech, and (b) noisy speech is reverberated. CI users were tested in different listening environments using IEEE sentences presented at 65 dB sound pressure level. No significant effect of ADRO processing on speech intelligibility was observed.

---

[10] The ADRO pre-processing strategy is available in CI devices manufactured by Cochlear Limited and many digital HAs (e.g., HAs manufactured by Interton, Siemens).

A.1 Introduction

Electric hearing possess challenges in terms of mapping the input dynamic range of the acoustic signal (~90 dB) to the limited output electric dynamic range (the range between threshold levels and the maximum comfort levels which could be as low as 5 dB). This emphasizes the need to perform intelligent compression to optimally place the characteristic features of speech in the available limited output range for better intelligibility and quality of coded sounds. Commonly used CI coding strategies such as Continuous Interleaved Sampling (CIS) [44] and Advanced Combination Encoder (ACE) [47, 48] use a global compression scheme (e.g., logarithmic compression) at the output level to compensate for the loudness growth. Adaptive Dynamic Range Optimization (ADRO), on the other hand, adaptively adjusts gains in each frequency band prior to the global compression to optimally utilize the limited output range based on the signal statistics.

ADRO is a multichannel signal equalization strategy to improve the audibility, comfort, and intelligibility of sounds for individuals who use CIs and/or hearing aids (HA) [60, 62]. The strategy uses statistical analysis of acoustic signal to select the most information-rich section of the input dynamic range in multiple frequency channels, and adaptively adjusts the channel gains based on a set of fuzzy logic rules to optimally place the signal in the users' available hearing range. Thus, ADRO aims to make soft sounds more audible and loud sounds more comfortable, and is used in conjunction with sound processing in clinical HA and CI processors as a pre-processing strategy. Clinical studies indicate preference for ADRO over alternative amplification strategies in quiet and various noisy conditions with HA and CI subjects (e.g., see Martin *et al.*, 2001 [140]; James *et al.*, 2002 [141]; Dawson *et al.*, 2004 [61]; Iwaki *et al.*, 2008 [142]).

James *et al.* [141] tested 9 adult cochlear implantees using ACE/SPEAK speech processing strategies (with and without ADRO) in quiet and in noise (multi-talker babble, SNR = 10 and 15 dB). Although significant speech perception improvement (16%) was observed using ADRO for low input level [50 dB sound pressure level (SPL)] in quiet, no significant improvement was seen in noise. Moreover, the environmental sound loudness tests indicated a 59% quality preference for the ADRO program in a majority of the conditions, where only 10% of the time the program without ADRO was preferred (31% of the time, sounds with and without ADRO programs were judged to have the same loudness level).

In a later study by Dawson *et al.* [61], children with CIs (mean age: 10.6 yr) were tested with and without ADRO to establish if young implantees benefit from ADRO preprocessing in the same way as adults. A smaller mean group improvement was observed when testing children with CIs in quiet (50 dB SPL) compared to adults studied by James *et al.* [141]. They concluded that differences in microphone sensitivities for the two groups could be a contributing factor for this observed difference. Although speech perception scores for sentences in noise were not significantly different with and without ADRO for adults [141], the speech perception scores of children improved significantly (single-digit percentage improvement) when using ADRO in noise [61]. This may be due to the wider dynamic ranges and consequently steeper mapping functions seen in children as compared to adults [143]. Children preferred sound coding with ADRO in 46% of the conditions, which is relatively smaller than the preference by adults (59% of conditions).

All studies conducted so far have evaluated ADRO in quiet and/or noisy (multi-talker babble) conditions. However, these two conditions do not represent the naturalistic everyday situations where CI users are challenged to understand speech in the presence of reverberation and noise, individually and in combination. Speech perception scores of CI users drop substantially in reverberant environments when early and late reflections of the direct sound are added to speech, thereby blurring both temporal and spectral characteristics of speech [144]. Unlike reverberation, noise is additive and affects speech in a different and complimentary fashion. Noise masks weak consonants to a greater degree than higher intensity vowels, but unlike reverberation this masking does not depend on the energy of the preceding segments [145]. Therefore, the combined effects of reverberation and noise affect speech intelligibility to a greater degree than either reverberation or noise alone [146].

In the afore-mentioned study, we compared speech intelligibility scores obtained from ten adult CI users in quiet, noisy, reverberant, and noisy + reverberant (where noise and reverberation are simultaneously present) conditions. The main goal of the study was to evaluate the effect of ADRO pre-processing on speech perception of CI users in adverse listening conditions in terms of intelligibility.

A.2 Methods

Subjects and Material

Ten adult post-lingually deafened CI recipients participated in this study. All participants were native speakers of American English who received no benefit from hearing aids pre-operatively. All subjects were paid for their participation. CI users were fitted with the Nucleus 24 multichannel implant devices manufactured by Cochlear Corporation. All listeners used their devices routinely and had a minimum of 3 years experience with their devices. All participants were experienced users of ADRO as it was locked into their everyday MAPs. The detailed biographical data of the CI participants is presented in Table A.1. All subjects had at least 20 active electrodes and a stimulation rate of 900 Hz per channel (except S5 and S6 with 1200 and 500 Hz stimulation rates, respectively).

Table A.1. Demographic data of CI participants in the ADRO study

Subjects	Gender	Age, years	Years implanted	CI processor	Etiology of hearing loss	Sens. level	Average electric dynamic range
S1	M	60	3	N5	Noise	9	38
S2	F	62	7	N5	Unknown	12	21
S3	F	54	4	N5	Unknown	12	48
S4	F	56	3	N5	Hereditary	12	39
S5	M	80	8	N5	Hereditary	12	30
S6	F	60	2	N5	Hereditary	10	10
S7	F	65	4	Freedom	Antibiotics	12	51
S8	M	61	3	N5	Meniere’s Disease	12	45
S9	M	65	3	N5	Hereditary	12	52
S10	M	70	8	N5	Unknown	12	5

IEEE sentences [111]were used as the speech stimuli for testing. The root-mean-square (RMS) level of all sentences was equalized and presented at 65 dB SPL. The reverberant stimuli were generated by convolving the clean signals with measured room impulse responses (RIR) recorded in a 227.46 m3 room [146] with a reverberation time equal to 0.6 s, which is allowable in U.S. classrooms according to ANSI S12.60 standard [147]. The direct-to-reverberant ratio (DRR) of the RIR was -1.8 dB. The distance between the single-source signal and the microphone was 5.5 m, which is beyond the critical distance ($\simeq$1 m).

Speech-shaped noise (SSN) with the same long-term spectrum as the test sentences from the IEEE corpus was used as a continuous (steady-state) masker to generate the noisy signals at a 10 dB SNR level.

The noisy reverberant stimuli were generated using the following model [the masker was added to the reverberant stimuli at a 10 dB reverberant speech signal to noise ratio (RSNR)[11]]:

$$y(n) = \{x(n) * h(n)\} + m(n)$$

where $y(n)$, $x(n)$, $h(n)$, and $m(n)$ denote corrupted signal (by noise and reverberation), anechoic clean signal, RIR, and additive noise, respectively.

The reverberant noisy speech stimuli were generated using the following model (the noise-masked speech at 10 dB SNR was reverberated):

$$y(n) = \{x(n) + m(n)\} + h(n)$$

[11] For generating noisy reverberant stimuli, the reverberant signal served as the target signal in the SNR computation. Hence, we refer to the SNR values in this condition as reverberant signal to noise ratios (RSNR)

Signal Processing

All CI participants used ACE speech coding strategy in their clinical processors (clinical processors were programmed with the users' clinical MAP and configured with and without ADRO for each listening condition). In the ACE coding strategy, the acoustic signal is split into 22 frequency bands by a combination of coefficients produced from an FFT analysis. ADRO dynamically applies channel gains to the output of the frequency bands every 2 ms. Next, "*n maxima*" (bands with highest energy, e.g., eight bands) are selected and compressed through a compression scheme (typically logarithmic compression) to generate current levels in the output dynamic range of the selected (active) electrodes.

ADRO uses four rules to continuously vary the input signal gain in each frequency band. The channel gain adjustments are conducted based on comfort, background noise, audibility, and maximum gain rules. The rules are applied based on the long-term calculated output levels (every 2 ms) using a percentile level estimator with a time constant of 20 dB/s. Three target levels (comfort, background, and audibility) define the dynamic range at each frequency band. The comfort rule reduces the gain if the 98th percentile of the long-term output level is greater than the target comfort level. The background noise rule decreases the gain if the 40th percentile of the long-term output level is greater than the background target level. If the 70th percentile of the long-term output is below the audibility target level, then the audibility rule increases the gain. Finally, the maximum gain rule limits the gain in order not to exceed a pre-determined maximum value (for more details on ADRO algorithm see James *et al.* [141]).

Procedure

Subjects were tested using a clinical CI processor in a double-wall sound-proof booth (Acoustic Systems, Inc.). Recorded sentences were presented in free field at 65 dB SPL. CI listeners were tested unilaterally using the ear with the best performance. Bilateral/bimodal listeners were asked to remove the CI/hearing aid of the contralateral ear during test. The clinical processor was programmed with each individual subject's everyday clinical MAP (e.g., stimulation rate, microphone sensitivity, comfort, and threshold levels) using Custom Sound software developed by Cochlear Limited, and was configured with and without ADRO. All CI listeners used similar compression function with a base level of 4 and Q value of 20. Participants selected their sensitivity settings based on experience with their processors. Institutional review board (IRB) approval and informed consent were obtained from all participants prior to testing.

Subjects participated in a total of ten listening conditions: (1) Anechoic quiet (T60 ≈ 0.0 s), (2) reverberant (T60 = 0.6 s), (3) noisy (SNR = 10 dB SSN), (4) noisy reverberant (T60 = 0.6 s, RSNR = 10 dB), and (5) reverberant noisy (SNR = 10 dB, T60 = 0.6 s) conditions (each with and without ADRO). Twenty IEEE sentences (two lists) were used per condition. None of the lists used was repeated across conditions. The sequence of test conditions was randomized across subjects to minimize any order effects. To achieve a balance test order, half the CI users were tested with ADRO (ACE + ADRO) first, and the other half without ADRO (standard ACE). Evaluations were blind so subjects were not aware which was the ADRO condition. For each testing condition, 20 training sentences (not used in the test sessions) were played to the listener in order to familiarize them with the new condition. Participants were instructed to repeat as many words as they could identify. The responses of each individual were collected and scored off-line based on the number of words correctly identified. All words were scored. The percent correct scores for each condition were calculated by dividing the number of words correctly identified by the total number of words. To avoid listener fatigue, participants were given a 15 min break every 60 min during the test session. The entire test duration for each subject was approximately 4 hours.

A.3 Results

Intelligibility listening tests were conducted in five different environments with and without ADRO. The individual as well as mean speech intelligibility scores for all five conditions are presented in Figure A.1. The intelligibility scores progressively declined with the level of difficulty in test condition, ranging from 96% for clean to 23% in reverberant noisy environment. The mean speech intelligibility difference between ADRO and non-ADRO test conditions varied from a minimum absolute value of 0.44% in clean (ADRO > non-ADRO) to a maximum absolute value of 4.76% in reverberant noisy condition (non-ADRO > ADRO). However, individual variations between ADRO and non-ADRO conditions ranged from −26% to +24%. On average, non-ADRO program performed slightly better (3.23%) than the ADRO program in the most challenging listening conditions (R, NR, and RN from Figure A.1).

Repeated-measures analysis of variance (ANOVA) was performed to assess the effect of environment type and program (ADRO/non-ADRO) on the intelligibility scores with an α factor set to 0.05. Subjects were considered a random (blocked) factor, while environment type and ADRO/non-ADRO conditions were used as the main analysis factors. No statistically significant difference in speech intelligibility was found between ADRO/non-ADRO conditions ($F[1,9] = 0.656$, $p = 0.439$). The interaction between the environment type and ADRO/non-

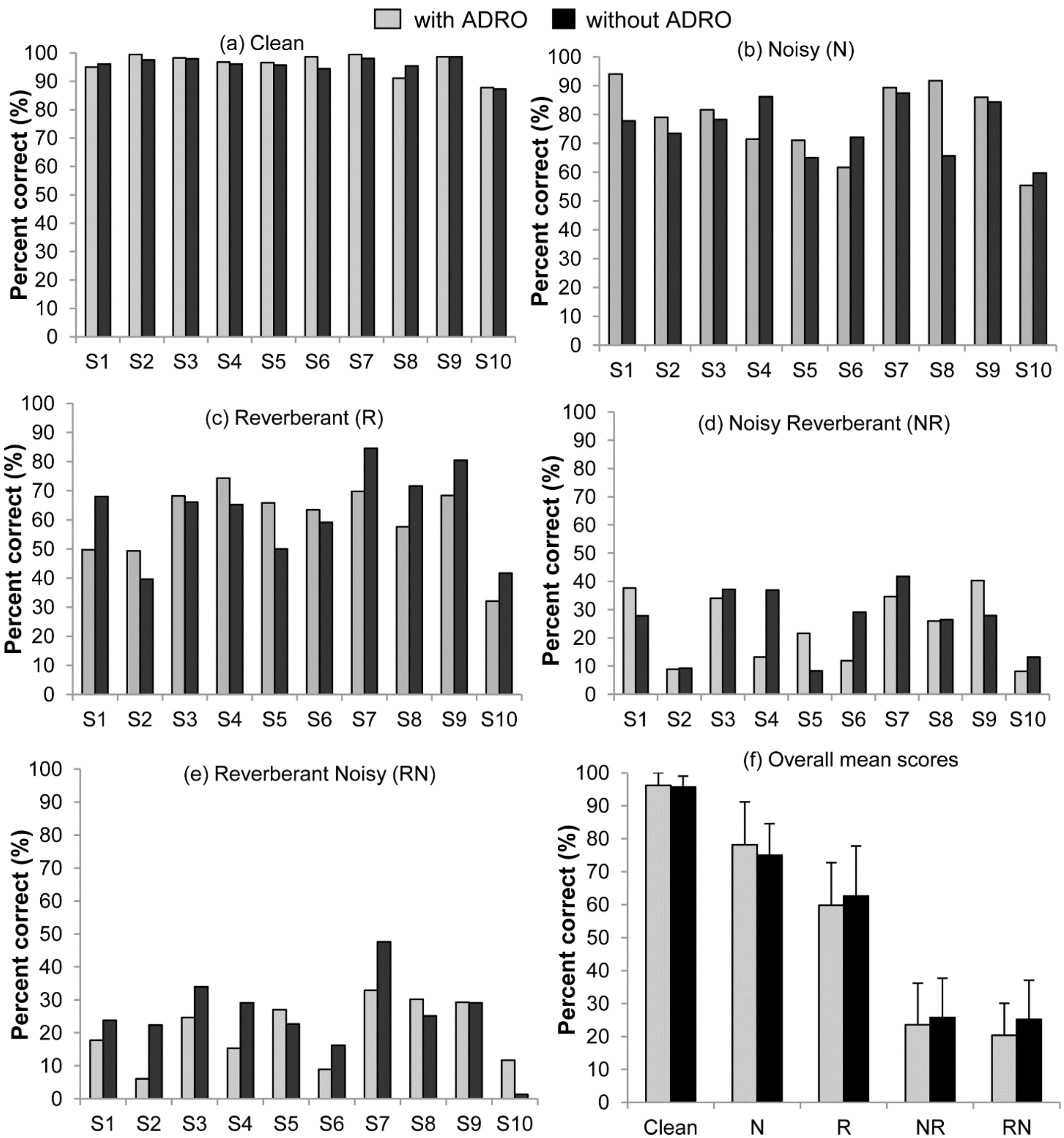


Figure A.1. Individual speech intelligibility scores of ten CI users in (a) anechoic quiet (clean), (b) noisy (N, SNR = 10 dB), (c) reverberant (R, T 60 = 0.6 s), (d) noisy reverberant (NR, T 60 = 0.6 s, RSNR = 10 dB), and (e) reverberant noisy (RN, SNR = 10 dB, T 60 = 0.6 s) conditions. Panel (f) demonstrates average scores in all conditions. The error bars in panel (f) indicate standard deviations.

Table A.2. Correlation coefficients between speech intelligibility and electric dynamic range of CI users in different listening conditions. “ACE” and “ACE + ADRO” stand for standard ACE strategy without and with ADRO program, respectively. Significant correlation values are marked with “*.”

Condition	Clean	N	R	NR	RN	**Mean**
ACE	0.76*	0.66*	0.83*	0.62*	0.85*	**0.74***
ACE + ADRO	0.36	0.82*	0.65-	0.83-	0.81*	**0.69***

ADRO conditions was not significant ($F[4,36] = 0.900$, $p = 0.474$). However, a significant main effect of environment type on speech intelligibility was observed ($F[4,36] = 333.937$, $p < 0.001$). The post hoc Bonferroni test for pairwise comparisons between the five environment types indicated significant differences between all, with the exception of reverberant-noisy and noisy-reverberant environments ($p = 1.000$).

In order to assess the effect of CI users' MAP parameters on speech intelligibility, correlations between the subjects' average electric dynamic range and speech intelligibility scores were computed for the five environment types for ADRO/non-ADRO programs. The results are presented in Table A.2. Speech intelligibility was positively correlated with average electric dynamic range in all five conditions.

A.4 Summary and Discussion

The main goal of this study was to assess the effect of ADRO pre-processing on speech intelligibility for CI users in various listening environments (anechoic quiet, noisy, reverberant, noisy reverberant, and reverberant noisy).

The ADRO strategy was initially developed for bimodal listening and has been previously validated for hearing aids and cochlear implants [60]. Studies by James *et al.* [141]

and Dawson *et al.* [61] indicated that sound quality and speech perception performance were improved using ADRO as compared to fixed channel gains in both adults and children. The later study with children suggested ADRO to be locked into the processor for young children whose MAPs have been stabilized and may be left as an option for the older ones. In line with previous studies, Iwaki *et al.* [142] reported significantly improved speech intelligibility with ADRO for six adult CI users in clean and noisy conditions using Japanese hearing in noise test (JHINT). However, all studies assessing the effect of ADRO pre-processing on speech intelligibility of CI users only considered anechoic quiet and noisy environments. The current study aimed to assess the potential ADRO benefit in everyday realistic environments where reverberation and/or noise can exist individually or in combination.

For all five environment types, our results indicate non-significant speech intelligibility benefit of ADRO over standard ACE program when speech material at 65 dB SPL were presented to CI users. Due to the subjective variability in scores, no clear trend in the pattern of results for either condition/program was found. On average, intelligibility scores for standard ACE program (non-ADRO) were only 1.23% higher than the ACE + ADRO program.

On average, the standard ACE program performed better than the ACE + ADRO program in R, NR, and RN conditions by 3.23%. Seven out of ten subjects had equal or better scores for the non-ADRO program in NR and RN conditions. One of the potential causes which could be attributed to this is that low energy late reflections of the reverberant sound may become amplified by the ADRO strategy as it tends to amplify low-intensity sounds. In such a scenario, ADRO programming may not be beneficial in reverberant environments. Further investigation into how late reflections of the sound are processed in ADRO is required to establish the exact explanation.

Eight out of ten subjects had similar sensitivity settings (12) in their processors. Because of the limited dataset, no relationship between subjects' intelligibility scores and sensitivity settings could be determined in the current study. Positive correlation between the subjects' electric dynamic range and intelligibility scores was observed in all tested conditions, indicating that subjects with a wider dynamic range could be expected to perform better in various listening conditions. This is in line with studies conducted by Loizou *et al.* [148] as well as Fu and Shannon [149].

The present study could not establish any significant benefit with ADRO preprocessing on speech intelligibility in the specific tested conditions. Due to the limited number of participants, their highly variable performance, and similarity in their MAP parameters, no clear trend between the intelligibility scores and their processing parameters (such as stimulation rate and sensitivity level) could be determined. Given that a CI user may or may not benefit with ADRO in different listening environments, ADRO may be left as an optional setting which could be turned on or off according to personal preference of the implant user. Further research is warranted to investigate long-term benefits of ADRO in practical listening environments (reverberation + noise) as well as the effect of ADRO strategy on intelligibility of reverberant speech at both soft and loud presentation levels.

APPENDIX B

SUPPLEMENTARY DETAILS ON STUDY 1: VOCODER-SIMULATIONS WITH NORMAL HEARING INDIVIDUALS

In Section 4.4.1, we presented a study on acoustic simulations of cochlear implants (using noise-band vocoder) with normal hearing listeners to assess the efficacy of customized frequency assignment strategy and compare it against the clinical mapping scheme acutely. The data and analysis provided in the main text only considered the mean results. In this section, we would provide supplementary details on each individual image map, and the resulting performance.

Here is a brief overview of study 1: 42 normal hearing individuals participated. A total of 14 image maps were tested. Each map was tested with three individuals and scores were averaged. Four frequency mapping conditions were evaluated:

Condition#1: Default frequency allocation with ideal electrode position: In this condition, same set of analysis and synthesis filters were used to simulate an ideal scenario in which acoustic frequencies were matched to cochlear place in a matched one-to one scenario. This condition used default frequency allocation table which is used in ACE coding strategy. Filter frequency characteristic for this condition are given in Table 4.1.

Condition#2: Default frequency allocation with true electrode position: In this condition, we try to mimic the actual listening perception experienced by CI users. This is achieved by using the default ACE filterbanks at the analysis stage, and filterbanks derived from DVF curves at the synthesis stage. The resulting signal is, thus, typically mismatched.

Condition#3: Custom frequency allocation with true electrode position: In this condition, custom frequency allocation were used at the analysis stage and filterbanks derived from DVF curves were used as synthesis filters to simulate the perceived sound. Custom filter-banks were designed according to each individual's DVF curve data using the methods from section 4.3.

Condition#4: Frequency allocation matched with true electrode position: Analysis and synthesis filter-banks were chosen identically from the DVF curves. This condition mimics if acoustic filterbanks were chosen to match exactly with electrode positions

All four conditions are summarized in the Table 4.2 and depicted graphically in Figure 4.6. Speech recognition was assessed on four speech material, namely 20-medial consonants, 12-medial vowels, IEEE sentences in quiet, and IEEE sentences in 10 dB SNR speech-shaped noise. The test order of speech material, and mapping conditions was randomized across subjects. This study aimed to assess acute performance with different frequency mapping configurations.

The scores and statistical analysis of the data indicated that performance with the proposed custom frequency assignment strategy was on average better than the clinical mapping. In the following section, frequency characteristics, performance, and statistical analysis of each image map is provided separately.

B.1. MAP01

Table B.1. Frequency allocation tables for MAP01. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	291	238	291	344
2	313	375	438	302	-	-	-
3	438	500	563	397	344	397	483
4	563	625	688	474	-	-	-
5	688	750	813	569	483	569	613
6	813	875	938	656	613	656	695
7	938	1000	1063	733	695	733	799
8	1063	1125	1188	864	799	864	964
9	1188	1250	1313	1063	964	1063	1132
10	1313	1438	1563	1201	1132	1201	1331
11	1563	1688	1813	1284	-	-	-
12	1813	1938	2063	1461	1331	1461	1745
13	2063	2188	2313	1628	-	-	-
14	2313	2500	2688	1834	-	-	-
15	2688	2875	3063	2028	1745	2028	2132
16	3063	3313	3563	2236	2132	2236	2521
17	3563	3813	4063	2559	2521	2806	3456
18	4063	4375	4688	2806	-	-	-
19	4688	5000	5313	3230	3456	4106	4555
20	5313	5688	6063	3795	4555	5003	5534
21	6063	6500	6938	4283	5534	6064	6692
22	6938	7438	7938	4995	6692	7320	7947

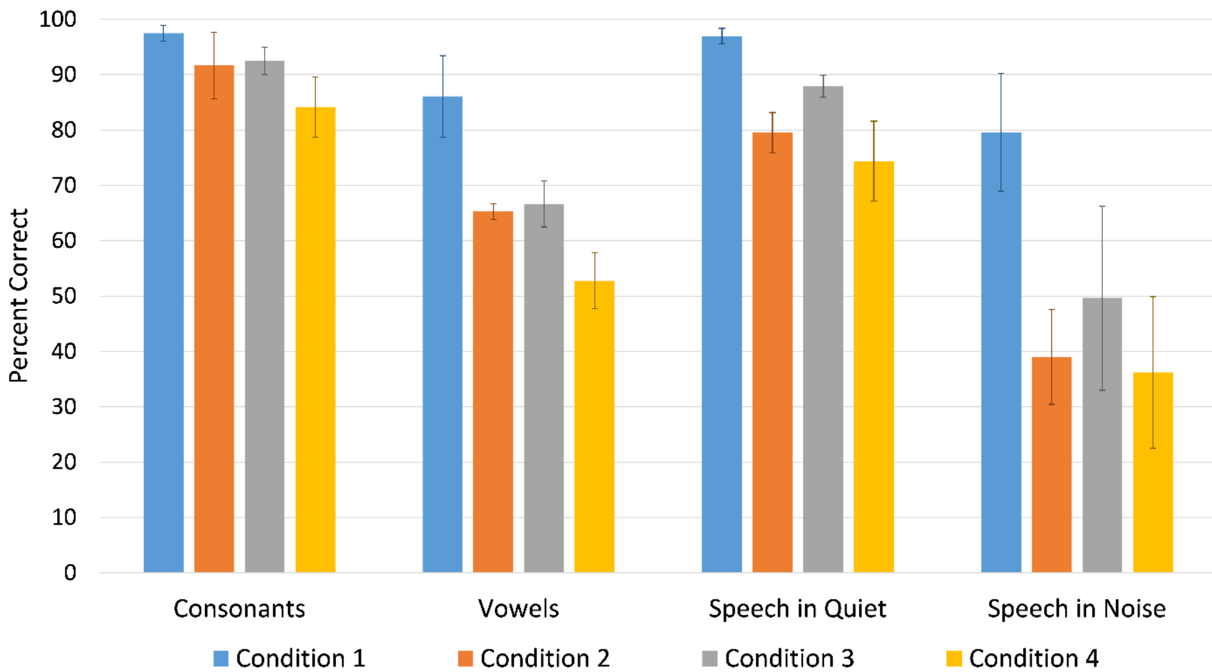


Figure B.1. MAP01 - Average speech recognition scores for MAP01 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	$F[3, 6] = 19.825$	$p = 0.002$
Effect of speech material:	Significant	$F[3, 6] = 10.590$	$p = 0.008$
Interaction (condition × speech material):	Significant	$F[9, 18] = 3.313$	$p = 0.015$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1				
Cond2				
Cond3				
Cond4				

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1				
Cond2				
Cond3				
Cond4				

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1			*	
Cond2				
Cond3	*			
Cond4				

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*			
Cond3				
Cond4	*			

Significantly different pair is marked with asterisk.

B.2. MAP02

Table B.2. Frequency allocation tables for MAP02. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	337	-	-	-
2	313	375	438	422	356	422	489
3	438	500	563	555	489	555	591
4	563	625	688	627	591	627	665
5	688	750	813	702	665	702	775
6	813	875	938	749	-	-	-
7	938	1000	1063	847	775	847	974
8	1063	1125	1188	916	-	-	-
9	1188	1250	1313	1101	974	1101	1182
10	1313	1438	1563	1263	1182	1263	1362
11	1563	1688	1813	1284	-	-	-
12	1813	1938	2063	1461	1362	1461	1803
13	2063	2188	2313	1484	-	-	-
14	2313	2500	2688	1943	-	-	-
15	2688	2875	3063	2145	1803	2145	2336
16	3063	3313	3563	2267	-	-	-
17	3563	3813	4063	2526	2336	2526	2685
18	4063	4375	4688	2843	2685	2843	3474
19	4688	5000	5313	3271	3474	4106	4555
20	5313	5688	6063	3842	4555	5003	5534
21	6063	6500	6938	4655	5534	6064	6692
22	6938	7438	7938	5233	6692	7320	7947

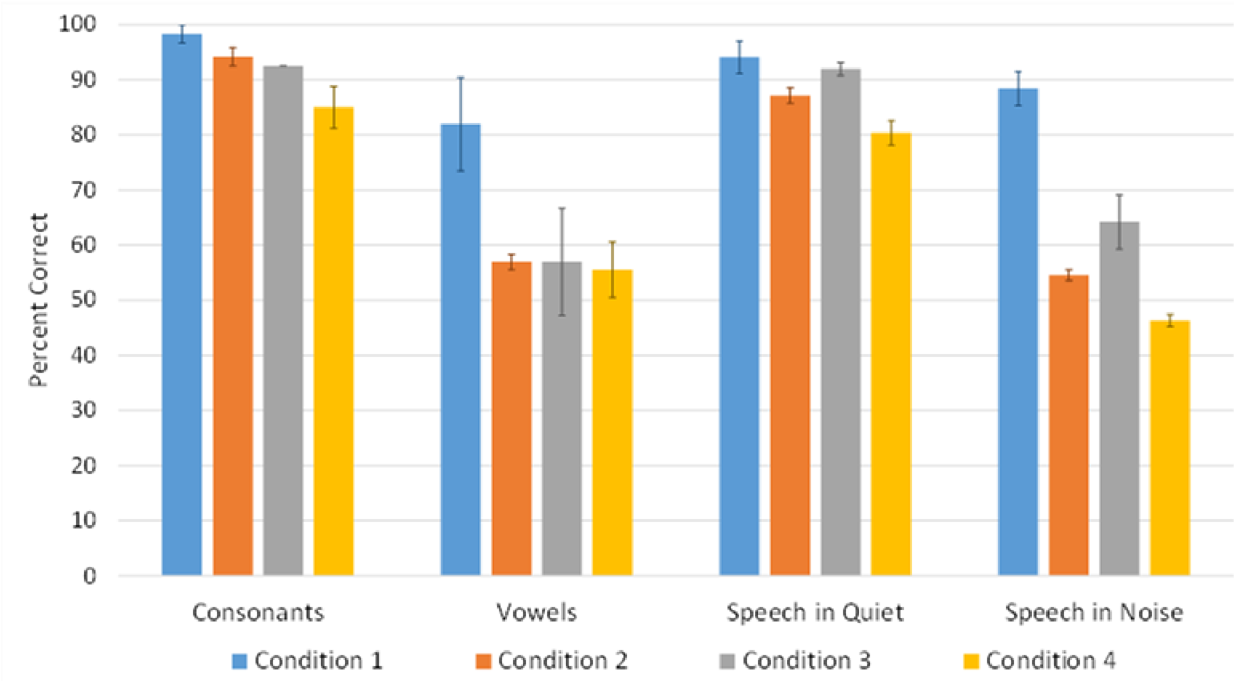


Figure B.2. MAP02 - Average speech recognition scores for MAP02 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 20.372	*p* = 0.002
Effect of speech material:	Significant	F[3, 6] = 58.157	*p* < 0.001
Interaction (condition × speech material):	Significant	F[9, 18] = 4.639	*p* = 0.003

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1				
Cond2				
Cond3				
Cond4				

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1			*	
Cond2				
Cond3	*			
Cond4				

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1				*
Cond2				
Cond3				
Cond4	*			

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			*
Cond3	*			*
Cond4	*	*	*	

Significantly different pair is marked with asterisk.

B.3. MAP03

Table B.3. Frequency allocation tables for MAP03. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	656	500	583	667
2	313	375	438	718	-	-	-
3	438	500	563	765	667	750	833
4	563	625	688	952	833	917	1038
5	688	750	813	1160	1038	1160	1244
6	813	875	938	1327	1244	1327	1429
7	938	1000	1063	1531	1429	1531	1710
8	1063	1125	1188	1888	1710	1888	2078
9	1188	1250	1313	2267	2078	2267	2431
10	1313	1438	1563	2594	2431	2594	2775
11	1563	1688	1813	2955	2775	2955	3265
12	1813	1938	2063	3482	3265	3576	3790
13	2063	2188	2313	3842	3790	4004	4238
14	2313	2500	2688	4232	-	-	-
15	2688	2875	3063	4995	4238	4473	4731
16	3063	3313	3563	5481	-	-	-
17	3563	3813	4063	5872	4731	4989	5272
18	4063	4375	4688	6958	5272	5555	5866
19	4688	5000	5313	8590	5866	6177	6518
20	5313	5688	6063	10450	6518	6859	7234
21	6063	6500	6938	12677	7234	7608	7983
22	6938	7438	7938	12813	-	-	-

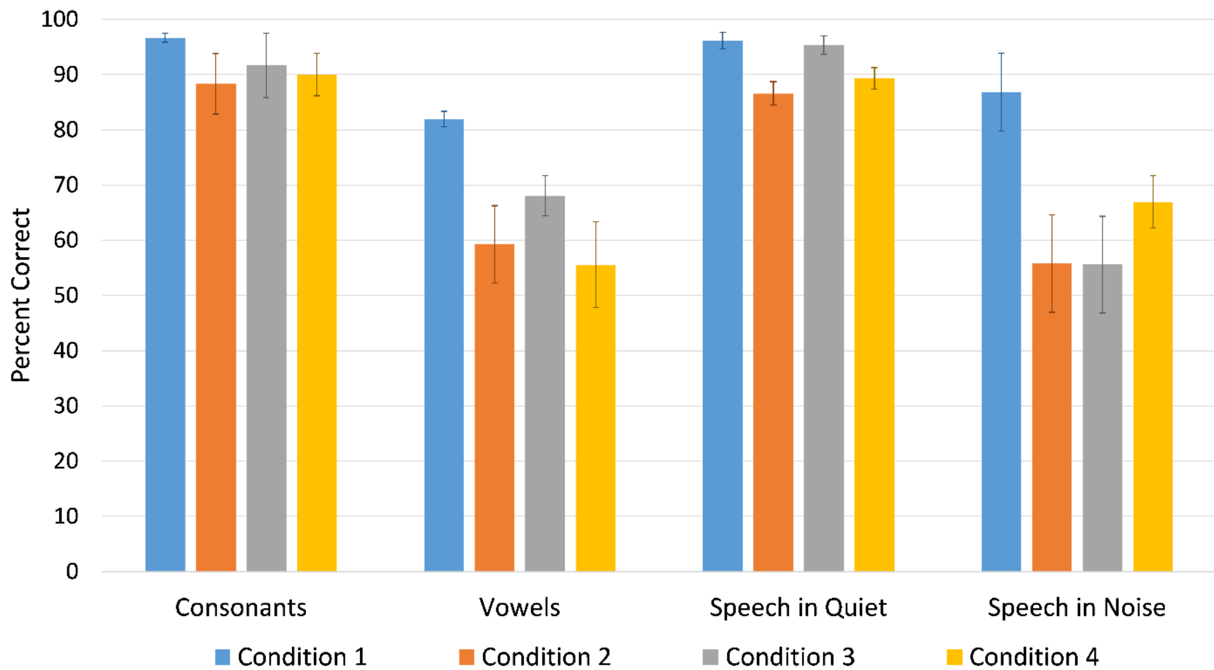


Figure B.3. MAP03 - Average speech recognition scores for MAP03 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	$F[3, 6] = 13.037$	$p = 0.005$
Effect of speech material:	Significant	$F[3, 6] = 13.258$	$p < 0.005$
Interaction (condition × speech material):	Significant	$F[9, 18] = 5.447$	$p = 0.001$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1				
Cond2				
Cond3				
Cond4				

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1			*	
Cond2				
Cond3		*		
Cond4				

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*			*
Cond3				
Cond4	*	*		

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*			
Cond3				
Cond4	*			

Significantly different pair is marked with asterisk.

B.4. MAP04

Table B.4. Frequency allocation tables for MAP04. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	702	500	583	667
2	313	375	438	749	667	750	833
3	438	500	563	952	833	917	990
4	563	625	688	1063	990	1063	1142
5	688	750	813	1221	1142	1221	1296
6	813	875	938	1371	1296	1371	1440
7	938	1000	1063	1508	1440	1508	1740
8	1063	1125	1188	1971	1740	1971	2119
9	1188	1250	1313	2267	2119	2267	2397
10	1313	1438	1563	2361	-	-	-
11	1563	1688	1813	2526	2397	2526	2703
12	1813	1938	2063	2880	2703	2880	3197
13	2063	2188	2313	3525	3197	3514	3703
14	2313	2500	2688	3842	3703	3892	4097
15	2688	2875	3063	4132	-	-	-
16	3063	3313	3563	4766	4097	4303	4526
17	3563	3813	4063	5355	4526	4750	4993
18	4063	4375	4688	5939	4993	5236	5500
19	4688	5000	5313	7693	5500	5765	6052
20	5313	5688	6063	10116	6052	6340	6653
21	6063	6500	6938	12147	6653	6965	7305
22	6938	7438	7938	13950	7305	7646	7986

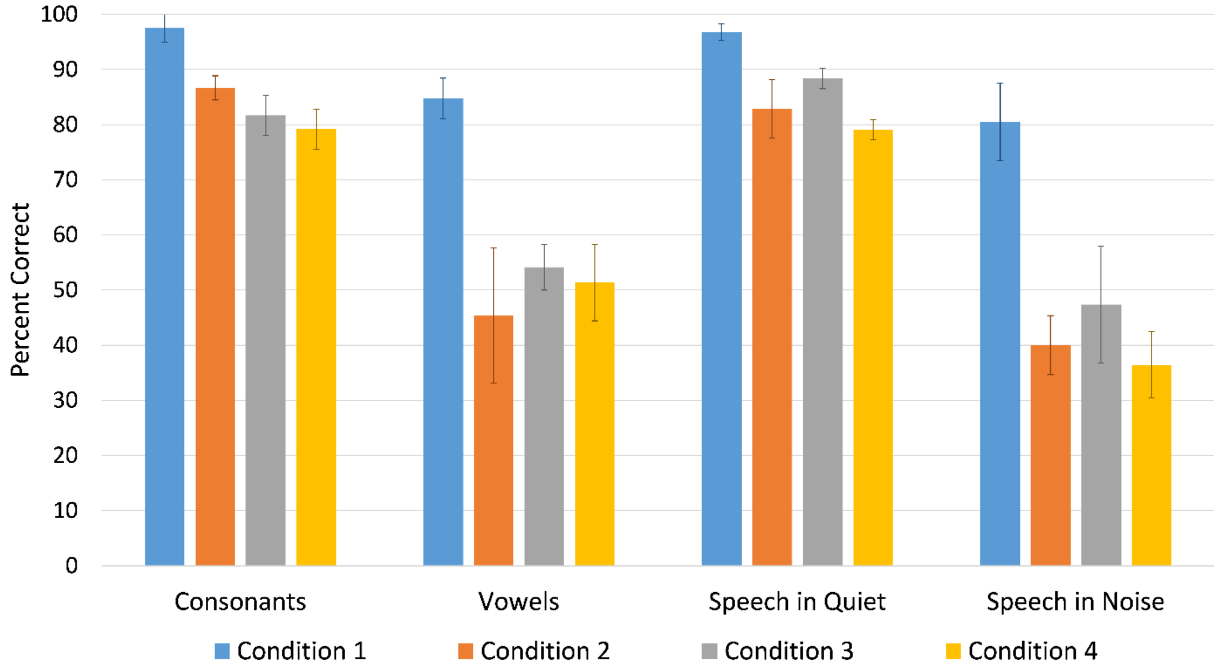


Figure B.4. MAP04 - Average speech recognition scores for MAP04 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 61.697	$p < 0.001$
Effect of speech material:	Significant	F[3, 6] = 13.217	$p = 0.005$
Interaction (condition × speech material):	Significant	F[9, 18] = 2.999	$p = 0.023$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*			
Cond3				
Cond4	*			

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1			*	*
Cond2				
Cond3	*			
Cond4	*			

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1				*
Cond2				
Cond3				*
Cond4	*		*	

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*			
Cond3				
Cond4	*			

Significantly different pair is marked with asterisk.

B.5. MAP05

Table B.5. Frequency allocation tables for MAP05. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	733	500	625	750
2	313	375	438	864	750	875	960
3	438	500	563	1044	960	1044	1143
4	563	625	688	1242	1143	1242	1318
5	688	750	813	1393	1318	1393	1462
6	813	875	938	1531	1462	1531	1710
7	938	1000	1063	1888	1710	1888	2157
8	1063	1125	1188	2115	-	-	-
9	1188	1250	1313	2426	2157	2426	2527
10	1313	1438	1563	2628	2527	2628	3008
11	1563	1688	1813	3071	3008	3389	3529
12	1813	1938	2063	3439	3529	3668	3817
13	2063	2188	2313	3842	3817	3966	4126
14	2313	2500	2688	4232	4126	4285	4454
15	2688	2875	3063	4937	4454	4624	4805
16	3063	3313	3563	5544	4805	4986	5179
17	3563	3813	4063	6146	5179	5372	5578
18	4063	4375	4688	8590	5578	5784	6003
19	4688	5000	5313	10564	6003	6223	6457
20	5313	5688	6063	12409	6457	6692	6942
21	6063	6500	6938	13803	6942	7192	7458
22	6938	7438	7938	15023	7458	7725	7991

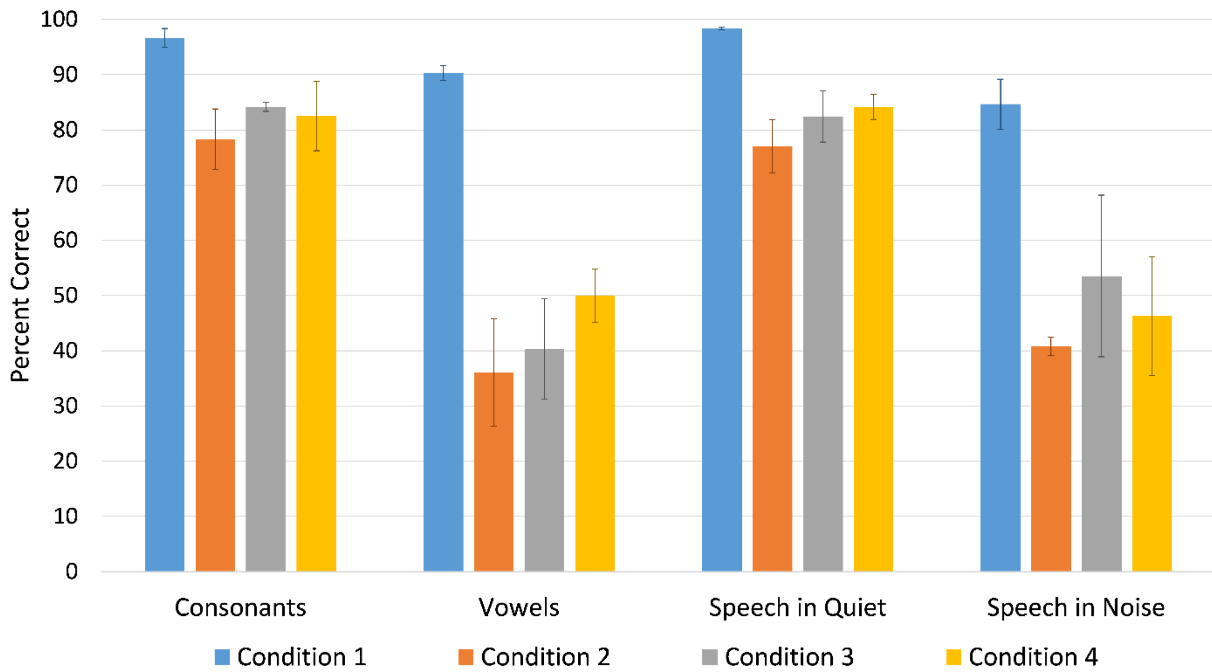


Figure B.5. MAP05 - Average speech recognition scores for MAP05 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	$F[3, 6] = 59.535$	$p < 0.001$
Effect of speech material:	Significant	$F[3, 6] = 33.405$	$p < 0.001$
Interaction (condition × speech material):	Not Significant	$F[9, 18] = 2.171$	$p = 0.077$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1			*	
Cond2				
Cond3	*			
Cond4				

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1				*
Cond2				
Cond3				
Cond4	*			

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*			
Cond3				
Cond4	*			

Significantly different pair is marked with asterisk.

B.6. MAP06

Table B.6. Frequency allocation tables for MAP06. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	765	500	625	750
2	313	375	438	899	750	875	998
3	438	500	563	1121	998	1121	1257
4	563	625	688	1221	-	-	-
5	688	750	813	1393	1257	1393	1523
6	813	875	938	1653	1523	1653	1771
7	938	1000	1063	1888	1771	1888	2002
8	1063	1125	1188	2115	2002	2115	2337
9	1188	1250	1313	2267	-	-	-
10	1313	1438	1563	2559	2337	2559	2683
11	1563	1688	1813	2806	2683	2806	3160
12	1813	1938	2063	3110	3160	3514	3703
13	2063	2188	2313	3525	-	-	-
14	2313	2500	2688	3985	3703	3892	4097
15	2688	2875	3063	4546	4097	4303	4526
16	3063	3313	3563	5172	4526	4750	4993
17	3563	3813	4063	5872	4993	5236	5500
18	4063	4375	4688	7276	5500	5765	6052
19	4688	5000	5313	9376	6052	6340	6653
20	5313	5688	6063	11269	6653	6965	7305
21	6063	6500	6938	12950	7305	7646	7986
22	6938	7438	7938	13950	-	-	-

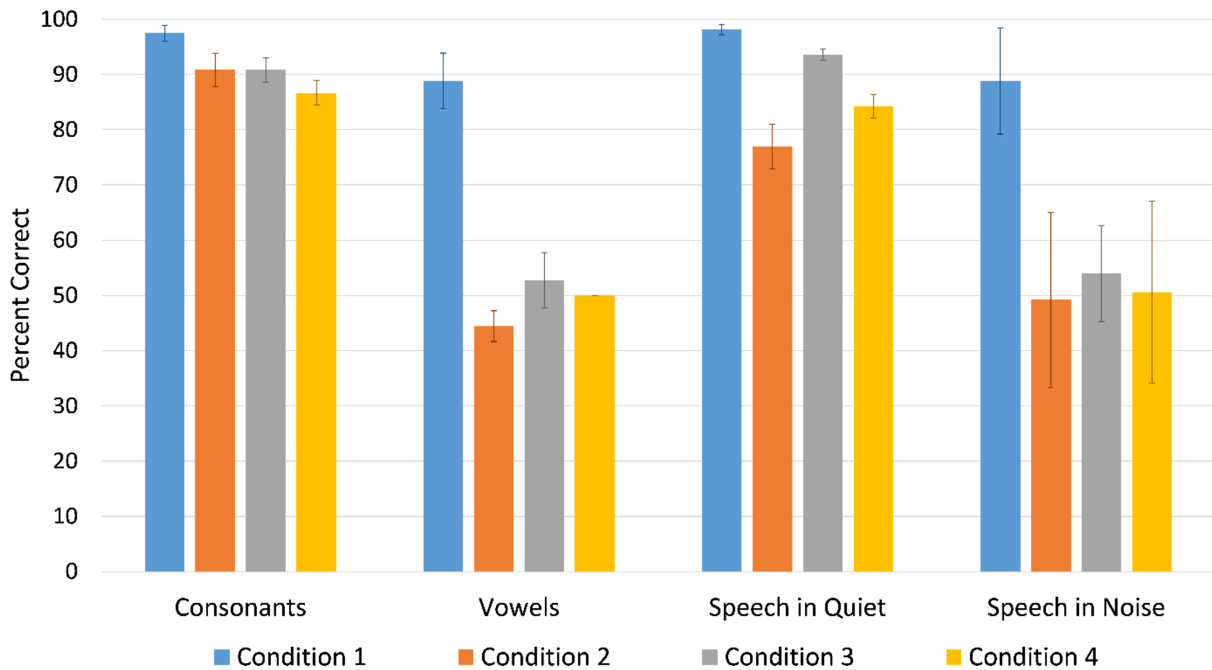


Figure B.6. MAP06 - Average speech recognition scores for MAP06 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	$F[3, 6] = 76.685$	$p < 0.001$
Effect of speech material:	Significant	$F[3, 6] = 8.797$	$p = 0.013$
Interaction (condition × speech material):	Significant	$F[9, 18] = 5.950$	$p = 0.001$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1				*
Cond2				
Cond3				
Cond4	*			

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*		*	
Cond3	*	*		*
Cond4	*		*	

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Significantly different pair is marked with asterisk.

B.7. MAP07

Table B.7. Frequency allocation tables for MAP07. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	797	500	625	750
2	313	375	438	916	750	875	969
3	438	500	563	1063	969	1063	1142
4	563	625	688	1221	1142	1221	1285
5	688	750	813	1349	1285	1349	1417
6	813	875	938	1484	1417	1484	1620
7	938	1000	1063	1755	1620	1755	1921
8	1063	1125	1188	2086	1921	2086	2256
9	1188	1250	1313	2426	2256	2426	2580
10	1313	1438	1563	2734	2580	2734	3061
11	1563	1688	1813	3110	3061	3389	3529
12	1813	1938	2063	3614	3529	3668	3817
13	2063	2188	2313	4232	3817	3966	4126
14	2313	2500	2688	4600	4126	4285	4454
15	2688	2875	3063	5294	4454	4624	4805
16	3063	3313	3563	5872	4805	4986	5179
17	3563	3813	4063	6880	5179	5372	5578
18	4063	4375	4688	8877	5578	5784	6003
19	4688	5000	5313	10678	6003	6223	6457
20	5313	5688	6063	12409	6457	6692	6942
21	6063	6500	6938	13803	6942	7192	7458
22	6938	7438	7938	14865	7458	7725	7991

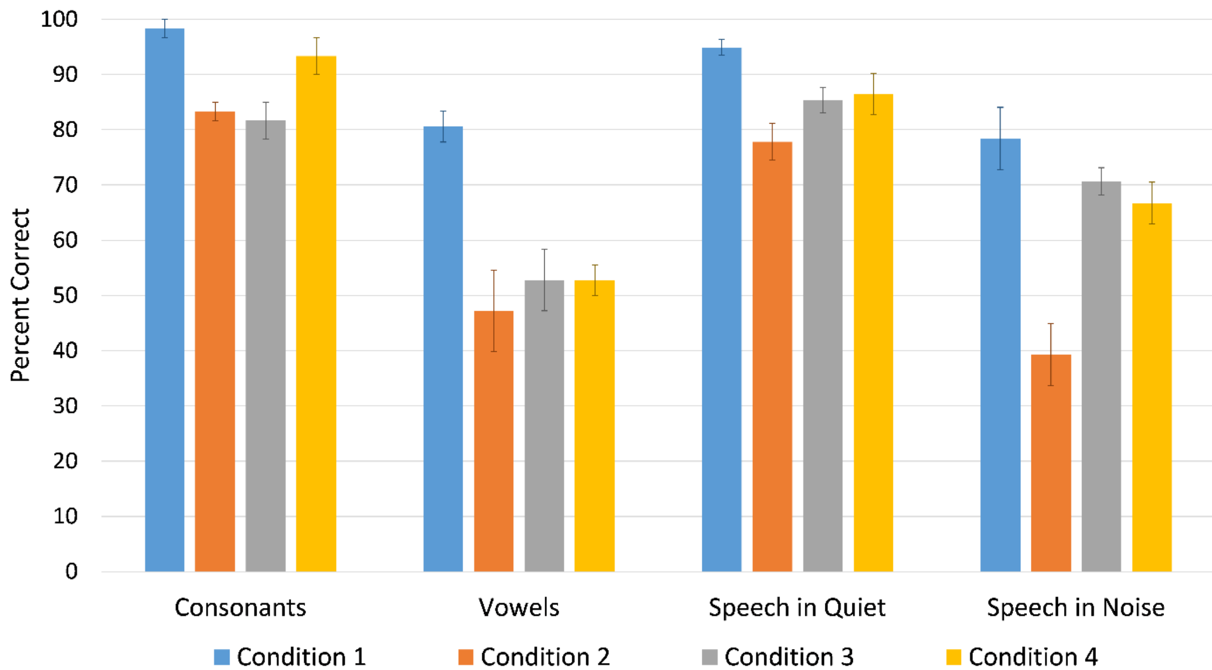


Figure B.7. MAP07 - Average speech recognition scores for MAP07 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	$F[3, 6] = 75.074$	$p < 0.001$
Effect of speech material:	Significant	$F[3, 6] = 105.812$	$p < 0.001$
Interaction (condition × speech material):	Significant	$F[9, 18] = 3.573$	$p = 0.010$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			*
Cond4	*		*	

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*		
Cond2	*		*	
Cond3		*		
Cond4				

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*		*	*
Cond3		*		
Cond4	*	*		

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*			
Cond3				
Cond4	*			

Significantly different pair is marked with asterisk.

B.8. MAP08

Table B.8. Frequency allocation tables for MAP08. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	847	500	625	750
2	313	375	438	970	750	875	979
3	438	500	563	1082	979	1082	1162
4	563	625	688	1140	-	-	-
5	688	750	813	1242	1162	1242	1318
6	813	875	938	1393	1318	1393	1427
7	938	1000	1063	1461	1427	1461	1570
8	1063	1125	1188	1678	1570	1678	1839
9	1188	1250	1313	1999	1839	1999	2165
10	1313	1438	1563	2330	2165	2330	2532
11	1563	1688	1813	2734	2532	2734	3079
12	1813	1938	2063	3110	3079	3423	3576
13	2063	2188	2313	3704	3576	3729	3893
14	2313	2500	2688	4387	3893	4058	4234
15	2688	2875	3063	5233	4234	4410	4599
16	3063	3313	3563	6007	4599	4787	4990
17	3563	3813	4063	7693	4990	5192	5409
18	4063	4375	4688	10227	5409	5627	5860
19	4688	5000	5313	11890	5860	6093	6342
20	5313	5688	6063	13513	6342	6592	6860
21	6063	6500	6938	14865	6860	7128	7415
22	6938	7438	7938	15836	7415	7703	7990

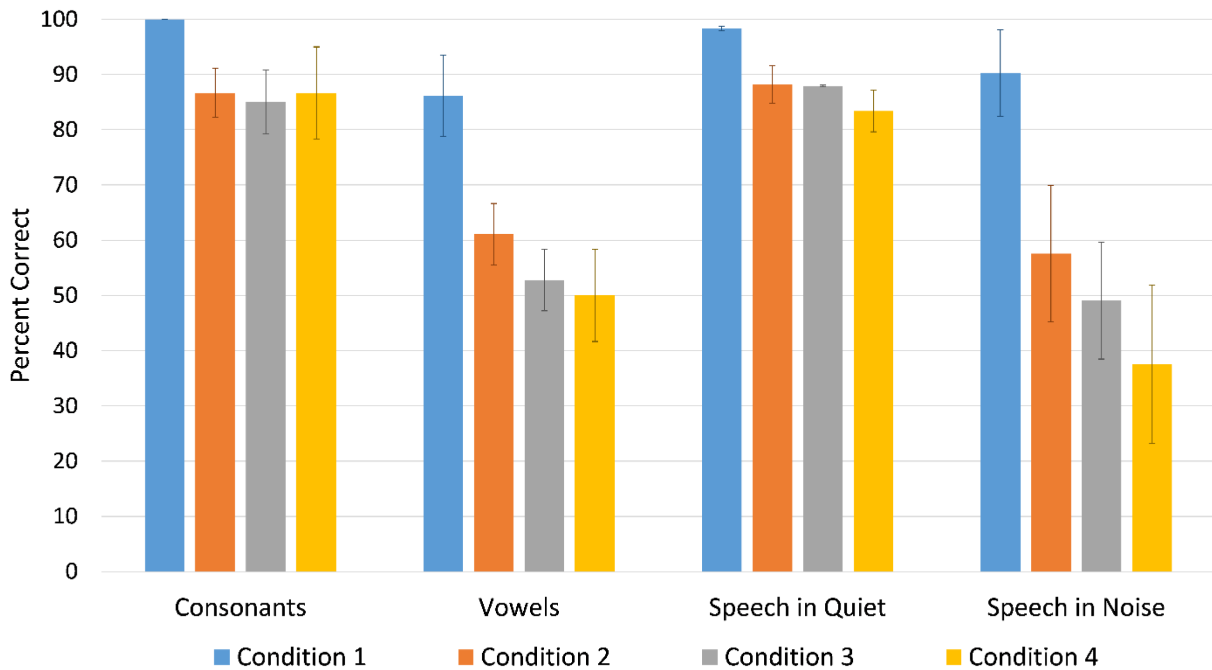


Figure B.8. MAP08 - Average speech recognition scores for MAP08 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 14.465	$p = 0.004$
Effect of speech material:	Significant	F[3, 6] = 35.955	$p < 0.001$
Interaction (condition × speech material):	Significant	F[9, 18] = 6.821	$p < 0.001$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1				
Cond2				
Cond3				
Cond4				

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	
Cond2	*			
Cond3	*			
Cond4				

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Significantly different pair is marked with asterisk.

B.9. MAP09

Table B.9. Frequency allocation tables for MAP09. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	847	500	625	750
2	313	375	438	1006	750	875	1000
3	438	500	563	1140	1000	1125	1250
4	563	625	688	1305	1250	1375	1453
5	688	750	813	1531	1453	1531	1630
6	813	875	938	1729	1630	1729	1893
7	938	1000	1063	2057	1893	2057	2258
8	1063	1125	1188	2298	-	-	-
9	1188	1250	1313	2459	2258	2459	3017
10	1313	1438	1563	2663	-	-	-
11	1563	1688	1813	2843	-	-	-
12	1813	1938	2063	3354	3017	3576	3790
13	2063	2188	2313	3704	-	-	-
14	2313	2500	2688	3889	-	-	-
15	2688	2875	3063	4132	3790	4004	4238
16	3063	3313	3563	4822	-	-	-
17	3563	3813	4063	5233	4238	4473	4731
18	4063	4375	4688	5805	4731	4989	5272
19	4688	5000	5313	6728	5272	5555	5866
20	5313	5688	6063	8975	5866	6177	6518
21	6063	6500	6938	11148	6518	6859	7234
22	6938	7438	7938	12813	7234	7608	7983

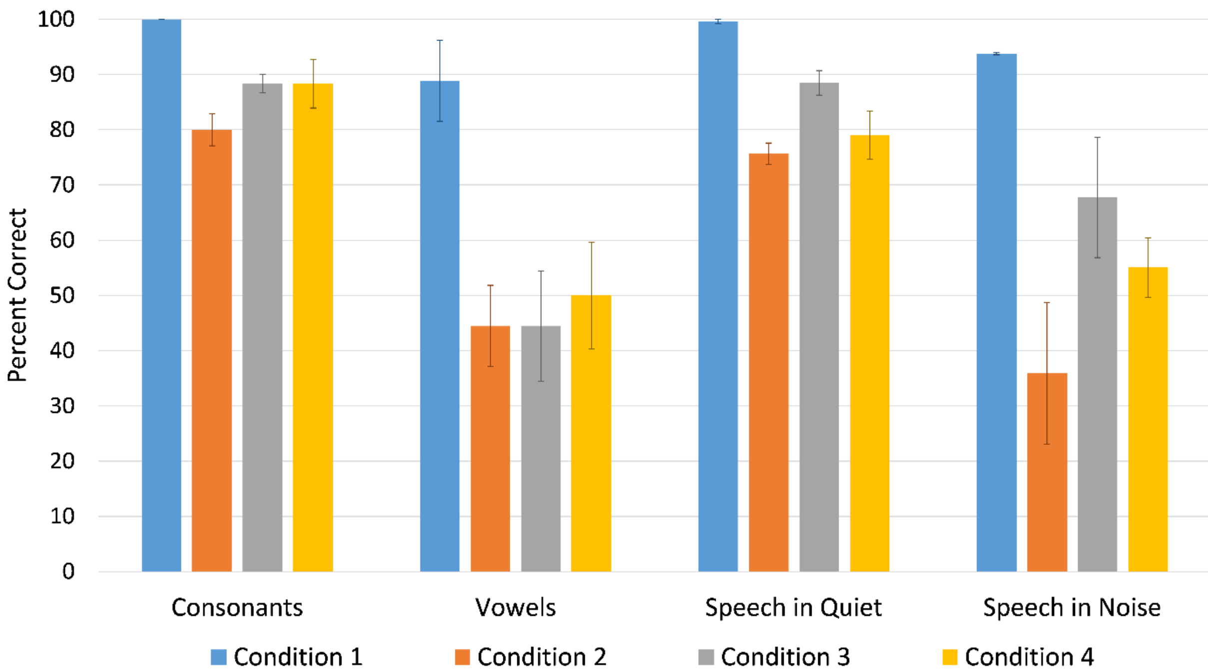


Figure B.9. MAP09 - Average speech recognition scores for MAP09 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 45.978	*p* < 0.001
Effect of speech material:	Significant	F[3, 6] = 31.278	*p* < 0.001
Interaction (condition × speech material):	Significant	F[9, 18] = 6.289	*p* < 0.001

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*			
Cond3				
Cond4	*			

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*		*	*
Cond3		*		
Cond4	*	*		

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*		*
Cond2	*		*	
Cond3		*		
Cond4	*			

Significantly different pair is marked with asterisk.

B.10. MAP10

Table B.10. Frequency allocation tables for MAP10. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	864	500	625	750
2	313	375	438	1082	-	-	-
3	438	500	563	1121	750	875	1063
4	563	625	688	1284	-	-	-
5	688	750	813	1484	1063	1250	1542
6	813	875	938	1834	1542	1834	2114
7	938	1000	1063	2086	-	-	-
8	1063	1125	1188	2394	2114	2394	2929
9	1188	1250	1313	2663	-	-	-
10	1313	1438	1563	3354	2929	3464	3633
11	1563	1688	1813	3704	-	-	-
12	1813	1938	2063	4232	3633	3802	3985
13	2063	2188	2313	4995	-	-	-
14	2313	2500	2688	5418	3985	4168	4365
15	2688	2875	3063	6007	4365	4562	4774
16	3063	3313	3563	7036	4774	4987	5216
17	3563	3813	4063	10450	5216	5446	5694
18	4063	4375	4688	12542	5694	5942	6209
19	4688	5000	5313	14708	6209	6476	6765
20	5313	5688	6063	16004	6765	7054	7365
21	6063	6500	6938	16173	-	-	-
22	6938	7438	7938	16691	7365	7677	7988

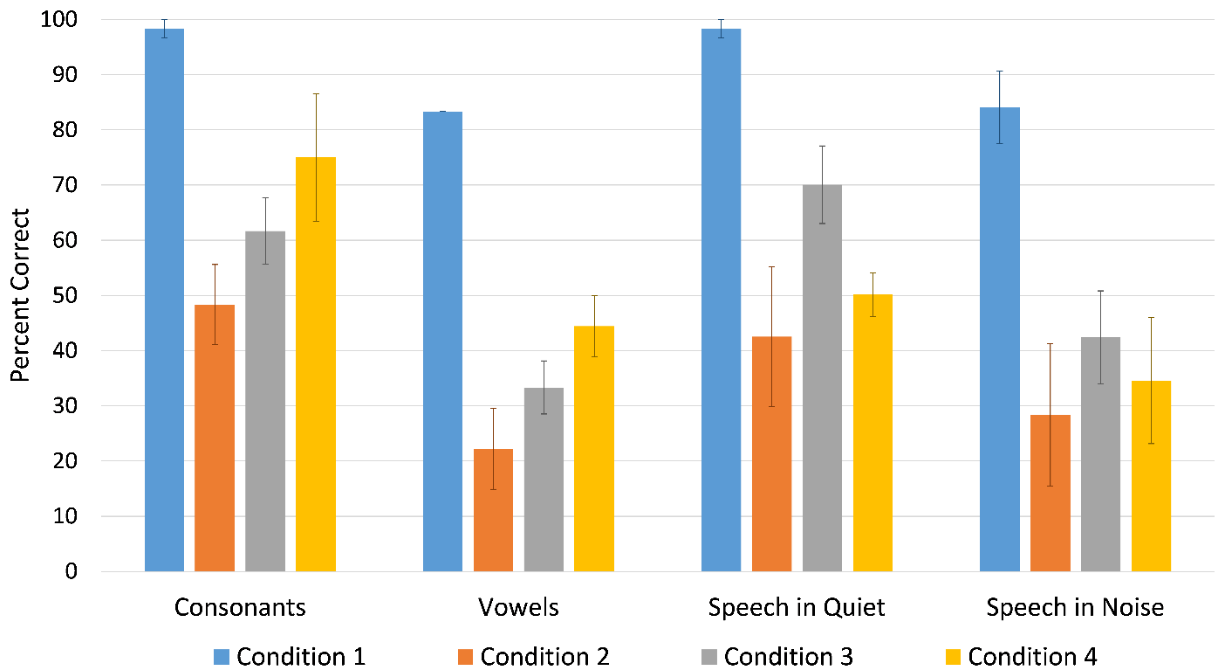


Figure B.10. MAP10 - Average speech recognition scores for MAP10 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 77.073	$p < 0.001$
Effect of speech material:	Significant	F[3, 6] = 48.609	$p < 0.001$
Interaction (condition × speech material):	Significant	F[9, 18] = 3.038	$p = 0.021$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	
Cond2	*			*
Cond3	*			
Cond4		*		

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			*
Cond3	*			
Cond4	*	*		

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*		*	
Cond3	*	*		
Cond4	*			

Significantly different pair is marked with asterisk.

B.11. MAP11

Table B.11. Frequency allocation tables for MAP11. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	864	500	625	750
2	313	375	438	970	750	875	979
3	438	500	563	1082	979	1082	1142
4	563	625	688	1201	1142	1201	1264
5	688	750	813	1327	1264	1327	1418
6	813	875	938	1508	1418	1508	1606
7	938	1000	1063	1703	1606	1703	1837
8	1063	1125	1188	1971	1837	1971	2119
9	1188	1250	1313	2267	2119	2267	2363
10	1313	1438	1563	2459	2363	2459	2633
11	1563	1688	1813	2806	2633	2806	3115
12	1813	1938	2063	3230	3115	3423	3576
13	2063	2188	2313	3937	3576	3729	3893
14	2313	2500	2688	4439	3893	4058	4234
15	2688	2875	3063	4937	4234	4410	4599
16	3063	3313	3563	5481	4599	4787	4990
17	3563	3813	4063	5939	4990	5192	5409
18	4063	4375	4688	7196	5409	5627	5860
19	4688	5000	5313	9478	5860	6093	6342
20	5313	5688	6063	11269	6342	6592	6860
21	6063	6500	6938	12542	6860	7128	7415
22	6938	7438	7938	13513	7415	7703	7990

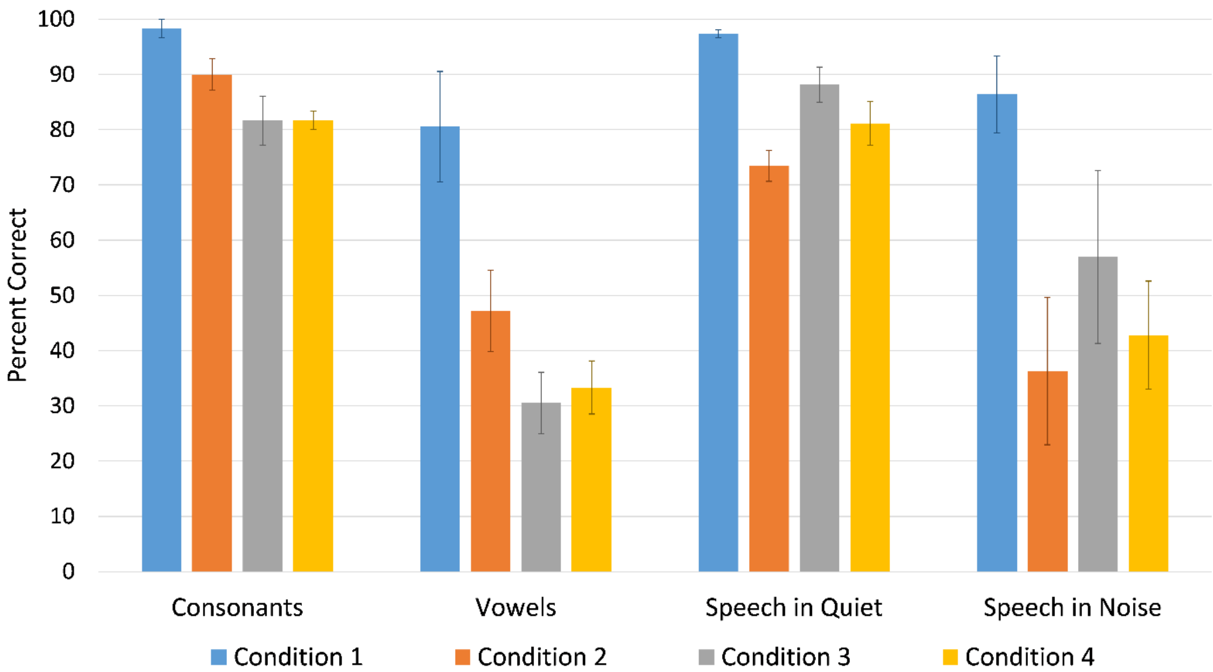


Figure B.11. MAP11 - Average speech recognition scores for MAP11 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 27.789	$p = 0.001$
Effect of speech material:	Significant	F[3, 6] = 21.110	$p = 0.001$
Interaction (condition × speech material):	Significant	F[9, 18] = 6.650	$p < 0.001$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1				*
Cond2				
Cond3				
Cond4	*			

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*		
Cond2	*			
Cond3				
Cond4				

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Significantly different pair is marked with asterisk.

B.12. MAP12

Table B.12. Frequency allocation tables for MAP12. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	934	500	625	750
2	313	375	438	1082	750	875	1063
3	438	500	563	1201	-	-	-
4	563	625	688	1371	1063	1250	1542
5	688	750	813	1461	-	-	-
6	813	875	938	1834	1542	1834	1975
7	938	1000	1063	1943	-	-	-
8	1063	1125	1188	2115	1975	2115	2223
9	1188	1250	1313	2330	2223	2330	2462
10	1313	1438	1563	2594	2462	2594	2700
11	1563	1688	1813	2806	2700	2806	3191
12	1813	1938	2063	3110	-	-	-
13	2063	2188	2313	3525	3191	3576	3790
14	2313	2500	2688	4132	3790	4004	4238
15	2688	2875	3063	4710	4238	4473	4731
16	3063	3313	3563	5544	4731	4989	5272
17	3563	3813	4063	6958	5272	5555	5866
18	4063	4375	4688	9274	5866	6177	6518
19	4688	5000	5313	11391	6518	6859	7234
20	5313	5688	6063	12950	7234	7608	7983
21	6063	6500	6938	13657	-	-	-
22	6938	7438	7938	14708	-	-	-

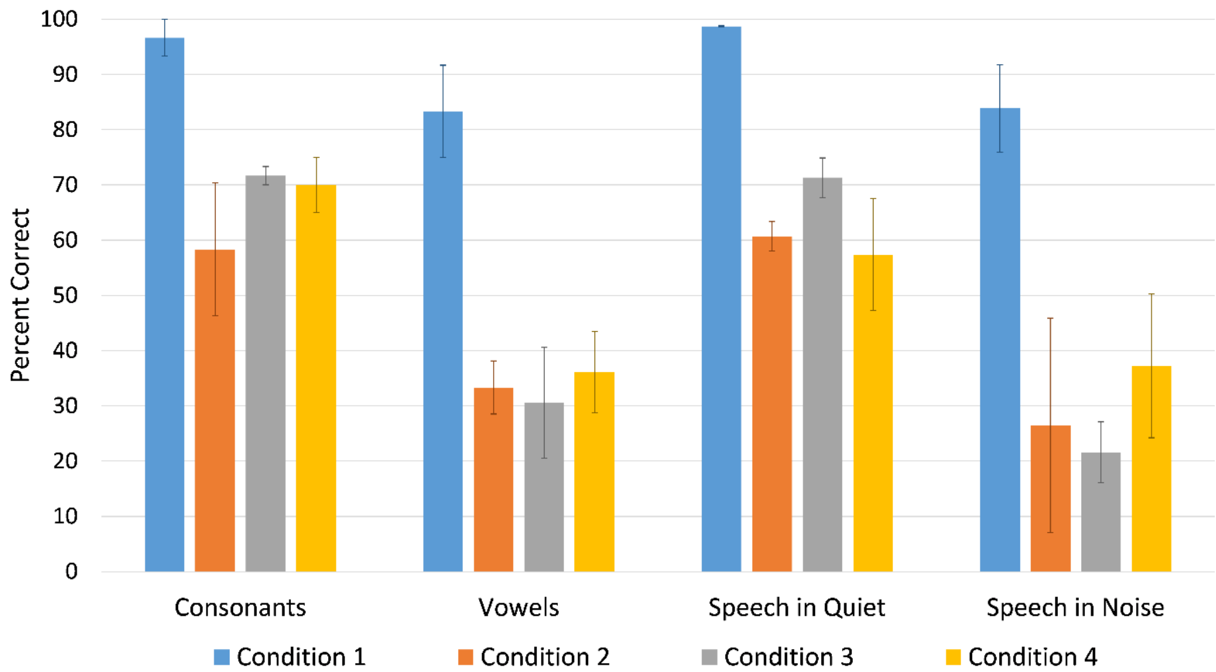


Figure B.12. MAP12 - Average speech recognition scores for MAP12 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 39.301	*p* < 0.001
Effect of speech material:	Significant	F[3, 6] = 78.190	*p* < 0.001
Interaction (condition × speech material):	Significant	F[9, 18] = 2.712	*p* = 0.034

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1			*	
Cond2				
Cond3	*			
Cond4				

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Significantly different pair is marked with asterisk.

B.13. MAP13

Table B.13. Frequency allocation tables for MAP13. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	1101	500	625	750
2	313	375	438	1305	750	875	1038
3	438	500	563	1508	1038	1200	1400
4	563	625	688	1703	1400	1600	1800
5	688	750	813	1999	1800	1999	2133
6	813	875	938	2267	2133	2267	2397
7	938	1000	1063	2526	2397	2526	2666
8	1063	1125	1188	2806	2666	2806	3083
9	1188	1250	1313	3150	3083	3360	3488
10	1313	1438	1563	3569	3488	3617	3753
11	1563	1688	1813	3842	3753	3890	4035
12	1813	1938	2063	4182	4035	4180	4334
13	2063	2188	2313	4710	4334	4488	4651
14	2313	2500	2688	5294	-	-	-
15	2688	2875	3063	5872	4651	4814	4988
16	3063	3313	3563	7116	4988	5161	5346
17	3563	3813	4063	8877	5346	5530	5726
18	4063	4375	4688	10678	5726	5921	6129
19	4688	5000	5313	12277	6129	6337	6557
20	5313	5688	6063	13803	6557	6778	7012
21	6063	6500	6938	14708	7012	7246	7495
22	6938	7438	7938	15670	7495	7744	7993

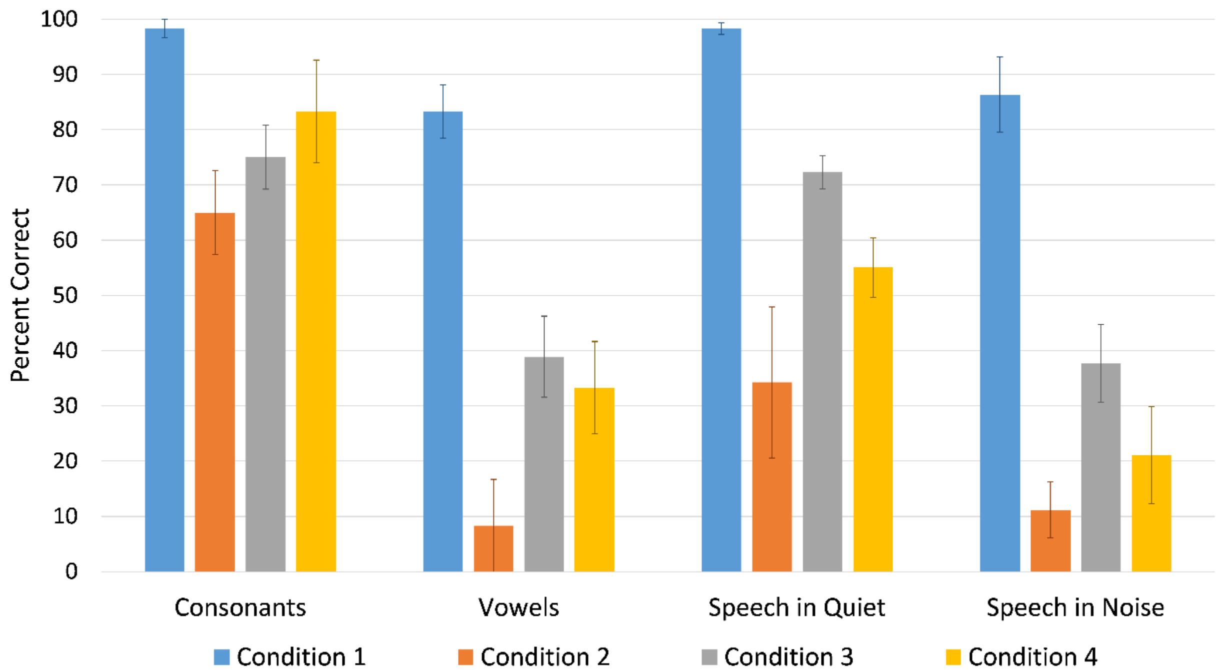


Figure B.13. MAP13 - Average speech recognition scores for MAP13 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 49.521	$p < 0.001$
Effect of speech material:	Significant	F[3, 6] = 46.897	$p < 0.001$
Interaction (condition × speech material):	Significant	F[9, 18] = 7.225	$p < 0.001$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	
Cond2	*			
Cond3	*			
Cond4				

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*		*	*
Cond3	*	*		
Cond4	*	*		

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			*
Cond3	*			
Cond4	*	*		

Significantly different pair is marked with asterisk.

B.14. MAP14

Table B.14. Frequency allocation tables for MAP14. All numbers in the table represent frequencies in Hz. LF: Low cut-off Frequency, CF: Center Frequency, and UF: Upper cut-off Frequency. Channels marked with hyphen were de-activated.

	Clinical			Image	Custom		
	LF	**CF**	**UF**	**CF**	**LF**	**CF**	**UF**
1	188	250	313	1201	500	625	750
2	313	375	438	1349	-	-	-
3	438	500	563	1438	-	-	-
4	563	625	688	1531	750	875	1138
5	688	750	813	1755	1138	1400	1686
6	813	875	938	1971	1686	1971	2183
7	938	1000	1063	2145	-	-	
8	1063	1125	1188	2394	2183	2394	2494
9	1188	1250	1313	2594	2494	2594	2719
10	1313	1438	1563	2843	2719	2843	3154
11	1563	1688	1813	3271	-	-	-
12	1813	1938	2063	3658	3154	3464	3633
13	2063	2188	2313	4033	-	-	-
14	2313	2500	2688	4439	3633	3802	3985
15	2688	2875	3063	4879	3985	4168	4365
16	3063	3313	3563	5481	4365	4562	4774
17	3563	3813	4063	6146	4774	4987	5216
18	4063	4375	4688	7196	5216	5446	5694
19	4688	5000	5313	9899	5694	5942	6209
20	5313	5688	6063	11269	6209	6476	6765
21	6063	6500	6938	12813	6765	7054	7365
22	6938	7438	7938	14401	7365	7677	7988

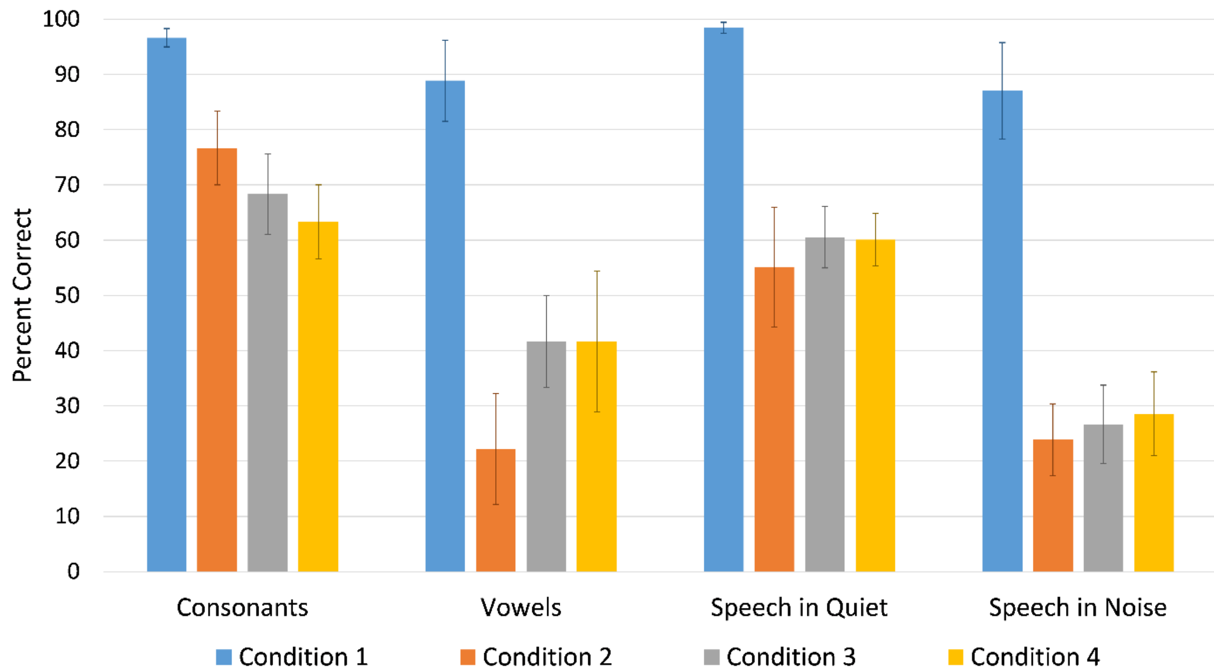


Figure B.14. MAP14 - Average speech recognition scores for MAP14 tested with 3 NH listeners.

Analysis: 2-way repeated measures ANOVA:

Effect of condition:	Significant	F[3, 6] = 256.664	$p < 0.001$
Effect of speech material:	Significant	F[3, 6] = 14.550	$p = 0.004$
Interaction (condition × speech material):	Significant	F[9, 18] = 3.724	$p = 0.008$

Pairwise LSD comparisons:

Consonants

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Vowels

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Sentences Quiet

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*		*	*
Cond3	*	*		
Cond4	*	*		

Sentences Noise

	Cond1	Cond2	Cond3	Cond4
Cond1		*	*	*
Cond2	*			
Cond3	*			
Cond4	*			

Significantly different pair is marked with asterisk.

REFERENCES

[1] A. Volta, "On the electricity excited by mere contact of conducting substances of different kinds," *Philosophical Transactions of the Royal Society of London,* vol. 90, pp. 403-431, January; 1, 1800.

[2] R. Finn, A. Hudspeth James, J. Zwislocki, E. Young and M. Merzenich, "Sound from silence: The development of cochlear implants," *Beyond Discovery: The Path from Research to Human Benefit,* pp. 1-8, 1998.

[3] F. Zeng, "Auditory prostheses: Past, present, and future," in *Cochlear Implants: Auditory Prostheses and Electric Hearing*, F. Zeng, A. Popper N. and R. Fay R., Eds. New York: Springer-Verlag, 2004, pp. 1-6.

[4] B. S. Wilson and M. F. Dorman, "Interfacing Sensors With the Nervous System: Lessons From the Development and Success of the Cochlear Implant," *IEEE Sensor,* vol. 8, pp. 131-147, 2008.

[5] M. D. Eisen, "The history of cochlear implants," in *Cochlear Implants Principles & Practices*, Second edition ed., J. K. Niparko, Ed. Philadelphia: Lippincott Williams & Wilkins, 2009, pp. 89-93.

[6] Anonymous "NIDCD fact sheet: Cochlear implants," NIDCD, NIH, Tech. Rep. NIH Publication No. 11-4798, 2013.

[7] Anonymous "NIDCD fact sheet: Cochlear implants," NIDCD, NIH, Tech. Rep. NIH Publication No. 11-4798, 2011.

[8] Fan-Gang Zeng, S. Rebscher, W. Harrison, Xiaoan Sun and Haihong Feng, "Cochlear Implants: System Design, Integration, and Evaluation," *Biomedical Engineering, IEEE Reviews In,* vol. 1, pp. 115-142, 2008.

[9] (October 17, 2015). *PUBMED - Annual number of scientific publications in the field of cochlear implants*. Available: http://www.pubmed.gov.

[10] B. S. Wilson and M. F. Dorman, "Cochlear implants: a remarkable past and a brilliant future," *Hear. Res.,* vol. 242, pp. 3-21, 06/22, 2008.

[11] B. S. Wilson and M. F. Dorman, "The design of cochlear implants," in *Cochlear Implants Principles & Practices*, Second edition ed., J. K. Niparko, Ed. Philadelphia: Lippincott Williams & Wilkins, 2009, pp. 95-135.

[12] R. V. Shannon, Q. J. Fu and J. Galvin 3rd, "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," *Acta Otolaryngol. Suppl.*, vol. (552), pp. 50-54, May, 2004.

[13] L. M. Friesen, R. V. Shannon, D. Baskent and X. Wang, "Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.*, vol. 110, pp. 1150-1163, Aug, 2001.

[14] M. F. Dorman and A. J. Spahr, "Speech perception by adults with mutlichannel cochlear implants," in , Second eitition ed.Anonymous New York: Thieme Medical Publishers, 2006, pp. 193-204.

[15] C. Garnham, M. O'Driscoll, R. Ramsden And and S. Saeed, "Speech understanding in noise with a Med-El COMBI 40+ cochlear implant using reduced channel sets," *Ear Hear.*, vol. 23, pp. 540-552, Dec, 2002.

[16] B. S. Wilson, C. C. Finley, D. T. Lawson and M. Zerbi, "Temporal representations with cochlear implants," *Am. J. Otol.*, vol. 18, pp. S30-4, Nov, 1997.

[17] B. R. Glasberg and B. C. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, vol. 47, pp. 103-138, Aug 1, 1990.

[18] B. C. Moore, "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants," *Otol. Neurotol.*, vol. 24, pp. 243-254, Mar, 2003.

[19] T. Balkany, A. Hodges, C. Menapace, L. Hazard, C. Driscoll, B. Gantz, D. Kelsall, W. Luxford, S. McMenomy, J. G. Neely, B. Peters, H. Pillsbury, J. Roberson, D. Schramm, S. Telian, S. Waltzman, B. Westerberg and S. Payne, "Nucleus Freedom North American Clinical Trial," *Otolaryngology -- Head and Neck Surgery*, vol. 136, pp. 757-762, May 01, 2007.

[20] D. B. Koch, M. J. Osberger, P. Segel and D. Kessler, "HiResolution^{TM} and Conventional Sound Processing in the HiResolution^{TM} Bionic Ear: Using Appropriate Outcome Measures to Assess Speech Recognition Ability," *Audiol Neurotol*, vol. 9, pp. 214-223, 2004.

[21] C. Arnoldner, D. Riss, M. Brunner, M. Durisin, W. Baumgartner and >. Hamzavi, "Speech and music perception with the new fine structure speech coding strategy: preliminary results," *Acta Otolaryngol.*, vol. 127, pp. 1298-1303, 01/01; 2015/10, 2007.

[22] A. J. Spahr, M. F. Dorman and L. H. Loiselle, "Performance of Patients Using Different Cochlear Implant Systems: Effects of Input Dynamic Range," *Ear Hear.*, vol. 28, 2007.

[23] J. B. Firszt, L. K. Holden, M. W. Skinner, E. A. Tobey, A. Peterson, W. Gaggl, C. Runge-Samuelson and P. A. Wackym, "Recognition of Speech Presented at Soft to Loud Levels by Adult Cochlear Implant Recipients of Three Cochlear Implant Systems," *Ear Hear.*, vol. 25, 2004.

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

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

[26] D. B. Pisoni, "Cognitive factors and cochlear implants: some thoughts on perception, learning, and memory in speech perception," *Ear Hear.*, vol. 21, pp. 70-78, Feb, 2000.

[27] A. Kral and J. Tillein, "Brain plasticity under cochlear implant stimulation," *Adv. Otorhinolaryngol.*, vol. 64, pp. 89-108, 2006.

[28] J. B. Fallon, D. R. Irvine and R. K. Shepherd, "Cochlear implants and brain plasticity," *Hear. Res.*, vol. 238, pp. 110-117, Apr, 2008.

[29] D. S. Lazard, C. Vincent, F. Venail, P. Van de Heyning, E. Truy, O. Sterkers, P. H. Skarzynski, H. Skarzynski, K. Schauwers, S. O'Leary, D. Mawman, B. Maat, A. Kleine-Punte, A. M. Huber, K. Green, P. J. Govaerts, B. Fraysse, R. Dowell, N. Dillier, E. Burke, A. Beynon, F. Bergeron, D. Baskent, F. Artieres and P. J. Blamey, "Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: a new conceptual model over time," *PLoS One*, vol. 7, pp. e48739, 2012.

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

[31] P. Blamey, F. Artieres, D. Baskent, F. Bergeron, A. Beynon, E. Burke, N. Dillier, R. Dowell, B. Fraysse, S. Gallego, P. J. Govaerts, K. Green, A. M. Huber, A. Kleine-Punte, B. Maat, M. Marx, D. Mawman, I. Mosnier, A. F. O'Connor, S. O'Leary, A. Rousset, K. Schauwers, H. Skarzynski, P. H. Skarzynski, O. Sterkers, A. Terranti, E. Truy, P. Van de Heyning, F. Venail, C. Vincent and D. S. Lazard, "Factors affecting auditory performance of

postlinguistically deaf adults using cochlear implants: an update with 2251 patients," *Audiol. Neurootol.*, vol. 18, pp. 36-47, 2013.

[32] M. Lenarz, G. Joseph, H. Sönmez, A. Büchner and T. Lenarz, "Effect of technological advances on cochlear implant performance in adults," *Laryngoscope,* vol. 121, pp. 2634-2640, 2011.

[33] G. M. Clark, P. J. Blamey, A. M. Brown, P. A. Gusby, R. C. Dowell, B. K. Franz, B. C. Pyman, R. K. Shepherd, Y. C. Tong and R. L. Webb, "The University of Melbourne--nucleus multi-electrode cochlear implant," *Adv. Otorhinolaryngol.*, vol. 38, pp. V-IX, 1-181, 1987.

[34] P. M. Seligman, J. F. Patrick, Y. C. Tong, G. M. Clark, R. C. Dowell and P. A. Crosby, "A signal processor for a multiple-electrode hearing prosthesis," *Acta Otolaryngol. Suppl.*, vol. 411, pp. 135-139, 1984.

[35] P. J. Blamey, R. C. Dowell, G. M. Clark and P. M. Seligman, "Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant," *J. Acoust. Soc. Am.*, vol. 82, pp. 38-47, Jul, 1987.

[36] R. C. Dowell, P. M. Seligman, P. J. Blamey and G. M. Clark, "Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis," *Annals of Otology, Rhinology & Laryngology,* vol. 96, pp. 132-134, 1987.

[37] J. F. Patrick and G. M. Clark, "The Nucleus 22-channel cochlear implant system," *Ear Hear.*, vol. 12, pp. 3S-9S, Aug, 1991.

[38] E. L. von Wallenberg and R. D. Battmer, "Comparative speech recognition results in eight subjects using two different coding strategies with the Nucleus 22 channel cochlear implant," *Br. J. Audiol.,* vol. 25, pp. 371-380, Dec, 1991.

[39] R. C. Dowell, P. W. Dawson, S. J. Dettman, R. K. Shepherd, L. A. Whitford, P. M. Seligman and G. M. Clark, "Multichannel cochlear implantation in children: a summary of current work at the University of Melbourne," *Am. J. Otol.,* vol. 12 Suppl, pp. 137-143, 1991.

[40] M. W. Skinner, L. K. Holden, T. A. Holden, R. C. Dowell, P. M. Seligman, J. A. Brimacombe and A. L. Beiter, "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.,* vol. 12, pp. 3-22, Feb, 1991.

[41] D. K. Eddington, "Speech discrimination in deaf subjects with cochlear implants," *J. Acoust. Soc. Am.*, vol. 68, pp. 885-891, Sep, 1980.

[42] M. F. Dorman, M. T. Hannley, K. Dankowski, L. Smith and G. McCandless, "Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant," *Ear Hear.*, vol. 10, pp. 44-49, Feb, 1989.

[43] H. W. Dudley, "The Vocoder," vol. 18, pp. 122-126, 1939.

[44] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature,* vol. 352, pp. 236-238, Jul 18, 1991.

[45] H. J. McDermott, C. M. McKay and A. E. Vandali, "A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant," *J. Acoust. Soc. Am.*, vol. 91, pp. 3367-3371, Jun, 1992.

[46] P. Seligman and H. McDermott, "Architecture of the Spectra 22 speech processor," *Ann. Otol. Rhinol. Laryngol. Suppl.*, vol. 166, pp. 139-141, Sep, 1995.

[47] J. Kiefer, S. Hohl, E. Sturzebecher, T. Pfennigdorff and W. Gstoettner, "Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system," *Audiology,* vol. 40, pp. 32-42, Jan-Feb, 2001.

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

[49] W. Noguiera, A. Buchner, T. Lenarz and B. and Edler, "A psychoacoustic ‘nofm’-type speech coding strategy for cochlear implants," *EURASIP J. Appl. Signal Process.*, vol. 18, pp. 3044-3059, 2005.

[50] A. Buchner, W. Nogueira, B. Edler, R. D. Battmer and T. Lenarz, "Results from a psychoacoustic model-based strategy for the nucleus-24 and freedom cochlear implants," *Otol. Neurotol.*, vol. 29, pp. 189-192, Feb, 2008.

[51] I. Hochmair, P. Nopp, C. Jolly, M. Schmidt, H. Schosser, C. Garnham and I. Anderson, "MED-EL Cochlear implants: state of the art and a glimpse into the future," *Trends Amplif,* vol. 10, pp. 201-219, Dec, 2006.

[52] D. Riss, J. S. Hamzavi, A. Selberherr, A. Kaider, M. Blineder, V. Starlinger, W. Gstoettner and C. Arnoldner, "Envelope versus fine structure speech coding strategy: a crossover study," *Otol. Neurotol.*, vol. 32, pp. 1094-1101, Sep, 2011.

[53] J. Muller, S. Brill, R. Hagen, A. Moeltner, S. J. Brockmeier, T. Stark, S. Helbig, J. Maurer, T. Zahnert, C. Zierhofer, P. Nopp and I. Anderson, "Clinical trial results with the MED-EL

fine structure processing coding strategy in experienced cochlear implant users," *ORL J. Otorhinolaryngol. Relat. Spec.*, vol. 74, pp. 185-198, 2012.

[54] B. S. Wilson, D. T. Lawson, M. Zerbi and C. C. Finley, "Speech processors for auditory prostheses: Virtual channel interleaved sampling (VCIS) processors-initial studies with subject SR2," Neural Prosthesis Program, National Institutes of Health (NIH), Bethesda (MD), Tech. Rep. First Quarterly Progress Report, NIH project N01-DC-2-2401, 1992.

[55] W. Nogueira, A. Buechner, J. Ostermann and B. Edler, "Signal processing strategies for cochlear implants using current steering," in *Voice Communication (SprachKommunikation), 2008 ITG Conference On*, 2008, pp. 1-4.

[56] G. S. Donaldson, P. K. Dawson and L. Z. Borden, "Within-subjects comparison of the HiRes and Fidelity120 speech processing strategies: speech perception and its relation to place-pitch sensitivity," *Ear Hear.*, vol. 32, pp. 238-250, Mar-Apr, 2011.

[57] F. Zeng, S. Rebsher, W. Harrison, X. Sun and H. Feng, "Cochlear Implants: System Design, Integration, and Evaluation," *IEEE Reviews in Biomedical Engineering*, vol. 1, pp. 115-142, 2008.

[58] J. H. L. Hansen, "Speech enhancement," in *Wiley Encyclopedia of Electrical and Electronics Engineering*Anonymous John Wiley & Sons, Inc., 2001, .

[59] P. C. Loizou, *Speech Enhancement: Theory and Practice, Second Edition.* CRC Press, 2013.

[60] P. J. Blamey, "Adaptive dynamic range optimization (ADRO): a digital amplification strategy for hearing aids and cochlear implants," *Trends Amplif*, vol. 9, pp. 77-98, 2005.

[61] P. W. Dawson, J. A. Decker and C. E. Psarros, "Optimizing dynamic range in children using the nucleus cochlear implant," *Ear Hear.*, vol. 25, pp. 230-241, Jun, 2004.

[62] P. Blamey, C. J. James, H. J. McDermott, L. Martin and K. Wildi, "Adaptive dynamic range optimization sound processor," International Patent Application PCT/ AU99/00076, U.S. Patent Application 09/478,022, 07/12, 1999.

[63] Anonymous "ClearVoice clinical results white paper," Advanced Bionics, 2012.

[64] H. Ali, O. Hazrati, E. A. Tobey and J. H. Hansen, "Evaluation of adaptive dynamic range optimization in adverse listening conditions for cochlear implants," *J. Acoust. Soc. Am.*, vol. 136, pp. EL242, Sep, 2014.

[65] D. D. Greenwood. Critical bandwidth and the frequency coordinates of the basilar membrane. *J. Acoust. Soc. Am. 33(10),* pp. 1344-1356. 1961. . DOI: http://dx.doi.org/10.1121/1.1908437.

[66] D. D. Greenwood. A cochlear frequency-position function for several species—29 years later. *J. Acoust. Soc. Am. 87(6),* pp. 2592-2605. 1990. . DOI: http://dx.doi.org/10.1121/1.399052.

[67] R. Glueckert, K. Pfaller, A. Kinnefors, H. Rask-Andersen and A. Schrott-Fischer, "The human spiral ganglion: new insights into ultrastructure, survival rate and implications for cochlear implants," *Audiol. Neurootol.,* vol. 10, pp. 258-273, Sep-Oct, 2005.

[68] I. Hochmair, W. Arnold, P. Nopp, C. Jolly, J. Muller and P. Roland, "Deep electrode insertion in cochlear implants: apical morphology, electrodes and speech perception results," *Acta Otolaryngol.,* vol. 123, pp. 612-617, Jun, 2003.

[69] J. Otte, H. F. Schunknecht and A. G. Kerr, "Ganglion cell populations in normal and pathological human cochleae. Implications for cochlear implantation," *Laryngoscope,* vol. 88, pp. 1231-1246, Aug, 1978.

[70] J. B. Nadol Jr, "Patterns of neural degeneration in the human cochlea and auditory nerve: implications for cochlear implantation," *Otolaryngol. Head. Neck. Surg.,* vol. 117, pp. 220-228, Sep, 1997.

[71] A. Kawano, H. L. Seldon and G. M. Clark, "Computer-aided three-dimensional reconstruction in human cochlear maps: measurement of the lengths of organ of Corti, outer wall, inner wall, and Rosenthal's canal," *Ann. Otol. Rhinol. Laryngol.,* vol. 105, pp. 701-709, Sep, 1996.

[72] E. Erixon, H. Hogstorp, K. Wadin and H. Rask-Andersen, "Variational anatomy of the human cochlea: implications for cochlear implantation," *Otol. Neurotol.,* vol. 30, pp. 14-22, Jan, 2009.

[73] S. Biedron, M. Westhofen and J. Ilgner, "On the number of turns in human cochleae," *Otol. Neurotol.,* vol. 30, pp. 414-417, Apr, 2009.

[74] O. Stakhovskaya, D. Sridhar, B. H. Bonham and P. A. Leake, "Frequency Map for the Human Cochlear Spiral Ganglion: Implications for Cochlear Implants," *J. Assoc. Res. Otolaryngol.,* vol. 8, pp. 220-233, Jun, 2007.

[75] W. R. TIFFANY and D. N. BENNETT, "Intelligibility of slow-played speech," *J. Speech Hear. Res.,* vol. 4, pp. 248-258, Sep, 1961.

[76] R. G. Daniloff, T. H. Shriner and W. R. Zemlin. Intelligibility of vowels altered in duration and frequency. *J. Acoust. Soc. Am. 44(3),* pp. 700-707. 1968. . DOI: http://dx.doi.org/10.1121/1.1911164.

[77] M. Nagafuchi, "Intelligibility of distorted speech sounds shifted in frequency and time in normal children," *Audiology,* vol. 15, pp. 326-337, Jul-Aug, 1976.

[78] C. M. Reed, B. L. Hicks, L. D. Braida and N. I. Durlach, "Discrimination of speech processed by low-pass filtering and pitch-invariant frequency lowering," *J. Acoust. Soc. Am.*, vol. 74, pp. 409-419, Aug, 1983.

[79] D. S. Beasley, N. L. Mosher and D. J. Orchik, "Use of frequency-shifted/time-compressed speech with hearing-impaired children," *Audiology,* vol. 15, pp. 395-406, Sep-Oct, 1976.

[80] L. D. Braida, N. I. Durlach, R. P. Lippmann, B. L. Hicks, W. M. Rabinowitz and C. M. Reed, "Hearing aids--a review of past research on linear amplification, amplitude compression, and frequency lowering," *ASHA Monogr.,* vol. (19), pp. 1-114, Apr, 1979.

[81] M. F. Dorman, P. C. Loizou and D. Rainey. Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J. Acoust. Soc. Am. 102(5),* pp. 2993-2996. 1997. . DOI: http://dx.doi.org/10.1121/1.420354.

[82] Q. J. Fu and R. V. Shannon, "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," *J. Acoust. Soc. Am.*, vol. 105, pp. 1889-1900, Mar, 1999.

[83] R. V. Shannon, F. G. Zeng and J. Wygonski, "Speech recognition with altered spectral distribution of envelope cues," *J. Acoust. Soc. Am.*, vol. 104, pp. 2467-2476, Oct, 1998.

[84] D. Baskent and R. V. Shannon, "Speech recognition under conditions of frequency-place compression and expansion," *J. Acoust. Soc. Am.,* vol. 113, pp. 2064-2076, Apr, 2003.

[85] D. Baskent and R. V. Shannon, "Frequency-place compression and expansion in cochlear implant listeners," *J. Acoust. Soc. Am.,* vol. 116, pp. 3130-3140, Nov, 2004.

[86] D. Baskent and R. V. Shannon, "Interactions between cochlear implant electrode insertion depth and frequency-place mapping," *J. Acoust. Soc. Am.,* vol. 117, pp. 1405-1416, Mar, 2005.

[87] D. Baskent and R. V. Shannon, "Combined effects of frequency compression-expansion and shift on speech recognition," *Ear Hear.,* vol. 28, pp. 277-289, Jun, 2007.

[88] ANSI, "American National Standard Methods for Calculation of the Speech Intelligibility Index," *ANSI S3. 5 American National Standards Institute, New York,* 1997.

[89] M. J. Goupell, B. Laback, P. Majdak and W. D. Baumgartner, "Effects of upper-frequency boundary and spectral warping on speech intelligibility in electrical stimulation," *J. Acoust. Soc. Am.,* vol. 123, pp. 2295-2309, Apr, 2008.

[90] S. Rosen, A. Faulkner and L. Wilkinson, "Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants," *J. Acoust. Soc. Am.,* vol. 106, pp. 3629-3636, Dec, 1999.

[91] C. L. De Filippo and B. L. Scott, "A method for training and evaluating the reception of ongoing speech," *J. Acoust. Soc. Am.,* vol. 63, pp. 1186-1192, Apr, 1978.

[92] J. D. Harnsberger, M. A. Svirsky, A. R. Kaiser, D. B. Pisoni, R. Wright and T. A. Meyer, "Perceptual "vowel spaces" of cochlear implant users: implications for the study of auditory adaptation to spectral shift," *J. Acoust. Soc. Am.,* vol. 109, pp. 2135-2145, May, 2001.

[93] Q. J. Fu, R. V. Shannon and J. J. Galvin 3rd, "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," *J. Acoust. Soc. Am.,* vol. 112, pp. 1664-1674, Oct, 2002.

[94] A. Faulkner, S. Rosen and D. Stanton, "Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth," *J. Acoust. Soc. Am.,* vol. 113, pp. 1073-1080, Feb, 2003.

[95] A. Faulkner, S. Rosen and C. Norman, "The right information may matter more than frequency-place alignment: simulations of frequency-aligned and upward shifting cochlear implant processors for a shallow electrode array insertion," *Ear Hear.,* vol. 27, pp. 139-152, Apr, 2006.

[96] C. M. Siciliano, A. Faulkner, S. Rosen and K. Mair, "Resistance to learning binaurally mismatched frequency-to-place maps: implications for bilateral stimulation with cochlear implants," *J. Acoust. Soc. Am.,* vol. 127, pp. 1645-1660, Mar, 2010.

[97] J. P. J. Pinel, *Biopsychology.* Pearson, 2014.

[98] T. Li, J. J. Galvin 3rd and Q. J. Fu, "Interactions between unsupervised learning and the degree of spectral mismatch on short-term perceptual adaptation to spectrally shifted speech," *Ear Hear.,* vol. 30, pp. 238-249, Apr, 2009.

[99] Q. J. Fu and J. J. Galvin 3rd, "Perceptual learning and auditory training in cochlear implant recipients," *Trends Amplif,* vol. 11, pp. 193-205, Sep, 2007.

[100] M. A. Svirsky, T. M. Talavage, S. Sinha, H. Neuburger and M. Azadpour, "Gradual adaptation to auditory frequency mismatch," *Hear. Res.,* vol. 322, pp. 163-170, Apr, 2015.

[101] G. Parikh and P. C. Loizou, "The influence of noise on vowel and consonant cues," *J. Acoust. Soc. Am.,* vol. 118, pp. 3874-3888, Dec, 2005.

[102] L. A. Reiss, C. W. Turner, S. R. Erenberg and B. J. Gantz, "Changes in pitch with a cochlear implant over time," *J. Assoc. Res. Otolaryngol.,* vol. 8, pp. 241-257, Jun, 2007.

[103] L. A. Reiss, B. J. Gantz and C. W. Turner, "Cochlear implant speech processor frequency allocations may influence pitch perception," *Otol. Neurotol.,* vol. 29, pp. 160-167, Feb, 2008.

[104] L. A. Reiss, M. W. Lowder, S. A. Karsten, C. W. Turner and B. J. Gantz, "Effects of extreme tonotopic mismatches between bilateral cochlear implants on electric pitch perception: a case study," *Ear Hear.,* vol. 32, pp. 536-540, Jul-Aug, 2011.

[105] J. H. Noble, R. F. Labadie, O. Majdani and B. M. Dawant, "Automatic segmentation of intracochlear anatomy in conventional CT," *IEEE Trans. Biomed. Eng.,* vol. 58, pp. 2625-2632, Sep, 2011.

[106] J. H. Noble, R. F. Labadie, R. H. Gifford and B. M. Dawant, "Image-guidance enables new methods for customizing cochlear implant stimulation strategies," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 21, pp. 820-829, Sep, 2013.

[107] J. H. Noble, R. H. Gifford, A. J. Hedley-Williams, B. M. Dawant and R. F. Labadie, "Clinical evaluation of an image-guided cochlear implant programming strategy," *Audiol. Neurootol.,* vol. 19, pp. 400-411, 2014.

[108] D. R. Ketten, M. W. Skinner, G. Wang, M. W. Vannier, G. A. Gates and J. G. Neely, "In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays," *Ann. Otol. Rhinol. Laryngol. Suppl.,* vol. 175, pp. 1-16, Nov, 1998.

[109] M. W. Skinner, D. R. Ketten, L. K. Holden, G. W. Harding, P. G. Smith, G. A. Gates, J. G. Neely, G. R. Kletzker, B. Brunsden and B. Blocker, "CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients," *J. Assoc. Res. Otolaryngol.,* vol. 3, pp. 332-350, Sep, 2002.

[110] L. T. Cohen, P. A. Busby, L. A. Whitford and G. M. Clark, "Cochlear implant place psychophysics 1. Pitch estimation with deeply inserted electrodes," *Audiol. Neurootol.,* vol. 1, pp. 265-277, Sep-Oct, 1996.

[111] Anonymous "IEEE Recommnded Pratice for Speech Quality Measurements," *Audio and Electroacoustics, IEEE Transactions On,* vol. 17, pp. 225-246, 1969.

[112] J. Hillenbrand, L. A. Getty, M. J. Clark and K. Wheeler. Acoustic characteristics of american english vowels. *J. Acoust. Soc. Am. 97(5),* pp. 3099-3111. 1995. . DOI: http://dx.doi.org/10.1121/1.411872.

[113] Emily Shannon Fu Foundation, "Internet-based Computer-Assisted Speech Testing Open Platform," .

[114] G. E. PETERSON and I. LEHISTE, "Revised CNC lists for auditory tests," *J. Speech Hear. Disord.,* vol. 27, pp. 62-70, Feb, 1962.

[115] A. J. Spahr, M. F. Dorman, L. M. Litvak, S. Van Wie, R. H. Gifford, P. C. Loizou, L. M. Loiselle, T. Oakes and S. Cook, "Development and validation of the AzBio sentence lists," *Ear Hear.,* vol. 33, pp. 112-117, Jan-Feb, 2012.

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

[117] Anonymous "MSTB: The new minimum speech test battery," 2011.

[118] R. H. Gifford, A. Hedley-Williams and A. J. Spahr, "Clinical assessment of spectral modulation detection for adult cochlear implant recipients: a non-language based measure of performance outcomes," *Int. J. Audiol.,* vol. 53, pp. 159-164, Mar, 2014.

[119] S. Gatehouse and W. Noble, "The Speech, Spatial and Qualities of Hearing Scale (SSQ)," *Int. J. Audiol.,* vol. 43, pp. 85-99, Feb, 2004.

[120] G. Fairbanks, "Voice and articulation drillbook," in *Voice and Articulation Drillbook*, New York: Harper & Row, Ed. 1969, pp. 124-139.

[121] R. S. Tyler, J. P. Preece, C. R. Lansing, S. R. Otto and B. J. Gantz, "Previous experience as a confounding factor in comparing cochlear-implant processing schemes," *J. Speech Hear. Res.,* vol. 29, pp. 282-287, Jun, 1986.

[122] J. Wouters, H. J. McDermott and T. Francart, "Sound Coding in Cochlear Implants: From electric pulses to hearing," *Signal Processing Magazine, IEEE,* vol. 32, pp. 67-80, 2015.

[123] L. Litvak, B. Delgutte and D. Eddington, "Improved neural representation of vowels in electric stimulation using desynchronizing pulse trains," *J. Acoust. Soc. Am.,* vol. 114, pp. 2099-2111, Oct, 2003.

[124] L. M. Litvak, B. Delgutte and D. K. Eddington, "Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains," *J. Acoust. Soc. Am.,* vol. 114, pp. 2079-2098, Oct, 2003.

[125] J. C. Middlebrooks, "Effects of cochlear-implant pulse rate and inter-channel timing on channel interactions and thresholds," *J. Acoust. Soc. Am.,* vol. 116, pp. 452-468, Jul, 2004.

[126] R. V. Shannon, R. J. Cruz and J. J. Galvin 3rd, "Effect of stimulation rate on cochlear implant users' phoneme, word and sentence recognition in quiet and in noise," *Audiol. Neurootol.,* vol. 16, pp. 113-123, 2011.

[127] L. M. Collins, T. A. Zwolan and G. H. Wakefield, "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects," *J. Acoust. Soc. Am.,* vol. 101, pp. 440-455, Jan, 1997.

[128] T. A. Zwolan, L. M. Collins and G. H. Wakefield. Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects. *J. Acoust. Soc. Am. 102(6),* pp. 3673-3685. 1997. . DOI: http://dx.doi.org/10.1121/1.420401.

[129] S. I. Duran, "Psychophysics-based electrode selection for cochlear implant listeners," *Ph. D. Dissertation, Duke University,* 2014.

[130] J. Wouters and J. Vanden Berghe, "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," *Ear Hear.,* vol. 22, pp. 420-430, Oct, 2001.

[131] P. C. Loizou, "Speech processing in vocoder-centric cochlear implants," *Adv. Otorhinolaryngol.,* vol. 64, pp. 109-143, 2006.

[132] R. J. van Hoesel and G. M. Clark, "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *J. Acoust. Soc. Am.,* vol. 97, pp. 2498-2503, Apr, 1995.

[133] P. C. Loizou, *Speech Enhancement – Theory and Practice.* CRC Press, 2012.

[134] Y. Hu, P. C. Loizou, N. Li and K. Kasturi, "Use of a sigmoidal-shaped function for noise attenuation in cochlear implants," *J. Acoust. Soc. Am.,* vol. 122, pp. EL128-34, Oct, 2007.

[135] Y. Hu and P. C. Loizou, "A new sound coding strategy for suppressing noise in cochlear implants," *J. Acoust. Soc. Am.,* vol. 124, pp. 498-509, Jul, 2008.

[136] R. C. Dowell, D. J. Mecklenburg and G. M. Clark, "Speech recognition for 40 patients receiving multichannel cochlear implants," *Arch. Otolaryngol. Head. Neck. Surg.,* vol. 112, pp. 1054-1059, Oct, 1986.

[137] P. C. Loizou, "Mimicking the human ear," *Signal Processing Magazine, IEEE,* vol. 15, pp. 101-130, 1998.

[138] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *Speech and Audio Processing, IEEE Transactions On,* vol. 11, pp. 466-475, 2003.

[139] H. Ali, A. P. Lobo and P. C. Loizou, "Design and Evaluation of a Personal Digital Assistant-based Research Platform for Cochlear Implants," *Biomedical Engineering, IEEE Transactions On,* vol. 60, pp. 3060-3073, 2013.

[140] L. F. A. Martin, P. J. Blamey, C. J. James, K. L. Galvin and D. Macfarlane, "Adaptive dynamic range optimization for hearing aids," *Acoust. Australia,* vol. 29, pp. 21-24, 2001.

[141] C. J. James, P. J. Blamey, L. Martin, B. Swanson, Y. Just and D. Macfarlane, "Adaptive dynamic range optimization for cochlear implants: a preliminary study," *Ear Hear.,* vol. 23, pp. 49S-58S, Feb, 2002.

[142] T. Iwaki, P. Blamey and T. Kubo, "Bimodal studies using adaptive dynamic range optimization (ADRO) technology," *Int. J. Audiol.,* vol. 47, pp. 311-318, Jun, 2008.

[143] M. L. Hughes, C. J. Brown, P. J. Abbas, A. A. Wolaver and J. P. Gervais, "Comparison of EAP thresholds with MAP levels in the nucleus 24 cochlear implant: data from children," *Ear Hear.,* vol. 21, pp. 164-174, Apr, 2000.

[144] O. Hazrati, "Development of dereverberation algorithms for improved speech intelligibility by cochlear implant users," *Ph. D. Dissertation, University of Texas, Dallas, TX,* 2012.

[145] A. K. Nabelek, T. R. Letowski and F. M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.,* vol. 86, pp. 1259-1265, Oct, 1989.

[146] O. Hazrati and P. C. Loizou, "The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners," *Int. J. Audiol.,* vol. 51, pp. 437-443, Jun, 2012.

[147] American National Standards Institute, "Acoustical Performance Criteria, Design Requirements and Guidelines for Schools," *ANSI S12. 60 American National Standards Institute, New York,* 2002.

[148] P. C. Loizou, M. Dorman and J. Fitzke, "The effect of reduced dynamic range on speech understanding: implications for patients with cochlear implants," *Ear Hear.,* vol. 21, pp. 25-31, Feb, 2000.

[149] Q. J. Fu and R. V. Shannon, "Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users," *Ear Hear.,* vol. 21, pp. 227-235, Jun, 2000.

VITA

Hussnain Ali was born in Sialkot, Pakistan in 1985. He received his Bachelor’s degree in Electrical Engineering from National University of Science & Technology (NUST) in April 2008. Thereafter, he worked in Center for Advanced Research in Engineering, Islamabad for one and a half years where he worked on the development of a cardiac telemedicine system. He joined the Electrical Engineering department at The University of Texas of Dallas in January 2010 to undertake graduate studies. He received his Master of Science degree in Electrical Engineering in 2012 under the supervision of Prof. Philipos C. Loizou. Since then he has been working towards his Ph.D. degree and working as a lab manager and project supervisor on projects funded by NIH. His research interests include implantable and wearable medical devices and systems, biomedical signal processing, cochlear implants, and emerging healthcare technologies to improve quality of life.