A FRAMEWORK FOR THE DEVELOPMENT AND VALIDATION OF PHENOMENOLOGICALLY DERIVED COCHLEAR IMPLANT STIMULATION STRATEGIES

by

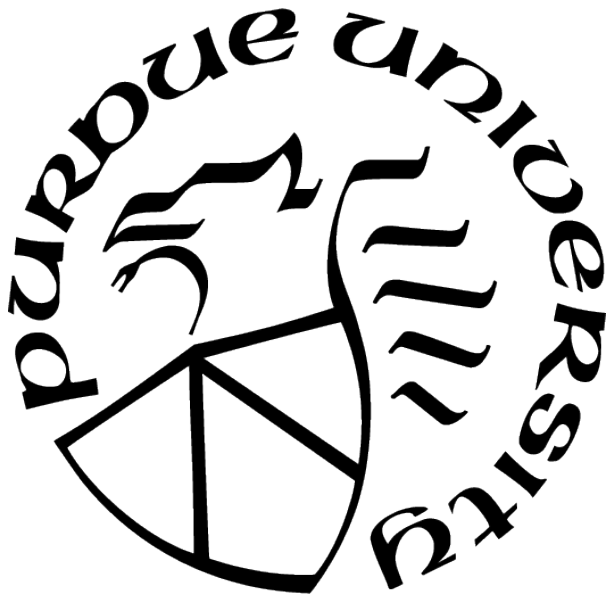
Andres F. Llico Gallardo

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Weldon School of Biomedical Engineering

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Thomas Talavage, Co-Chair

Weldon School of Biomedical Engineering

Dr. George Wodicka, Co-Chair

Weldon School of Biomedical Engineering

Dr. Michael Heinz

Weldon School of Biomedical Engineering

Dr. Shannon Van Hyfte

School of Speech, Language, and Hearing Sciences

Approved by:

Dr. Andrew Brightman

ACKNOWLEDGMENTS

This work was partially funded by the Leslie A. Geddes Fellowship Endowment. I wish to thank my committee members Dr. Talavage, Dr. Wodicka, Dr. Heinz, and Dr. Van Hyfte for their support and insight during my time as a graduate student at Purdue University. I would also like to thank Satyabrata Parida, Ravinderjit Singh, William Salloom, Andrew Sivaprakasam, François Deloche, Ikbeom Jang, Art Terlep, and Brad Fitzgerald for their intellectual discussions and feedback on presentations.

I would like to acknowledge the unconditional support from family and friends through the years of graduate school. Specifically, thank you to my friends Simon Asenjo, Ariel Cornejo, Manuel Diaz, and Fabian Rubilar for being there for me when I needed a break from graduate school. Thank you also to my parents and grandparents for their unconditional love and support. Lastly, I would like to thank my fiancée Mattea Scanlan for proofreading, editing, and loving me through my best and my worst during graduate school.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Cochlear implants (CI) are sensory neuroprostheses capable of partially restoring hearing loss by electrically stimulating the auditory nerve to mimic normal hearing conditions. Despite their success and ongoing advances in both hardware and software, CI patients can still struggle to understand speech, most notably in complex auditory settings, also referred to as the cocktail party problem. Efforts to develop new CI algorithms to overcome this challenge rely on CI simulators and vocoders to test with normal hearing (NH) patients. However, recent studies have suggested that these tools fail to reproduce the stimuli perceived by CI patients. It is therefore critical to develop tools capable of producing better representations of the stimuli as perceived by CI patients. Thus, this work proposes a framework that incorporates physiological models of the peripheral auditory nerve. Using these models, the framework generates stimulations that elicit a neural response at the auditory nerve closer to that observed in NH conditions. Stimulations generated by the framework were evaluated by performing a vowel identification task. However, the task was performed by a classifier trained using deep learning techniques instead of a CI patient. These results give insight into how the framework could be applied for the development and validation of CI stimulation strategies.

1. INTRODUCTION

Hearing loss is one of the most prevalent disabilities in the United States, and different treatments are available based on its nature and severity. Among them, cochlear implants are arguably the most successful machine-brain interface capable of partially restoring auditory perception in those with moderate to profound sensorineural hearing loss, where more than 300,000 people have received implants worldwide [149]. Cochlear implants deliver electrical pulses to the cochlea inside the ear to directly stimulate the auditory nerve. Over the years, technology has advanced to provide better hardware and software to enhance speech quality; however, cochlear implants users still struggle to understand speech, most notably in complex auditory settings. Moreover, studies have shown that vocoders, a commonly used tool in the development of cochlear implant stimulation algorithms, misrepresent acoustic stimuli as perceived by cochlear implant users [37], [87]. Therefore, there is a need for developing tools that better represent the percept of sound that cochlear-implant users receive from the electrical stimuli.

To address this need, the present thesis uses a framework that incorporates physiological aspects of hearing rather than focusing on only the phenomenon itself [1], [86]. Therefore, the proposed framework seeks to match neural responses elicited by electrical stimulation of the auditory nerve to those observed from acoustic stimulation. To do so, computational models of the auditory nerve are chosen to represent both electrical and acoustic stimulation, and a cochlear implant simulator is used to recreate their processing algorithms. For the purpose of this work, the performance of the framework was evaluated using a deep learning-derived classifier performing a vowel identification task.

The remainder of this thesis is organized as follows. Chapter 2 presents an overview of the literature covering the anatomy of the peripheral auditory system, types of hearing loss, a description of cochlear implant technologies and their processing algorithms, computational models of the periphery auditory system, cochlear implant simulators, and performance of current technologies on speech perception. Chapter 3 presents the methods used, including the corpus used in this work, a detailed description of the proposed framework, and the classifier chosen to evaluate its performance. Chapter 4 presents the results of the functionality

of the framework and the performance of the classifier for a vowel identification task. Chapter 5 discusses the findings of this work and its shortcomings. Lastly, Chapter 6 proposes possible future directions to improve upon the work presented.

2. BACKGROUND

2.1 Hearing loss

Hearing loss is one of the most prevalent disabilities in the United States, with roughly 20 percent of the population, aged 12 or older, reporting unilateral or bilateral hearing loss [93]. The ability to communicate with others plays an important role in cultural and social life, and as such, efforts have been made for the past decades towards restoring hearing. Different technologies are available depending on the type and severity of hearing loss. However, the focus of this dissertation is on cochlear implant devices.

The following sections give an overview of the anatomy and physiology of the peripheral auditory system, as well as technologies developed for restoring hearing. In particular, components and signal processing algorithms used in cochlear implants are reviewed, as well as case studies illustrating their performance in speech perception and simulators used for evaluating signal processing algorithms. Following this, computational models of the peripheral auditory system are presented. Lastly, a framework using computational models for the development of signal processing algorithms on cochlear implants is introduced.

2.1.1 Normal hearing

The peripheral auditory system is comprised of three sections: outer ear, middle ear, and inner ear. The outer ear includes the pinna that helps with spatial sound localization, and the external auditory meatus (ear canal). The tympanic membrane or eardrum sits at the end of the ear canal and serves as a barrier separating the outer and middle ear. On the other side of the eardrum, a series of small bones or auditory ossicles (malleus, incus, and stapes) connect the tympanic membrane to the oval window in the cochlea. Lastly, the inner ear includes the cochlea that is connected to the cochlear nerve that goes to the brain. A diagram of the auditory peripheral system is shown in Figure 2.1.

The cochlea contains three chambers: the scala tympani, the scala media, and the scala vestibuli (see Figure 2.2). All three chambers are fluid-filled, where the scala vestibuli and scala tympani contain a sodium rich fluid called perilymph, and the scala media a potassium

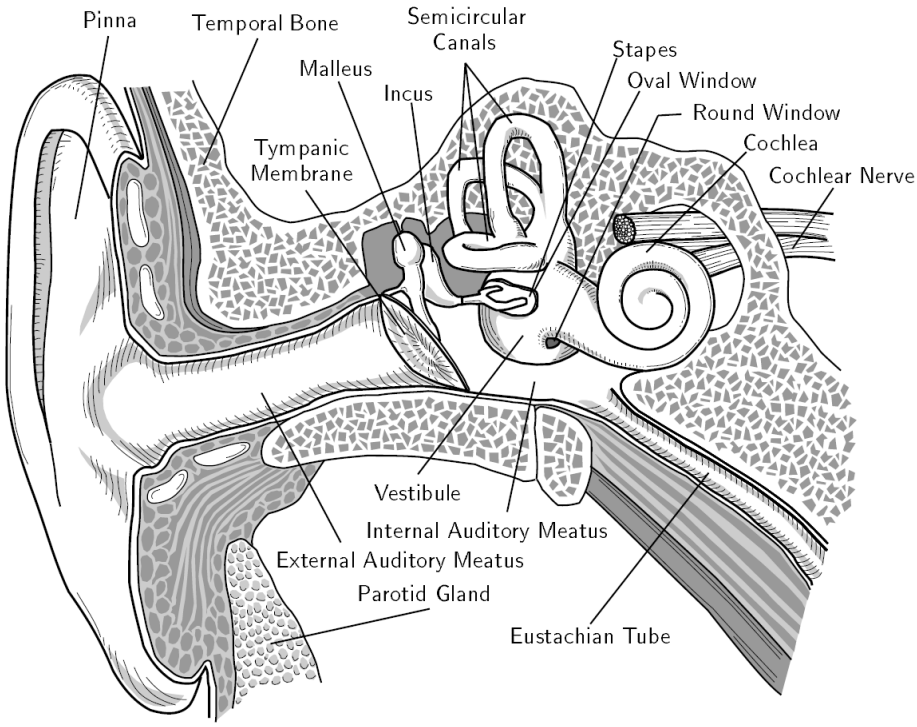


Figure 2.1. Anatomy of the auditory peripheral system. Image from Watts (1993) [144]

rich fluid called endolymph. The membrane separating scalae vestibuli and media is called Reissner's membrane, and separating scalae tympani and media is the basilar membrane (BM). If the cochlea is uncoiled, the area closer to the oval and round windows is referred to as base, and the tip where the scalae vestibuli and tympani meet is referred to as apex.

Inside the scala media and attached to the BM is the Organ of Corti, which contains the hair cells. There are two types of hair cells: outer hair cells (OHC), and inner hair cells (IHC); and humans have three and one rows of each, respectively, as shown in Figure 2.3. Atop the hair cells there are fine filaments called stereocillia, which in OHCs are attached to the tectorial membrane (TM), but not for IHCs. Each type of hair cell serves a different purpose and therefore their innervation differs as well. OHCs innervation is predominately of efferent fibers to receive information from the medial nucleus of the trapezoid body (MNTB) in the central nervous system (CNS). In contrast, IHCs innervation is predominately of afferent fibers from the spiral ganglion, specifically, type I neurons. This innervation allows IHCs to function as auditory sensory transducers [137]. Additionally, OHCs contain a motor

protein called prestin that allows them to transduce electrical stimulation into mechanical movement. As a result, when OHCs are stimulated by efferent fibers, they can modify their length to change the movement of the BM, acting as amplifiers.

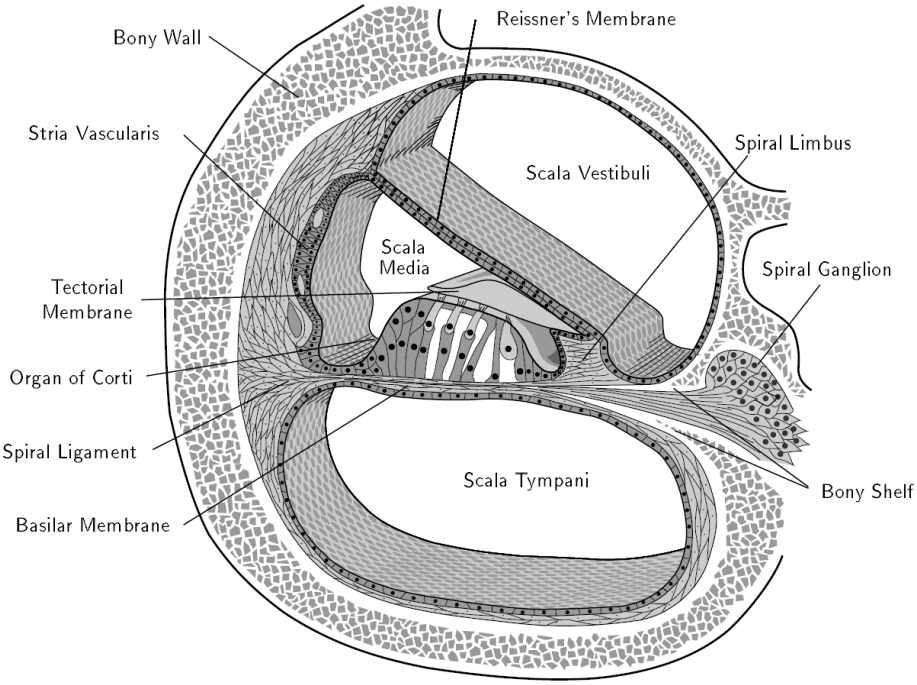


Figure 2.2. Cross section of the cochlea. Image from Watts (1993) [144]

When an auditory stimuli is presented to the auditory system, sound waves travel through the ear canal causing vibrations in the air. When these vibrations reach the tympanic membrane, the auditory ossicles transform it into mechanical energy and transmits it to the cochlea. The movement of the stapes connected to the oval window causes the fluid inside the cochlea to move allowing waves to travel from the scala vestibuli to the scala tympani. As the traveling waves propagate through the perilymph, the vibrations cause a displacement of the BM. Subsequently, the stereocillia of OHCs are pressed against the TM causing them to bend, and the stereocillia of IHCs bend as well, due to the drag force caused by the movement of the fluid between the TM and the hair cells. The movement of the stereocillia triggers an influx of ionic currents into the hair cells, causing depolarization of the membrane and the release of neurotransmitters. This increase of neurotransmitters facilitates the firing of an action potential or spike by the spiral ganglion cells (SGC) connected to the hair cells.

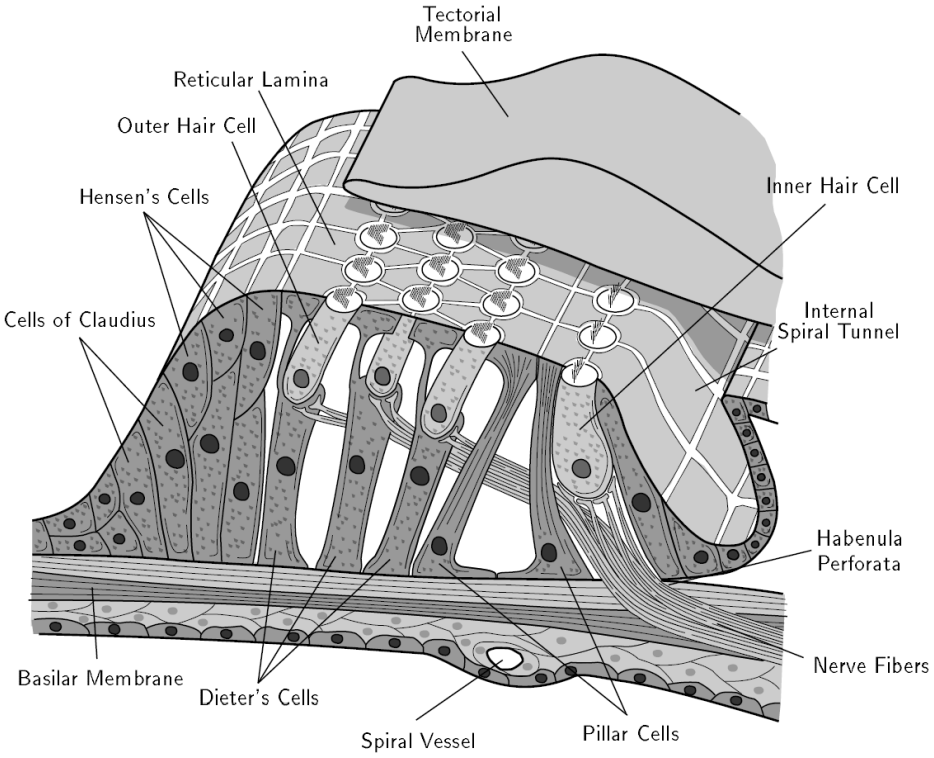


Figure 2.3. Organ of Corti. Image from Watts (1993) [144]

The pattern of neural activity generated by all IHCs going up the auditory nerve fibers (ANF) is referred to as a neural activation pattern (NAP), representing an estimate of the probability of the auditory nerve evoking an action potential as a function of time and location along the cochlea [124]. Subsequently, spikes generated by the SGCs are transmitted to the cochlear nuclei, where information is collected from each cochlea. This corresponds to the first relay of the primary auditory pathway, followed by stages at the superior olivary complex, inferior colliculus, medial geniculate body, and lastly, the auditory cortex.

As described above, hair cells are an essential component in the process of firing action potentials in response to the vibration of the BM from traveling sound waves inside the cochlear duct. The mass and stiffness of the BM varies from base to apex, producing maximum vibration amplitudes at the base for high frequency sounds, and maximum displacement at the apex for low frequencies [100]. The correlation between cochlear location

and stimulus frequency is known as tonotopy, and has been characterized by the Greenwood function shown in Equation 2.1 [61], [62].

$$f = A(10^{ax} - k) \tag{2.1}$$

Constants A, a, and k are determined based on experimental data and vary between species. In humans, $A = 165.4$ (to yield frequencies in Hz), $a = 0.06$ if x is expressed in millimeters, or $a = 2.1$ if normalized with respect to basilar length, and $k = 0.88$ (to yield a lower frequency limit of 20 Hz).

2.1.2 Impaired hearing

Reduction or loss of the ability to perceive sounds in one or both ears is referred to as hearing loss. The American Speech-Language-Hearing Association (ASHA) distinguishes three types of hearing loss based on their origin: (a) conductive hearing loss related to a malfunction in the outer or middle ear; sensorineural hearing loss (SNHL) associated with damage to the inner ear or auditory nerve pathway; and mixed hearing loss involving a combination of the previously mentioned conditions. Hearing loss can be classified in up to six categories based on the severity of the loss: (1) slight (16–25 dB HL), (2) mild (26–40 dB HL), (3) moderate (41–55 dB HL), (4) moderately severe (56–70 dB HL), (5) severe (71–90 dB HL), and (6) profound (91+ dB HL) [22]. Decibel (dB) is a logarithmic unit used to express the magnitude of a signal relative to a reference value. In audiology, dB HL (dB hearing level) refers to the loudness of a sound in dB relative to the quietest sound an average normal hearing (NH) listener can perceive [130].

Depending on the type and severity of hearing loss, different technologies are available. For a mild to moderate hearing loss, a hearing aid (HA) can help restore hearing. The HA amplifies the incoming sound and can be programmed to do so at specific frequency bands according to the need of the patient. HAs are typically worn behind the ear (BTE) or in the ear.

If the patient experiences a moderate to profound SNHL, then a cochlear implant (CI) is a viable alternative. CIs are medical devices designed to electrically stimulate the auditory

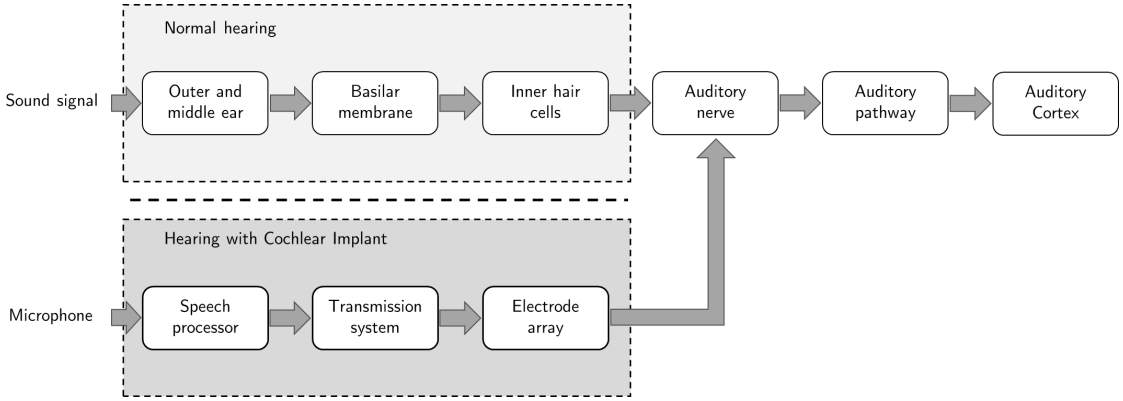


Figure 2.4. Diagram with stages of the peripheral auditory system for normal (upper) and electrical (bottom) stimulation.

nerve fibers inside the cochlea, bypassing the functions of outer, middle, and inner ear as shown in Figure 2.4. Various stimulation strategies have been developed in the past decades to deliver stimuli that best reproduce normal hearing. A more in depth discussion will be covered in the following sections.

In cases where patients qualify for a CI but have residual hearing at low frequency bands, then an electric-acoustic stimulation (EAS) implant could help. This type of system combines both technologies described above: a HA processes low frequency sounds such as barks, honks, or vowels; and a CI processes mid and high frequency sounds such as whistles, chirps, or consonants. EAS systems have shown to provide significant benefits when compared to electric stimulation alone by preserving and utilizing the residual hearing of the patient [74].

2.2 Cochlear Implants

The most common causes of SNHL involve damage or loss of cochlear hair cells in the inner ear due to natural aging of the auditory system, also known as presbycusis, or prolonged exposure to high-level of noise. Under these circumstances, the central auditory system remains intact and restoration of some percept of hearing is possible, by using an alternative source of auditory nerve stimulation: cochlear implants [100]. CIs are arguably the most successful machine-brain interfaces for restoring auditory perception in those with moderate

to profound SNHL, and more than 300,000 people have received CIs worldwide [149]. These implantable devices can benefit both children and adults with prelingual and post-lingual deafness. There are currently three CI manufacturers with Food and Drug Administration (FDA) approval to distribute devices in the United States: Cochlear Inc., MED-EL, and Advanced Bionics. Although implantation criteria vary among manufacturers, it is generally agreed that adults with bilateral severe-to-profound hearing loss are eligible candidates, as well as children ages 12 to 14 months with bilateral profound hearing loss [55].

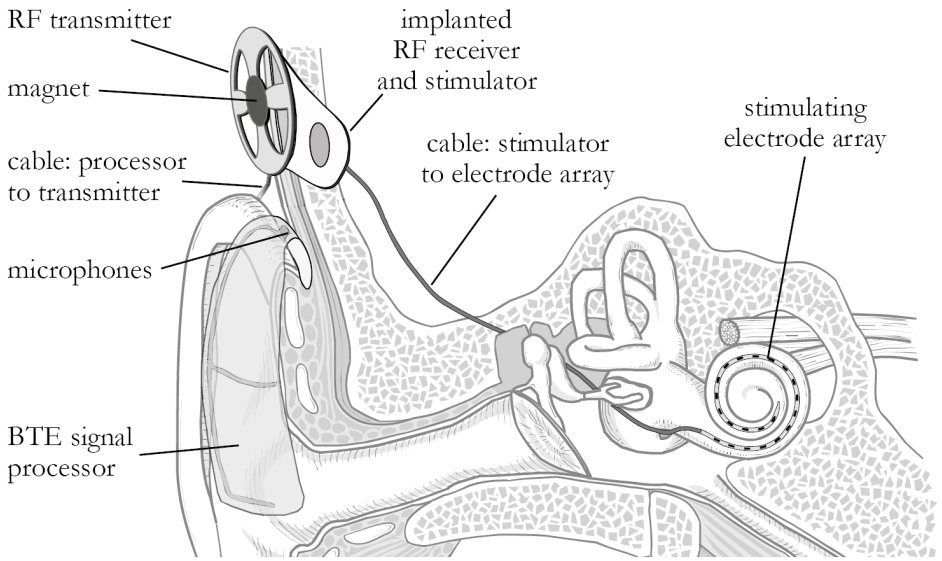


Figure 2.5. Schematic of a cochlear implant's core components. Image from Harczos (2015) [63]

CIs have evolved since the first implementation of an auditory prosthesis in 1957 [28]–[30], [150]. However, a common architecture can be identified among all developed devices: (1) an external unit, (2) a transmission link, and (3) an internal unit. Figure 2.5 shows a typical block diagram of a modern CI. The external unit consists of a microphone to capture sound waves, a digital signal processor (DSP) unit to extract features from the audio signal and generate a stream of data, and a power amplifier to drive energy to the transmission link. The transmission link can be either a percutaneous connection or a transcutaneous connection. In systems with a percutaneous connection, no electronics other than the electrodes are inserted inside the skull, leaving a plug to connect the external unit. In contrast, systems with a transcutaneous connector insert an electronic receiver that

decodes the signal transmitted through a radio frequency (RF) link from the external unit. Percutaneous implantation has the advantage of not being constrained on the information that can be decoded by an implanted receiver under the skull, making for a flexible platform for researching new stimulation strategies [94]. However, the Ineraid implant, a percutaneous cochlear implant developed at the University of Utah, never received FDA approval in the United States. As of 2008, all commercially available devices with FDA approval have opted for a transcutaneous connector [94].

In systems with a transcutaneous connector, the internal unit is powered by the transmission link using coils. The stream of data is then decoded to determine the sequence of electrodes to be stimulated in the electrode array, located inside the cochlea. Modern CIs have electrode arrays with up to 22 electrodes. Additionally, some devices have incorporated a back-telemetry circuit for monitoring and evaluating purposes [88], [150]. Detailed descriptions of CI components and their functionality have been well documented in the literature [38], [94], [146], [150].

The most commonly used stimulation strategies include: (a) continuous interleaved sampling (CIS), (b) advanced combination encoder (ACE), (c) MP3000$^{\text{TM}}$, (d) fine structure processing (FSP), and (e) HiRes 120 (HiResolution). These approaches have been developed and implemented by the top three CI manufacturers in the international market: Cochlear Inc.(CIS, ACE, MP3000$^{\text{TM}}$), Med-El (CIS, FSP), and Advanced Bionics (HiRes 120).

Recent studies have developed new algorithms to enhance CI strategies by addressing improved processing performance [3] and channel selection [4], [5], [113], as well as incorporating temporal fine structures [90], [111], across-frequency delays [140], and rapid temporal adaptation [50]. Other attempts have tried less traditional approaches based on neural networks (NN) [12], wavelet transform [27], [59], [115], bionic wavelet transform [26], and auditory models [1], [60], [64], [80], [86], [87]. Some of these stimulation strategies are discussed in the following section.

2.3 Stimulation Algorithms

Multiple stimulation strategies have been developed over the past decades which vary in the number of channels used to code the input signal, the number of electrodes stimulated, the type of stimulation, and the extracted spectral features. The audio signal is divided into sub-bands to extract spectral information, where a large number of bands provides better spectral resolution, i.e., the ability to carry pitch information from the acoustic source to the brain. In most cases, the number of bands used is directly related to the number of electrodes available. Multiple algorithms have been developed to decide what features are relevant to produce a high-quality representation of the input signal [94].

In 1961, House and Doyle implanted two subjects with an electrode in the inner ear and reported some auditory percept when electrically stimulated. Similar results were reported by Simmons at Stanford in 1965 [6], [135], [150]. Their efforts led to the development of single-channel CIs to stimulate the membrane at fixed locations inside the cochlea [45], [73], [104]. Despite their low spectral resolution, subjects were able to recognize phonemes and words; however, their performance was highly variable and inconsistent between English [24] and German [143] speaking subjects. During the late 1970s and early 1980s, multi-channel CIs were introduced, providing wider coverage of the cochlea to exploit its place-frequency encoding mechanism, thereby increasing the spectral information transferred to the auditory nerve from the audio signal [39], [103], [131]. Another important aspect to consider is the type of stimulation used, which can be analog or pulsatile. Early CIs used a compressed analog processor which presented analogue waveforms as stimuli to the electrode array, limiting implants to simultaneous excitation of all electrodes. Modern CIs are digital and programmed to perform multiple stimulation strategies while delivering both analog and pulsatile stimuli.

As previously mentioned, the number of electrodes is closely related to the number of bands processing the input signal. However, recent developments have introduced the concept of virtual channels where two or more neighboring electrodes are stimulated to create virtual channels between physical electrodes, thus increasing the spectral resolution [42], [81]. CI Manufacturer Advanced Bionics implemented this approach in their devices to create the

HiRes Fidelity 120^{TM} stimulation strategy, which demonstrated improved speech perception performance in both quiet and noisy conditions [14], [20], [82].

As for strategies with fixed channels, two different approaches have been used based on the number of electrodes to be stimulated. While Continuous Interleaved Sampling (CIS) uses all available electrodes to stimulate the cochlea, others (Advanced Combination Encoder (ACE) and Spectral Peak Coding (SPEAK)) solely use channels with the highest envelope amplitudes. This is referred to as 'n-of-m' coding where only 'n' (typically four to eight electrodes) out of 'm' available electrodes are stimulated. While the strategies mentioned so far seek to better represent spectral information carried in the audio signal, other strategies, such as FSP or MP3000^{TM}, aim to better represent the fine temporal structure of the audio source, which is usually masked by higher magnitude spectral components.

These strategies have been well documented in other studies, and a brief description of each strategy is presented in the following subsections [95], [149], [150].

2.3.1 Continuous Interleaved Sampling (CIS)

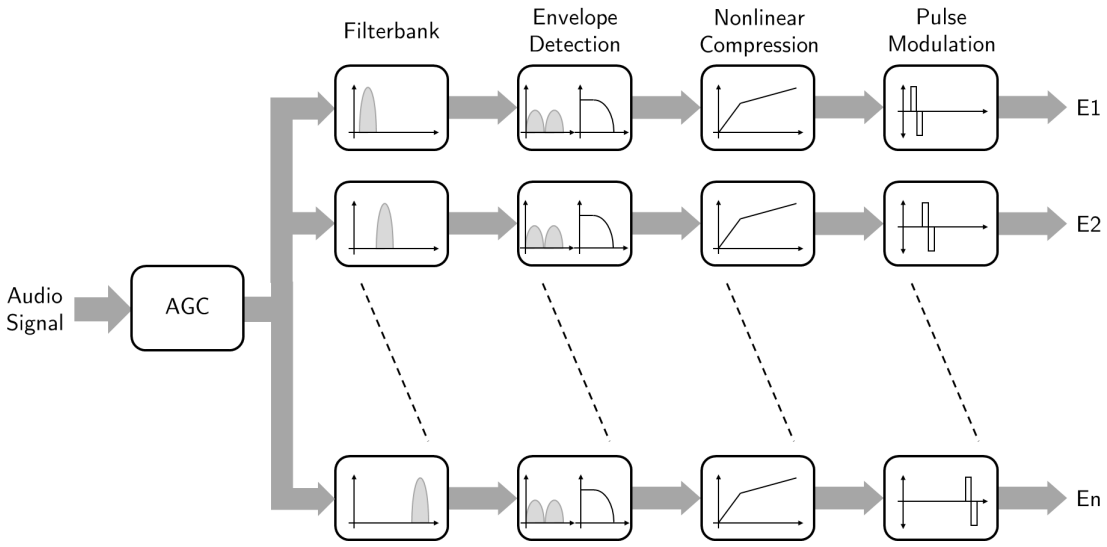


Figure 2.6. Block diagram of Continuous Interleaved Sampling algorithm. Based on Ahmad et al. (2009) [3]

CIS was developed by researchers at the Research Triangle Institute (RTI) as a solution to the channel interaction problem observed in Compressed Analog (CA) processors where all channels are simultaneously stimulated [148]. CIS is a multi-channel strategy that extracts spectral information from the audio signal. As result, interleaved, non-simultaneous pulses are delivered to each electrode, thereby avoiding interference [147], [148].

As shown in Figure 2.6, the audio signal is passed through an automatic gain circuit (AGC) to pre-emphasize high frequencies resulting in a flatter spectrum, which facilitates extraction of spectral cues. Then, a filter bank, composed of band-pass filters, is applied to divide the spectral information into sub-bands. At the output of the filter bank, signals are rectified and low-pass filtered with a typical cutoff of 200 Hz. Each resulting amplitude is compressed to match the dynamic range of the corresponding electrode. These ranges are specific to each CI user and can vary among channels. To compress the signal, two non-linear transformations are typically used: (a) power-law compression function (Equation 2.2) and (b) logarithm compression function (Equation 2.3).

$$y = Ax^p + B \tag{2.2}$$

$$y = A\log(x) + B \tag{2.3}$$

In these equations, A and B are constants, x is the uncompressed amplitude, and y is the compressed amplitude. For power-law transformation, $p < 1$. Parameters A and B are defined using two acoustic measures: T-Level and C-Level, and the relationships are as follows:

$$A = \frac{C_{level} - T_{level}}{x_{max} - x_{min}} \tag{2.4}$$

$$B = T_{level} - Ax_{min} \tag{2.5}$$

T-Level is the threshold in dB for the softest sound the CI user can detect, and C-Level is the highest comfortable loudness level that the CI user can tolerate. These parameters are patient-dependent and can vary from channel to channel [3], [95].

Once the amplitudes to each electrode are compressed to the corresponding dynamic range, a biphasic pulse modulates the output for time multiplexing, and therefore avoids inter-channel interference. Stimulation rates range from 500 to 2000 pulses per second (pps) per channel, and 12 to 24 channels are typically programmed in this scheme.

2.3.2 Advanced Combination Encoder (ACE)

Developed as a successor to the Spectral Maxima Sound Processor (SMSP) and Spectral Peak coding (SPEAK), ACE coding is a multi-channel stimulation strategy sharing many similarities with CIS coding. However, only frequency bands with the largest amplitudes are selected for determining which electrodes must be stimulated during each cycle. In contrast with SMSP where electrodes are stimulated in order of descending amplitude, ACE presents them in tonotopic order [132]. This scheme is referred to as 'n-of-m', where 'n' electrodes with the largest envelope amplitudes are selected out of 'm' electrodes available. ACE is the most commonly used scheme by Cochlear Inc. in their Nucleus implants, where four to eight electrodes are stimulated out of the total electrodes available. SPEAK and ACE follow the same scheme of stimulation, however, ACE is capable of achieving higher stimulation rates, between 600 to 1800 pps, compared to a limited 200-300 pps range in the SPEAK scheme [79].

2.3.3 Fine Structure Processing (FSP)

Previously described algorithms rely on the amplitude of band-pass filter outputs extracted with an envelope detector to determine which electrodes to stimulate, although no frequency information is recovered. However, using the Hilbert transformation, the signal can be decomposed into its envelope and temporal fine structure (TFS) as shown in Figure 2.7.

$$y(t) = A_c \cos\left(2\pi f_c t + 2\pi f_\Delta \int_0^t x_m(\tau) d\tau\right) \tag{2.6}$$

TFS shown in Figure 2.7 is referred to as a frequency modulated (FM) signal, defined in Equation 2.6. The carrier's base frequency, f_c, corresponds to the central frequency of

the band-pass filters used for processing the auditory stimulus, f_Δ is the frequency deviation away from f_c, and $x_m(\tau)$ characterizes the modulation over time. TFS has been shown to play an important role in pitch and speech perception, and efforts have been made to incorporate it into improving CI performance [107], [108]. MED-EL developed the FineHearing^{TM} technology based on FSP, where a burst of pulses is delivered to more apical electrodes (i.e., to the low frequency range), while the remaining electrodes follow a CIS-like scheme. These bursts are aligned with the zero-crossing locations in the band-pass filter output to preserve the fine structure information [7], [71], [126].

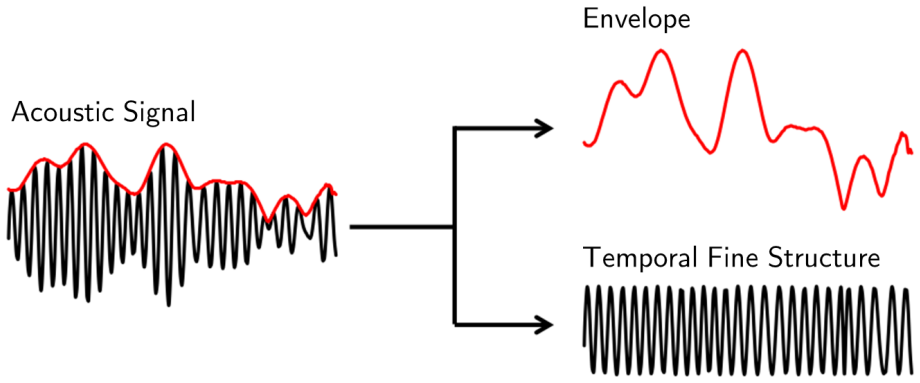


Figure 2.7. Acoustic signal decomposed into its envelope and temporal fine structure.

2.3.4 MP3000

MP3000^{TM} is a stimulation strategy that introduces a psychoacoustic masking model to reduce the number of active channels and to improve speech perception [114]. The concept underlying masking models describes excitation at a given location in the cochlea spreading out to adjacent regions and stimulating adjacent frequencies based on tonotopic coding. Therefore, low-energy frequency components close to a stronger adjacent component are masked, and there is no need to stimulate those regions, increasing spacing between stimuli in the cochlea and reducing electrode interference [18], [149]. This concept has previously been used to compress audio signal under the MP3 standard. Following an 'n-of-m' strategy, the highest 'n' components, relative to an estimate of the spread of masking, are selected.

While MP3000^{TM} is used exclusively for implants manufactured by Cochlear Inc., a similar strategy, Psychoacoustic Advanced Combination Encoder (PACE), is widely available [114].

2.3.5 HiRes (High Resolution)

Advanced Bionics introduced the HiRes (High Resolution) processing scheme to incorporate TFS of the audio signal. This system is based on the use of virtual channels to increase the number of excitable locations lying along the cochlea. Using independent current sources to drive each electrode, one or two adjacent electrodes can be simultaneously activated and current can be adjusted to control the location at which the maximum stimulation is delivered between the two physical electrodes.

The above concept, dubbed denominated current steering, was studied by Donaldson et al. (2005) who conducted a study on six post-lingually deafened adults [32]. Results showed that subjects could successfully discriminate different pitches depending on the level of current delivered to each pair of electrodes. Koch et al. (2007) showed that CI recipients using HiRes 90k, a processing strategy implemented in Advanced Bionics implants with 16 electrodes and a high stimulation rate (83,000 pps), were able to discriminate between an average of 93 potential channels [81]. Subsequently, an updated version of the standard HiRes 90k introduced eight additional stimulation sites between each pair of electrodes, creating a total of 120 potential spectral channels, hence, HiRes 120.

2.4 Speech understanding performance

Audiological evaluation is used to assess both pre- and post-operative performance of CI patients. In 1996, the American Academy of Otolaryngology-Head and Neck Surgery (AAO-HNS), along with the American Academy of Audiology (AAA), and CI manufacturers recommended the usage of a Minimum Speech Test Battery (MSTB) for clinical and research assessment [105]. This battery set includes monosyllabic consonant-nucleus-consonant (CNC) words [91], [122] to assess open-set word recognition, and utterances from the hearing in noise test (HINT) [112] to assess open-set sentence recognition in quiet and speech-shaped noise. A revised version of the MSTB was issued in 2011, replacing HINT sentences by the

AzBio, and Bamford-Kowal-Bamford Speech-in-Noise (BKB-SIN) tests. No changes were issued to CNC word tests [105].

In open-set tests, patients are asked to listen to auditory stimuli and repeat back what they heard without providing them feedback. Speech-shaped noise is used to simulate more everyday life situations where the patient tries to understand speech while other acoustic sources are present in the background. For example, having a conversation at a crowded bar or restaurant. This problem is referred to as the cocktail party.

A review of case studies evaluating the speech performance of the CI stimulation strategies described in Section 2.3 is presented below.

2.4.1 Continuous Interleaved Sampling (CIS)

Pelizzone et al. (1995) conducted a study comparing CI users using a Ineraid implant (CA strategy) to the CIS scheme. Two patients were evaluated during a six month period using a portable CIS processor prototype. Speech perception was evaluated using consonant and vowel identification tests in the absence of background noise. Results reported a 5-11% improvement in hearing performance during experimental sessions, as well as in everyday life usage with CIS compared to the CA scheme [121].

Kompis et al. (1999) conducted a similar study comparing both strategies in three experienced Ineraid users who participated in three sessions over a period of three weeks. Five different levels of signal-to-noise ratios (SNR) with broadband noise were evaluated. Results indicated that consonant and sentences are better transmitted by CIS when no noise is present (average score differences of CIS with respect to CA for the consonant tests yielded +7.8% correct at 15 dB SNR, and -6.8% at 5 dB SNR). However, vowel identification tests favored CA over CIS (average differences ranging from -5% to -20%), most likely due to a longer training period when switching between schemes [84].

2.4.2 Advanced Combination Encoder (ACE)

In a study comparing ACE and SPEAK conducted by Pasanisi et al. (2002), nine congenitally deaf children, wearing a Nucleus CI24M implant, were evaluated over a period of

three months [116]. At the end of the study, subjects reported significant improvements on open-set word and sentence recognition tasks with the ACE strategy, noting the greatest improvement in the presence of background noise at SNR of +10 dB (mean score differences of +20% in quiet, and +40% in noise).

In a separate study conducted by Psarros et al. (2002), seven children participated in an A-B-A experiment comparing SPEAK and ACE strategies over a period of 16 weeks [123]. The SPEAK strategy was used at the first and final sections of the study (A) corresponding to 2 and 4 weeks respectively. The ACE strategy was used during the middle of study (B) for 10 weeks. Using 2-way analysis of variance (ANOVA), significant improvements were reported for open-set monosyllabic words in quiet conditions, and sentence improvements were reported in competing noise. Both cases revealed mean differences ranging from +5% to +10% when switching from SPEAK to ACE. However, SPEAK outperformed ACE in sentences scores when subjects switched back to SPEAK. Improvements in the production of medial consonant sounds were also observed using the ACE strategy (mean score difference of +5.2%).

2.4.3 Fine Structure Processing (FSP)

In a study by Magnusson (2011), 20 experienced CI users were evaluated to compare an FSP strategy to CIS variations (CIS+ and HDCIS, both implemented in MED-EL devices) in speech and music perception. Statistical analysis using a repeated measures ANOVA showed significant within-subject differences in mean scores. Bonferroni adjusted pairwise comparisons between initial and final mean scores revealed no significant differences between strategies, although participants tended to prefer HDCIS over FSP for both speech and music [99].

In a different study by Muller and Mertins (2012), 46 adult CI users tested the same three strategies over a period of four months. Pairwise comparison tests yielded statistically significant non-inferior performance of FSP compared to HDCIS and CIS+ in speech understanding tests, except for monosyllable scores, where FSP outperformed CIS+. A significantly lower pitch perception and preference for FSP was also reported [110].

Two additional coding strategies based on FSP, FS4 and FS4-p, have been introduced by MED-EL to extend the maximum number of fine structure channels from three to four. While no significant differences regarding speech perception were observed among the three strategies, most participants preferred either FS4 or FS4-p over FSP. This outcome is believed to be influenced by the last strategy used by the participants [127].

2.4.4 MP3000

The study by Büchner et al. (2011) compared MP3000^{TM} coding to the SPEAK/ACE strategies in 221 Nucleus implant recipients following an A-B-A-B-A design over a time period of 14 weeks [18]. Strategies SPEAK/ACE (A) and MP3000 (B) were alternated following 4, 4, 2, 2, and 2 week periods respectively. Individuals were scored on identification performance in quiet conditions (depending on their native language), and sentences in the presence of noise. ANOVA tests showed no significant differences among strategies for both speech perception and coding preference in quiet and in noisy conditions. Fixed stimulation rates were used for all three schemes; however, a battery duration increase was observed when MP3000^{TM} was used due to the lower number of active channels in MP3000^{TM} compared to SPEAK/ACE (4 to 6 vs. 6 to 14).

A similar study conducted by Lai et al. (2008) compared music perception in two experienced CI users but yielded no significant differences between strategies [89]. Nogueira et al. (2005) conducted a similar study to compare PACE and ACE strategies in eight adult users of Nucleus 22. Results showed significant improvements in speech perception when four electrodes were stimulated, but no significant differences were observed when eight electrodes were selected [114].

2.4.5 HiRes (High Resolution)

Studies comparing both standard HiRes and HiRes 120 have reported no significant differences in speech recognition tests in quiet and in +10 dB SNR conditions. However, subjects using HiRes 120 reported higher ratings on music quality in terms of pleasantness

and distinctness, as well as an overall preference for HiRes 120 at the conclusion of the studies [19], [41].

2.4.6 Summary

Despite the improvements achieved by each individual stimulation strategy mentioned above, overall speech perception scores are not significantly different among them. In practice, most post-lingually deafened users of CIs performing open-set CNC word identification tasks, the gold standard in CI testing [10], achieve around 60% correct rates in quiet conditions [8], [56], [72]. However, this performance drops significantly in common social conditions of background noise [40], [46], [70].

2.5 Cochlear Implant Simulators

For the past 20 years, noise-band vocoders have been used in research to simulate the output delivered to CI users [46]–[48], [96], [133], [134]. These simulators have been used to evaluate speech perception in NH subjects, increasing control over experiments and the size of potential subject pools. In this process, acoustic signals are first filtered using simulations of CI processors. Next, amplitude profiles are extracted from outputs of each channel and used to modulate limited-band noise or sine waves centered in the middle of each analysis band. Resulting signals are then delivered as acoustic stimuli to NH subjects.

Although simulators cannot provide the exact stimulation experienced by CI users, multiple studies have been conducted to compare speech perception between NH subjects and CI users. These studies have been done to evaluate the effect of channel number [33], [34], [36], [46], insertion depth [35], intensity resolution [97], and information transfer [2]. While stimuli can be generated offline, most studies have used computer-driven CI simulators to provide real-time assessment of acoustic stimulation [43], [44], [78], [109], [138].

The study conducted by Svirsky et al. (2013) evaluated speech intelligibility in a CI user with single-sided-deafness [138]. Stimuli were presented directly to the implant (with the microphone deactivated) and through a loudspeaker with the goal of assessing similarities between stimuli when different CI simulators were used. Results from this case study showed

that most CI simulators sounded better than the sound provided by the CI, suggesting that results from NH subjects cannot be directly extrapolated to CI users.

In a separate study, Laflen et al. (2002) computed neural responses using auditory models in response to an acoustic stimulus and its representation using either a vocoder or a CIS strategy. Results showed that the neural response from the vocoder was closer to that from normal hearing than the one observed using a CI stimulation strategy (CIS or n-of-m), suggesting that they may not be a good representation of implant stimulation [87].

2.6 Computational Models

Stimulation strategies used in CI processing as described in Section 2.3 use a phenomenological approach focus on better characterizing acoustic stimuli rather than taking advantage of the underlining physiological process involved. Therefore, physiological computational models are useful resources to describe the auditory system, thus their development can help to better understand how the peripheral auditory system codes sound in normal and impaired hearing conditions. In 1975, Biondi et al. proposed a mathematical model to characterize the behavior of the peripheral auditory nerve [11]. Since then, advances in computational power have made possible the development of more complex mathematical models to compare and contrast predictions with experimental data.

Since models of the auditory pathway to the auditory cortex are beyond the scope of this dissertation, only models of the peripheral auditory system, from the outer ear to the auditory nerve, will be discussed. The complexity of the auditory system makes it challenging to combine components into one unique model. Therefore, multiple models have been developed over the past decades, which focus on specific stages of audition that can be coupled together in cascade [125]. Some of these models have been integrated into frameworks, such as the Auditory Image Model (AIM), and the Auditory-Nerve Fibers models from Carney Lab. A brief description of these frameworks is presented below.

2.6.1 Auditory Image Model (AIM)

Developed by Patterson and Holdsworth, the Auditory Image Model (AIM) seeks to analyze complex sounds (everyday sounds, music, speech, etc.) and transform them into 'auditory images' representing our initial impression of sound in the brain [119]. Early implementations of AIM used a functional model; however, growing interest in the scientific community to compare results more directly with physiological models led to the development of AIM as shown in Figure 2.8 [118].

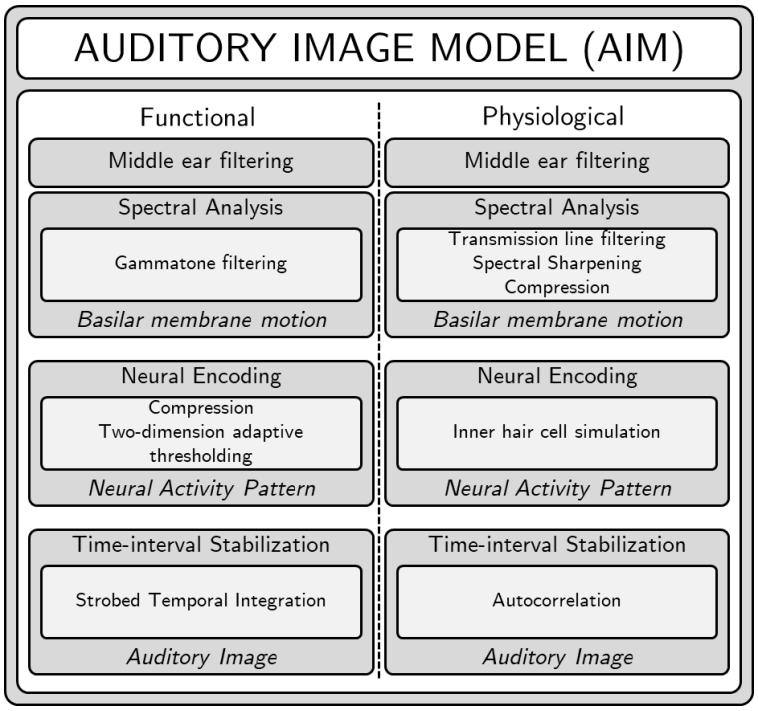


Figure 2.8. The three-stage structure of AIM. Left-hand column: functional path; right-hand column: physiological path. Based on Patterson et al. (1995) [118]

AIM is structured in three stages: spectral analysis, neural encoding, and time-interval stabilization, with a pre-processing stage where the input signal is filtered to simulate the middle ear transfer function. Various approaches have been used to model the middle ear transfer function: analog electrical circuits [58], [85], [106], [117], biomechanics and finite element methods [49], [83], and linear digital filters [128], with the latter being most commonly used.

On the spectral analysis stage, the functional path uses a gammatone function to convert sound waves into basilar membrane motion (BMM). This function is known to account for physiological characteristics observed in the impulse response of the auditory filter in mammals, and also psychological behaviors comparable to functions commonly used to characterize the human auditory filter [119]. The gammatone filter is a linear filter with a symmetric frequency response, but BM responses are non-linear and asymmetric depending on their location along the membrane and the amplitude of the stimulus [125], [129]. Gammatone function variants have been designed to account for these discrepancies and contribute to a more physiologically-relevant filter. Examples of these efforts include the all-pole gammatone filter (APGF) [98] and the gammachirp filter [75], [76], both producing an asymmetric gammatone-like filter. On the other hand, the physiological approach uses a transmission line model to represent cochlear hydrodynamics obtaining time-domain numerical solutions by using a technique called wave digital filtering [57].

On the neural encoding stage, the output from the BMM is converted into a NAP, simulating the transduction process at the IHCs. In the functional path, two-dimensional adaptive thresholding is applied simultaneously in time and frequency domains, introducing compression, rectification, adaptation, and enhancement of the stimulus. Adaptive thresholding causes low activity areas to be suppressed by higher activity areas, creating a masking effect that helps sharpen formants and reduce noise. The thresholding output generates a NAP which represents afferent neural activity at the auditory nerve. In the physiological path, individual IHC models [102] are coupled to each output of the BMM to simulate the flow of neurotransmitters from IHCs to ANFs. This model accounts for the non-linearities occurring at the junction between hair cells and ANFs.

NAPs are not representations of sounds perceived by humans, but temporal integration of NAPs can produce auditory images. However, temporal integration removes fine structure information contained in the NAPs, which is known to play an important role in assessing sound quality and source identification [118]. Therefore, AIM implements a module to preserve fine structure information during integration thereby producing better representations of auditory images [118]. In the functional path, strobed temporal integration uses a bank of delay lines to store the NAP [120]. On the other hand, the physiological path implements

an auto-correlation function between each channel of the NAP to extract periodicity and preserve fine structure information [136].

2.6.2 Auditory-Nerve-Fiber Model by Carney Lab

Carney and colleagues from Carney Lab at the University of Rochester developed a composite model of the auditory nerve response from the auditory stimulus at the eardrum to the spikes transmitted to the cochlear nucleus via the auditory nerve. An early version of this model was proposed by Carney to model the auditory-nerve fiber response in cats [21]. This model incorporates a narrow-band filter to represent the BM, and a feedback loop for the OHCs' response properties. A low-pass filter was used to model the effect of the IHCs, followed by a non-linear adaptation stage between the IHCs and the auditory nerve that affects the discharge rates of the auditory fibers.

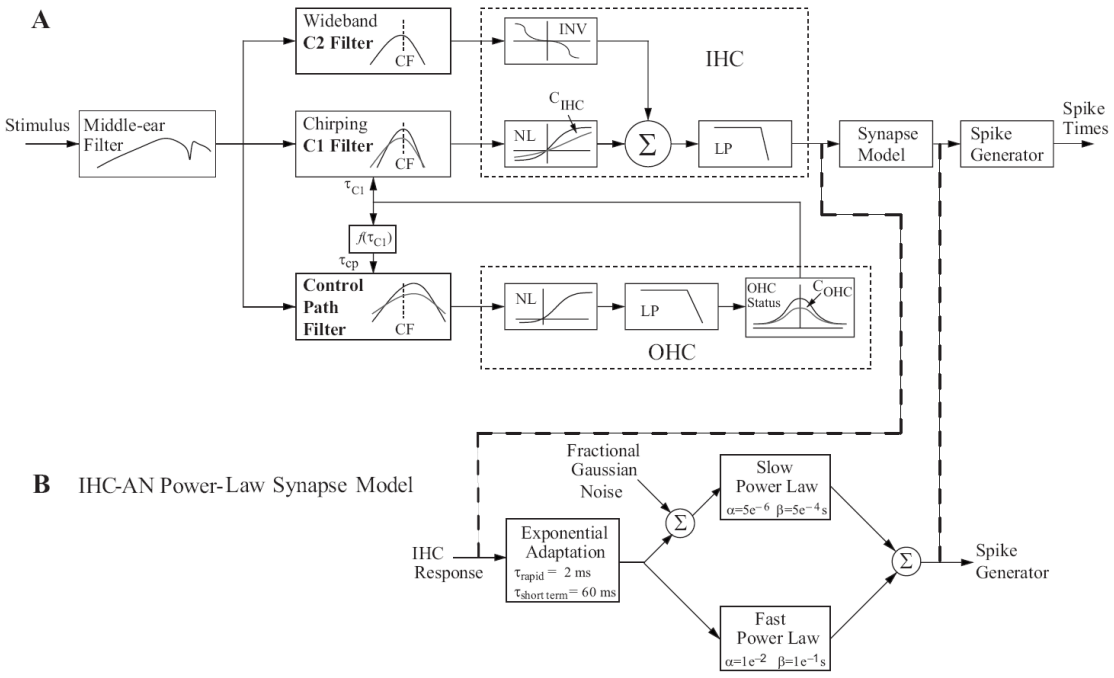


Figure 2.9. Diagram of the computational model of the auditory-nerve response developed at Carney Lab. Image from Zilany et al. (2009) [156]

Zhang et al. (2001) extended this work by replacing the feedback loop for a wide-band, feed-forward control path, adding non-linear response properties as well as two-tone sup-

pression [152]. However, this model was limited to responses of high-spontaneous-rate ANFs. This was then expanded by Heinz et al. (2001) to include low- and medium-spontaneous-rate ANFs. Heinz et al. also incorporated a time-varying discharge rate of the ANFs rather than using a modified Poisson renewal process. Additionally, frequency resolution was modified to fit human data, which allowed evaluation of normal and impaired human psychophysical performance [68]. Subsequent studies explored performance limits achieved by the model [66], [67] and added level-dependencies to provide more complete response features [142].

Bruce et al. (2003) used Zhang's model to study how impaired OHCs and IHCs would impact the auditory nerve response [16]. Results yielded degraded tonotopy in the cochlea, showing potential to predict the effects of frequency modulations on the auditory nerve. Zilany and Bruce (2006) extended this model to incorporate high-level responses of ANFs, with previously determined low and mid-level responses [153]. This version of the model is represented in the upper section of Figure 2.9. In this approach, an acoustic signal measured in Pascals is first filtered through the middle-ear (ME) filter. This filter is adapted from Bruce et al. (2003), but simplified from eleventh-order to fifth-order to ensure stability [16]. C1 and C2 filters are presented in Bruce et al. (2003) [16] and Zhang et al. (2001) [152], where C1 models the interaction between OHCs and taller rows of IHCs, and C2 models the shorter IHCs independent of neighboring hair cells [92]. Next, the feed-forward control path introduced by Zhang et al. (2001) models the level-dependent properties in the cochlea by regulating the gain and bandwidth of the C1 component. The outputs of the C1 and C2 filters are low-pass filtered by the IHC membrane, and then pass through the IHC-AN synapse model and discharge generator. The synapse model is the same as in previous versions and uses a time-varying three-store diffusion model [21], [145]. The model was used to predict vowel responses in cats and compared predictions to physiological data. The researchers observed a qualitative and quantitative match between their predictions and the physiological data [154].

Despite these promising results, the synapse model did not account for offset adaptation after the stimuli. Therefore, Zhang and Carney (2005) proposed a modified model based on models by Meddis (1986) [101] and Westerman and Smith (1988) [145] to account for both onset and offset adaptations and yield a more physiologically accurate synapse [151].

First, a shift is introduced to the synapse output that preserves the onset response of the original model, but generates a faster offset adaptation. However, this shifting generates an unnatural modulation in the auditory nerve response.

In a more recent study, a different approach was taken to incorporate both exponential and power-law dynamics and the physiological behavior observed within the IHCs and their interactions with ANFs. These modifications were developed by Zilany et al. (2009) as shown in the lower section of Figure 2.9, where two parallel power-law functions are added after the exponential functions used previously [156]. This new approach successfully describes responses to amplitude-modulated tones, noise, pure tones, and spontaneous activity that was not achievable with previous models. The most recent implementation of this model has addressed minor issues to better represent physiological data while preserving its structure and core components [155].

2.7 Physiological-based stimulation strategy

Auditory models have been integrated into speech coding algorithms to mimic human performance in speech recognition tasks [31], [51]–[54]. Furthermore, some of these models have been used in the development of physiologically based stimulation strategies that could be implemented on CI processors [60], [63], [64], [80]. While these approaches have integrated auditory models to better mimic normal hearing auditory processing, the resulting neural response elicited at the auditory nerve is not guaranteed to be closer to that of NH conditions since there is not feedback loop or optimization implemented into their designs. Therefore, and following the findings in Laflen et al. (2002) described in Section 2.5, Laflen (2003) proposed solving the inverse problem of finding the optimal electrical stimulation such that the neural response elicited at the auditory nerve by a CI processor best matches that observed in NH conditions [86]. It is worth mentioning that Bondy et al. (2004) proposed and implemented a similar approach to that from Laflen (2003) but for HA devices. Using the auditory periphery model from Bruce et al. (2003) [16], a neurocompensator is proposed to fit a HA such that its neural response best matches that of normal hearing [13]. However,

these approaches are usually complex and computationally expensive, thus limiting their use in practical applications.

This section describes the framework proposed by Laflen (2003) [86], later implemented by Aguiar (2012) [1], including preliminary results of an unpublished case study using the proposed framework. Their findings serve as the basis of the present dissertation work to be covered in the following chapters.

2.7.1 Framework

The proposed framework presents an acoustic stimulus to models of acoustic and electrical stimulation. For acoustic stimulation, the model of the auditory periphery system from Zilany and Bruce (2006) [153] is used to represent normal hearing conditions (top section of Figure 2.10). For electrical stimulation, the model from Bruce et al. (1999) [15], [17] is used to transform electrical pulses generated by a CI processor into neural responses at the auditory nerve (bottom section of Figure 2.10). The CI processor simulator can implement either a CIS or ACE stimulation strategy.

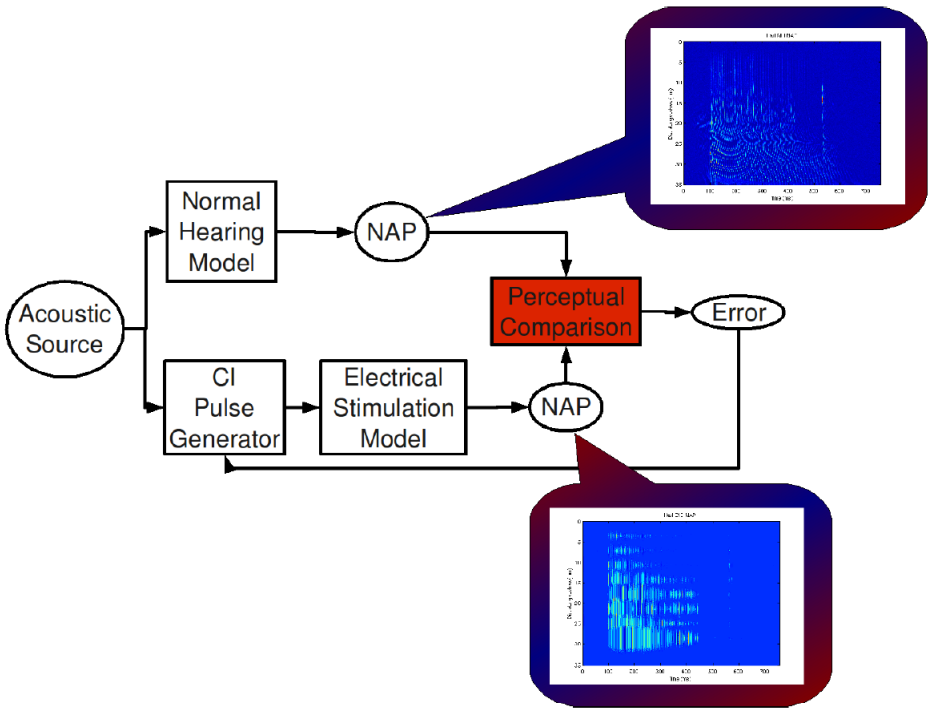


Figure 2.10. Architecture of the optimization framework proposed by Aguiar (2012) [1]

NAPs generated by both models are compared using a perceptual distance metric (PDM) to assess the degree of similarity between the response observed in CI patients to that from NH conditions. The PDM is used to adjust the output of the CI simulator to solve the optimization problem:

$$\hat{\theta}^* = \arg\max_{\hat{\theta}} D(NAP_{CI}(\hat{\theta}), NAP_{NH}) \tag{2.7}$$

In equation 2.7, NAP_{CI} and NAP_{NH} are the outputs of the electrical and acoustic stimulation models respectively, $D()$ is the function used to compute the PDM, $\hat{\theta}$ is the sequence of electrodes stimulated by the CI processor, and $\hat{\theta}^*$ is the optimal sequence of electrodes that maximizes the similarity between NAPs. Aguiar implements $D()$ to be a correlation function.

Aguiar proposed solving the optimization problem by using a state machine architecture, where each state represents a possible electrode to be stimulated. All states are connected forming a graph called trellis. Then, the Viterbi algorithm is used to find the path that maximises a given objective function. In its implementation, the PDM is used as the objective function. Therefore, the problem is reduced to find the Viterbi path that produces the NAP_{CI} that best correlates with the target NAP_{NH}. However, as the number of channels increases, so does the number of combinations to be evaluated, resulting in a computationally expensive optimization process requiring up to 48 hours to produce less than a second of stimulus.

2.7.2 Case Study

The following section describes an unpublished case study using the framework described above.

A 56-year old female post-lingually deafened subject, implanted unilaterally for 8 years, following a period of at least 5 years of bilateral profound hearing loss, was recruited for the study.

The experiment was conducted over four 60-minute sessions to evaluate two electrical stimulation algorithms: (A) the subject's preferred stimulation strategy, and (B) the opti-

mized sequence of electrode stimulation generated by the proposed framework. These algorithms were evaluated following an A-B-A-B paradigm, with each session corresponding to a single electrical stimulation algorithm. In each session, the subject performed two identification tasks: a 9-alternative-forced-choice (9AFC) vowel identification task, and an open-set word identification task. For the first task, 9 /hVd/ utterances (had, hawed, head, heard, heed, hid, hood, hud, who'd) were presented to the subject 15 times in a randomized order using a MATLAB-based graphical interface, providing feedback after each trial. The task was performed twice to test two acoustic conditions: (1) "quiet", in which the stimulus was presented without noise, and (2) "noise", in which the stimulus was presented with speech-shaped noise added such that the SNR was +10 dB. For the second task, the subject was asked to transcribe on the computer 24 unique words drawn from a phonetically-balanced (PB) word list, where the first half were presented in quiet, and the second half in noise. The level of noise used in these experiments was empirically derived from results presented in Aguiar (2012) [1].

For analysis purposes, two random variables, X and Y, were defined corresponding to the presented and perceived stimuli from the vowel identification tasks, respectively. Data were organized into confusion matrices for each background noise condition and stimulation strategy. Then, confusion rates were normalized such that all perceptual outcomes for any given presentation sum to one, and then divided by the total number of stimuli. The resulting matrices correspond to the joint probability between X and $P(X, Y)$, i.e., the probability of a subject identifying stimulus y when presented with stimulus x. Additionally, marginal probabilities for X and Y are computed from the joint probability matrices.

Using the probabilities described above, both entropy (H) and mutual information (I) are computed. While the former conveys how much information is contained in a random variable, the latter is a measure of how much information from the input is preserved in the output of the system and is computed as shown in equation 2.8.

$$I(X;Y) = H(Y) - H(Y|X) \tag{2.8}$$

$H(Y)$ is defined as the entropy of the perceived stimuli, and $H(Y|X)$ is the conditional entropy of the perceived stimuli given the presented stimuli. These terms are defined in equations 2.9 and 2.10 respectively.

$$H(Y) = -\sum_{y \in Y} p_Y(y) \log_2 p_y(y) \tag{2.9}$$

$$H(Y|X) = -\sum_{x \in X, y \in Y} p(x,y) \log_2 \frac{p(x,y)}{p(x)} \tag{2.10}$$

Speech perception scores on the vowel identification tasks ranged between 74.8% and 85.2% under quiet conditions, with a mean score of 80.0% for the preferred stimulation strategy, and 80.4% for the optimized stimulation strategy (Table 2.1). Perception scores under +10 dB SNR noise conditions ranged between 62.2% and 80.0%, with a mean score of 68.9% and 79.6% for the preferred and optimized stimulation strategies respectively (Table 2.2).

Table 2.1. Speech perception scores on the vowel identification tasks without added noise. Each /hVd/ stimulus was presented 15 times, and the number of correct answers is shown in this table for each utterance. Average correct rates and standard deviation for each day are shown at the bottom. Mutual information (measured in bits) is computed for each day under the preferred and optimized stimulation strategies

Stimuli in quiet	Preferred		Optimized	
	Day 1	Day 2	Day 1	Day 2
/had/	12/15	14/15	15/15	15/15
/hawed/	14/15	15/15	15/15	14/15
/head/	4/15	13/15	10/15	13/15
/heard/	15/15	15/15	14/15	15/15
/heed/	13/15	15/15	15/15	15/15
/hid/	15/15	15/15	1/15	3/15
/hood/	15/15	13/15	15/15	12/15
/hud/	4/15	0/15	9/15	5/15
/who'd/	9/15	15/15	15/15	15/15
Avg correct rate	$74.8 \pm 30.1\%$	$85.2 \pm 32.5\%$	$80.7 \pm 31.9\%$	$80.0 \pm 29.2\%$
Mutual information (bits)	2.363	2.769	2.650	2.585

Table 2.2. Speech perception scores on the vowel identification tasks with added speech-shaped noise (+10 dB SNR). Each /hVd/ stimulus was presented 15 times, and the number of correct answers is shown in this table for each utterance. Average correct rates and standard deviation for each day are shown at the bottom. Mutual information (measured in bits) is computed for each day under the preferred and optimized stimulation strategies.

Stimuli in noise	Preferred		Optimized	
	Day 1	Day 2	Day 1	Day 2
/had/	12/15	15/15	15/15	15/15
/hawed/	15/15	15/15	15/15	15/15
/head/	3/15	10/15	4/15	4/15
/heard/	15/15	15/15	15/15	15/15
/heed/	12/15	15/15	14/15	15/15
/hid/	8/15	12/15	14/15	12/15
/hood/	3/15	5/15	15/15	15/15
/hud/	1/15	0/15	0/15	2/15
/who'd/	15/15	15/15	15/15	15/15
Avg correct rate	$62.2 \pm 38.2\%$	$75.6 \pm 36.4\%$	$79.3 \pm 38.1\%$	$80.0 \pm 34.8\%$
Mutual information (bits)	2.027	2.441	2.692	2.695

Table 2.3 shows speech perception scores on the open-set word identification tasks. Results are presented as total number of correct phonemes per session for each condition and stimulation strategy evaluated. Scores ranged between 42.1% and 59.5% when stimuli were presented in quiet, and between 2.5% and 37.5% in noise.

Table 2.3. Speech perception scores on the open-set word identification tasks. Total number of correct phonemes per session are displayed for each assessed condition (quiet and noise) under the preferred and optimized stimulation strategies.

	Preferred		Optimized	
	Day 1	Day 2	Day 1	Day 2
Quiet	16/38 42.1%	22/40 55.0%	25/42 59.5%	17/39 43.6%
Noise	1/40 2.5%	7/34 11.8%	15/40 37.5%	9/35 25.7%

2.8 Summary

This chapter introduced the anatomy of the peripheral auditory system, as well as the processes involved in hearing and the different technologies developed to restore hearing when some of these processes are impaired. Among these technologies, the focus of this dissertation is on CIs, medical devices for restoring hearing in those with moderate to profound SNHL. Since their introduction in the 1970s, various algorithms have been developed to process the incoming auditory stimulus and determine how the electrodes inserted inside the cochlea should be stimulated. Over the years, CI simulators have been used in the development and assessment of new algorithms, giving researchers a controlled environment and access to a larger subject pool, i.e., NH subjects. However, recent studies have shown that current CI simulators might not be delivering an stimuli that elicit a neural response close to that from CI patients.

Computational models of the peripheral auditory system have been developed to help better understand how hearing works. However, these models are typically complex and computationally expensive to implement for practical applications. Therefore, some attempts have been made to incorporate approximations or simplified versions of these models into CI signal processors. Following this approach, Laflen proposed the use of physiological models to compute an optimized electrical stimulation that could elicit a neural response closer to

that observed under NH conditions. A framework was proposed, and later implemented by Aguiar as described in Section 2.7.

Results of an unpublished case study using the proposed framework were presented in Section 2.7.2. Although a simple objective function was used for the optimization, these preliminary results yielded increased and more consistent vowel identification rates in quiet and noisy conditions. The present dissertation seeks to further expand on the results obtained in this case study using an updated version of the proposed framework.

3. METHODS

3.1 Acoustic Stimuli

As described in Section 2.4, CNC words are commonly used in vowel identification tasks to evaluate the performance of CI stimulation strategies among CI patients. Expanding on the case study described in Section 2.7.2, the present work used a speech corpus with CNC words of the form /hVd/. However, the number of utterances available in the aforementioned corpus was limited, therefore, a different corpus was used instead, the Nationwide Speech Project (NSP) corpus [23].

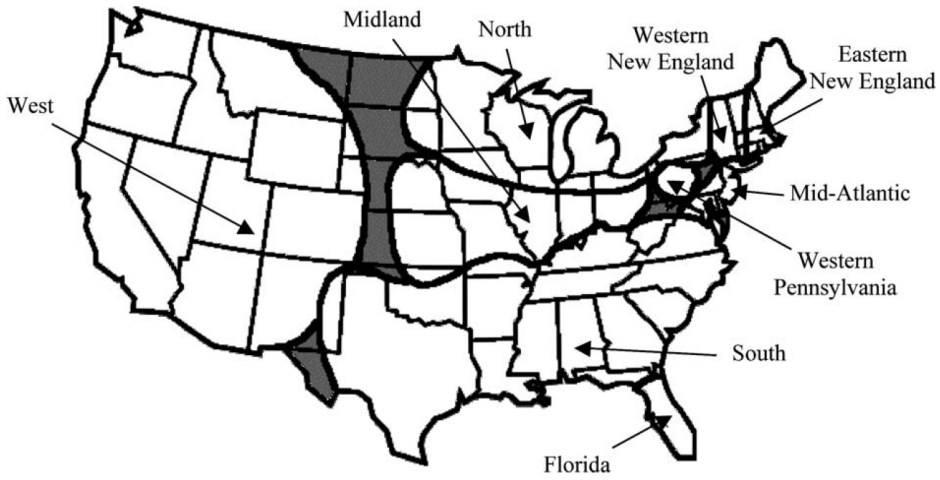


Figure 3.1. Major dialects and regions of American English language (Image from [23])

The NSP corpus includes speech samples from a total of 60 American English speakers from six regions across the United States: West, North, Midland, South, New England, and Mid-Atlantic (see Figure 3.1). Speech samples include isolated words, sentences, passages, and interview speech. A set of 10 /hVd/ words (had, hayed, head, heed, hid, hod, hoed, hood, hudd, who'd) from a total of 15 male speakers across three regions (West, North, and Midland) were used for the purpose of the present work. However, loudness of speech samples varied across utterances and speakers, creating potential bias in the analysis. Therefore, samples were processed to create homogeneous conditions across all stimuli using the algorithms described below.

3.1.1 Signal conditioning

Each speech recording was individually processed and went through four stages: cropping, low-pass filtering, normalization, and noise reduction. Each stage of processing was performed using scripts developed in MATLAB as described below.

Cropping

Speech recordings from the NSP corpus are approximately 2-seconds long with silence segments at the beginning and end. To reduce processing time in later stages, the silence segments were cropped out using a MATLAB graphic user interface (GUI) (see Figure 3.2). The GUI allows the user to select the start and end points of the segment, as well as listening and saving the resulting cropped recording. Audio files were save in .WAV format with a sampling rate of 44,100 kHz.

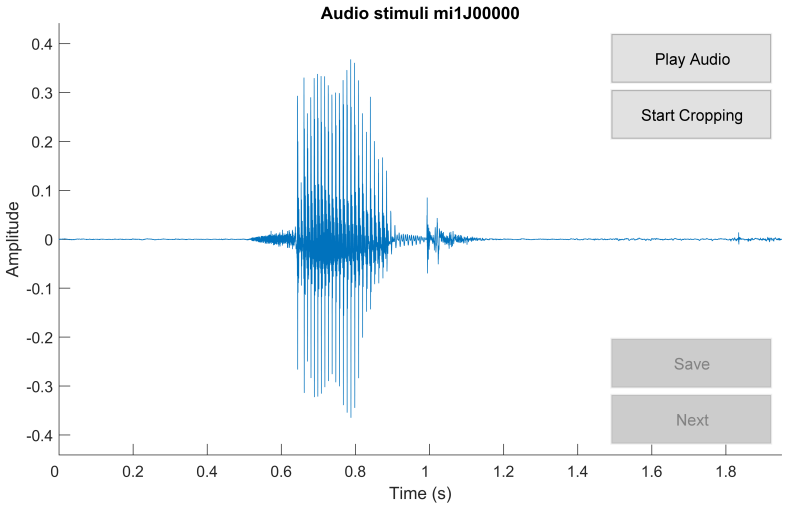


Figure 3.2. MATLAB GUI developed for cropping speech recordings.

Low-pass filtering

CIs have a limited frequency range, and in most cases electrodes are mapped to a maximum frequency of 8,000 Hz. Therefore, speech recordings were low-pass filtered using a fourth order Butterworth filter with cut-off frequency at 7,300 Hz. Filtering was applied to each recording using the MATLAB GUI shown in Figure 3.3. The input cropped speech recording from the previous step is shown in the first row at the top, and the low-pass fil-

tered signal is shown in the second row. The GUI displays the temporal waveform (left) and spectrogram (right) of each signal.

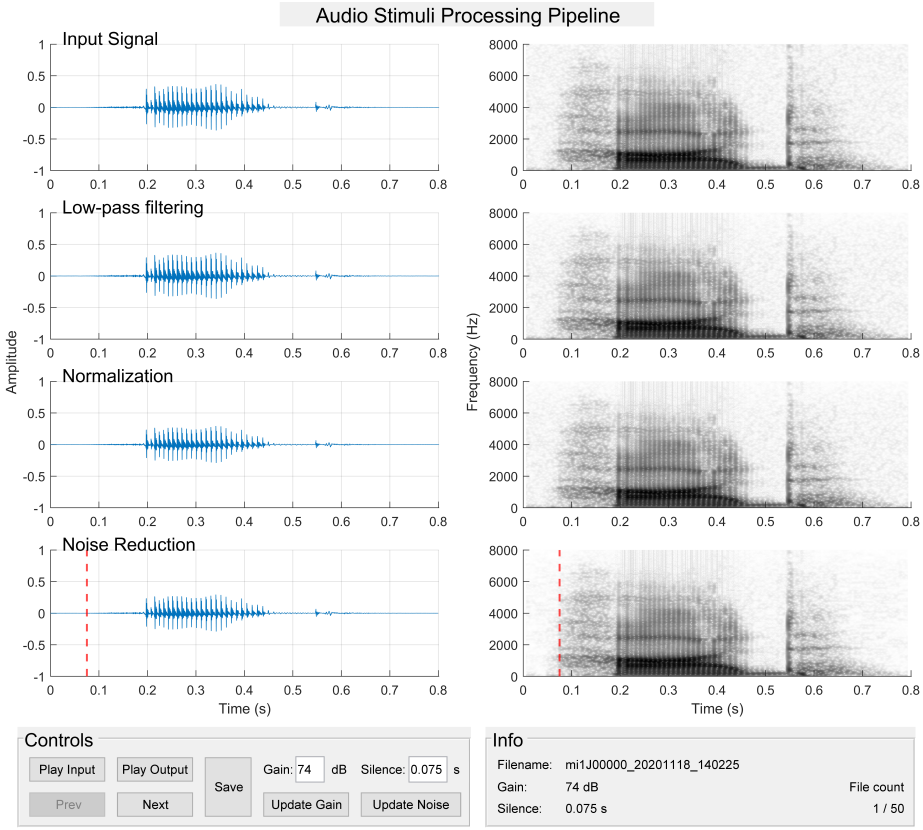


Figure 3.3. MATLAB GUI developed for low-pass filtering (second row), normalizing (third row), and noise reduction (bottom row) of speech recordings. A red dashed vertical line is included in the bottom panels to visualize the silence segment used for noise reduction. Temporal waveforms are shown in the left panels, and spectrograms in the right panels.

Normalization

Speech recordings were normalized using a non-linear gain function to amplify voiced segments only and increase their SNR. The moving root mean square (RMS) value was calculated for each sample in the recording and used to compute the gain (g) as the ratio between the target RMS and the measured RMS value. The gain function used depends

on whether the sample needs to be amplified ($g > 1$) or attenuated ($g \leq 1$) as shown in Equation 3.1.

$$y(x) = \begin{cases} x \times (1 + f(x)), & \text{if } g > 1 \\ x \times g, & \text{otherwise} \end{cases} \tag{3.1}$$

where $f(x)$ is a logistic function of the form:

$$f(x) = \frac{L}{1 + \mathrm{e}^{-k(x - x_0)}} \tag{3.2}$$

Parameter L is the curve's maximum value and is calculated as $(g-1)$ because $f(x) \in [0, L]$. Parameter x_0 is the sigmoid's middle point and is calculated as the average between the lower and upper amplification thresholds. The lower amplification threshold (T_l) is the point at which samples are amplified at $T\%$ of the gain, and the upper amplification threshold (T_u) is the point at which the gain reaches $(1 - T)\%$ of its value. Lastly, parameter k is the steepness of the curve and is calculated as:

$$k = -\frac{\log(\frac{1-T}{T})}{T_l - x_0} \tag{3.3}$$

Parameters used for the processing were empirically chosen and are shown in Table 3.1

Table 3.1. Parameters used by the MATLAB GUI for normalization.

Parameter	Description	Value
-	Target Amplitude (dB)	74
T_l	Lower amplitude threshold (dB)	58
T_u	Upper amplitude threshold (dB)	64
T	Sigmoid's threshold	0.05
x_0	Sigmoid's middle point (dB)	61

Noise Reduction

Background noise intrinsic to the recordings can be amplified despite the use of a non-linear gain function for normalization in the previous step. Therefore, a noise reduction stage was incorporated to reduce background noise. For this purpose, the algorithm for speech

enhancement proposed by Berouti et al. (1979) [9] was implemented in MATLAB and its output is shown at the bottom row of the GUI in Figure 3.3. The algorithm uses a technique called spectral subtraction where an estimate of the noise power spectrum is subtracted from the speech power spectrum. The estimated noise power spectrum is taken from the silence segment at the beginning of each utterance. A red dashed vertical line in the bottom panels of the GUI marks the end of the silence segment and its location can be modified using the controls at the bottom of the GUI.

3.1.2 Measurement of SNR improvement

The improvement achieved by the signal conditioning was measured by computing the SNR before and after processing for each stimuli. The RMS value of the noise was computed from the first 0.05 seconds corresponding to the silence portion of each stimuli. The RMS value of the signal, however, was computed for the entire stimuli using a moving RMS, and its peak value was used to compute the SNR as shown in Equation 3.4.

$$SNR = 20\log_{10}\left(\frac{\text{signal}_{RMS}}{\text{noise}_{RMS}}\right) \tag{3.4}$$

3.2 Optimization Framework

As discussed in Section 2.8, the present work uses an updated version of the framework implemented by Aguiar (2012) [1]. The underlying concept behind the framework was described in Section 2.7.1. Figure 3.4 depicts the updated framework and the implementation of its components is presented in the following sections.

3.2.1 Acoustic stimulation models

The current iteration of the framework implements the AN model developed by Zilany et al. (2014)[155] to model the behavior of the peripheral auditory system under NH conditions (see Section 2.6.2 for a description of the model). The model provides the response of the auditory nerve at a given location inside the cochlea. As described in Section 2.1.1, the

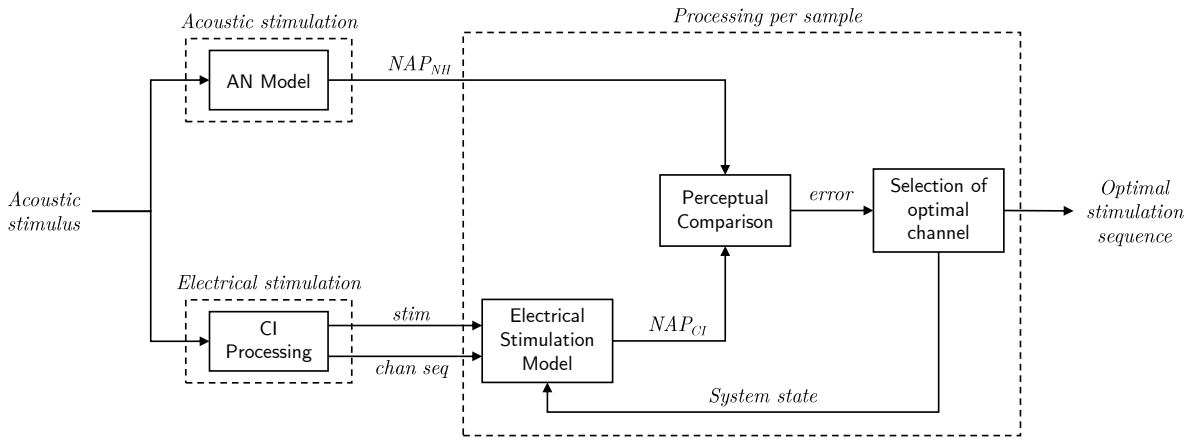


Figure 3.4. Schematic diagram of the updated framework for the optimization of CI electrode selection based on matching acoustic and electrical neural responses at the AN.

cochlea is tonotopically organized, and the frequency at which the AN is more sensitive to is known as characteristic frequency (CF). For the purposes of this framework, the cochlea was modeled using 300 fibers linearly distributed. Then, the location of the fibers was mapped to their corresponding CF using Greenwood's equation (see Equation 2.1) and limited to the frequency range from 125 to 20,000 Hz. The lower boundary is imposed by the model itself, while the upper boundary corresponds to the highest frequency humans can perceive.

Each fiber creates a time series with spike activity in response to the acoustic stimulus. Spikes are grouped together in what is known as a peri-stimulus time histogram (PSTH) to provide information of how many spikes are fired per time unit. Then, the probability of firing spikes can be estimated by computing the average of PSTHs across multiple iterations of the model. The probabilities of firing spikes from each fiber are grouped together to form what was previously introduced as a NAP.

The acoustic stimulus was presented to the model multiple times ($N = 100$) and their PSTHs were used to compute a NAP. Given the stochastic nature of the model, each acoustic stimulus was presented multiple times to generate a total of 5 different NAPs per stimulus. The NAPs generated by the model (NAP_{NH}) were used in later stages of the framework as the target neural response to be mimicked by the CI processor (see Figure 3.4).

3.2.2 Electrical stimulation models

The framework incorporates two modules involved in modeling electrical stimulation: a CI simulator, and a model of electrical stimulation of the AN. The former was implemented using the Nucleus Matlab Toolbox (NMT), provided by Cochlear, to recreate the processing performed by their CI processors. And the latter was implemented using Bruce's model of electrical stimulation of the AN [15], [17].

The NMT was set up to emulate the processing performed by a Nucleus processor. For this purpose, the following modules of processing were included: (1) front-end/microphone signal scaling, (2) auto-gain control (AGC), (3) filterbank filtering, (4) rectification, (5) loudness growth compression, (6) uniform resampling, and (7) pulse mapping. The NMT produces a frequency-time matrix (FTM) containing the energy at the frequency band of each electrode, for each analyzed sample in time. The FTM is processed using the stimulation strategy of choice to transform it into a sequence of electrodes to be stimulated and their respective amplitudes. Electrodes available and their dynamic range are patient dependent and are adjusted accordingly. For the purpose of this dissertation, two stimulation strategies were used: (1) ACE to serve as control, and (2) the optimized sequence generated by the present framework. Details of the optimization algorithm are described below in Section 3.2.3.

The sequence of electrodes and amplitudes is used as input to stimulate Bruce's model of the AN as shown in the bottom section of Figure 3.4. The model first generates a pulse with amplitude specified by the input sequence and duration according to the parameters set for the implant itself. Depending on the state of the system, which is influenced by past and current stimulation pulses, the model decides whether or not a spike should be generated. Due to the stochastic nature of the model, a total of 250 repetitions were simulated to estimate the probability of firing a spike in response to the input stimulus.

The total number of AN fibers simulated was the same as for acoustic stimulation and was selected as described above in Section 3.2.1. The probability of firing spikes by each fiber were grouped together to produce the NAP in response to electrical stimulation (NAP_{CI}). For the control case, the NAP_{CI} was generated in response to a fixed electrode sequence

produced by the ACE strategy. Therefore, the subsequent stages shown in Figure 3.4 were not applicable to the control case.

3.2.3 Optimization algorithm

The optimization algorithm takes as input the outputs from the acoustic stimulation model, NAP_{NH}, and the electrical stimulation model, NAP_{CI}. While NAP_{NH} is computed upfront for the whole stimulus, NAP_{CI} is computed one segment at a time. Each segment of NAP_{CI} is the result of stimulating the AN with one of the values contained in the FTM, where each row corresponds to one of the available electrodes $n \in [1, N]$, and each column corresponds to the sample being processed $m \in [1, M]$. The sampling rate for the FTM is computed based on the number of electrodes set to be stimulated and the pulse rate per electrode.

Then, each segment of NAP_{CI} is compared against the corresponding segment in NAP_{NH} using a PDM. As introduced in Section 2.7.1, the PDM chosen for the optimization algorithm was the cross-correlation at lag zero ($n = 0$) between NAPs. The processing is repeated for each available electrode for a given sample, and the optimal electrode to be stimulated, $\hat{\theta^*}$, is selected such that:

$$\hat{\theta^*} = \arg\max_{\hat{\theta}} \left(NAP_{CI}(\hat{\theta}) \star NAP_{NH} \right)[0] \tag{3.5}$$

As shown in Figure 3.4, the optimal electrode selected is stored as part of the system state and used in the subsequent iterations of the algorithm. For each sample, the search for the optimal electrode is repeated until all electrodes to be stimulated are selected. Once all samples are processed, the framework generates a sequence with all selected electrodes and their corresponding amplitudes, along with a NAP_{CI} for the completed stimulus.

3.2.4 Framework Validation

As described in Section 3.1, 10 /hVd/ words from 15 male speakers were selected for testing the framework. For each utterance, there were five recordings available per speaker, accounting for a total of 750 acoustic stimuli. To reduce computational time, only a segment

of the vowel in each utterance was used. Each segment was processed by (1) the framework to generate the optimal stimulation sequence and its corresponding NAP_{CI}, and (2) the ACE strategy as the control case. To assess similarities between NAP_{NH} and NAP_{CI}, the following metrics were computed: (1) cross-correlation, (2) mean square error (MSE), (3) peak SNR (pSNR), and (4) mutual information. Metrics were calculated for both the optimized sequence and ACE strategy.

Simulations were run on a desktop computer with a 6-cores Intel Core i5-9600 @ 3.10 Ghz, 32 GB of RAM, and running Ubuntu 18.04.5 LTS. The acoustic stimulation model and the CI processor were both implemented using Matlab, and the electrical stimulation model and the optimization algorithm were implemented in C++. All modules in the framework incorporated parallel computing, with Matlab scripts using the Parallel Computing Toolbox, and the C++ scripts using OpenMP.

Implant-specific parameters were chosen to mimic those from the patient who participated in the case study described in Section 2.7.2. A full list of parameters used by each module is shown in Table 3.2.

Table 3.2. Parameters used for testing the optimization framework.

Parameter	Description	Value
`numFibers`	Total number of AN fibers simulated	300
`MonteCarloNum`	Number of Monte Carlo repetitions	250
`cohc`	AN OHC condition	1 (normal)
`cihc`	AN IHC condition	1 (normal)
`species`	AN model tuning	2 (humans from Shera et al.)
`noiseType`	AN type of fractional Gaussian noise (fGn)	1 (variable)
`fiberType`	AN spontaneous spike rate	3 (high)
`implnt`	AN implementation of power-law function	0 (approximation)
`stimdb`	AN stimulus intensity (dB SPL)	74
`CF_range`	AN CF range (Hz)	125–20000
`nrep`	AN number of stimulus repetitions	100
`psthbinwidth`	AN PSTH binwidth (ms)	0.14
`electconfig`	CI electrode configuration	mp (monopolar)
`pulserate`	CI stimulation rate (pulses / sec)	7200
`pulsewidth`	CI pulse width (μs)	37
`numChan`	CI total number of electrodes	22
`num_bands`	CI number of electrodes available	12
`implantMaxima`	CI number of electrodes stimulated (maxima)	6
`insertDepth`	CI normalized insertion depth from apex	0.2857

3.3 NAP Classification

As described in Section 2.4, vowel identification tasks are commonly used to assess speech performance in CI patients. In these assessments, the stimulation can be presented through a loudspeaker, or by directly stimulating the implant through connection with a computer. In both cases the acoustic stimulus evokes a neural response at the AN fibers that is then interpreted by higher brain processes to be identified as one word or another. One of the goals of this work is to recreate the identification task performed by the CI user, which is implemented with a classifier using machine learning techniques. This section describes the classifier implementation, the inputs needed for its training and testing, and the experiments conducted to validate its functionality.

3.3.1 Classifier

The classifier implemented a closed-set vowel identification task. In this type of task the number of possible classes is limited to a reduced set of options, which is one of the 10 /hVd/ words used in this work. The elements to be used by the classifier were the NAPs generated using the models described in Section 3.2 that are two-dimensional matrices and can be treated as images. To classify these images, a solution using machine learning techniques from the well studied field of image classification was implemented. The solution used transfer learning to retrain a convolutional neural network (CNN) to classify the images of interest.

The CNN chosen was a 50-layer residual neural network, commonly referred to as ResNet-50, developed by He et al. (2016) [65]. The CNN was pre-trained using ImageNet, a large dataset of natural images developed by Deng et al. (2009) [25] that over the years has grown to have more than 10 million images. Using a Matlab script, the last layers of the CNN were replaced and trained to identify the new classes, i.e., one of the 10 /hVd/ words.

3.3.2 Segmentation

The ResNet-50 requires images of size 224x224x3 as inputs, corresponding to width, height, and number of color channels, respectively. However, NAPs generated using acoustic and electrical models are larger both in the time and frequency domain. Therefore, the NAPs had to be resized before used with the CNN.

Using a Matlab script, a segment of the vowel was extracted from each NAP_{NH} and the corresponding acoustic stimuli was saved as an audio file in WAV format. Two segment sizes, 224 and 448 samples, were chosen to evaluate its impact in the performance of the classifier. Then, each audio file was used to stimulate the CI simulator and generate a NAP_{CI} using both the optimized framework and the ACE strategy. Segmented NH NAPs were normalized between $[0, 1]$ to increase contrast by limiting the probability range to $[0.05, 0.25]$, clipping any values outside the range, and saved as 8-bit grayscale images in PNG format. No normalization was performed on CI NAPs. Resulting NAPs were resized to match the input size required by the CNN to account for the number of AN fibers modeled

(frequency domain), the duration of the segment extracted (time domain), and the number of color channels (grayscale to RGB).

3.3.3 Testing

Using a Matlab script, the ResNet-50 was trained with segmented NAP_{NH} to recreate higher brain processes used for identifying words under NH conditions. Then, NAP_{CI} were evaluated using the trained network to recreate the scenario of a post-lingually deafened CI patient performing the task. NAP_{NH} were generated using acoustic stimuli from all 15 subjects available, each one repeated five times, for a total of 3730 NAPs (four stimuli were discarded due to poor quality). The dataset was divided so that data from 12 subjects (2980 NAPs) were used for training and validation of the CNN, split 80% and 20%, respectively. The data from the remaining three subjects were used for testing the classifier under NH conditions (750 NAPs), and to generate NAP_{CI} for testing CI conditions. However, each stimulus was repeated only one time due to computational time constraints, accounting for a total of 150 NAP_{CI}. Lastly, the CNN trained with NAP_{NH} was evaluated using NAP_{CI} from both the optimization framework and the ACE strategy.

The training, validation, and testing of the CNN was repeated a total of 50 times. Training and validation was performed in 6 epochs with 208 iterations each.

4. RESULTS

4.1 Acoustic stimuli

Measurements of SNR pre and post signal conditioning for each utterance are shown in Figure 4.1. Mean SNR values varied between 39.7 and 44.6 dB pre-signal conditioning and between 50.6 and 52.3 dB post conditioning. Averaging over all stimuli, SNR improved by 8.4 dB after signal conditioning. Performing a two-sample t-test showed that the SNR improvement was statistically significant ($p < 0.0001$).

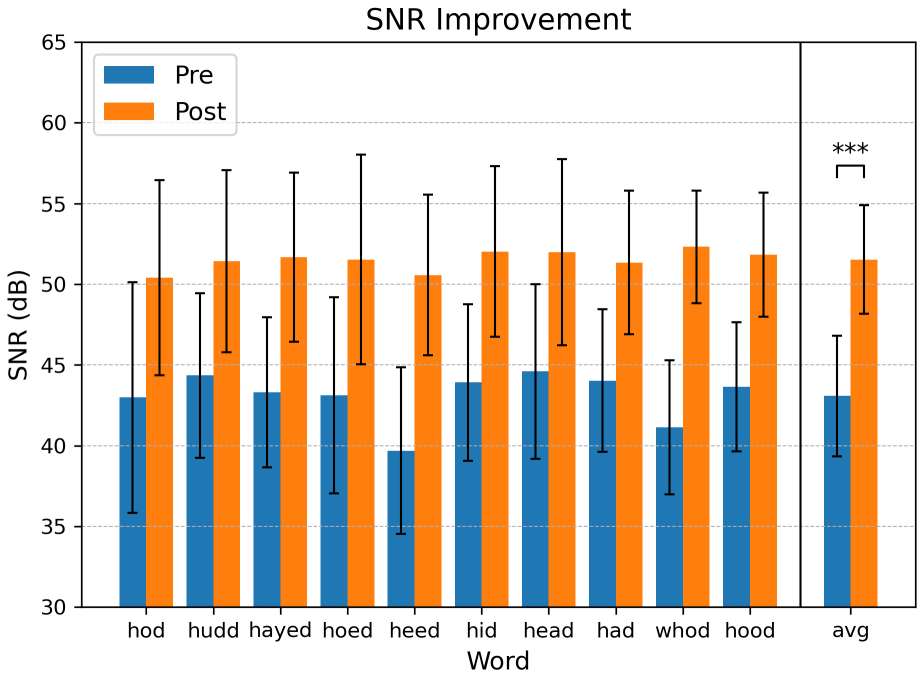
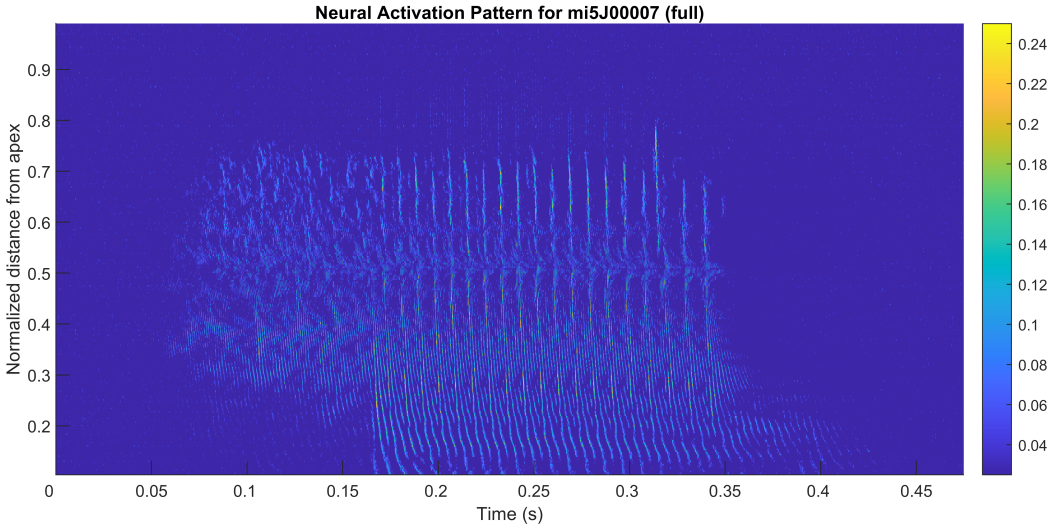
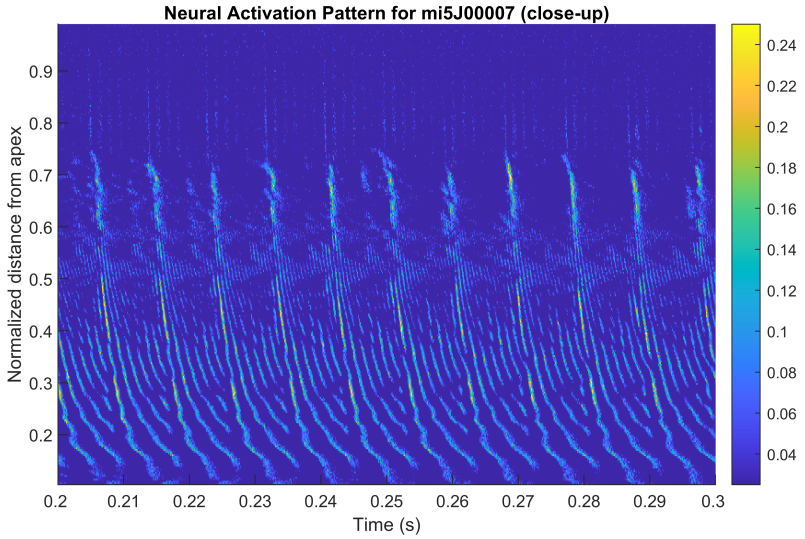


Figure 4.1. Measurements of SNR pre (blue) and post (orange) signal conditioning. Mean values and the standard deviation are shown for each individual utterance. Average values are included in the last bars on the right.

4.2 Framework



(a) NH NAP Full stimulus

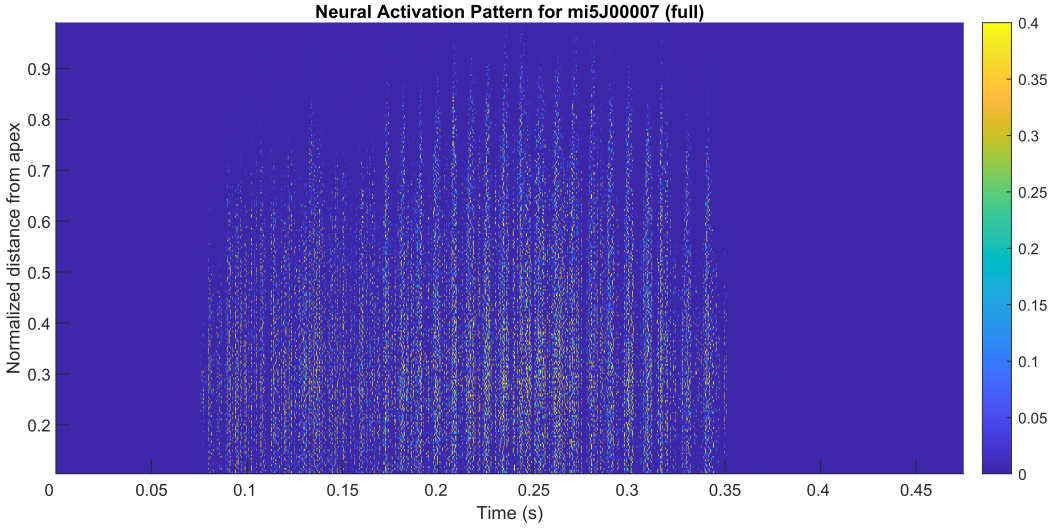


(b) NH NAP Close-up

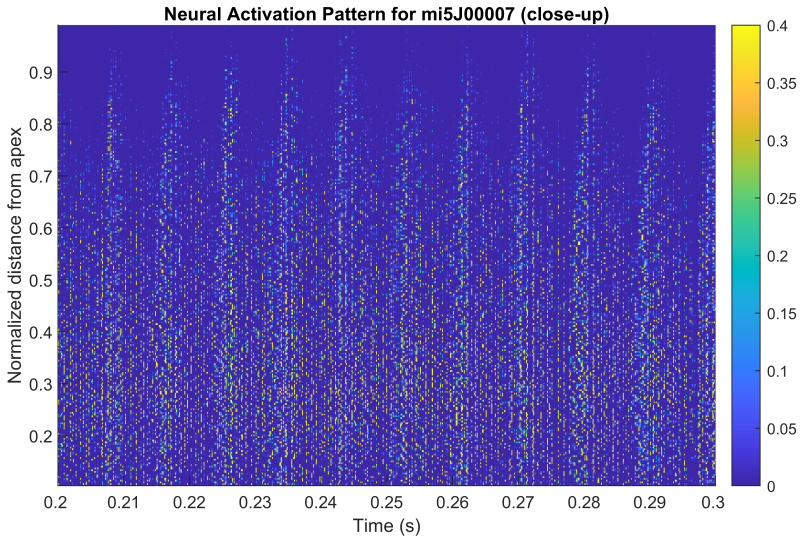
Figure 4.2. NAP generated using the AN model corresponding to NH conditions. (a) Full stimulus of word /had/. (b) Close-up of the vowel between 0.2 and 0.3 seconds.

A NAP generated using the AN model under NH conditions for the word /had/ is shown in Figure 4.2. The full utterance is shown in panel (a), and a close-up of the vowel portion is shown in panel (b). For each of the panels, the x-axis shows time in seconds, and the y-axis shows the normalized distance measured from the apex inside the cochlea. For the

latter, values closer to zero are mapped to low frequencies (apex), and values closer to one are mapped to high frequencies (base). Lastly, the color scale represents the probability of firing an action potential by the AN fiber at each time and location. Probability values were clipped between 0.025 and 0.25 for visualization purposes.



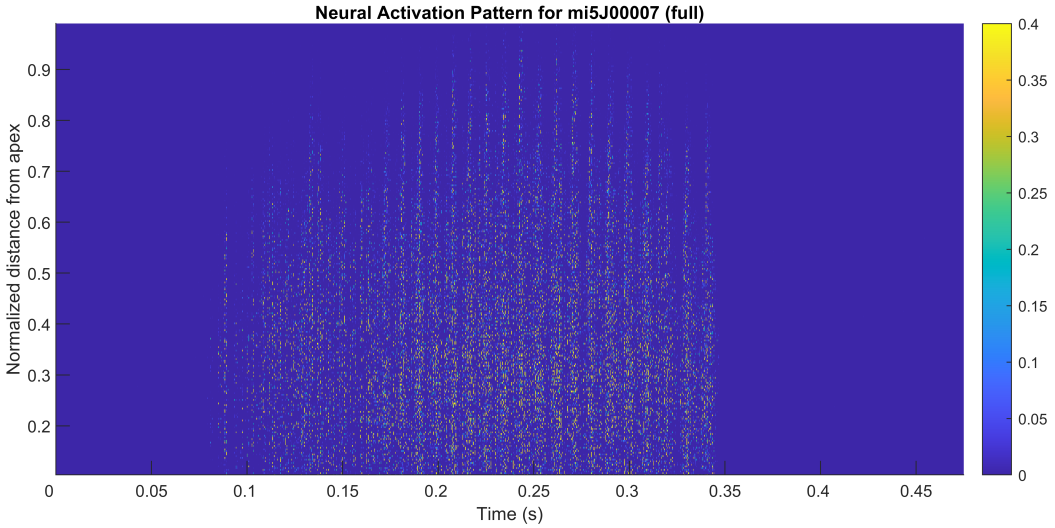
(a) CI NAP (ACE) Full stimulus



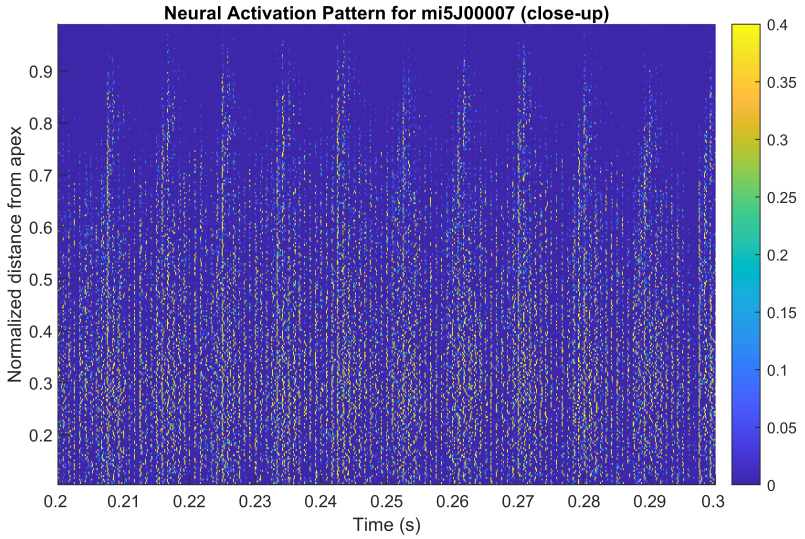
(b) CI NAP (ACE) Close-up

Figure 4.3. NAP generated using the electrical stimulation model and the ACE coding strategy. (a) Full stimulus of word /had/. (b) Close-up of the vowel between 0.2 and 0.3 seconds.

Similarly, NAPs generated with the electrical stimulation model and the CI simulator using the ACE strategy (control) and the optimization framework are shown in Figure 4.3 and Figure 4.4, respectively. As before, panel (a) shows the full stimulus, and panel (b) a close-up of the vowel portion of the utterance /had/. The probability of firing an action potential spans the full range between 0 and 1.



(a) CI NAP (Optimized) Full stimulus



(b) CI NAP (Optimized) Close-up

Figure 4.4. NAP generated using the electrical stimulation model and the optimization framework. (a) Full stimulus of word /had/. (b) Close-up of the vowel between 0.2 and 0.3 seconds.

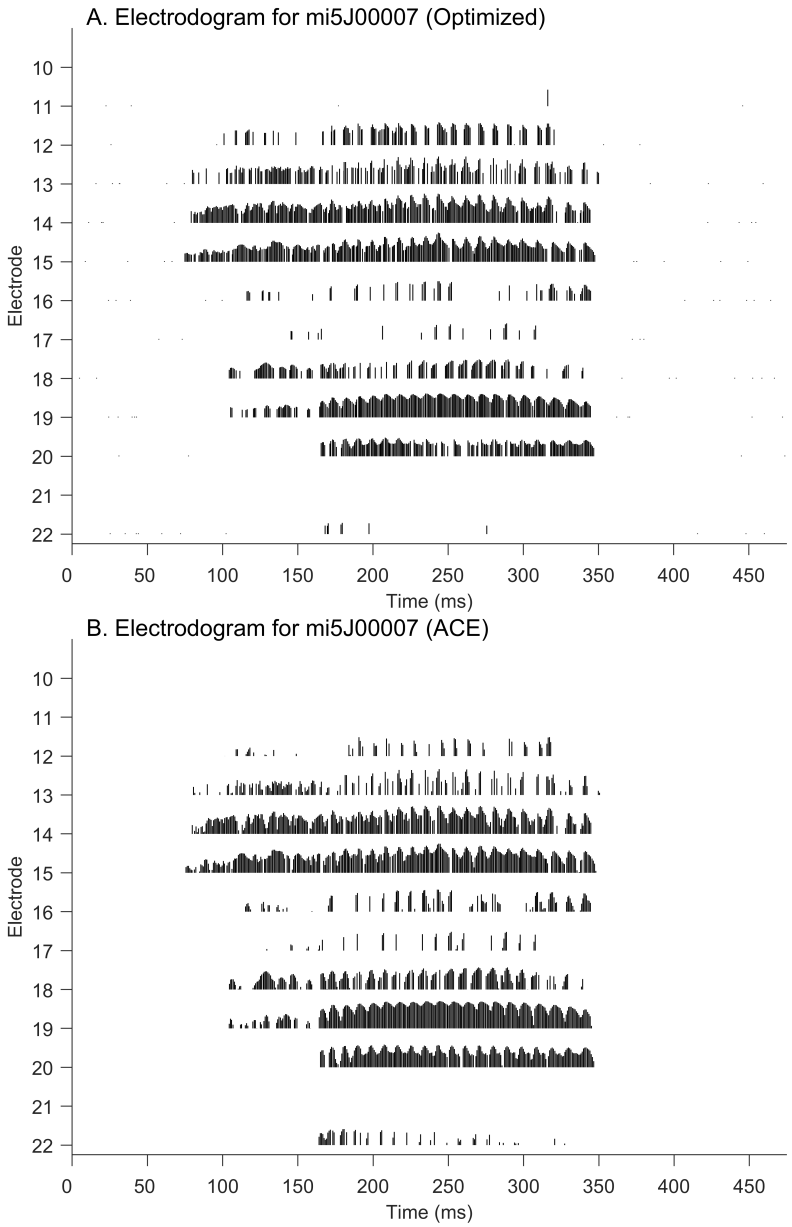


Figure 4.5. Electrodograms showing electrodes and amplitudes stimulated using the optimization framework (A) and the ACE stimulation strategy (B).

Electrode sequences used to generate NAP_{CI} are shown in Figure 4.5 as electrodograms. An electrodogram shows the channel being stimulated and the amplitude of the pulse being delivered to the electrodes. Electrodes' numbers increase from base to apex, thus, electrode 10 stimulates high frequencies and electrode 22 stimulates low frequencies. Panel (a) shows the electrodogram generated using the optimization framework, and panel (b) shows the

electrodogram using the ACE strategy for comparison. As before, both sequences were generated in response to the utterance /had/.

Overall, most electrodes were stimulated similarly by both approaches, most notably electrodes 14, 15, 18, 19, and 20. However, the optimization framework provided a cleaner stimulation at mid- and low-frequencies (electrodes 16, 17, and 22) and an emphasis of high frequencies (electrodes 12 and 13) when compared to the ACE strategy.

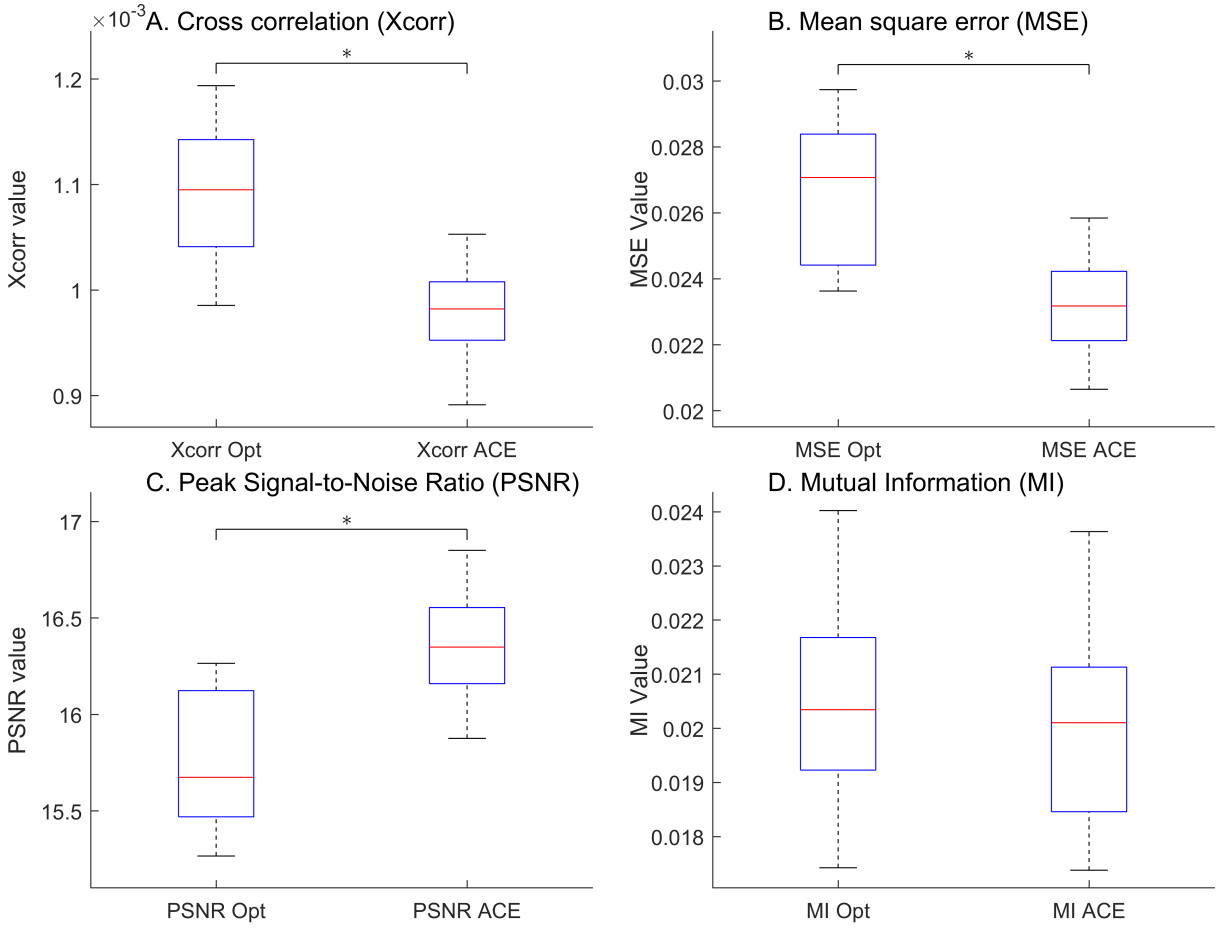
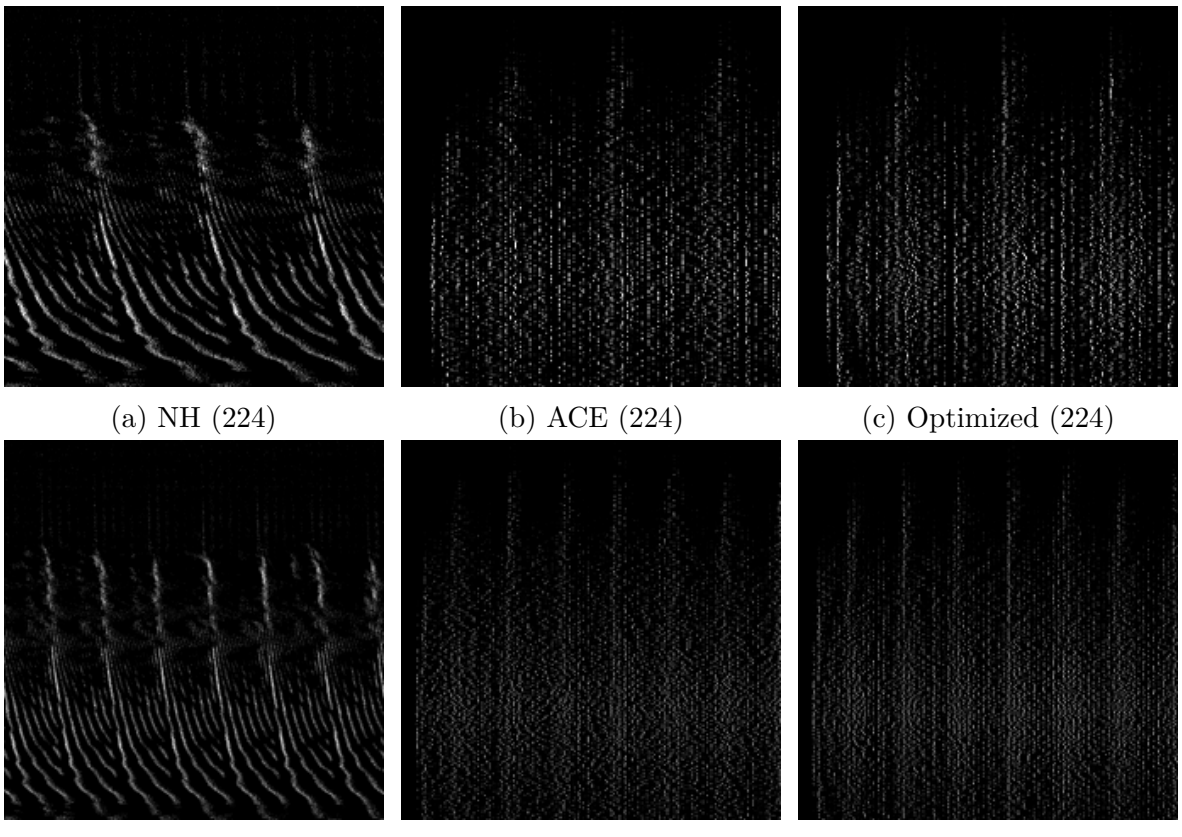


Figure 4.6. Performance metrics comparing CI NAPs generated using the ACE strategy and the optimization framework, computed with respect to NH NAPs. (A) Cross correlation. (B) Mean Square Error (MSE). (C) Peak signal-to-noise (PSNR). (D) Mutual Information (MI)

Lastly, results of the performance metrics are shown in Figure 4.6. Panel (a) shows the cross correlation computed for the NAP_{CI} using the optimization framework and the ACE strategy, both with respect to the NAP_{NH}. Subsequent panels show the same comparison

for the mean square error (panel B), peak signal-to-noise (panel C), and mutual information (panel D). Results shown were computed using NAPs from five different instances of the word /had/. Cross correlation scores showed statistically significant differences favoring the optimization framework ($p < 0.05$). However, mean square error and peak signal-to-noise scores showed statistically significant differences favoring the ACE strategy ($p < 0.05$). Lastly, mutual information scores showed better scores for the optimization framework, however, differences were not statistically different.

4.3 Classifier



(a) NH (224) (b) ACE (224) (c) Optimized (224)

(d) NH (448) (e) ACE (448) (f) Optimized (448)

Figure 4.7. Examples of images of NAPs used for training, validation, and testing of the CNN. All images are 224x224 in size, where those in the first row were taken from segments that were 224 samples long, and those in the second row from segments that were 448 samples long. (a & d) Segment of a NAP_{NH}. (b & e) Segment of a NAP_{CI} (ACE). (c & f) Segment of a NAP_{CI} (Optimized).

Examples of the segmented NAPs used for training the CNN classifier are shown in Figure 4.7. Segments that were 224 samples long in duration (time) are shown in the top row, and segments with 448 samples are shown in the bottom row. NAP_{NH} used for training and validation of the CNN are shown on the left column (a & d panels). NAP_{CI} used for testing the CNN are shown in the center (b & e panels) and right (c & f panels) columns for the ACE strategy and the optimized stimulation, respectively.

Results for the trained CNN identifying NAP_{NH} (blue), NAP_{CI} using ACE (orange), and NAP_{CI} using optimization framework (green) are shown in Figure 4.8. Identification scores are shown for each individual utterance and the overall average (last bars on the right), displaying mean values and standard deviation across all 50 instances of the CNN. Identification scores of the CNN trained and tested with 224-samples long segments are shown in panel A (left), and with 448-samples long segments are shown in panel B (right).

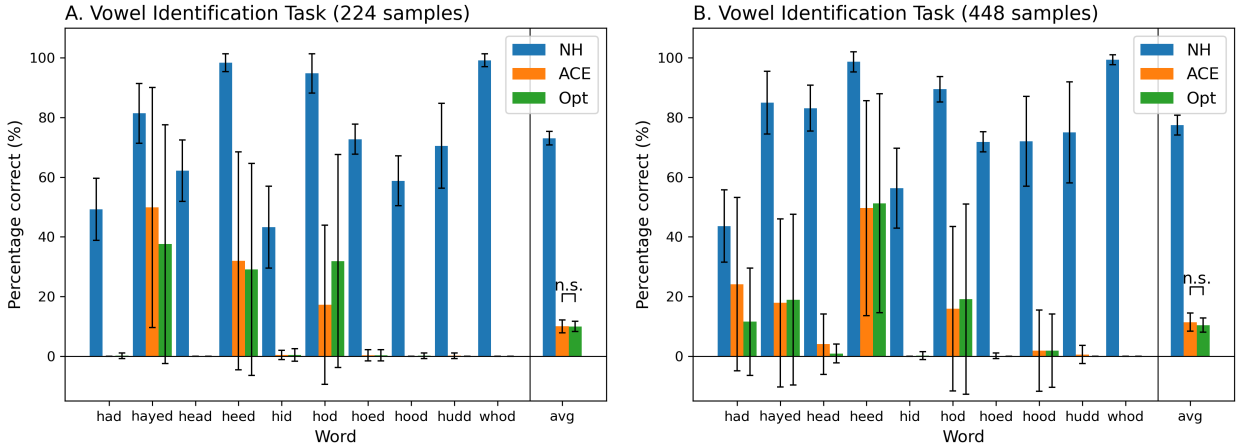


Figure 4.8. Recognition scores of the vowel identification task performed by the trained CNN to recognize NAP_{NH} and NAP_{CI} (ACE and optimized). Mean values and the standard deviation are shown for each individual utterance. Average values are included in the last bars on the right. Panel A shows results for 224-samples long NAPs, and panel B for 448-samples.

Overall recognition scores achieved by each type of NAP for the two segment sizes used for training and testing the CNN are shown in Figure 4.9. Identification scores using 448-samples long segments were higher across all NAPs; however, the differences were not statistically significant.

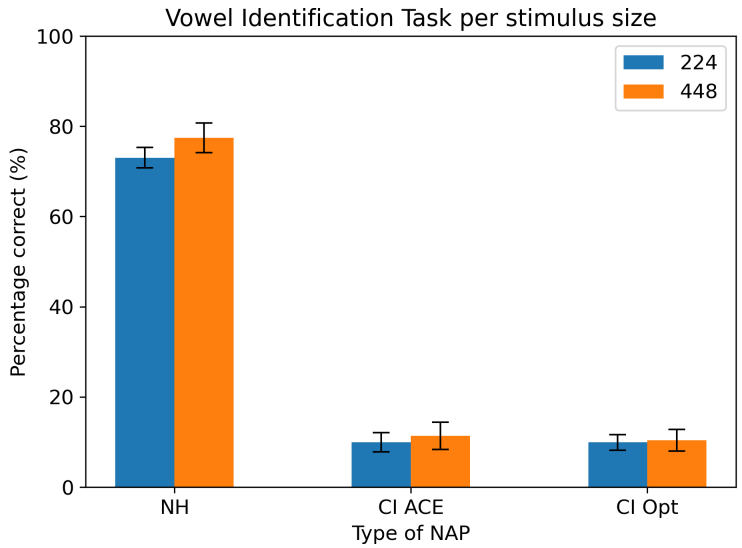
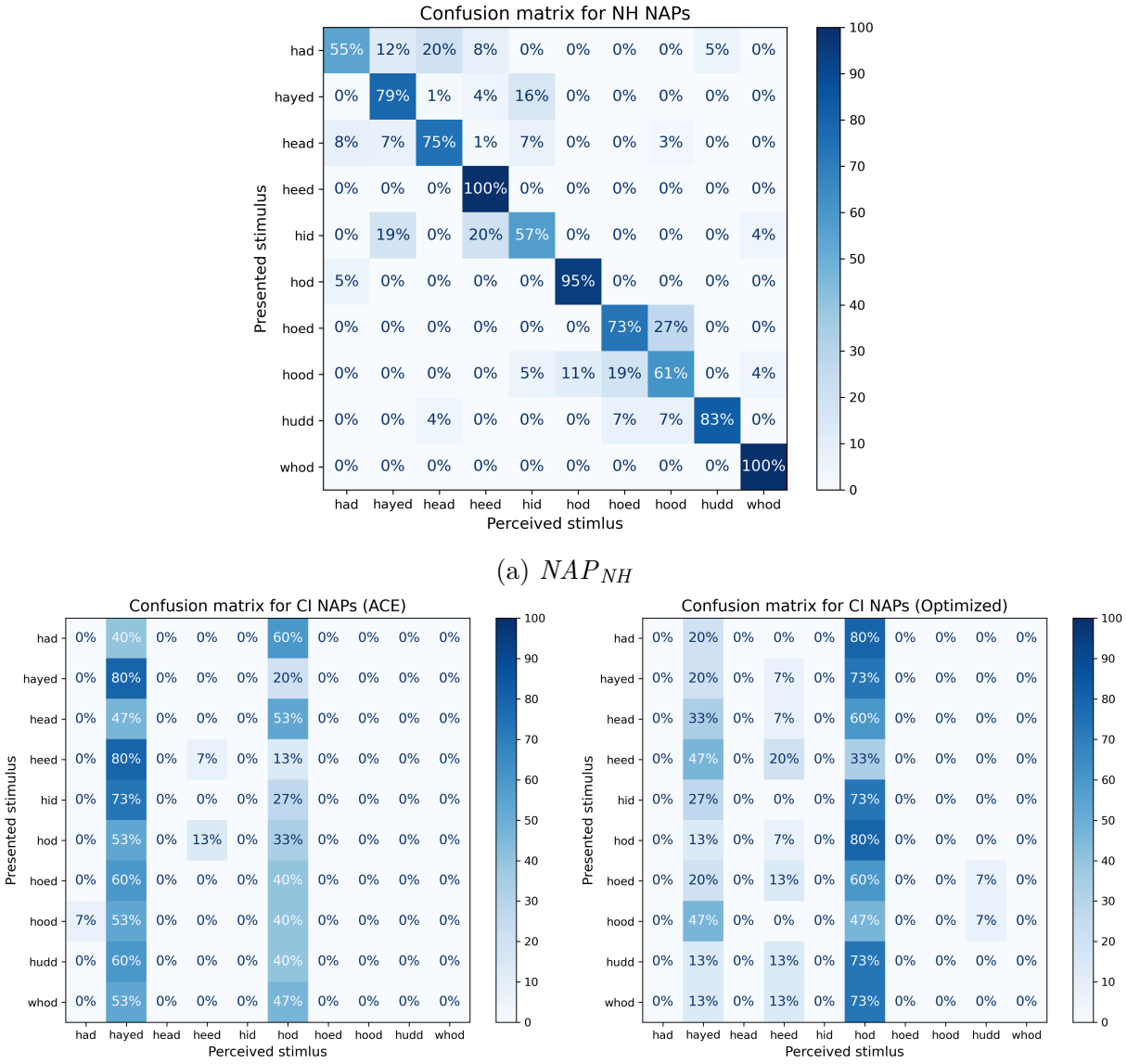


Figure 4.9. Overall recognition scores of the vowel identification task for each segment size.

Confusion matrices corresponding to one out of the 50 instances of the CNN are shown in Figures 4.10 and 4.11 for segment sizes 224 and 448, respectively. Each element of the matrix shows what percentage of the presented stimulus (rows) is classified as one class or another (columns). Elements along the diagonal correspond to correctly classified stimuli, and elements outside the diagonal represent misclassifications. Classification results of testing the CNN with NAP_{NH} are shown in panel A, and classification results for NAP_{CI} using the ACE strategy and the optimization framework are shown in panels B and C, respectively.

Lastly, examples of the NAP_{NH} and NAP_{CI} used for training and validation of the classifier are shown in Figures 4.12 (224-sample segments) and 4.13 (448-sample segments). Panels A-J show a NAP_{NH} for each of the utterances to be identified by the classifier. Panels K and L show NAP_{CI} computed by either the ACE strategy or the optimization framework. These NAPs serve as an example to visualize that the fine details observed in NAP_{NH} are absent in NAP_{CI}; however, the periodicity at which some features are repeated over time is preserved.

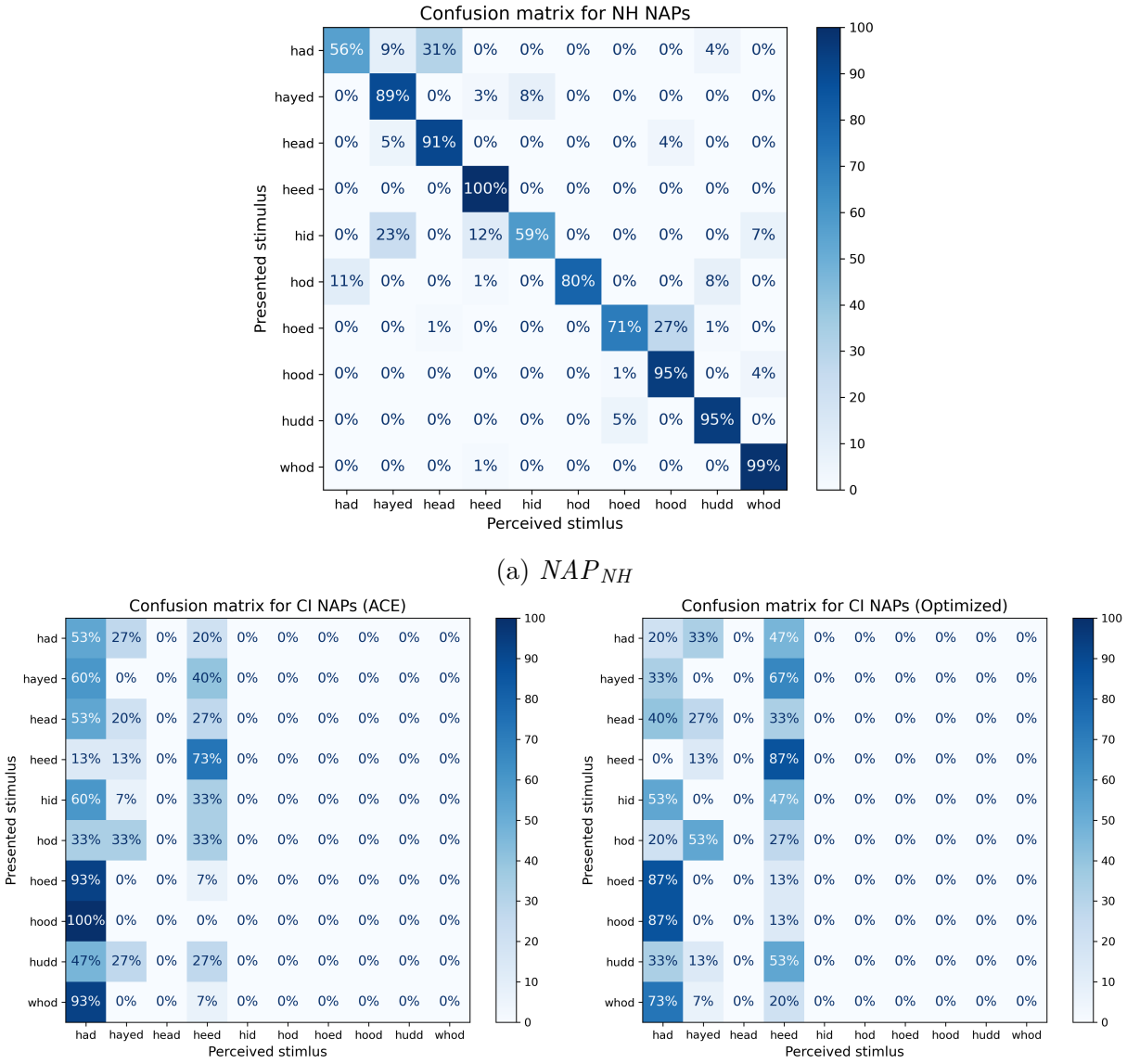


(a) NAP_{NH}

(b) NAP_{CI} (ACE)

(c) NAP_{CI} (Optimized)

Figure 4.10. Confusion matrices of the vowel identification task using 224 sample segments performed with NAP_{NH} (A), NAP_{CI} ACE (B), and NAP_{CI} Optimization framework (C). Each row corresponds to the presented stimulus, and each column to the perceived stimulus.

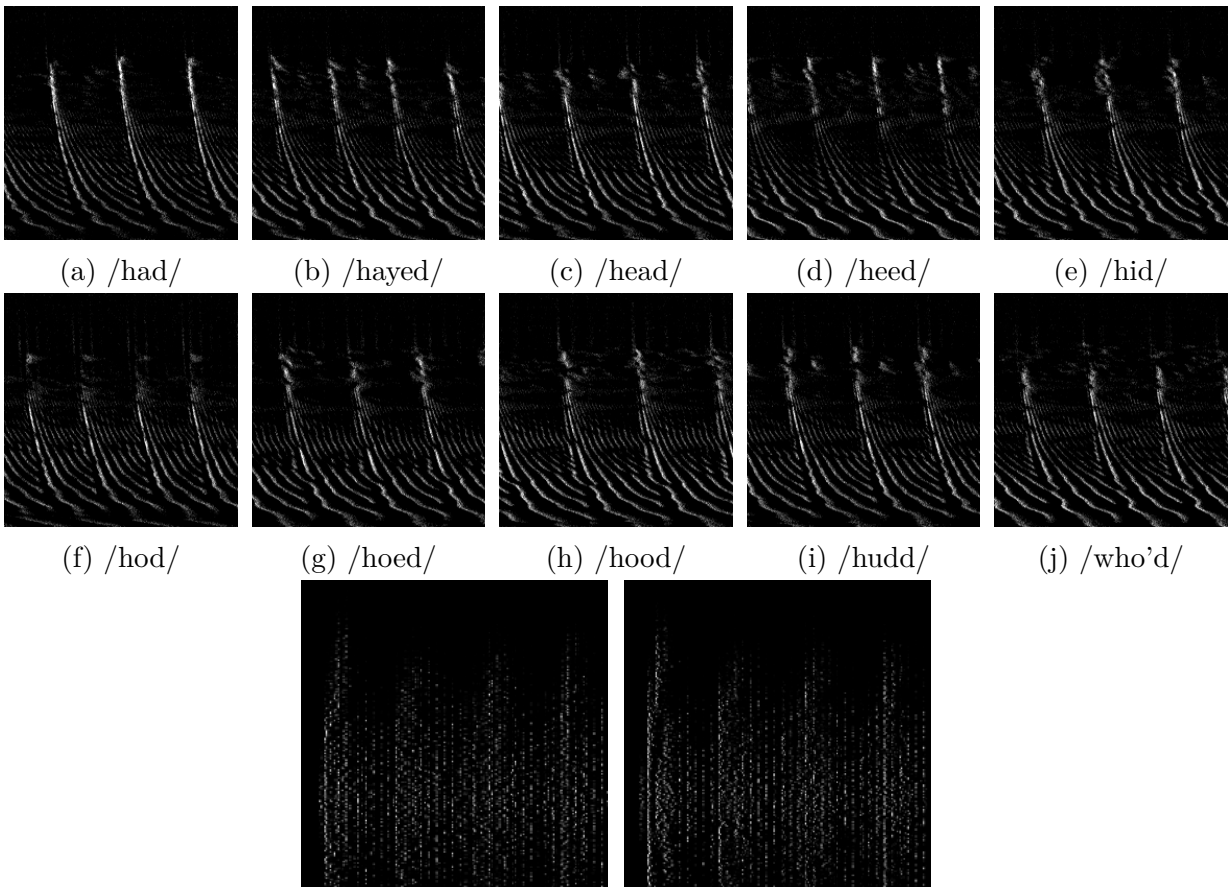


(a) NAP_{NH}

(b) NAP_{CI} (ACE)

(c) NAP_{CI} (Optimized)

Figure 4.11. Confusion matrices of the vowel identification task using 448 sample segments performed with NAP_{NH} (A), NAP_{CI} ACE (B), and NAP_{CI} Optimization framework (C). Each row corresponds to the presented stimulus, and each column to the perceived stimulus.

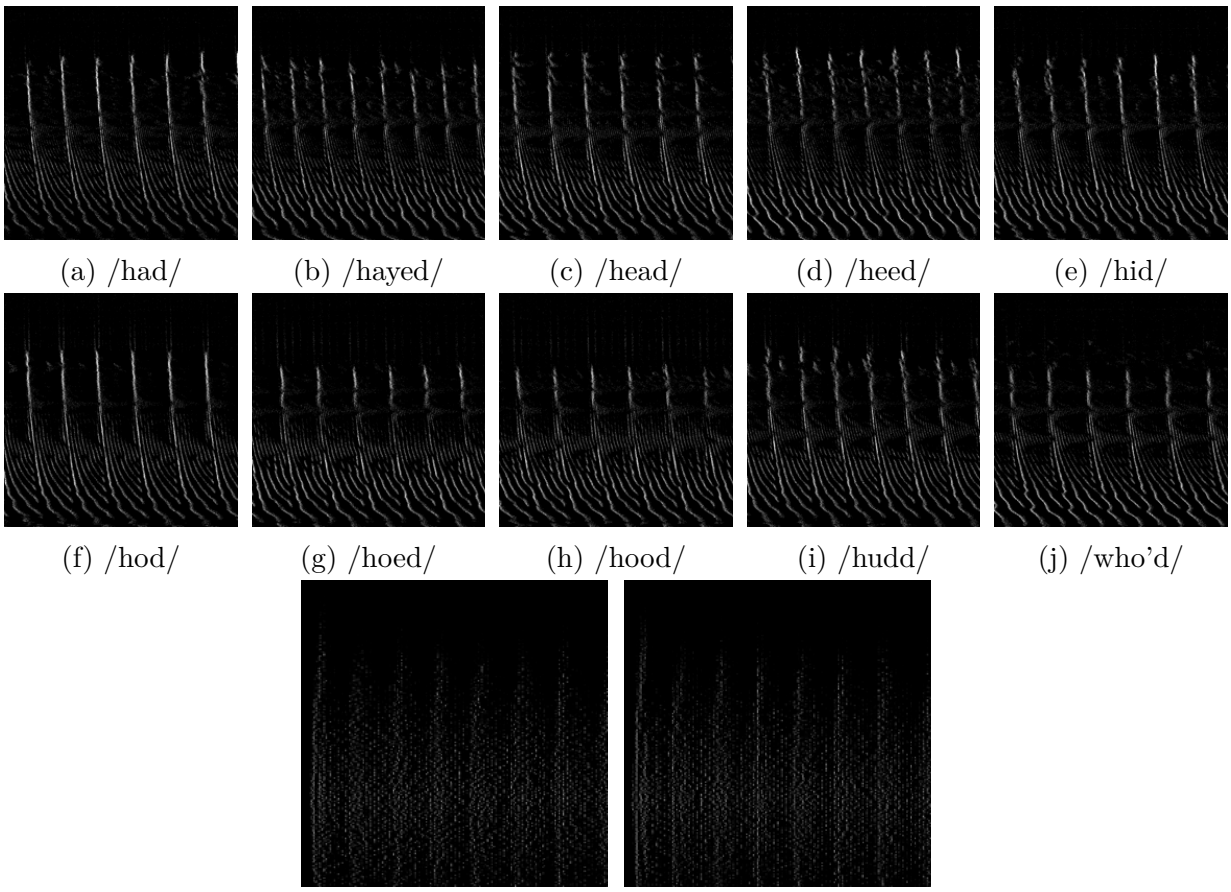


(a) /had/ (b) /hayed/ (c) /head/ (d) /heed/ (e) /hid/

(f) /hod/ (g) /hoed/ (h) /hood/ (i) /hudd/ (j) /who'd/

(k) /who'd/ (ACE) (l) /who'd/ (Optimized)

Figure 4.12. Examples of NAP_{NH} and NAP_{CI} (224 samples) used for training and testing of the classifier. NAP_{NH} of each utterance are shown in panels a-j. NAP_{CI} of the word /who'd/ using the ACE and optimization framework are shown in panels k and l, respectively.



(a) /had/ (b) /hayed/ (c) /head/ (d) /heed/ (e) /hid/

(f) /hod/ (g) /hoed/ (h) /hood/ (i) /hudd/ (j) /who'd/

(k) /who'd/ (ACE) (l) /who'd/ (Optimized)

Figure 4.13. Examples of NAP_{NH} and NAP_{CI} (448 samples) used for training and testing of the classifier. NAP_{NH} of each utterance are shown in panels a-j. NAP_{CI} of the word /who'd/ using the ACE and optimization framework are shown in panels k and l, respectively.

5. DISCUSSION

5.1 Optimization Framework

The cross correlation performance results show that the framework, as described in Section 3.2.3, was successfully implemented and maximized the similarity between NAP_{CI} and NAP_{NH}. However, results from other commonly used metrics to compare image quality (MSE and PSNR) suggest that NAP_{CI} produced using the ACE strategy better match NAP_{NH} than those produced by the optimization framework. Lastly, mutual information scores suggest that NAP_{CI} produced by the optimization framework carry more information than those generated by the ACE strategy; however, differences were not statistically significant.

The mixed results described above could be partially attributed to the limitations of the chosen metrics. The metrics PSNR and MSE are known to be susceptible to misalignment errors, meaning that delays introduced by the models used in the framework could lead to a mismatch between NAP_{CI} and NAP_{NH}. Therefore, a NAP_{CI} that is more similar to the target NAP_{NH} but not properly aligned would score lower in these metrics, putting more weight on the error due to the mismatch rather than that from their differences. Additionally, misalignment errors can be enhanced due to the more sparse nature of NAP_{CI} (see Figures 4.3 and 4.4) when compared to NAP_{NH} (see Figure 4.2). With fewer pixels to match between NAP_{CI} and NAP_{NH}, a misalignment between NAPs could have a significant impact on the scores of these metrics. Consequently, introducing a processing stage to better align NAP_{CI} and NAP_{NH} before computing these metrics would be suggested.

Another factor contributing to the observed results is the number of samples used in this work. Due to computational limitations, performance metrics were computed from 100-ms NAP segments corresponding to the vowel portion of five different instances of the word /had/ performed by the same speaker. Therefore, the results described above could be biased by the utterance or speaker chosen to compute the metrics. It is unclear if these results are generalizable due to the fact that only one utterance was used instead of the ten total utterances available. Similarly, the observed results do not take into account the variability introduced by different speakers. Consequently, the performance metrics should be computed over a larger number of speakers using all utterances available to remove bias.

Despite the cross correlation results favoring the optimization framework, the resulting NAP_{CI} still lack some of features observed in NAP_{NH}. For instance, the neural activity seen in NAP_{NH} creates traces that run vertically across locations along the cochlea (see Figure 4.2). The bottom section of these traces are delayed as they get closer to the apical region. It is believed that these observed behaviors are attributed to the effect of the traveling wave. Due to the tonotopicity of the cochlea, high frequency stimuli are encoded first at the basal region (top section of a NAP), and low frequency stimuli are encoded last at the apical region (bottom section of a NAP). However, the effect of the traveling wave is missing in NAP_{CI} from both the ACE strategy and the optimization framework (see Figures 4.3 and 4.4). On the other hand, the traces observed in NAP_{NH} repeat themselves in a periodic manner. Even though the fine details of the traces is missing in NAP_{NH}, the periodic aspect of the traces is still captured by NAP_{CI} in both the ACE strategy and the optimization. However, no measurements were performed to assess the accuracy of the periodicity observed in NAP_{CI} with respect to NAP_{NH}.

It was hypothesized that the cross correlation function would help recreate the neural response seen in NAP_{NH}, including the effect of the traveling wave. However, results so far seem to only support this hypothesis up to some extent. It is believed that the sparse nature of NAP_{CI} might be contributing to the observed results. Additionally, the resolution achieved with a limited number of electrodes in CIs might not be sufficient to recreate the traces as seen in NAP_{NH}. Even though the optimization framework selects the sequence of electrodes that maximizes similarity with respect to NAP_{NH}, it is possible that none of the neural responses produced during the optimization process contained portions of the traveling wave. Consequently, it is believed that the chosen electrical stimulation model might not be capable of recreating the fine details observed in NAP_{NH}. Efforts have been made to develop models that better capture the phenomena of electrically stimulated auditory nerve fibers as described in the review by Takanen et al. (2016) [141]. For instance, models proposed by Joshi et al. (2017) [77] or Tabibi et al. (2021) [139] could be explored as alternatives to the electrical stimulation model currently used by the optimization framework.

5.2 Classifier

Despite the limited dataset available for training and validation, the classifier identified NAP_{NH} with average recognition scores ranging between 74% and 77% as shown in Figure 4.9. These recognition scores were higher than those of NAP_{CI} most likely due to training the classifier with only NAP_{NH}. Although identification scores were higher when using NAP_{NH} that carried more information (448 samples vs 224 samples), the differences were not statistically significant (see Figure 4.9). When observing average recognition scores per word, data show that the classifier scored lower for words /had/ and /hid/, independent of the number of samples used (see Figure 4.8). Furthermore, results shown in Figures 4.10a and 4.11a, corresponding to one out of 50 instances of the CNN, suggest systemic errors in classification. For example, the word /had/ is confused in most cases by /head/, and the word /hid/ is confused by either /heed/ or /hayed/, regardless of sample size. It is believe that these confusions could be attributed to the similarity in frequency of their first (F1) and second (F2) formants. Using the data from Hillenbrand et al. (1995) [69], we observe that the vowel sounds /æ/ in /had/ and /ɛ in /head/ have their F1 at 580 and 588 Hz, and their F2 at 1799 and 1952 Hz for male speakers, respectively. It is important to mention that these frequencies are average values and they can vary across speakers; therefore, they have to be interpreted as part of a vicinity where each vowel sound can exist as shown in Figure 3 in Hillenbrand et al. (1995) [69]. Similarly, the vowel sounds /ɪ/ in /hid/ (F1: 427 Hz, F2: 2034 Hz), /i/ in /heed/ (F1: 342 Hz, F2: 2322 Hz), and /eɪ/ in /hayed/ (F1: 476 Hz, F2: 2089 Hz) also have formants in the vicinity of each other. Words /hid/ and /hayed/ have more similar formant frequencies than /hid/ and /heed/, which is reflected in more tokens misclassified as /hayed/ than /heed/.

On the other hand, recognition scores for NAP_{CI} were closer to chance level regardless of the stimulation strategy used. As shown in Figure 4.9, average recognition scores for NAP_{CI} stayed within the vicinity of 10% regardless of the sample sizes used. While average recognition scores per word shown in Figure 4.8 might suggest that words such as /had/, /hayed/, /heed/, or /hod/ are being labeled correctly, individual results suggest otherwise. As shown in Figures 4.10 and 4.11 for one out of 50 instances of the CNN, all utterances

are being classified in two or three classes only. These results suggest that the classifier cannot use the features learned from NAP_{NH} to correctly identify NAP_{CI}. As discussed in Section 5.1, NAP_{CI} lack the fine details observed in NAP_{NH} while still capturing their periodicity. Therefore, it is believed that the classifier is identifying utterances based mainly on the number of periodic segments present in the NAP rather than other features. In the case shown in Figure 4.12, we observe that the NAP_{CI} for the word /who'd/ has four cycles with prominent peaks at the beginning of each of them. Then, out of all 10 classes, only the NAP_{NH} for words /hayed/ and /hod/ also have four full cycles. As a result, the classifier preferentially selected /hayed/ and /hod/ when presented with the word /who'd/ as shown in Figure 4.10. While the example is given for only one of the utterances, it is believed that this behavior might help explain the similar trends observed for other tokens in Figure 4.10. Similarly, when the number of samples used per segment is increased to 448, we observed that the number of cycles in the NAP_{CI} for the word /who'd/ matches that of the NAP_{NH} for words /had/ and /heed/ (see Figure 4.13). These results are in line with the trends observed in Figure 4.11.

Another potential factor influencing the low recognition scores observed in NAP_{CI} is the information encoded in their periodicity. It is believed that the frequency at which the cycles repeat in each NAP might correlate with the fundamental frequency (F0) associated with the speaker. While a more extensive assessment should be conducted to validate this hypothesis, then it would imply that the performance of the classifier is speaker dependent. Given that utterances from different speakers would have different fundamental frequencies, then NAPs would be classified in a way that favors matching the fundamental frequency rather than the utterance being modeled.

A final factor impacting the low recognition scores seen in NAP_{CI} is the type of images that the classifier was originally trained on. As described in Section 3.3, the classifier was trained to distinguish between natural images, such as a dog from a tree. Since the NAPs are more similar to each other than a dog is to a tree, it is believed that the set of features needed to discriminate between NAPs is different from that used for natural images. Additionally, the performance of the classifier might be prone to errors due to location of the NAP segment chosen for classification. Traditionally, the location of the object of interest within the frame

in natural images would not affect the performance of the classifier. However, in this work the classifier uses a segment of the NAP and selecting different starting points would result in different segments. As discussed previously, the classifier might be relying on the number of periods seen in NAP_{CI}; therefore, selecting a segment that leaves out part of a period might have an impact on the classification of the utterances.

The initial hypothesis that motivated the usage of a classifier stated that it would help model the scenario of a post-lingually deafened CI user. Under this scenario, the classifier was trained with NAP_{NH} to represent the existing brain connections in the user pre-implantantion. Using the proposed framework, NAP_{CI} were optimized to best match NAP_{NH}. Therefore, it was expected that the features learned from NAP_{NH} would be sufficient to correctly identify NAP_{CI}; however, that was not the case. Instead, it could be argued that the low recognition scores achieved by the classifier on NAP_{CI} resemble that of CI users when their implant is activated for the first time. It is known that CI users can perform poorly at first and require time and training before their brains adapt to the new type of stimulation. To account for this adaptation period, most studies recruit CI users when their performance has plateaued, usually three to six months after implantation. Similarly, the classifier trained with NAP_{NH} lacks the knowledge to correctly identify NAP_{CI} and might require additional training to learn a new set of features to achieve that. Therefore, training the classifier with NAP_{CI} could lead to better recognition scores in the same way CI users improve their speech understanding upon training.

The incorporation of NAP_{CI} into the training of the classifier might not only help increase its overall performance but also model the improvement experienced by CI users upon training. The current implementation of the classifier was trained 100% with NAP_{NH}, but this percentage could be reduced in favor of adding NAP_{CI}. For instance, the ratio between NAP_{NH} and NAP_{CI} could be changed in increments of 10% from 100/0 to 90/10 and so on, to assess the impact of the types of NAPs in the overall performance of the classifier. The results of this experiment would help better understand how CI users improve their performance over time, having the potential for predicting experimental data. However, a longitudinal study following the progress of a CI population post-implantation would be necessary to validate the findings of the proposed experiment. Having access to experimental

data would not only help address the shortcomings of the classifier but also some of the limitations of the optimization framework discussed in Section 5.1.

A validated optimization framework and classifier capable of predicting behavioral data would provide insight into explaining behavioral results beyond what the classifier was trained on. For instance, it could be used for studying and designing experiments that would highlight specific results observed with CI users. If CI users are displaying difficulties at recognizing specific words, then this framework could be used to model that specific behavior. Currently, counselors lack the data to support how CI users' performance improves upon training. Consequently, having this type of data available for new CI users during their training could help provide better and personalized counseling.

6. CONCLUSION

CIs have been successfully used as sensory neuroprostheses to partially restore hearing for those with moderate to profound SNHL. Despite advances in both hardware and software, CIs are unable to allow their users to fully understand speech, achieving around 60% word recognition in standardized tests [8], [56], [72], but dropping significantly in complex acoustic scenarios [40], [46], [70]. Traditionally, the design of algorithms for CI processing has focused on characterizing acoustic stimuli using phenomenological approaches. However, newer technologies with more computational power allow physiological approaches to be explored. Therefore, this work focused on exploring the second approach and incorporated computational models of the peripheral auditory system.

In this work, a framework was proposed to best match neural responses elicited by electrical stimulation of the auditory nerve to those elicited by acoustic stimulation. The framework implements Zilany et al. (2014) [155] model to represent acoustic stimulation, Bruce et al. (1999) [15], [17] model for electrical stimulation, and a CI simulator. The optimal neural activation pattern, or NAP, is computed by solving the optimization problem of finding the sequence of electrodes that needs to be stimulated to produce the desired response. Resulting NAPs were evaluated using a vowel identification task performed by a classifier developed using deep learning techniques. The classifier was trained to identify NAPs generated by acoustic stimulation (NAP_{NH}), and evaluated using NAPs generated by electrical stimulation (NAP_{CI})

Results suggest that the framework generates NAP_{CI} that correlate better to NAP_{NH} when compared to those generated by a more traditional stimulation strategy (ACE). However, while the classifier was successful at identifying NAP_{NH}, it performed poorly at identifying NAP_{CI}. It is believed that the observed behavior might resemble that of a post-lingually deafened CI user whose implant is activated for the first time and their brain has yet to adapt to the new type of stimulation. Additionally, NAP_{CI} lack some of the fine details observed in NAP_{NH}, suggesting that the electrical model used in this framework or the electrical stimulation itself might not fully capture what happens at the auditory nerve.

Despite promising results, the current framework has shown some shortcomings that would need to be addressed in a future iteration. For instance, it would be worth exploring alternative models of electrical stimulation to evaluate whether or not they can produce NAPs that better represent the fine details observed in those from acoustic stimulation. Additionally, performance metrics (cross correlation, MSE, PSNR, and MI) should be computed for all available stimuli to assess to what extent results shown in this work could be generalized.

Similarly, improvements can be made to the evaluation of the framework. ResNet-50, the CNN used to develop the classifier, is one of many CNNs available and it would be worth exploring other alternatives such as GoogleNet, AlexNet, or VGG. In the same way CI users' brains have to adapt to the new type of stimulation, it would be reasonable to train the classifier with NAP_{CI}. In the case of post-lingually deafened CI users, their brains had been trained to recognize NAP_{NH} and has to learn to extract a new set of features to recognize NAP_{CI}. Therefore, it would be worth evaluating how the ratio between NAP_{CI} and NAP_{NH} used for training the classifier affects identification scores.

Ultimately, the framework needs to be evaluated with CI users to assess its performance on speech recognition tasks. Additionally, data collected from subjects would help validate to what extent the proposed classifier can predict recognition scores. The present work lays out the foundations of a new approach for developing and validating cochlear implant stimulation strategies. However, further research is needed to generalize the findings of this dissertation.

REFERENCES

[1] Aguiar, D. E., "Use of Computational Modeling of Auditory Nerve Activation Patterns for the Optimization of Cochlear Implant Electrical Stimulation Patterns," PhD thesis, Purdue University, 2012.

[2] Aguiar, D. E., Taylor, N. E., Li, J., Gazanfari, D. K., Talavage, T. M., Laflen, J. B., Neuberger, H., and Svirsky, M. A., "Information theoretic evaluation of a noiseband-based cochlear implant simulator," *Hearing Research*, vol. 333, pp. 185–193, 2016, ISSN: 18785891. DOI: 10.1016/j.heares.2015.09.008.

[3] Ahmad, T. J., Ali, H., Ajaz, M. A., and Khan, S. A., "Efficient algorithm development of CIS speech processing strategy for cochlear implants.," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, pp. 1270–1273, 2009, ISSN: 1557170X.

[4] Ali, H., Hong, F., Hansen, J. H. L., and Tobey, E., "Improving channel selection of sound coding algorithms in cochlear implants," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 905–909, ISBN: 9781479928927. DOI: 10.1109/ICASSP.2014.6853728.

[5] Ali, H., Noble, J. H., Gifford, R. H., Labadie, R. F., Dawant, B. M., Hansen, J. H. L., and Tobey, E., "Image-guided customization of frequency-place mapping in cochlear implants," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2015-Augus, IEEE, Apr. 2015, pp. 5843–5847, ISBN: 9781467369978. DOI: 10.1109/ICASSP.2015.7179092. [Online]. Available: http://ieeexplore.ieee.org/document/7179092/.

[6] American Speech-Language-Hearing Association (ASHA), "Cochlear Implants," Tech. Rep., 2004. DOI: 10.1044/policy.TR2004-00041. [Online]. Available: http://www.asha.org/policy/TR2004-00041/.

[7] Arnoldner, C., Riss, D., Brunner, M., Durisin, M., Baumgartner, W.-D., and Hamzavi, >.-S., "Speech and music perception with the new fine structure speech coding strategy: preliminary results," *Acta Oto-Laryngologica*, vol. 127, no. 12, pp. 1298–1303, Jan. 2007, ISSN: 0001-6489. DOI: 10.1080/00016480701275261. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/00016480701275261.

[8] Bassim, M. K., Buss, E., Clark, M. S., Kolln, K. A., Pillsbury, C. H., Pillsbury, H. C., and Buchman, C. A., "MED-EL Combi40+ cochlear implantation in adults," *Laryngoscope*, vol. 115, no. 9, pp. 1568–1573, 2005, ISSN: 0023852X. DOI: 10.1097/01.mlg.0000171023.72680.95.

[9] Berouti, M., Schwartz, R., and Makhoul, J., "Enhancement of speech corrupted by acoustic noise," in *ICASSP '79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Washington, DC, USA: Institute of Electrical and Electronics Engineers, Apr. 1979. DOI: 10.1109/ICASSP.1979.1170788. [Online]. Available: https://doi.org/10.1109/ICASSP.1979.1170788.

[10] Bierer, J. A., Spindler, E., Bierer, S. M., and Wright, R., "An Examination of Sources of Variability Across the Consonant-Nucleus-Consonant Test in Cochlear Implant Listeners," *Trends in Hearing*, vol. 20, pp. 1–8, 2016, ISSN: 23312165. DOI: 10.1177/2331216516646556.

[11] Biondi, E., Dacquino, G., and Grandori, F., "Compound action potential and single acoustic nerve fibres activity generation: An equivalent neuron approach," *International Journal of Bio-Medical Computing*, vol. 6, no. 3, pp. 157–166, 1975. DOI: 10.1016/0020-7101(75)90001-X. [Online]. Available: https://doi.org/10.1016/0020-7101(75)90001-X.

[12] Bolner, F., Goehring, T., Monaghan, J., Dijk, B. v., Wouters, J., and Bleeck, S., "Speech enhancement based on neural networks applied to cochlear implant coding strategies," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, IEEE, 2016, pp. 6520–6524, ISBN: 978-1-4799-9988-0. DOI: 10.1109/ICASSP.2016.7472933. [Online]. Available: http://ieeexplore.ieee.org/document/7472933/.

[13] Bondy, J., Becker, S., Bruce, I., Trainor, L., and Haykin, S., "A novel signal-processing strategy for hearing-aid design: Neurocompensation," *Signal Processing*, vol. 84, no. 7, pp. 1239–1253, Jul. 2004, ISSN: 01651684. DOI: 10.1016/j.sigpro.2004.04.006.

[14] Bosco, E., D'Agosta, L., Mancini, P., Traisci, G., D'Elia, C., and Filipo, R., "Speech perception results in children implanted with Clarion® devices: Hi-resolution™ and standard resolution modes," *Acta Oto-Laryngologica*, vol. 125, no. 2, pp. 148–158, Feb. 2005, ISSN: 0001-6489. DOI: 10.1080/00016480410023010. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/00016480410023010.

[15] Bruce, I. C., Irlicht, L. S., White, M. W., O'Leary, S. J., Dynes, S., Javel, E., and Clark, G. M., "A stochastic model of the electrically stimulated auditory nerve: Single-pulse response," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 6, pp. 617–629, 1999, ISSN: 0018-9294. DOI: 10.1109/10.764938.

[16] Bruce, I. C., Sachs, M. B., and Young, E. D., "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," *Journal of the Acoustical Society of America*, vol. 113, no. 1, pp. 369–388, Jan. 2003, ISSN: 0001-4966. DOI: 10.1121/1.1519544. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1519544.

[17] Bruce, I. C., White, M. W., Irlicht, L. S., O'Leary, S. J., Dynes, S., Javel, E., and Clark, G. M., "A stochastic model of the electrically stimulated auditory nerve: Pulse-train response," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 6, pp. 630–637, 1999, ISSN: 0018-9294. DOI: 10.1109/10.764939.

[18] Büchner, A., Beynon, A., Szyfter, W., Niemczyk, K., Hoppe, U., Hey, M., Brokx, J., Eyles, J., Van de Heyning, P., Paludetti, G., Zarowski, A., Quaranta, N., Wesarg, T., Festen, J., Olze, H., Dhooge, I., Müller-Deile, J., Ramos, A., Roman, S., Piron, J.-P., Cuda, D., Burdo, S., Grolman, W., Vaillard, S. R., Huarte, A., Frachet, B., Morera, C., Garcia-ibáñez, L., Abels, D., Walger, M., Müller-Mazotta, J., Leone, C. A., Meyer, B., Dillier, N., Steffens, T., Gentine, A., Mazzoli, M., Rypkema, G., Killian, M., Smoorenburg, G., Heyning, P. V. D., Frachet, B., Morera, C., Garcia-ibáñez, L., Abels, D., Meyer, B., Dillier, N., Steffens, T., and Gentine, A., "Clinical evaluation of cochlear implant sound coding taking into account conjectural masking functions, MP3000™.," *Cochlear implants international*, vol. 12, no. 4, pp. 194–204, 2011, ISSN: 1754-7628. DOI: 10.1179/1754762811Y0000000009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3175094&tool=pmcentrez&rendertype=abstract.

[19] Büchner, A., Lenarz, T., Boermans, P.-P., Frijns, J. H., Mancini, P., Filipo, R., Fielden, C., Cooper, H., Eklöf, M., Freijd, A., Lombaard, S., Meerton, L., Pickerill, M., Vanat, Z., Wesarg, T., Aschendorff, A., Kienast, B., Boyle, P., Arnold, L., Meyer, B., Sterkers, O., Müller-Deile, J., Ambrosch, P., Helbig, S., Frachet, B., Gallego, S., Truy, E., Jeffs, E., Morant, A., and Marco, J., "Benefits of the HiRes 120 coding strategy combined with the Harmony processor in an adult European multicentre study," *Acta Oto-Laryngologica*, vol. 132, no. 2, pp. 179–187, Feb. 2012, ISSN: 0001-6489. DOI: 10.3109/00016489.2011.630015. [Online]. Available: http://www.tandfonline.com/doi/full/10.3109/00016489.2011.630015.

[20] Buechner, A., Frohne-Buechner, C., Gaertner, L., Lesinski-Schiedat, A., Battmer, R.-D., and Lenarz, T., "Evaluation of Advanced Bionics high resolution mode," *International Journal of Audiology*, vol. 45, no. 7, pp. 407–416, Jan. 2006, ISSN: 1499-2027. DOI: 10.1080/14992020600625155. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/14992020600625155.

[21] Carney, L. H., "A model for the responses of low-frequency auditory-nerve fibers in cat.," *The Journal of the Acoustical Society of America*, vol. 93, no. 1, pp. 401–17, Jan. 1993, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/8423257.

[22] Clark, J. G., "Uses and abuses of hearing loss classification.," eng, *ASHA*, vol. 23, no. 7, pp. 493–500, Jul. 1981, ISSN: 0001-2475 (Print).

[23] Clopper, C. G. and Pisoni, D. B., "The Nationwide Speech Project: A new corpus of American English dialects," *Speech Communication*, vol. 48, no. 6, pp. 633–644, 2006, ISSN: 01676393. DOI: 10.1016/j.specom.2005.09.010.

[24] Danhauer, J. L., Ghadialy, F. B., Eskwitt, D. L., and Mendel, L. L., "Performance of 3M/house cochlear implant users on tests of speech perception.," eng, *Journal of the American Academy of Audiology*, vol. 1, no. 4, pp. 236–239, Oct. 1990, ISSN: 1050-0545.

[25] Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2009, ISBN: 978-1-4244-3992-8. DOI: 10.1109/CVPR.2009.5206848.

[26] Derbel, A., Ghorbel, M., Samet, M., and Ben Hamida, A., "New algorithm of bionic wavelet transform to DSP-implementation for cochlear implant," in *2012 16th IEEE Mediterranean Electrotechnical Conference*, IEEE, Mar. 2012, pp. 1025–1029, ISBN: 978-1-4673-0784-0. DOI: 10.1109/MELCON.2012.6196602. [Online]. Available: http://ieeexplore.ieee.org/document/6196602/.

[27] Derbel, A., Ghorbel, M., Samet, M., Hamida, A. B., and Tomas, J., "Cochlear implant stimulation strategy based on wavelet transform: Toward DSP real time implementation," in *2011 IEEE 9th International New Circuits and systems conference*, IEEE, Jun. 2011, pp. 117–121, ISBN: 978-1-61284-136-6. DOI: 10.1109/NEWCAS.2011.5981233. [Online]. Available: http://ieeexplore.ieee.org/document/5981233/.

[28] Djourno, A. and Eyries, C., "Auditory prosthesis by means of a distant electrical stimulation of the sensory nerve with the use of an indwelt coiling," fre, *Presse médicale*, vol. 65, no. 63, p. 1417, 1957, ISSN: 0032-7867. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/13484817.

[29] Djourno, A., Eyries, C., and Vallancien, B., "Electric excitation of the cochlear nerve in man by induction at a distance with the aid of micro-coil included in the fixture," *C. R. Seances Soc. Biol. Fil.*, vol. 151, no. 3, pp. 423–425, 1957, ISSN: 0037-9026. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/13479991.

[30] Djourno, A., Eyries, C., and Vallancien, B., "Preliminary attempts of electrical excitation of the auditory nerve in man, by permanently inserted micro-apparatus," fre, *Bulletin de l'Académie nationale de médecine*, vol. 141, no. 21-23, pp. 481–3, 1957, ISSN: 0001-4079. [Online]. Available: http://dx.doi.org/%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/13500039.

[31] Doh-Suk Kim, Soo-Young Lee, and Kil, R., "Auditory processing of speech signals for robust speech recognition in real-world noisy environments," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 1, 1999, ISSN: 10636676. DOI: 10.1109/89.736331. [Online]. Available: https://doi.org/10.1109/89.736331.

[32] Donaldson, G. S., Kreft, H. A., and Litvak, L., "Place-pitch discrimination of single- versus dual-electrode stimuli by cochlear implant users," *The Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 623–626, Aug. 2005, ISSN: 0001-4966. DOI: 10.1121/1.1937362. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1937362.

[33] Dorman, M. F. and Loizou, P. C., "Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants.," *The American journal of otology*, vol. 18, no. 6 Suppl, pp. 113–4, Nov. 1997, ISSN: 0192-9763. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9391623.

[34] Dorman, M. F. and Loizou, P. C., "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels.," *Ear and hearing*, vol. 19, no. 2, pp. 162–6, Apr. 1998, ISSN: 0196-0202. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9562538.

[35] Dorman, M. F., Loizou, P. C., and Rainey, D., "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding.," *The Journal of the Acoustical Society of America*, vol. 102, no. 5 Pt 1, pp. 2993–6, Nov. 1997, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9373986.

[36] Dorman, M. F., Loizou, P. C., and Rainey, D., "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs.," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2403–2411, Oct. 1997, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9348698.

[37] Dorman, M. F., Natale, S. C., Baxter, L., Zeitler, D. M., Carlson, M. L., Lorens, A., Skarzynski, H., Peters, J. P. M., Torres, J. H., and Noble, J. H., "Approximations to the Voice of a Cochlear Implant: Explorations With Single-Sided Deaf Listeners," *Trends in Hearing*, vol. 24, pp. 1–12, Jan. 2020, ISSN: 2331-2165. DOI: 10.1177/2331216520920079. [Online]. Available: https://dx.doi.org/10.1177%2F2331216520920079.

[38] Dorman, M. F. and Wilson, B. S., "The design and function of cochlear implants," *American Scientist*, vol. 92, no. 5, pp. 436–445, 2004, ISSN: 00030996. DOI: 10.1511/2004.5.436. [Online]. Available: http://www.americanscientist.org/issues/feature/2004/5/the-design-and-function-of-cochlear-implants.

[39] Eddington, D. K., Brackmann, D. E., Dobelle, W. H., and Mladejovsky, M. G., "Auditory Prostheses Research with Multiple Channel Intracochlear Stimulation in Man," *Annals of Otology, Rhinology & Laryngology*, vol. 87, no. 6_suppl, pp. 5–39, 1987. DOI: 10.1177/00034894780870S602.

[40] Fetterman, B. L. and Domico, E. H., "Speech recognition in background noise of cochlear implant patients," *Otolaryngology - Head and Neck Surgery*, vol. 126, no. 3, pp. 257–263, 2002, ISSN: 01945998. DOI: 10.1067/mhn.2002.123044.

[41] Firszt, J. B., Holden, L. K., Reeder, R. M., and Skinner, M. W., "Speech recognition in cochlear implant recipients: comparison of standard HiRes and HiRes 120 sound processing," *Otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, vol. 30, no. 2, p. 146, 2009.

[42] Firszt, J. B., Koch, D. B., Downing, M., and Litvak, L., "Current Steering Creates Additional Pitch Percepts in Adult Cochlear Implant Recipients," *Otology & Neurotology*, vol. 28, no. 5, pp. 629–636, Aug. 2007, ISSN: 1531-7129. DOI: 10.1097/01.mao.0000281803.36574.bc. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17667771%20http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00129492-200708000-00007.

[43] Fitzgerald, M. B., Morbiwala, T. A., and Svirsky, M. A., "Customized selection of frequency maps in an acoustic simulation of a cochlear implant," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, Aug. 2006, pp. 3596–3599, ISBN: 1-4244-0032-5. DOI: 10.1109/IEMBS.2006.260462. [Online]. Available: http://ieeexplore.ieee.org/document/4462575/.

[44] Fitzgerald, M. B., Sagi, E., Morbiwala, T. A., Tan, C.-T., and Svirsky, M. A., "Feasibility of real-time selection of frequency tables in an acoustic simulation of a cochlear implant.," *Ear and hearing*, vol. 34, no. 6, pp. 763–72, 2013, ISSN: 1538-4667. DOI: 10.1097/AUD.0b013e3182967534. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23807089%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3899943.

[45] Fretz, R. J. and Fravel, R. P., "Design and Function: A Physical and Electrical Description of the 3M House Cochlear Implant System," *Ear and Hearing*, vol. 6, no. 3, 14S–19S, 1985, ISSN: 0196-0202.

[46] Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X., "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *The Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 1150–1163, Aug. 2001, ISSN: 0001-4966. DOI: 10.1121/1.1381538. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1381538.

[47] Fu, Q.-J. and Shannon, R. V., "Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners.," *The Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 2570–7, Nov. 1998, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9821336.

[48] Fu, Q.-J. and Shannon, R. V., "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," *The Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1889–1900, 1999, ISSN: 00014966. DOI: 10.1121/1.426725. [Online]. Available: http://dx.doi.org/10.1121/1.426725%20http://asa.scitation.org/toc/jas/105/3.

[49] Gan, R. Z., Sun, Q., Dyer, R. K., Chang, K.-H., and Dormer, K. J., "Three-dimensional modeling of middle ear biomechanics and its applications.," *Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, vol. 23, no. 3, pp. 271–80, May 2002, ISSN: 1531-7129. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11981381.

[50] Geurts, L. and Wouters, J., "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants.," *The Journal of the Acoustical Society of America*, vol. 105, no. 4, pp. 2476–84, Apr. 1999, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10212428.

[51] Ghitza, O., "Auditory models and human performance in tasks related to speech coding and speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 1, Jan. 1994, ISSN: 1063-6676. DOI: 10.1109/89.260357. [Online]. Available: https://doi.org/10.1109/89.260357.

[52] Ghitza, O., "Auditory nerve representation as a front-end for speech recognition in a noisy environment," *Computer Speech & Language*, vol. 1, no. 2, Dec. 1986, ISSN: 08852308. DOI: 10.1016/S0885-2308(86)80018-3. [Online]. Available: https://doi.org/10.1016/S0885-2308(86)80018-3.

[53] Ghitza, O., "Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment," *Journal of Phonetics*, vol. 16, no. 1, Jan. 1988, ISSN: 00954470. DOI: 10.1016/S0095-4470(19)30469-3. [Online]. Available: https://doi.org/10.1016/S0095-4470(19)30469-3.

[54] Ghitza, O., "Adequacy of auditory models to predict human internal representation of speech sounds," *The Journal of the Acoustical Society of America*, vol. 93, no. 4, Apr. 1993, ISSN: 0001-4966. DOI: 10.1121/1.406679. [Online]. Available: https://doi.org/10.1121/1.406679.

[55] Gifford, R. H., "Who is a cochlear implant candidate?" *The Hearing Journal*, vol. 64, no. 6, pp. 16–22, 2011, ISSN: 07457472. DOI: 10.1097/01.HJ.0000399149.53245.b1. [Online]. Available: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00025572-201106000-00007.

[56] Gifford, R. H., Shallop, J. K., and Peterson, A. M., "Speech Recognition Materials and Ceiling Effects: Considerations for Cochlear Implant Programs," eng, *Audiology and Neurotology*, vol. 13, no. 3, pp. 193–205, 2008, ISSN: 1420-3030. DOI: 10.1159/000113510.

[57] Giguère, C. and Woodland, P. C., "A computational model of the auditory periphery for speech and hearing research. I. Ascending path.," *The Journal of the Acoustical Society of America*, vol. 95, no. 1, pp. 331–42, Jan. 1994, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/8120244.

[58] Goode, R. L., Killion, M., Nakamura, K., and Nishihara, S., "New knowledge about the function of the human middle ear: development of an improved analog model.," *The American journal of otology*, vol. 15, no. 2, pp. 145–54, Mar. 1994, ISSN: 0192-9763. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/8172293.

[59] Gopalakrishna, V., Kehtarnavaz, N., and Loizou, P. C., "Real-time implementation of wavelet-based advanced combination encoder on PDA platforms for cochlear implant studies," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, IEEE, 2010, pp. 1670–1673, ISBN: 9781424442966. DOI: 10.1109/ICASSP.2010.5495509.

[60] Grayden, D. B., Burkitt, A. N., Kenny, O. P., Clarey, J. C., Paolini, A. G., and Clark, G. M., "A cochlear implant speech processing strategy based on an auditory model," in *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004.*, IEEE, 2004, pp. 491–496, ISBN: 0-7803-8894-1. DOI: 10.1109/ISSNIP.2004.1417510. [Online]. Available: http://ieeexplore.ieee.org/document/1417510/.

[61] Greenwood, D. D., "Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane," *The Journal of the Acoustical Society of America*, vol. 33, no. 10, pp. 1344–1356, 1961, ISSN: 0001-4966. DOI: 10.1121/1.1908437. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1908437.

[62] Greenwood, D. D., "A cochlear frequency-position function for several species–29 years later.," *The Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2592–605, Jun. 1990, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2373794.

[63] Harczos, T., "Cochlear Implant Electrode Stimulation Strategy Based on a Human Auditory Model," PhD thesis, Technische Universität Ilmenau, Jul. 2015. [Online]. Available: https://www.db-thueringen.de/receive/dbt_mods_00026207.

[64] Harczos, T., Chilian, A., and Husar, P., "Making use of auditory models for better mimicking of normal hearing processes with cochlear implants: The SAM coding strategy," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 7, no. 4, pp. 414–425, Aug. 2013, ISSN: 19324545. DOI: 10.1109/TBCAS.2012.2219530. [Online]. Available: http://ieeexplore.ieee.org/document/6384826/.

[65] He, K., Zhang, X., Ren, S., and Sun, J., "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.90.

[66] Heinz, M. G., Colburn, H. S., and Carney, L. H., "Evaluating Auditory Performance Limits: I. One-Parameter Discrimination Using a Computational Model for the Auditory Nerve," *Neural Computation*, vol. 13, no. 10, pp. 2273–2316, 2001, ISSN: 0899-7667. DOI: 10.1162/089976601750541804. [Online]. Available: http://www.mitpressjournals.org/doi/10.1162/089976601750541804.

[67] Heinz, M. G., Colburn, H. S., and Carney, L. H., "Evaluating Auditory Performance Limits: II. One-Parameter Discrimination with Random-Level Variation," *Neural Computation*, vol. 13, no. 10, pp. 2317–2338, 2001, ISSN: 0899-7667. DOI: 10.1162/089976601750541804. [Online]. Available: http://www.mitpressjournals.org/doi/10.1162/089976601750541804.

[68] Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H., "Auditory nerve model for predicting performance limits of normal and impaired listeners," *Acoustics Research Letters Online*, vol. 2, no. 3, pp. 91–96, Jul. 2001, ISSN: 1529-7853. DOI: 10.1121/1.1387155. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1387155.

[69] Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K., "Acoustic characteristics of American English vowels," *The Journal of the Acoustical Society of America*, vol. 97, no. 5, May 1995, ISSN: 0001-4966. DOI: 10.1121/1.411872. [Online]. Available: https://doi.org/10.1121/1.411872.

[70] Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S., "Effects of noise and noise suppression onspeech perception by cochlear implantusers," *Ear and Hearing*, vol. 13, no. 4, pp. 263–271, 1992, ISSN: 15384667. DOI: 10.1097/00003446-199208000-00008.

[71] Hochmair, I., Nopp, P., Jolly, C., Schmidt, M., Schösser, H., Garnham, C., and Anderson, I., "MED-EL Cochlear implants: state of the art and a glimpse into the future.," *Trends in amplification*, vol. 10, no. 4, pp. 201–19, Dec. 2006, ISSN: 1084-7138. DOI: 10.1177/1084713806296720. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17172548%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4111377.

[72] Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., Heydebrand, G., and Skinner, M. W., "Factors affecting open-set word recognition in adults with cochlear implants," *Ear and Hearing*, vol. 34, no. 3, pp. 342–360, 2013, ISSN: 01960202. DOI: 10.1097/AUD.0b013e3182741aa7.

[73] House, W. F. and Urban, J., "Long term results of electrode implantation and electronic stimulation of the cochlea in man.," *The Annals of otology, rhinology, and laryngology*, vol. 82, no. 4, pp. 504–517, 1973, ISSN: 0003-4894. DOI: 10.1177/000348947308200408. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/4721186%0Ahttp://journals.sagepub.com/doi/pdf/10.1177/000348947308200408.

[74] Incerti, P. V., Ching, T. Y., and Cowan, R., "A systematic review of electric-acoustic stimulation: Device fitting ranges, outcomes, and clinical fitting practices," *Trends in Amplification*, vol. 17, no. 1, pp. 3–26, 2013, ISSN: 19405588. DOI: 10.1177/1084713813480857.

[75] Irino, T. and Patterson, R. D., "A Time Domain, Level-Dependent Auditory Filter: The Gammachirp," *The Journal of the Acoustical Society of America*, vol. 101, no. 1, pp. 412–419, 1997, ISSN: 00014966. DOI: 10.1121/1.417975.

[76] Irino, T. and Patterson, R. D., "A compressive gammachirp auditory filter for both physiological and psychophysical data," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2008–2022, May 2001, ISSN: 0001-4966. DOI: 10.1121/1.1367253. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1367253.

[77] Joshi, S. N., Dau, T., and Epp, B., "A Model of Electrically Stimulated Auditory Nerve Fiber Responses with Peripheral and Central Sites of Spike Generation," *JARO - Journal of the Association for Research in Otolaryngology*, vol. 18, no. 2, pp. 323–342, 2017, ISSN: 14387573. DOI: 10.1007/s10162-016-0608-2.

[78] Kaiser, A. R. and Svirsky, M. A., "Using a Personal Computer to Perform Real-Time Signal Processing in Cochlear Implant Research," in *9th Digital Signal Processing Workshop. IEEE Signal Processing Society*, Hunt, Texas, 2000.

[79] Kiefer, J., Hohl, S., Stürzebecher, E., Pfennigdorff, T., and Gstöettner, W., "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:Comparación del reconocimiento del len," *International Journal of Audiology*, vol. 40, no. 1, pp. 32–42, 2001, ISSN: 1499-2027. DOI: 10.3109/00206090109073098. [Online]. Available: http://informahealthcare.com/doi/abs/10.3109/00206090109073098.

[80] Kitazawa, S., Muramoto, K., and Ito, J., "Acoustic simulation of auditory model based speech processor for cochlear implant system," in *Proc. International Conference on Spoken Language Processing ICSLP'94*, 1994, pp. 2043–2046.

[81] Koch, D. B., Downing, M., Osberger, M. J., and Litvak, L., "Using Current Steering to Increase Spectral Resolution in CII and HiRes 90K Users," *Ear and Hearing*, vol. 28, no. Supplement, 38S–41S, Apr. 2007, ISSN: 0196-0202. DOI: 10.1097/AUD.0b013e31803150de. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17496643%20http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00003446-200704001-00009.

[82] Koch, D. B., Osberger, M. J., Segel, P., and Kessler, D., "HiResolution and Conventional Sound Processing in the HiResolution Bionic Ear: Using Appropriate Outcome Measures to Assess Speech Recognition Ability," *Audiology & Neuro-Otology*, vol. 9, no. 4, pp. 214–223, 2004, ISSN: 1420-3030. DOI: 10.1159/000078391. [Online]. Available: http://www.karger.com/DOI/10.1159/000078391.

[83] Koike, T., Wada, H., and Kobayashi, T., "Modeling of the human middle ear using the finite-element method," *The Journal of the Acoustical Society of America*, vol. 111, no. 3, pp. 1306–1317, Mar. 2002, ISSN: 0001-4966. DOI: 10.1121/1.1451073. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1451073.

[84] Kompis, M., Vischer, M. W., and Häusler, R., "Performance of compressed analogue (CA) and continuous interleaved sampling (CIS) coding strategies for cochlear implants in quiet and noise," *Acta Oto-Laryngologica*, vol. 119, no. 6, pp. 659–64, 1999. DOI: 10.1080/00016489950180595. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10586998?dopt=AbstractPlus.

[85] Kringlebotn, M., "Network model for the human middle ear.," *Scandinavian audiology*, vol. 17, no. 2, pp. 75–85, 1988, ISSN: 0105-0397. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/3187377.

[86] Laflen, J. B., "Measurement and Analysis of Perceptual Coding in the Human Auditory System: Multi-Modal Studies Using Neural Activation Patterns," PhD thesis, Purdue University, 2003.

[87] Laflen, J. B., Talavage, T. M., Thirukkonda, P. M., and Svirsky, M. A., "Physiologically based analysis of cochlear implant representations," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, vol. 3, 2002.

[88] Lai, W. K., Aksit, M., Akdas, F., and Dillier, N., "Longitudinal behaviour of neural response telemetry (NRT) data and clinical implications," *International Journal of Audiology*, vol. 43, no. 5, pp. 252–263, Jan. 2004, ISSN: 1499-2027. DOI: 10.1080/14992020400050034. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/14992020400050034.

[89] Lai, W. K. and Dillier, N., "Investigating the MP3000 coding strategy for music perception," Tech. Rep., 2008, pp. 1–4. DOI: 10.5167/uzh-13612.

[90] Lai, Y.-H., Wang, S.-S., Su, Y.-T., Han-Che, C., Fu, F. K., and Tsao, Y., "Improving the performance of speech perception in noisy environment based on an FAME strategy," in *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, IEEE, Oct. 2016, pp. 1–5, ISBN: 978-1-5090-4294-4. DOI: 10.1109/ISCSLP.2016.7918430. [Online]. Available: http://ieeexplore.ieee.org/document/7918430/.

[91] Lehiste, I. and Peterson, G. E., "Linguistic Considerations in the Study of Speech Intelligibility," *The Journal of the Acoustical Society of America*, vol. 31, no. 3, pp. 280–286, Mar. 1959, ISSN: 0001-4966. DOI: 10.1121/1.1907713. [Online]. Available: https://doi.org/10.1121/1.1907713.

[92] Liberman, M. C. and Kiang, N. Y., "Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions.," *Hearing research*, vol. 16, no. 1, pp. 75–90, Oct. 1984, ISSN: 0378-5955. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/6511674.

[93] Lin, F. R., "Hearing Loss Prevalence in the United States," *Archives of Internal Medicine*, vol. 171, no. 20, p. 1851, Nov. 2011, ISSN: 0003-9926. DOI: 10.1001/archinternmed.2011.506. [Online]. Available: http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinternmed.2011.506.

[94] Loizou, P. C., "Mimicking the human ear," *IEEE Signal Processing Magazine*, vol. 15, no. 5, pp. 101–130, 1998, ISSN: 1558-0792. DOI: 10.1109/79.708543. [Online]. Available: https://doi.org/10.1109/79.708543.

[95] Loizou, P. C., "Signal-processing techniques for cochlear implants: A review of progress in deriving electrical stimuli from the speech signal," *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 3, pp. 34–46, 1999, ISSN: 07395175. DOI: 10.1109/51.765187. [Online]. Available: http://ieeexplore.ieee.org/document/765187/.

[96] Loizou, P. C., "Speech Processing in Vocoder-Centric Cochlear Implants," in, 2006. DOI: 10.1159/000094648.

[97] Loizou, P. C., Dorman, M. F., Poroy, O., and Spahr, T., "Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution.," *The Journal of the Acoustical Society of America*, vol. 108, no. 5 Pt 1, pp. 2377–2387, Nov. 2000, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11108378.

[98] Lyon, R. F., "All-pole models of auditory filtering," *Diversity in auditory mechanics*, pp. 205–211, 1997.

[99] Magnusson, L., "Comparison of the fine structure processing (FSP) strategy and the CIS strategy used in the MED-EL cochlear implant system: Speech intelligibility and music sound quality," *International Journal of Audiology*, vol. 50, no. 4, pp. 279–287, Apr. 2011, ISSN: 1499-2027. DOI: 10.3109/14992027.2010.537378. [Online]. Available: http://www.tandfonline.com/doi/full/10.3109/14992027.2010.537378.

[100] May, B. J., "Auditory Physiology and Perception," in *Cochlear Implants: Principles and practices*, Niparko, J. K., Kirk, K. I., Robbins, A. M., Tucci, D. L., and Wilson, B. S., Eds., Philadelphia: Lippincott Williams & Wilkins, 2000, pp. 9–28, ISBN: 0-7817-1782-5.

[101] Meddis, R., "Simulation of mechanical to neural transduction in the auditory receptor.," *The Journal of the Acoustical Society of America*, vol. 79, no. 3, pp. 702–11, Mar. 1986, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2870094.

[102] Meddis, R., "Simulation of auditory–neural transduction: Further studies," *The Journal of the Acoustical Society of America*, vol. 83, no. 3, pp. 1056–1063, 1988, ISSN: 0001-4966.

[103] Merzenich, M. M., "UCSF Cochlear Implant Device," *Cochlear Implants*, pp. 121–129, 1985.

[104] Michelson, R. P., "The Results of Electrical Stimulation of the Cochlea in Human Sensory Deafness.," *The Annals of Otology, Rhinology & Laryngology*, vol. 80, no. 6, pp. 914–919, 1971, ISSN: 0003-4894. DOI: 10.1177/000348947108000618.

[105] "Minimum Speech Test Battery (MSTB) For Adult Cochlear Implant Users 2011," Tech. Rep., Jun. 2011. [Online]. Available: http://www.auditorypotential.com/MSTBfiles/MSTBManual2011-06-20%20.pdf.

[106] Møller, A. R., "Network Model of the Middle Ear," *The Journal of the Acoustical Society of America*, vol. 33, no. 2, pp. 168–176, Feb. 1961, ISSN: 0001-4966. DOI: 10.1121/1.1908610. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1908610.

[107] Moon, I. J. and Hong, S. H., "What is temporal fine structure and why is it important?" *Korean journal of audiology*, vol. 18, no. 1, pp. 1–7, Apr. 2014, ISSN: 2092-9862. DOI: 10.7874/kja.2014.18.1.1. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24782944%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4003734.

[108] Moore, B. C. J., "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people.," *Journal of the Association for Research in Otolaryngology : JARO*, vol. 9, no. 4, pp. 399–406, Dec. 2008, ISSN: 1525-3961. DOI: 10.1007/s10162-008-0143-x. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18855069%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2580810.

[109] Morbiwala, T. A., Svirsky, M. A., El-Sharkway, M., and Rizkalla, M., "A PC-based speech processor for cochlear implant fitting that can be adjusted in real-time," in *48th Midwest Symposium on Circuits and Systems, 2005.*, IEEE, 2005, pp. 1310–1313, ISBN: 0-7803-9197-7. DOI: 10.1109/MWSCAS.2005.1594350. [Online]. Available: http://ieeexplore.ieee.org/document/1594350/.

[110] Müller, J., Brill, S., Hagen, R., Moeltner, A., Brockmeier, S.-J., Stark, T., Helbig, S., Maurer, J., Zahnert, T., Zierhofer, C., Nopp, P., and Anderson, I., "Clinical Trial Results with the MED-EL Fine Structure Processing Coding Strategy in Experienced Cochlear Implant Users," *ORL*, vol. 74, no. 4, pp. 185–198, 2012, ISSN: 0301-1569. [Online]. Available: http://www.karger.com/DOI/10.1159/000337089.

[111] Nie, K., Stickney, G., and Zeng, F.-G., "Encoding Frequency Modulation to Improve Cochlear Implant Performance in Noise," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 1, pp. 64–73, Jan. 2005, ISSN: 0018-9294. DOI: 10.1109/TBME.2004.839799. [Online]. Available: http://ieeexplore.ieee.org/document/1369589/.

[112] Nilsson, M., Soli, S. D., and Sullivan, J. A., "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," *Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, 1994. DOI: 10.1121/1.408469. [Online]. Available: https://asa.scitation.org/doi/10.1121/1.408469.

[113] Noble, J. H., Labadie, R. F., Gifford, R. H., and Dawant, B. M., "Image-Guidance enables new methods for customizing cochlear implant stimulation strategies," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 5, pp. 820–829, Sep. 2013, ISSN: 1558-0210. DOI: 10.1109/TNSRE.2013.2253333. [Online]. Available: http://ieeexplore.ieee.org/document/6482251/.

[114] Nogueira, W., Büchner, A., Lenarz, T., and Edler, B., "A Psychoacoustic "NofM" - Type Speech Coding Strategy for Cochlear Implants," *EURASIP Journal on Applied Signal Processing*, vol. 18, pp. 3044–3059, 2005, ISSN: 1687-6172. DOI: 10.1155/ASP.2005.3044.

[115] Paglialonga, A., Tognola, G., Baselli, G., Parazzini, M., Ravazzani, P., and Grandori, F., "Speech processing for cochlear implants with the discrete wavelet transform: feasibility study and performance evaluation.," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 1, IEEE, Aug. 2006, pp. 3763–3766, ISBN: 1424400325. DOI: 10.1109/IEMBS.2006.259281. [Online]. Available: http://ieeexplore.ieee.org/document/4462618/.

[116] Pasanisi, E., Bacciu, A., Vincenti, V., Guida, M., Berghenti, M. T., Barbot, A., Panu, F., and Bacciu, S., "Comparison of speech perception benefits with SPEAK and ACE coding strategies in pediatric Nucleus CI24M cochlear implant recipients," *International Journal of Pediatric Otorhinolaryngology*, vol. 64, no. 2, pp. 159–163, Jun. 2002, ISSN: 01655876. DOI: 10.1016/S0165-5876(02)00075-7. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0165587602000757.

[117] Pascal, J., Bourgeade, A., Lagier, M., and Legros, C., "Linear and nonlinear model of the human middle ear.," *The Journal of the Acoustical Society of America*, vol. 104, no. 3 Pt 1, pp. 1509–16, Sep. 1998, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9745735.

[118] Patterson, R. D., Allerhand, M. H., and Giguère, C., "Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform.," *The Journal of the Acoustical Society of America*, vol. 98, no. 4, pp. 1890–1894, Oct. 1995, ISSN: 0001-4966. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/7593913.

[119] Patterson, R. D. and Holdsworth, J., "A functional model of neural activity patterns and auditory images," in *Advances in speech, hearing and language processing*, Part B, Ainsworth, W. A., Ed., vol. 3, London: JAI Press, 1996, pp. 547–563.

[120] Patterson, R. D., Holdsworth, J., and Allerhand, M., "Auditory Models as Preprocessors for Speech Recognition," in *The Auditory Processing of Speech*, Berlin, Boston: DE GRUYTER MOUTON, 1992, pp. 67–83. DOI: 10.1515/9783110879018.67. [Online]. Available: https://www.degruyter.com/view/books/9783110879018/9783110879018.67/9783110879018.67.xml.

[121] Pelizzone, M., Boëx-Spano, C., Sigrist, A., François, J., Tinembart, J., Degive, C., and Montandon, P., "First Field Trials with a Portable CIS Processor for the Ineraid Multichannel Cochlear Implant," *Acta Oto-Laryngologica*, vol. 115, no. 5, pp. 622–628, Jan. 1995, ISSN: 0001-6489. DOI: 10.3109/00016489509139377. [Online]. Available: http://www.tandfonline.com/doi/full/10.3109/00016489509139377.

[122] Peterson, G. E. and Lehiste, I., "Revised CNC Lists for Auditory Tests," *Journal of Speech and Hearing Disorders*, vol. 27, no. 1, Feb. 1962, ISSN: 0022-4677. DOI: 10.1044/jshd.2701.62.

[123] Psarros, C. E., Plant, K. L., Lee, K., Decker, J. A., Whitford, L. A., and Cowan, R. S. C., "Conversion from the SPEAK to the ACE Strategy in Children Using the Nucleus 24 Cochlear Implant System: Speech Perception and Speech Production Outcomes," *Ear and Hearing*, vol. 23, no. 1, 18S–27S, 2002, ISSN: 0196-0202. [Online]. Available: http://journals.lww.com/ear-hearing/Fulltext/2002/02001/Conversion_from_the_SPEAK_to_the_ACE_Strategy_in.3.aspx.

[124] Purves, D. and Williams, S. M., "Chapter 13: The Auditory System," in *Neuroscience*, 2nd, Sunderland, MA: Sinauer Associates, 2001, ISBN: 0878937420. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK10799/.

[125] Ray Meddis, Enrique A. Lopez-Poveda, Richard R. Fay, and Arthur N. Popper, *Computational Models of the Auditory System*, 1st ed., Meddis, R., Lopez-Poveda, E. A., Fay, R. R., and Popper, A. N., Eds., ser. Springer Handbook of Auditory Research. Boston, MA: Springer, 2010, ISBN: 978-1-4419-1370-8. DOI: 10.1007/978-1-4419-5934-8. [Online]. Available: http://link.springer.com/10.1007/978-1-4419-5934-8.

[126] Riss, D., Arnoldner, C., Baumgartner, W.-D., Kaider, A., and Hamzavi, J.-S., "A New Fine Structure Speech Coding Strategy: Speech Perception at a Reduced Number of Channels," *Otology & Neurotology*, vol. 29, no. 6, pp. 784–788, Sep. 2008, ISSN: 1531-7129. DOI: 10.1097/MAO.0b013e31817fe00f. [Online]. Available: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00129492-200809000-00009.

[127] Riss, D., Hamzavi, J.-S., Blineder, M., Honeder, C., Ehrenreich, I., Kaider, A., Baumgartner, W.-D., Gstoettner, W., and Arnoldner, C., "FS4, FS4-p, and FSP: a 4-month crossover study of 3 fine structure sound-coding strategies," *Ear and Hearing*, vol. 35, no. 6, e272–e281, 2014, ISSN: 0196-0202. DOI: 10.1097/AUD.0000000000000063. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25127325%20http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00003446-201411000-00020.

[128] Robert, A. and Eriksson, J. L., "A composite model of the auditory periphery for simulating responses to complex sounds," *Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1852–1864, Aug. 1999. DOI: 10.1121/1.427935. [Online]. Available: http://asa.scitation.org/doi/abs/10.1121/1.427935.

[129] Robles, L. and Ruggero, M. A., "Mechanics of the mammalian cochlea.," *Physiological reviews*, vol. 81, no. 3, pp. 1305–52, Jul. 2001, ISSN: 0031-9333. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11427697%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3590856.

[130] Roeser, R. J., Valente, M., and Hosford-Dunn, H., *Audiology: Diagnosis*, 2nd, Ross J. Roeser, Michael Valente, and Holly Hosford-Dunn, Eds. New York: Thieme Medical Publishers, Inc, 2007, ISBN: 9781604066326.

[131] Seligman, P. M., Patrick, J. F., Tong, Y. C., Clark, G. M., Dowell, R. C., and Crosby, P. A., "A signal processor for a multiple-electrode hearing prosthesis," *Acta Oto-Laryngologica*, vol. 97, no. S413, pp. 135–139, 1984, ISSN: 00016489. DOI: 10.3109/00016488409121684.

[132] Seligman, P. and McDermott, H. J., "Architecture of the spectra 22 speech processor," *Annals of Otology, Rhinology and Laryngology*, vol. 104, no. 9 II SUPPL. Pp. 139–141, May 1995, ISSN: 00034894. [Online]. Available: http://hdl.handle.net/11343/27441%20file:///bitstream/handle/11343/27441/119423_vol8_761.pdf?sequence=1.

[133] Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M., "Speech recognition with primarily temporal cues.," *Science*, vol. 270, no. 5234, pp. 303–304, Oct. 1995, ISSN: 0036-8075. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/7569981.

[134] Shannon, R. V., Zeng, F.-G., and Wygonski, J., "Speech recognition with altered spectral distribution of envelope cues," *The Journal of the Acoustical Society of America*, vol. 104, no. 4, Oct. 1998, ISSN: 0001-4966. DOI: 10.1121/1.423774. [Online]. Available: https://doi.org/10.1121/1.423774.

[135] Simmons, F. B., Epley, J. M., Lummis, R. C., Guttman, N., Frishkopf, L. S., Harmon, L. D., and Zwicker, E., "Auditory Nerve: Electrical Stimulation in Man.," *Science (New York, N.Y.)*, vol. 148, no. 3, pp. 104–106, 1965, ISSN: 0036-8075. DOI: 10.1126/science.148.3666.104.

[136] Slaney, M. and Lyon, R. F., "A perceptual pitch detector," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE, 1990, pp. 357–360. DOI: 10.1109/ICASSP.1990.115684. [Online]. Available: http://ieeexplore.ieee.org/document/115684/.

[137] Spoendlin, H., "Anatomy of Cochlear Innervation," *American Journal of Otolaryngology*, vol. 6, pp. 453–467, 1985. DOI: 10.1016/S0196-0709(85)80026-0. [Online]. Available: https://doi.org/10.1016/S0196-0709(85)80026-0.

[138] Svirsky, M. A., Ding, N., Sagi, E., Tan, C.-T., Fitzgerald, M. B., Glassman, E. K., Seward, K., and Neuman, A. C., "Validation of Acoustic Models of Auditory Neural Prostheses," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP (Conference)*, NIH Public Access, May 2013, pp. 8629–8633. DOI: 10.1109/ICASSP.2013.6639350. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25435816%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4244817.

[139] Tabibi, S., Boulet, J., Dillier, N., and Bruce, I. C., "Phenomenological model of auditory nerve population responses to cochlear implant stimulation," *Journal of Neuroscience Methods*, vol. 358, Jul. 2021, ISSN: 1872678X. DOI: 10.1016/j.jneumeth.2021.109212.

[140] Taft, D. A., Grayden, D. B., and Burkitt, A. N., "Across-Frequency Delays Based on the Cochlear Traveling Wave: Enhanced Speech Presentation for Cochlear Implants," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 596–606, Mar. 2010, ISSN: 0018-9294. DOI: 10.1109/TBME.2009.2034014. [Online]. Available: http://ieeexplore.ieee.org/document/5290084/.

[141] Takanen, M., Bruce, I. C., and Seeber, B. U., "Phenomenological modelling of electrically stimulated auditory nerve fibers: A review," *Network: Computation in Neural Systems*, vol. 27, no. 2-3, pp. 157–185, Jul. 2016, ISSN: 0954-898X. DOI: 10.1080/0954898X.2016.1219412. [Online]. Available: https://doi.org/10.1080/0954898X.2016.1219412.

[142] Tan, Q. and Carney, L. H., "A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide," *The Journal of the Acoustical Society of America*, vol. 114, no. 4, pp. 2007–2020, Oct. 2003, ISSN: 0001-4966. DOI: 10.1121/1.1608963. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1608963.

[143] Tyler, R. S., "Open-set word recognition with the 3M/Vienna single-channel cochlear implant.," eng, *Archives of otolaryngology–head & neck surgery*, vol. 114, no. 10, pp. 1123–1126, Oct. 1988, ISSN: 0886-4470 (Print).

[144] Watts, L., "Cochlear Mechanics: Analysis and Analog VLSI," PhD thesis, California Institute of Technology, Oct. 1993. [Online]. Available: http://www.lloydwatts.com/thesis.html.

[145] Westerman, L. A. and Smith, R. L., "A diffusion model of the transient response of the cochlear inner hair cell synapse," *The Journal of the Acoustical Society of America*, vol. 83, no. 6, pp. 2266–2276, Jun. 1988, ISSN: 0001-4966. DOI: 10.1121/1.396357. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.396357.

[146] Wilson, B. S., "Cochlear Implant Technology," in *Cochlear Implants: Principles and practices*, Niparko, J. K., Kirk, K. I., Robbins, A. M., Tucci, D. L., and Wilson, B. S., Eds., Philadelphia: Lippincott Williams & Wilkins, 2000, pp. 109–127, ISBN: 0-7817-1782-5.

[147] Wilson, B. S., Finley, C. C., Lawson, D. T., and Wolford, R. D., "Speech processors for cochlear prostheses," *Proceedings of the IEEE*, vol. 76, no. 9, pp. 1143–1154, 1988, ISSN: 00189219. DOI: 10.1109/5.9660. [Online]. Available: http://ieeexplore.ieee.org/document/9660/.

[148] Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M., "Better speech recognition with cochlear implants," *Nature*, vol. 352, no. 6332, pp. 236–238, 1991, ISSN: 0028-0836. DOI: 10.1038/352236a0. [Online]. Available: http://www.nature.com/doifinder/10.1038/352236a0.

[149] Wouters, J., McDermott, H. J., and Francart, T., "Sound coding in cochlear implants: From electric pulses to hearing," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 67–80, Mar. 2015, ISSN: 10535888. DOI: 10.1109/MSP.2014.2371671. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7038260.

[150] Zeng, F.-G., Rebscher, S., Harrison, W., Sun, X., and Feng, H., "Cochlear Implants: System Design, Integration, and Evaluation," *IEEE Reviews in Biomedical Engineering*, vol. 1, pp. 115–142, Jan. 2008, ISSN: 19411189. DOI: 10.1109/RBME.2008.2008250. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2782849/.

[151] Zhang, X. and Carney, L. H., "Analysis of models for the synapse between the inner hair cell and the auditory nerve," *The Journal of the Acoustical Society of America*, vol. 118, no. 3, pp. 1540–1553, 2005, ISSN: 0001-4966. DOI: 10.1121/1.1993148. [Online]. Available: http://scitation.aip.org/content/asa/journal/jasa/118/3/10.1121/1.1993148.

[152] Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H., "A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression," *The Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 648–670, Feb. 2001, ISSN: 0001-4966. DOI: 10.1121/1.1336503. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1336503.

[153] Zilany, M. S. A. and Bruce, I. C., "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," *The Journal of the Acoustical Society of America*, vol. 120, no. 3, pp. 1446–1466, Sep. 2006, ISSN: 0001-4966. DOI: 10.1121/1.2225512. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.2225512.

[154] Zilany, M. S. A. and Bruce, I. C., "Representation of the vowel /ϵ/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats," *The Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 402–417, Jul. 2007, ISSN: 0001-4966. DOI: 10.1121/1.2735117. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.2735117.

[155] Zilany, M. S. A., Bruce, I. C., and Carney, L. H., "Updated parameters and expanded simulation options for a model of the auditory periphery," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 283–286, 2014, ISSN: 0001-4966. DOI: 10.1121/1.4837815. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.4837815.

[156] Zilany, M. S. A., Bruce, I. C., Nelson, P. C., and Carney, L. H., "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2390–2412, 2009, ISSN: 0001-4966. DOI: 10.1121/1.3238250. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.3238250.

VITA

Andres F. Llico Gallardo received his Ph.D. in Biomedical Engineering in 2021 from Purdue University, and his B.S. and M.S. degrees in Electrical Engineering from Universidad Tecnica Federico Santa Maria, Chile, in 2010 and 2013, respectively. He is an active member of the Association for Research in Otolaryngology and has presented his work at two national conferences. He was founding member of the Auditory Neuroscience Association at Purdue.