

Psychophysics-Based Electrode Selection
for Cochlear Implant Listeners

by

Sara Ingrid Durán

Department of Electrical and Computer Engineering
Duke University

Date: _____

Approved:

Leslie M. Collins, Supervisor

Lianne Cartee

Jeffrey L. Krolik

Loren W. Nolte

Douglas P. Nowacek

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Electrical and Computer Engineering
in the Graduate School of Duke University

2014

ABSTRACT

Psychophysics-Based Electrode Selection
for Cochlear Implant Listeners

by

Sara Ingrid Durán

Department of Electrical and Computer Engineering
Duke University

Date: _____

Approved:

Leslie M. Collins, Supervisor

Lianne Cartee

Jeffrey L. Krolik

Loren W. Nolte

Douglas P. Nowacek

An abstract of a dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Electrical and Computer Engineering
in the Graduate School of Duke University
2014

Copyright © 2014 by Sara Ingrid Durán
All rights reserved except the rights granted by the
Creative Commons Attribution-Noncommercial License

Abstract

Cochlear implant listeners are presented with a time and frequency-quantized version of speech signals. In the frequency domain, resolution is limited by the number of electrodes in each listener's array. Current cochlear implant speech processing strategies implicitly assume that the information presented to each one of these electrodes is perceived as unique and independent. However, previous research suggests that stimuli presented on different electrodes can be indiscriminable (e.g. [1], [2], [3]). Additional studies suggest that stimuli presented on one electrode can influence the perception of stimuli on neighboring electrodes (e.g. [4], [5], [6]). Removing this redundant or occluded information could cause more distinct or perceivable information to be presented to the listener and possibly result in improved speech recognition.

Previous studies have used psychophysical data to identify the electrodes with the highest potential to confound speech recognition [1], [6], [7]. In order to minimize electrode interactions and maximize the amount of perceivable information, each of these studies used a single psychophysical metric to deactivate the electrodes across all time windows of the speech processing strategy. For some listeners, these reduced electrode sets resulted in improved speech recognition over using the of the electrodes in their array. These studies did not compare the results of using different psychophysical metrics to exclude electrodes for a group of listeners nor did they investigate speech recognition performance as a function of the number of electrodes

excluded from the array.

In this work, three different psychophysical metrics were used to obtain a multi-dimensional estimate of the potential “usefulness” of each electrode. These results were then used to inform two different methods of psychophysics-motivated electrode selection. The first method incorporated individual data into each listener’s energy-driven speech processing strategy. For each time window, the electrodes with the highest energy that were also most likely to be perceived, according to the psychophysical data, were selected for stimulation. The second method sequentially excluded the electrodes with the highest potential to confound from the array across all time windows, resulting in a group of psychophysics-motivated electrode sets for each metric. Evaluating each of these electrode sets exhaustively would require a prohibitive amount of experimental time. To mitigate this problem, an adaptive procedure was developed to estimate performance as a function of cochlear implant parameters in a time-efficient manner. For each metric, the procedure estimated the set with the highest estimated probability of correct phoneme identification. Listeners’ speech recognition performance using this electrode set was then compared to their performance using their full electrode array. For both electrode selection methods, listeners’ speech recognition scores were generally comparable to those obtained in the clinical condition. This finding supports the hypothesis that listeners were not perceiving all the information presented to them using their clinical speech processing strategy and their complete set of electrodes. Additionally, these results suggest that improvements to the proposed electrode selection strategies should be investigated in order to increase the amount of perceivable information presented to cochlear implant listeners.

Contents

Abstract	iv
List of Tables	ix
List of Figures	x
List of Abbreviations and Symbols	xiii
Acknowledgements	xv
1 Introduction	1
2 Background	7
2.1 Cochlear Implants	7
2.2 Dropping Electrodes According to their Psychophysical Measures . .	11
2.2.1 Electrode Discrimination	12
2.2.2 Forward Masking	13
2.2.3 Modulation Detection	15
2.3 Motivation	17
3 Psychophysical Measures of An Electrode’s Potential to Confound Speech Recognition	19
3.1 Subjects	19
3.2 Stimuli and Equipment	20
3.3 Threshold and Maximum Comfortable Loudness Measurements . . .	20
3.4 Experimental Tasks	22

3.4.1	Forward Masking Task	23
3.4.2	Electrode Discrimination Task	25
3.4.3	Modulation Detection Task	26
3.5	Results	29
3.6	Discussion	39
4	Dynamic Electrode Selection Using Psychophysical Measures	43
4.1	Dynamic Electrode Selection Algorithms	44
4.1.1	Electrode Discrimination Based Selection	45
4.1.2	Forward Masking Discrimination Based Selection	46
4.2	Experimental Task	50
4.2.1	Subjects	50
4.2.2	Stimuli and Equipment	50
4.2.3	Procedure	51
4.2.4	Speech Recognition Task	52
4.2.5	Preference Task	53
4.3	Results	53
4.4	Discussion	61
5	Adaptive Parameter Selection Procedure	66
5.1	Parameter Estimation	67
5.2	Sample Placement	70
5.2.1	Curve Fit Mode	70
5.2.2	Maximum Performance Find Mode	73
5.3	Adaptive Procedure Operation	76
5.4	Verification Using Simulated Data	76
5.5	Listener Verification Task	79

5.5.1	Subjects	80
5.5.2	Stimuli and Equipment	80
5.5.3	Procedure	81
5.6	Results	82
5.7	Discussion	84
6	Static Electrode Selection Using Psychophysical Measures	91
6.1	Subjects	92
6.2	Stimuli and Equipment	92
6.3	Procedure	94
6.4	Results	95
6.5	Discussion	97
7	Conclusions	103
	Bibliography	108
	Biography	113

List of Tables

3.1	Demographic information for implanted subjects.	21
3.2	Correlation between normalized psychophysical metrics.	39

List of Figures

2.1	Diagram (not to scale) of the human ear	8
2.2	Position of maximum vibration on the basilar membrane	8
2.3	Diagram of an implanted ear.	9
2.4	Block diagram of the ACE processing strategy.	10
2.5	Example pulse train (not to scale).	11
2.6	Synthetic forward masking pattern for a masker presented on electrode 10.	14
3.1	Graphical user interface used for forward masking task.	24
3.2	Stimuli presented during forward masking task (not to scale).	24
3.3	Pulse width-modulated pulse trains used for modulation detection task.	27
3.4	Effect of modulation depth on the value pulse width as a function of time.	28
3.5	Threshold shift values for the forward masking task.	30
3.6	Example threshold shift values, Gaussian curve fit and area under the masking curve for S4.	31
3.7	Gaussian fits to the threshold shift values for the forward masking task.	32
3.8	Area under the masking curve results for the forward masking task. . .	33
3.9	Results for the electrode discrimination task.	34
3.10	Average distance to the closest discriminable electrode results.	35
3.11	Results for the modulation detection task.	37

3.12	Normalized electrode discrimination, forward masking and modulation detection threshold data.	38
4.1	Electrode discrimination based-electrode selection.	47
4.2	Forward masking based-electrode selection.	48
4.3	Graphical user interface used for psychophysics-based N-of-M speech recognition task.	52
4.4	Graphical user interface used for psychophysics-based N-of-M preference task.	54
4.5	Speech recognition results for psychophysics-based dynamic electrode selection.	56
4.6	Preference results for psychophysics-based dynamic electrode selection versus the clinical selection algorithm.	58
4.7	Number of pulses presented on each electrode for dynamic electrode selection and clinical algorithms in quiet, noise and reverberation.	60
4.8	Cluster lengths across all words presented in quiet, noise and reverberation.	61
5.1	Graphical representation of the adaptive procedure's operation.	77
5.2	Simulated curve fit and maximum performance find results.	79
5.3	Statistics for curve fit and maximum performance find adaptive procedures across 100 repetitions.	87
5.4	Electrode sets used for adaptive procedure verification task.	88
5.5	Adaptive electrode selection procedure verification task results.	89
5.6	Adaptive procedure verification task results for even and exhaustive sampling.	90
6.1	Static psychophysics-motivated electrode sets.	96
6.2	Speech recognition results for the clinical condition, the ED,FM and MDT electrode sets selected by the adaptive procedure.	97
6.3	Adaptive electrode selection procedure outcomes for S1.	98
6.4	Adaptive electrode selection procedure outcomes for S2.	99
6.5	Adaptive electrode selection procedure outcomes for S3.	100

6.6 Adaptive electrode selection procedure outcomes for S4. 101

List of Abbreviations and Symbols

CI	Cochlear implant
CL	Current level
CNC	Consonant-nucleus-consonant
DR	Dynamic Range
ED	Electrode Discrimination
e	Rlectrode
FM	Forward Masking
$f_p(p)$	Probability density function of random variable p .
$F_p(p)$	Cumulative distribution function of random variable p .
GUI	Graphical user interface
i	Iteration index
k	Number speech units
KLD	KullbackLeibler divergence
M	Number of alternatives in a forced choice task
$M(e)$	Forward masking threshold
N	Number of electrodes selected per time window
n	Parameter set index
m	Modulation depth
MCL	Maximum comfortable loudness
MD	Modulation detection

MDT	Modulation detection threshold
NH	Normal hearing
p_n	Estimated probability of discrimination for parameter set n
p_n^t	True underlying probability of discrimination for parameter set n
PC	Personal computer
PW	Pulse width
r	Number of correct responses out of k speech units presented
RT_{60}	Reverberation time, time required for sound level to drop 60 dB from the onset level
s	Number of distinct parameter sets
S	Collection of sets of indiscriminable electrodes. S_i
S_i	Set of indiscriminable electrodes determined using e_i as a reference.
SNR	Signal-to-noise ratio
SSN	Speech-shaped noise
T	Threshold
T_{shift}	Threshold shift

Acknowledgements

If there's one thing I've learned over the past six years is that I couldn't have possibly made it this far alone. There are many people I must thank for helping me get to and cross the finish line. So, grab a drink and pull up a chair because this is going to take a while.

This work wouldn't have been possible without the generosity and dedication of the research subjects who gave their time and their ears to further the development of cochlear implants. Thank you for sitting through all those testing sessions, listening to beeps and corrupted speech, while maintaining your smiles and a great dispositions. It has been a privilege to work with you professionally and an honor to get to know you personally.

I would also like to thank the organizations that funded my research and allowed me to live a comfortable graduate student life: the Duke University Endowment, James B. Duke 100th University Fund, the National Institutes of Health, and the National Science Foundation Graduate Research Fellowship Program.

I must express my gratitude to the inspiring women responsible for securing said funding and advising me throughout this process, Leslie Collins and Sandy Throckmorton. Leslie, if not for your persuasive Southern charm I probably wouldn't have started a Ph.D. program in the first place. During my time here, you've provided me with everything I could have ever wanted: a well-equipped laboratory, a cool and smart group of colleagues, the nicest office a graduate student could hope for, and

permission to maintain a healthy work-life balance. You've created an environment where I could have big ideas, pursue them freely and feel comfortable enough to be myself, that is, swear when things didn't work out. Thank you for all your support, advice, and for keeping me laughing. Sandy, I couldn't have possibly finished this work without your guidance. Thank you for answering an infinite number of my knocks on your door with a cheerful "yees?" and thoughtfully answering just as many of my questions. Of everyone I've encountered in the last six years, you've taught me the most about how to conduct research and present my findings ("make your own figures, think about your transitions, explain the process with a flow chart, 'Gopenize' your writing[...]!"). I'm going to miss bouncing ideas off of you tremendously. Thank you both for making me feel like I could be honest with you when I was struggling. I'll forever be grateful that instead of kicking me to the curb you leaned into me even harder.

While being advised by Leslie and Sandy, I've been very fortunate to learn from a knowledgeable group of research scientists: Stacy Tantum, Pete Torrione, Kenny Morton and Mary Knox. Stacy, thank you for bringing the same intense engagement and attention to detail to our discussions about my research as you bring to your own work. Pete and Kenny, thank you for explaining your Bayesian ways in a language I could understand. And, in retrospect, I'm glad you occasionally rained on my parade; it only made me want to work harder.

My time at Duke wouldn't have been nearly as enjoyable without the company of my fellow SSPACISS students, both past and present: Josh Stohl, Phil Brown, Kyle Bradbury, Chris Ratto, Jill Desmond, Ken Colwell, Achut Manandhar, Rayn Sakaguchi, Jordan Malof, Boyla Mainsah, Patrick Wang, Kedar Prabhudesai, Dima Kalika, Nick Czarnek, Joe Camilo, Jillian Clements, Daniël Reichman. Josh and Phil, thank you for mentoring me when I first arrived. I know you both had more important things to do but you always made time to kindly answer all of my questions. Chris,

thank you for being the perfect example of how undramatic getting a Ph.D. can be and for teaching me how to love America. Kyle, thank you for infecting me with your joy and for being my loyal theater companion. Jill, thank you for being a great office mate, for keeping me caffeinated, and for listening to me rant whenever I needed to. Ken, thank you for making me laugh loudly and often with your random emails, pranks and googly-eye activism. Rayn, thank you for teaching me how to make a killer loaf of bread and for showing me the ways of bike commuting so I could work off all those carbohydrates. Patrick, thank you for solving so many of my technical issues without making me feel completely inept. Achut, Jordan, Patrick, Boyla, Kedar, Dima, Nick, Joe, Jillian and Daniel, thank you for your insightful feedback on my work, for letting ramble on and on when I needed break and for being the most fun group of colleagues I could ask for.

Outside of Duke, I've had the pleasure of making many other friends in Durham: Allison Ratto, Alicia Morelli, Zana Devlin, Gillian Locke, Beth Perry, Jo Supernaw, Donya Smith, Elizabeth Schuler, Kevin McCabe, Emily and Steve Anderson, Todd Quenneville, and Ryan Melsert. Thank you all for listening to me gripe, for cheering me on and for knowing when I needed a good meal or a stiff drink.

I also must thank my dear friend Ana Martín-Romo who has supported me from across the Atlantic. Reina, gracias por hacerme sentir tan cerca de ti y de la tierra patria aún estando tan lejos.

I'm very lucky to have an amazing family who have always helped accomplish anything I set my mind to. My maternal grandmother, Olive Kuniholm, started the engineering recruitment process early by taking me on tours of Worcester Polytechnic Institute when I was twelve. Years later she sent me to a summer program there and when I came back I knew I wanted to be an engineer. Grammie, thank you for your encouragement, for providing me with the resources I needed to achieve my goals and for never thinking any of them were too lofty. Wherever you are, I hope you can

see me when I don the wizard robes. My paternal grandmother, Sara Puebla, has taught me the value of hard work and straight talk. Abuela, gracias por enseñarme a trabajar duro, siempre con la cabeza bien alta y la lengua afilada.

My parents, Juan Ramón Durán and Donna Kuniholm, instilled the importance of education in me at a very early age. They've always been an invaluable source of support, encouragement and love. Papa, your confidence in me has always been so unlimited I often thought it was misguided. But I guess you were right: I made it! Mom, thanks for being the parent more rooted in reality. I've always felt that no matter what issue I came to you with you would listen to me with an open mind and tell me what I needed, not what I wanted, to hear. I couldn't have made it to this point without the considerable financial and emotional sacrifices you've both made to ensure my success. Thank you for giving me the courage to move far away, to set my sights high, and to believe I can have it all.

I'd also like to thank my partner's parents, Bob Vaillancourt and Jackie Nowell. Thank you for all the pep talks and for celebrating my successes as if I were one of your own.

Finally, I owe the greatest debt of gratitude to my partner, Austin Vaillancourt. Shortly after I started at Duke, he left the security of his friends, family and a stable job in Connecticut to come live an uncertain North Carolinian adventure with me. For the first couple years he had to work long hours, nights, and weekends, and he did so without ever complaining or resenting me. Austin, I know you're not one for keeping score but I owe you a big one. Thank you for being calm when I was insane, strong when I was weak, for being a believer when all I had was doubt, and for loving me when I was most unpleasant. Without you as my motivator, I'm certain I would've given up years ago. Since I can't give you the part of this degree I feel you deserve, I bestow unto you a doctorate of unwavering support, diffusion of stress, and selfless dedication.

Introduction

Most cochlear implant (CI) listeners can achieve high speech recognition scores in benign listening conditions. As listening conditions become more challenging, speech recognition is adversely affected, in part due to device limitations. In the time domain, the sampling rate of the implant is limited by the rate at which the device can process and transmit incoming speech as well as the physiological impact of sampling rate (e.g. [8], [9], [10]). In the frequency domain, CI users are only presented with information at discrete locations along the cochlea. Due to limits in size and selectivity of physical electrodes, this means that, at best, listeners can resolve as many distinct pitches as there are electrodes along the array. Methods of increasing the amount of temporal and frequency information provided to CI listeners, such as increasing the pulse rate to more finely sample the envelope or steering the current to increase the number of pitch percepts, have had limited success in improving speech recognition (e.g. [11], [12]).

An alternate approach to increase the amount of information CI listeners perceive is to minimize the amount of confounding or redundant stimuli. Speech processing algorithms assume stimuli presented on different electrodes are perceived as unique

and independent. However, previous work suggests that stimuli presented on different electrodes can be indiscriminable (e.g. [1], [2], [3]). Additional studies indicate that stimuli presented on one electrode can influence the perception of stimuli on neighboring electrodes (e.g. [4], [5], [6]). The physiological conditions that can cause these electrode interactions can be estimated using psychophysical data. These predictions of the conditions around each stimulation site can be used to assess the potentially positive or negative impact of each electrode on speech understanding. Previous studies gathered a single type of psychophysical data – such as a measure of electrode discrimination, modulation detection or forward masking– and used it to identify electrodes to be excluded from listeners’ signal processing algorithms [1], [6], [7]. When compared to performance using listeners’ full electrode set, speech recognition scores for some subjects improved significantly when using the reduced set of electrodes. This outcome suggests that presenting less information that can be better perceived, by using psychophysical data to guide electrode selection, can result in better speech understanding for CI listeners.

Previous studies excluded electrodes based on a single psychophysical measure [1] [6] [7]. Assuming that each measure can estimate different underlying physiological conditions, gathering a variety of these measures may allow for a more complete characterization of each electrode’s potential to confound. In this work, three different psychophysical measures were collected for every electrode in the array: electrode discrimination, forward masking and modulation detection. Electrode discrimination measures indicated if different information was presented on the same perceptual channel (e.g. [8]). Forward masking measurements indicated when presenting a stimulus on one electrode impeded the perception of stimuli subsequently presented to adjacent electrodes (e.g. [5],[2]) . Lastly, measuring sensitivity to modulation provided a predictor of the temporal acuity of the nerve fibers closest to each electrode (e.g. [13], [14]). The combination of these three measures provided a multidimen-

sional estimate of each electrode’s potential to confound speech recognition across a group of listeners.

Previous psychophysics-based electrode selection methods were limited in scope. Boëx et al. [6] only studied the forward masking effect of apical electrodes. Garadat et al. [7] compared listeners’ performance using the electrodes with the highest and lowest sensitivity to modulation, however, performance was not compared to that of the full electrode array. Additionally, none of these studies considered selecting electrodes for each time window or the impact of varying the number of electrodes excluded from the array for each listener [1] [6] [7]. This work investigates these unanswered questions, proposing two different psychophysics-based electrode selection methods. The first method is a modification to the energy-driven speech processing strategy most CI listener’s current use. The proposed strategy uses each individual’s psychophysical data to maximize the amount of perceivable information presented for each time window. The second method first forms a group of psychophysics-motivated electrode sets by sequentially excluding electrodes with the highest potential to negatively impact speech recognition. These electrodes were excluded from the array across all time windows. An adaptive procedure was then used to select the psychophysics-based electrode set that results in the highest estimated speech recognition scores.

In order to maximize the amount of information listeners perceived, electrodes were selected dynamically using a psychophysics-based speech processing strategy. For each time window, the electrodes with the highest energy that were also most likely to be perceivable, as determined previously collected electrode discrimination and forward masking data, were selected for stimulation. Previous speech processing strategies have attempted to maximize the amount of information being perceived within a given time window by selecting electrodes based on across-channel interactions estimated from normal hearing listeners or by limiting the number of adjacent

electrodes stimulated ([15], [16]). In these studies, the same processing strategy was implemented across all listeners, regardless of potential underlying differences in electrode placement, spread of excitation or other physiological factors. In this work, electrode selection was based on subject-specific psychophysical data. It was hypothesized that this individual approach would increase speech recognition by preserving the energy-driven selection of listeners' everyday processors while maximizing the amount of information each listener perceived. Because no electrodes were permanently deactivated, the frequencies corresponding to each electrode remained unchanged thereby avoiding any potential negative impact of frequency reallocation (e.g. [17], [18], [19]).

In addition to using psychophysical data to select electrodes for each time window, electrodes were also selected across all time windows using a static selection method. It was hypothesized that excluding the electrodes would the highest potential to confound speech recognition would allow electrodes that present more distinct or perceivable information to be selected more frequently, thereby increasing the amount of information available to the listener. For this task, the previously collected electrode discrimination, forward masking and modulation detection data were used to determine three groups or perceptually relevant electrode sets. For each metric the unique electrode sets were determined by sequentially excluding the electrode with the highest potential to negatively impact speech recognition thereby progressively increasing the aggressiveness of the psychophysics-based selection method. This resulted in a large number of electrode sets to consider. Because evaluating each electrode set individually would require a prohibitive amount of time, a time-efficient adaptive procedure was used to determine the set with the highest estimated performance for each metric.

Previous research indicates that the speech processor parameters that maximize speech recognition are subject dependent [20], [21], [22]. Exhaustively evaluating

listeners performance as a function of these parameters to find those that maximize speech recognition can be very time consuming. To allow for more efficient parameter selection, several strategies have been implemented for CIs [23] and hearing aids (e.g. [24], [25]). These strategies attempted to maximize performance over a range of parameters using a genetic algorithm that used listener's subjective evaluation of speech as a measure of fitness. In contrast, this work proposes an adaptive procedure driven by listener's objective speech recognition. For each iteration a speech token is presented to the listener using one of the parameter sets considered and the response is then graded by the experimenter. Given the previously collected responses for all parameter sets, the procedure decides which set to use to present the next word based on one of two different operation modes. The first method, termed curve fit mode, samples the parameter space to obtain equal confidence estimates across conditions. This mode allows for a performance comparison across parameters that is difficult to obtain with a genetic algorithm. The second mode of operation, termed find maximum performance mode, samples the parameter space to determine the parameter that results in the highest speech recognition. It presents the next word using the parameter that is most likely to change the maximum performance estimate. The proposed adaptive procedure a time-efficient method of optimizing any of the cochlear implant speech processing parameters. In this work, the adaptive process was used to estimate the performance of each psychophysics-motivated electrode set in order to find the set with the highest probability of outperforming the clinical set.

The remainder of the document is organized as follows. Chapter 2 includes background information experiments. Chapter 3 presents the the results of collecting three three different psychophysical metrics across the electrode array. Chapter 4 presents the results of using two of the psychophysical metrics to dynamically select the electrodes with the highest energy that also have the highest likelihood

of being perceived. Chapter 5 describes the adaptive parameter selection procedure and includes two verification tasks, one using simulated data and another conducted with CI listeners. Chapter 6 describes the method for selecting electrodes across all time windows using the previously collected psychophysical data. Additionally, it includes the results of using the adaptive procedure was used to select the psychophysics-motivated electrode set with the highest estimated speech recognition. Lastly, Chapter 7 summarizes the contributions of the aforementioned studies.

2

Background

2.1 Cochlear Implants

Cochlear implants are auditory prostheses for individuals whose hearing loss is due to an impairment in the mechanical to electrical transduction of stimuli, but for whom the neural population is still intact (Figure 2.1). Inner hair cell demise is the most common condition leading to implantation. Hair cells are responsible for translating the mechanical vibrations of the basilar membrane, located inside the cochlea, into neural information. When bent by the movement of the basilar membrane, they release neurotransmitters that cause neurons to fire. These neural firings are relayed to the brain via the auditory nerve, resulting in the perception of sound. The apex vibrates maximally at low frequencies and the base does so at high frequencies (see Figure 2.2). The basilar membrane can thus be thought of as a frequency analyzer. This tonotopic arrangement is maintained at higher levels of the auditory system.

The death of hair cells causes the loss of this mechanical to chemical transduction and is one of the most common causes of deafness. Sensorineural hearing loss can have a variety of causes, including but not limited to aging, toxicity caused by

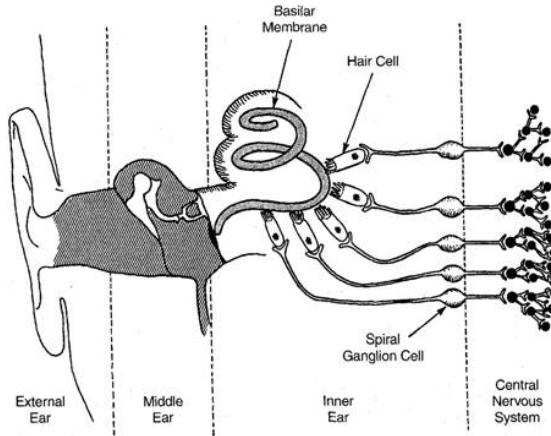


FIGURE 2.1: Diagram of the human ear, not to scale. The trajectory of sound, from the external ear to the central nervous system, is shown. Image reprinted from [26] with permission of the IEEE. ©1988 IEEE.

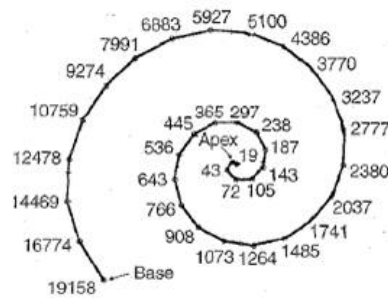


FIGURE 2.2: The position of maximum vibration on the basilar membrane in response to sinusoids at indicated frequencies is indicated from low, at the apex, to high, at the base. Image reprinted from [27] with permission of the IEEE. ©1999 IEEE.

certain medications, autoimmune illness, genetics or physical trauma. The purpose of a cochlear implant is to replace the function of damaged hair cells and stimulate the nerve fibers directly. A diagram of an implanted ear is provided in Figure 2.3. The speech processor is typically located in a small, hook-shaped package that sits behind the ear, shown in Figure 2.3. The speech processor specifies which electrodes to stimulate depending on the spectral and temporal characteristics of the sound

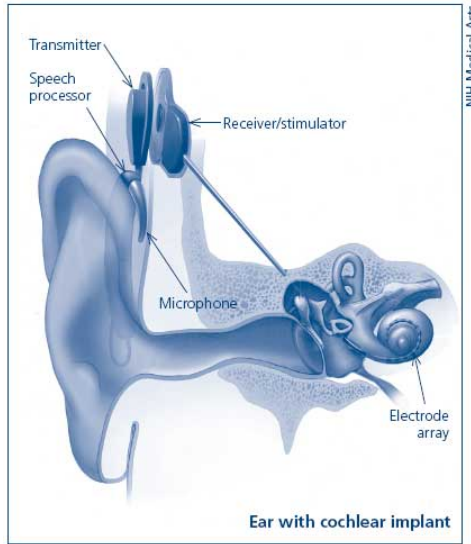


FIGURE 2.3: Diagram of a typical cochlear implant system. The sound impinges on the microphone located on the speech processor behind the ear. The processed signal is transmitted transcutaneously via radio frequency to the internal part of the device. This internal receiver/transmitter device sends the signal to the implanted electrode array which directly stimulates the auditory nerve fibers. Image source: Medical illustrations by NIH, Medical Arts and Photography Branch.

arriving at the microphone and on the speech processing strategy of the particular device. This information travels from the external processor to the external antenna which in turn transmits the information to the internal part of the device via radio frequency communication. The internal receiver/transmitter of the implant transmits this information to the electrode array which stimulates the auditory nerve fibers directly.

As previously stated, it is the speech processing strategy that determines how the electrodes are stimulated given a specific incoming sound. A variety of strategies have been developed and their use depends on the device manufacturers. The subjects in this study used the Advanced Combinational Encoder (ACE) strategy [28] [29]. ACE is a variation on both the Continuous Interleaved Sampling (CIS) strategy [30] and the Spectral Maxima Sound Processor (SMSP) strategy [31]. Figure 2.4 is a

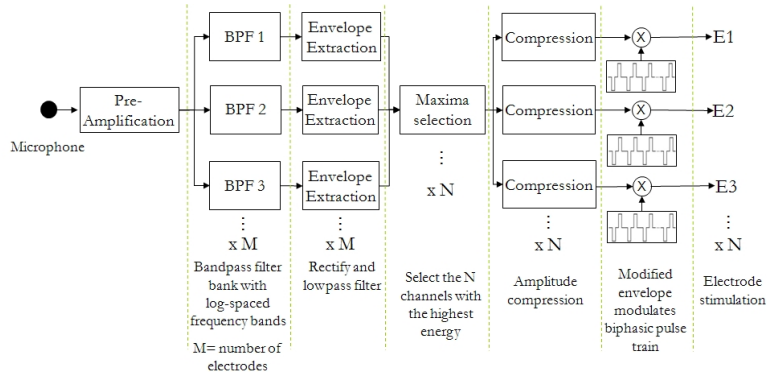


FIGURE 2.4: Block diagram the Advanced Combinational Encoder (ACE) strategy. Each processing stage is indicated.

diagram indicating each of the algorithm’s processing stages. For the purposes of this document only the ACE algorithm will be discussed.

In ACE, the audio signal arrives at the microphone and is then pre-amplified to simulate a high-pass filtering process that the human auditory system performs naturally [32]. This signal is then sent through a bank of bandpass filters which approximates the tonotopic arrangement of the cochlea. Each filter corresponds to an electrode: lower frequency brands correspond to high numbered electrodes located towards the apex and higher frequency bands correspond to lower numbered electrodes located towards the base. The frequency range of the bank of filters approximately spans 200 Hz to 8000 Hz. The lower limit was chosen to capture the lower frequency components of speech, such as the information encoded in vowels [33], while the upper limit is bound by the sampling frequency of the codec. The envelope of the signal in each channel is calculated by rectifying and low pass filtering the signal. Once the envelope has been extracted, the energy content of each channel is calculated for a given analysis window. The N highest energy channels out of the M available channels are selected for stimulation, in what is referred to as an N-of-M strategy. Before the electrodes are stimulated, amplitude compression is necessary to account for the reduced dynamic range in electric hearing compared to normal

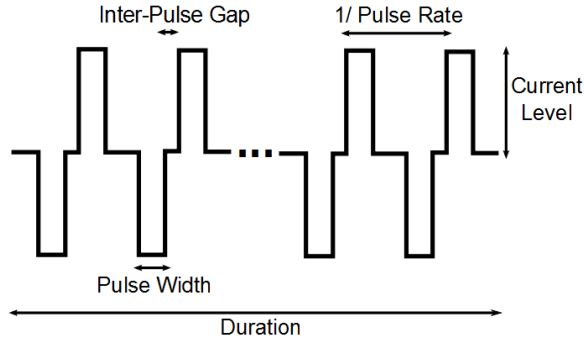


FIGURE 2.5: Example pulse train. Duration, pulse width, inter-pulse gap and current level are indicated.

hearing [34]. The compressed envelope corresponding to each channel is then used to amplitude modulate a biphasic pulse train. A schematic of a typical pulse train and the parameters that characterize it is included in Figure 2.5. These modulated pulse trains are sent to the electrodes in an interleaved fashion such that only one electrode is stimulated at a given time in order to minimize interactions between electrodes [30].

The collection of all the stimulation parameters for a particular CI user is referred to as a map. The representation of speech signals presented to CI listeners is defined by the map parameters, e.g. electrodes used for stimulation, first hearing and maximum comfortable stimulation levels for each of the electrodes, number of maxima selected for each time window, pulse rate, pulse width, and inter-pulse gap.

2.2 Dropping Electrodes According to their Psychophysical Measures

The stimuli presented to CI users are much lower in temporal and frequency resolution than the stimuli NH listeners receive acoustically. Even though they are presented with a reduced amount of information, CI users are not always able to perceive all temporal and frequency cues present in the electrically-coded stimuli. Depending on the physiological characteristics that define each electrode site, the presented

information could be perceived as redundant or confounding. Several studies have attempted to mitigate these effects by selecting electrodes for stimulation according to electrode-specific psychophysical metrics such as electrode discrimination (ED), forward masking (FM) and modulation detection thresholds (MDTs).

2.2.1 Electrode Discrimination

The first study of this kind examined the relationship between ED and speech recognition [1]. To measure how discriminable an electrode was from its neighbors, loudness-balanced pulse trains were presented on a reference and a target electrode. The distance between the reference and the target was varied adaptively until listeners could discriminate between stimuli presented on the two electrodes. Zwolan et al. [1] used this data to select only discriminable electrodes to include in listeners' maps. They hypothesized each of these electrodes would generate a unique percept, primarily in the pitch dimension. The underlying hypothesis is that discriminable electrodes stimulate unique populations of nerve fibers. The construction of the experimental maps started by including the most basal electrode in the map, for example electrode 1. Then the closest discriminable electrode to electrode 1 in the apical direction would also be included in the map. This sequential selection process continued across the array until the reference electrode was not discriminable from any of the electrodes in the apical direction. Consequently, the number of electrodes included in each subject's map depended on each individual's ED pattern. For both words and sentences presented in quiet, performance using subject-specific experimental maps was compared to performance using each subject's clinical map. Of the subjects who could not perfectly discriminate all electrodes across the array, more than half performed significantly better for at least one type of speech material when listening with their experimental map. This result suggests the potential benefits of maps tuned to a subject's perception.

2.2.2 Forward Masking

While electrode discriminability is primarily a frequency domain interaction, an overlap in nerve fiber populations can also be a confounding factor in the time domain. FM occurs when stimulation on one electrode, referred to as the masker, causes an elevation of the thresholds of subsequently stimulated electrodes, or probes. Elevated thresholds can result in the stimuli presented on the probe electrode not being perceived. FM effects occur when a stimulus presented on the masker electrode stimulates a subset of the nerve fiber population corresponding to the probe electrode [8]. The nerve fibers corresponding to the probe enter an absolute or relative refractory period, during which an action potential cannot be produced or a higher intensity stimulus is needed to produce an action potential (e.g. [35]). The greater the FM effect an electrode has on its neighboring electrodes, the more likely stimuli presented on the masker are to obscure the perception of subsequent stimuli presented on adjacent electrodes which may adversely affect speech recognition [2]. One way to measure the effect of the masker on a particular probe electrode is to compare the threshold (T), the level of first hearing, for the probe alone to the T measured after presenting a stimulus on the masker electrode. The difference between the probe alone T and the T measured after presenting the masker is called the T_{shift} . Larger values of T_{shift} indicate greater FM. The effect of a given masker on all electrodes across the array can be visualized with a masking pattern. Figure 2.6 includes an example masking pattern using synthetic data for a masker presented on electrode 10. The larger T_{shift} values for electrodes closer to electrode 10 indicate a greater effect of the masker on those sites. The Gaussian fit to the threshold shifts approximates the extent of the masking pattern, assuming that threshold shifts declines monotonically as the distance between masker and probe increases. The extent of effect of a given electrode on neighboring probes can be quantified by calculating

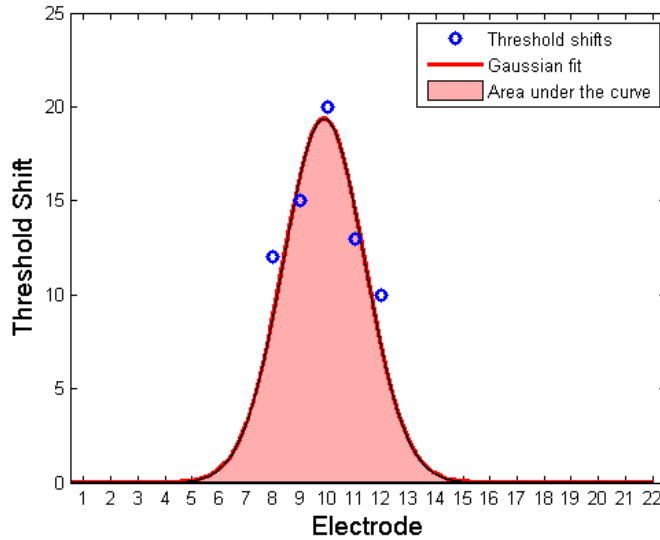


FIGURE 2.6: Synthetic forward masking pattern for a masker presented on electrode 10. Threshold shift is shown as a function of probe electrode. The Gaussian fit to the threshold shift values and the area under the masking curve are also indicated.

area under the estimated masking masking pattern.

Throckmorton and Collins [2] investigated the relationship between ED, FM, and speech recognition. In their experiment, FM patterns were measured for an electrode located in the center of the array. For the same subjects, the number of discriminable electrode pairs was determined. Subjects then participated in a speech recognition task using different types of speech material: consonants, vowels, words and sentences. For some types of speech material, performance was significantly correlated with both the number of discriminable electrode pairs and the average amount of FM. However, for other types of speech material performance was only correlated with one of these psychophysical metrics. These results suggest that the correlation between FM and speech recognition may be due to a combination of frequency and temporal interactions.

Boëx et al. [6] investigated the effect of excluding electrodes based on FM data. Electrode 1, located at the basal end of the array, was selected as a masker and probe

thresholds were measured for several other electrodes across the array. Four subjects presenting two different FM trends participated in a speech recognition task. For two of the subjects, T_{shift} values measured on electrodes 1 and 2 were large and close in magnitude, suggesting that the two electrodes were stimulating overlapping populations of nerve fibers. For the other two subjects, T_{shift} values measured on electrodes 1 and 2 were small, suggesting the 2 electrodes stimulate distinct groups of nerve fibers. For each subject, two experimental maps were constructed: one excluded electrode 1 (7-electrode map) and the other excluded both electrodes 1 and 2 (6-electrode map). It was hypothesized that for the subjects with higher levels of FM caused by electrode 1, eliminating these electrodes would allow the stimuli presented on neighboring electrodes to be more easily perceived. Subjects' speech recognition was evaluated after using the experimental maps on a daily basis for two weeks. The performance using these experimental maps was compared to each subjects clinical map using all 8 electrodes. As expected, for the subjects with smaller levels of FM, performance worsened when using the experimental maps. The subjects with higher FM levels performed better with the experimental maps. One of these listeners performed better using the 7-electrode strategy, although the improvement was not statistically significant. The other subject performed significantly better using the 6-electrode map. Both subjects decided to permanently adopt the experimental maps.

2.2.3 Modulation Detection

In a more recent study, Garadat et al. [7] investigated the effects of both masking and sensitivity to changes in modulation depth. Since speech is coded as a set of modulated biphasic pulse trains, sensitivity to changes in modulation of pulse trains is considered a potential predictor of speech recognition in CI listeners. Lower MDTs indicate higher sensitivity to these changes and are therefore more desirable. Garadat et al. [7] measured both MDTs and masked MDTs. Masked MDTs were collected

while an interleaved pulse train was presented on an adjacent apical electrode. The authors hypothesized that masked MDTs measure both the temporal and spatial sensitivity for a particular site, using stimuli which are more representative of speech signals. As expected, the masked MDTs were, on average, higher than the unmasked MDTs. However, the masked and unmasked MDTs were strongly correlated. To create the experimental maps, the electrode array was first divided into five sections, each with approximately the same number of electrodes. Each subject was assigned two experimental maps, each with 10 electrodes; one included the 2 electrodes with the highest masked MDTs from each section and the other included the 2 electrodes with the lowest masked MDTs for each section. Sentence recognition for the two experimental maps was then compared. On average, subjects performed better using the map including the electrodes with the lower masked MDTs in both quiet and in noise. Results from this experiment suggest that selecting sites with better temporal acuity may result in better perception of the amplitude modulations that encode speech. However, this experiment does not compare performance using the map with lower masked MDTs to performance using subjects' clinical map.

Garadat et al. [7] used the same frequency table for both experimental maps and assigned each of the frequency bands to the included electrodes in tonotopic order. The studies that dropped electrodes using ED and FM data did not explicitly state which method was used to reassign the frequencies corresponding to the dropped electrodes [1] [6]. The relationship between the frequency tables assigned to the electrodes in the experimental map and the frequency allocations subjects are used to listening to through their clinical map could affect speech recognition results using the experimental map. For this reason, it is important to reallocate the frequency band corresponding to the dropped electrodes using a method that minimizes the potentially detrimental effect said reallocation may have on speech recognition.

2.3 Motivation

The aforementioned studies have shown that selecting electrodes according to the potential to confound speech recognition predicted by psychophysical measures can improve speech recognition for some CI listeners [1] [6] [7]. However, each of these studies selected electrodes using a single psychophysical metric. Assuming that each psychophysical metric can predict different underlying physiological conditions, it is possible that the metric that is best suited for electrode selection may vary across listeners. This hypothesis was investigated in this work by collecting three different psychophysical metrics for each electrode: ED, FM and MDT. Together, these three metrics provided multidimensional estimate of each electrode’s “usefulness”.

Further, each of these psychophysics-based electrode selections strategies only used a single method for selecting electrodes across all time windows of the speech processing strategy [1] [6] [7]. These studies did not investigate the effect of selecting electrodes for each time window or the impact of varying numbers of selected electrodes. In contrast, this work explores two different methods for selecting electrodes using psychophysical data. The first method is a modification of listener’s clinical speech processing strategy. For each time window, the electrodes with the highest energy that also are most likely to present perceivable or distinct information are selected for stimulation. The second method selects electrodes across all time windows. For each psychophysical metric, a group of electrode sets is determined by sequentially excluding the electrode with the highest potential to confound speech recognition. This resulted in a large number of electrode sets. Ideally, each electrode set would be evaluated exhaustively to find the set resulting in the highest speech recognition score. However, this would require a prohibitive amount of experimental time. To make selection more tractable, an adaptive procedure was developed to estimate listeners’ speech recognition as function of each psychophysics-motivated

electrode set.

Psychophysical Measures of An Electrode's Potential to Confound Speech Recognition

Previous studies have investigated the use of different psychophysical metrics to determine electrodes that could potentially present confounding information: ED, FM, and MDT [1] [6] [7]. In this experiment, all three measures were collected from the same subjects. The ED task measured the similarities in perception associated with stimulation on neighboring electrodes. The MDT task measured the sensitivity of each electrode to changes in loudness modulation, similar to the modulations of pulse trains used to encode speech. Lastly, the FM task measured the impact that stimulating one electrode had on the perception of stimuli presented on neighboring electrodes. Measuring these three psychophysical outcomes provides a multi-dimensional characterization of each electrode and its potential to confound.

3.1 Subjects

Psychophysical data was collected for four postlingually deaf subjects. Table 3.1 contains demographic information for the four subjects who participated in the study.

All subjects were users of Cochlear Corporation’s CI24 family of devices and used monopolar 1+2 (MP1+2) stimulation mode, where both extra-cochlear electrodes (numbered 1 and 2) were used as ground for stimulation. The experiment was completed in six to eight sessions lasting three to four hours each. All subjects were compensated for their participation. The use of human subjects in the experiments described in the following sections was approved by the Institutional Review Board of Duke University.

3.2 Stimuli and Equipment

All stimuli consisted of biphasic pulse trains with $8\mu s$ inter-phase gaps. The pulse width was set to $25\mu s$, as specified by subjects’ clinical maps, in the FM and ED experiments. For the MDT task, the pulse width was sinusoidally modulated at 10 Hz around a mean pulse width (PW) of $50\mu s$. The minimum and maximum pulse widths were 25 and $75\mu s$, respectively, determined by the device’s limitations. For all testing, stimulus pulse trains were created using each subjects’ clinical stimulation rate (see Table 3.1). The Nucleus Implant Communicator (NIC v2) was used to stream all stimuli from a PC. Custom designed graphical user interfaces (GUIs) designed in MATLAB were used for all of the psychophysical tasks.

3.3 Threshold and Maximum Comfortable Loudness Measurements

Prior to testing, each subjects’ levels of first hearing or thresholds (Ts) and maximum comfortable loudness levels (Cs) were measured under three different conditions. First, Ts and Cs were measured using pulse trains presented at subjects’ clinical pulse width with a $300ms$ duration. This duration was selected because, for most CI users, increasing stimulus duration beyond $300ms$ does not result in an increase in loudness [8]. The inter-stimulus duration was $500ms$. These T and C levels were referred to as the clinical levels. Secondly, T levels were measured using pulse trains

Table 3.1: Demographic information for implanted subjects

Subject ID	Gender	Age (years)	Age at onset of deafness (years)	Age at implantation (years)	Implant type	Clinical Pulse Rate (pps)	Number of Maxima
S1	M	59	48	49	CI24R	900	8
S2	F	50	10	41	CI24RE	1200	12
S3	M	61	15	53	CI24RE	900	10
S4	M	74	15	67	CI24RE	900	8

with a duration of $30ms$. These levels were used as a reference point for measuring threshold shifts in the FM task. Lastly, Ts and Cs were measured using $1s$ -long pulse trains with pulse widths equal to $50\mu s$. These stimuli swept across all active electrodes, from the lowest frequency to the highest frequency electrodes. Each active electrode was stimulated for the same length of time. Starting at the clinical T level, listeners were asked to decrease the level of the sweep stimuli until they reached the level of first hearing, thereby moving the T curve as a whole, instead of adjusting the level for each electrode independently. The difference between the initial and final T measurements was denoted as the current level offset. A second repeat of this measurement was obtained starting below the first $50\mu s$ T curve. The average of the two current level offsets was added to the clinical T level to obtain the final $50\mu s$ T curve. The same process was repeated to obtain the $50\mu s$ C curve. Two level offsets measurements were obtained by adding a negative offset to the clinical C curve and allowing the listeners to increase the level until they reached their maximum comfortable loudness level. Once both estimated T and C curves were obtained for the $50\mu s$ stimuli, the dynamic range (DR), defined as C-T, was calculated for each electrode. Although the sweep approach may be less accurate than adjusting each electrode independently, it produced an estimate of the 50 % point of the DR for each electrode which was used to set the level of the stimuli in the MDT task.

3.4 Experimental Tasks

While the three experiments estimated different psychophysical percepts, FM, ED and MDT, the methods used to obtain them were similar. All estimates were obtained using an adaptive procedure that varied the psychophysical parameter of interest until 12 reversals were obtained [36]. For the first 4 reversals, following one correct response the perceptual distance between the target and the reference stim-

ulus was decreased by a predetermined step size. If the subject gave one incorrect response, the perceptual distance increased by the same step size. This is referred to as a 1-down/1-up rule. For the last 8 reversals, a 2-down/1-up rule was used. Two correct responses were required to decrease the perceptual distance between the target and the reference stimuli by a given step size; one incorrect response was required to increase the perceptual distance by that same step size. For each psychophysical measure, the estimate of the perceptual metric was taken to be the mean of the last 8 reversals, corresponding to the 70.7% probability of correct detection [36]. Subjects selected the target interval by clicking on the corresponding box using a computer mouse. Button flashes provided a visual cue during the stimulus presentation. A screen capture of the graphical user interface presented to subjects for the FM task is included in Figure 3.1. Feedback was provided by changing the color of the box corresponding to the correct interval. The number of intervals varied from one task to another but the inter-stimulus interval was always $500ms$. The remaining details of each experiment are included in the following sections.

3.4.1 Forward Masking Task

Subjects were presented with a two-alternative forced-choice task. One stimulus interval contained only the masker stimulus and the other contained both the masker and a probe stimulus. Subjects were asked to select the interval that contained two pulse trains. The masker had a duration of $300ms$. The probe was $30ms$ long and was presented $10ms$ after the masker. A schematic of the stimuli used in this task is provided in Figure 3.2.

The level of the masker was the same across all electrodes in the array, and set to a level corresponding to 70 – 90% of the clinical DR for each electrode. The exact level was selected by ensuring that the stimulus was comfortable and within the DR of all electrodes. The probe level was initially set to a level that was easily



FIGURE 3.1: Screen capture of graphical user interface used for the FM task. Subjects clicked on the Ready button to start a trial. Subjects indicated which interval contained two stimuli by clicking on button 1 or 2. Button color changes provided a visual cue during the stimulus presentation.

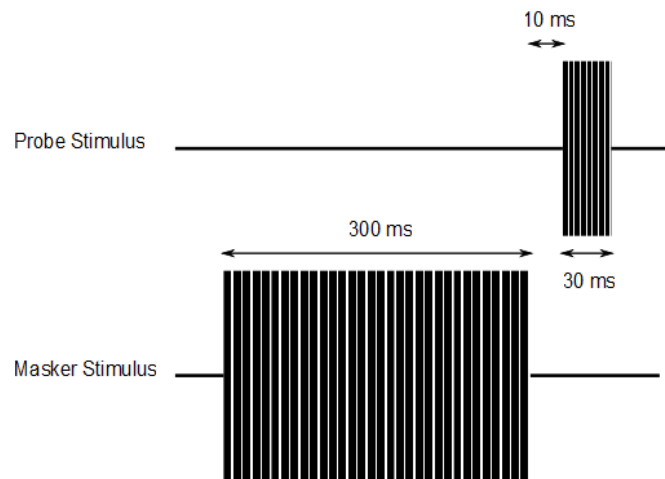


FIGURE 3.2: Stimuli presented during forward masking task. Masker duration, probe duration and masker-probe delay are indicated.

detectable but was varied adaptively thereafter. For the first 4 reversals, the step size was equal to 2 current steps. For the last 8 reversals the step size was 1 current step. A current step is a unit logarithmically related to current defined by Cochlear Corporation. The exact conversion between current in μA and current level (CL or number of current steps) for the CI24M/R and the CI24RE implants are provided in Equations 3.1 and 3.2. CL ranges from 0 to 255.

$$I(\mu A) = 10e^{\frac{CL \cdot \ln(175)}{255}} \quad (3.1)$$

$$I(\mu A) = 17.5 \cdot 100^{\frac{CL}{255}} \quad (3.2)$$

Masked probe thresholds were measured for all electrodes in the array, such that each electrode acted as a masker. For each masker electrode in the array, a single FM threshold measurement was obtained for each of 2 probes in the apical direction, 2 probes in the basal direction and for collocated masker and probe. In total 5 FM threshold measurements were obtained for each electrode as the masker. For electrodes at the edge of the array (i.e. 1, 2, 21 and 22) the measurements were taken for the available neighboring electrodes within the spatial limits defined above. For each probe, the estimate of masked probe threshold was subtracted from the average probe only threshold measurements (see Section 3.3) in order to estimate the threshold shift (T_{shift}). A larger T_{shift} value for a particular probe suggests the masker has a greater FM effect.

3.4.2 *Electrode Discrimination Task*

The goal of this task was to determine the closest discriminable target electrode for a given reference electrode. The experimental procedure was based on the task described in [1]. Subjects were presented with a four-interval two-alternative forced choice task and were asked to identify which interval contained the stimulus that was

different from the rest. The first and fourth interval always contained the reference stimulus, a pulse train presented at the reference electrode. The target stimulus consisted of a pulse train presented on a different electrode from the reference electrode and was presented randomly in either the second or third interval. All stimuli had a duration of $300ms$. The loudness for each stimulus was selected at random within the upper 60% of the DR. When stimuli loudness is varied in this way, ED results have been shown to correlate more closely to speech recognition results than ED results obtained using loudness balanced stimuli [3]. Subjects were instructed to ignore loudness cues.

For a given reference electrode, the target electrode was varied adaptively. Two estimates of the closest discriminable electrode were obtained for each direction along the array, corresponding to higher numbered and lower numbered electrodes, for a total of 4 estimates per reference electrode. When measuring in the apical direction the target stimulus was initially presented on electrode 22 and in the basal direction the target stimulus was initially presented on electrode 1. For electrodes 1 and 22 estimates were only obtained in the apical and basal directions respectively. For the first 4 reversals a step size of 2 electrodes was used. For the last 8 reversals a step size of 1 electrode was used.

3.4.3 Modulation Detection Task

The MDT task aimed to determine the minimum detectable modulation depth (m) both in the absence and presence of an interleaved masker. These measures are referred to as unmasked and masked MDT, respectively. The task was based on the experiments described in [37] and [7]. For both the unmasked and masked MDT tasks, subjects were presented with a four-interval two-alternative forced choice task and were asked to identify the stimulus that was different from the rest, ignoring loudness cues. The first and fourth interval contained the reference stimulus, a

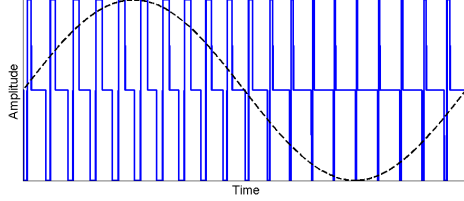


FIGURE 3.3: Synthetic example of the PW-modulated stimuli used for the modulation detection task. The PW is sinusoidally modulated at 40 Hz around a mean PW of $50\mu s$.

300ms pulse train with $50\mu s$ PW and an inter-pulse gap of $8\mu s$. The target stimulus was presented randomly in either the second or third interval. It consisted of a 300ms pulse train with pulse width sinusoidally modulated at 40 Hz around a mean PW of $50\mu s$. An illustrative example of a PW-modulated stimulus can be seen in Figure 3.3. PW was chosen over pulse amplitude as the parameter to modulate because, due to the hardware properties, it offers finer control over loudness. The modulation depth (m) was determined by the following formula [38]:

$$m = \frac{PW_{max} - PW_{min}}{PW_{max} + PW_{min}}, 0 < m < 1 \quad (3.3)$$

The effect of modulation depth, m , on the maximum and minimum pulse width (PW_{max} and PW_{min} , respectively) is illustrated in Figure 3.4. As modulation depth increases, the absolute value of the difference between the mean PW and PW_{max} and PW_{min} increases and the modulation becomes more detectable by the listener.

In the masked MDT case, all stimulus intervals contained an interleaved masker presented on the closest electrode in the apical, or lower frequency, direction. These stimuli containing both a target probe and an interleaved masker more closely resembles the interleaved stimuli listeners encounter in day-to-day CI use. Additionally, MDTs measured in this way provide a joint measure of the FM effects of the masker electrode and the temporal acuity of the probe electrode [7]. As in the study pre-

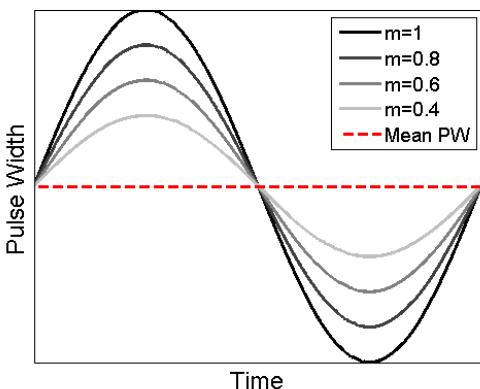


FIGURE 3.4: Effect of modulation depth on the value pulse width as a function of time. As the modulation depth increases, the absolute difference between the maximum and minimum pulse width and the mean pulse width increase.

sented in [7], the interleaved masker had a duration of $300ms$ and was presented at 50 % of the DR of the masker.

Due to the hardware considerations, the minimum m value was 0.008, which corresponded to a minimum PW of $0.2\mu s$. This value was also the smallest amount by which m could be increased or decreased. At the start of each run, m was set to 0.504, a multiple of the minimum value of the PW, or 50.4% modulation. For subsequent trials m was varied adaptively. The step sizes for the adaptive procedure are expressed in dB with respect to $m = 1$:

$$m_{dB} = 20\log(m) \tag{3.4}$$

The step size was 6 dB for the first four reversals and 1 dB for the last 8 reversals. All values of m were rounded to the nearest multiple of 0.008. The modulation detection threshold (MDT) was defined as the mean of the last 8 reversals. Two MDT measurements were obtained per electrode. If the difference between the 2 measurements was greater than 7 dB a third measurement was taken and the outlier was discarded [38]. The task was conducted for stimulus levels equal to 50% of the

DR.

3.5 Results

In order to estimate each electrode’s potential to confound speech recognition, FM, ED and MDT measurements were collected. For ED and FM, the results were a series of values that defined the perceptual effect of each reference electrode or masker. Because the goal of this task was to determine a single numerical measure of each electrode’s ”badness” according to each psychophysical metric, the FM and ED measures were each consolidated in to a single value for each electrode.

For the FM task, two estimates of the T_{shift} values were calculated for each of the two adjacent probes. If these values were less than 0, they were set to 0. This occurred rarely, and was likely a result of variability in the probe only or masked probe estimates. Figure 3.5 shows the T_{shift} values in CL for all subjects. Higher T_{shift} values indicate a higher effect of the masker on the neighboring probes. Generally the highest T_{shift} values were measured when the masker and the probe were co-located. As the distance between the masker and the probe increased, the T_{shift} values decreased, indicating a decreasing effect of the masker electrode.

In order to quantitatively estimate the effect of the masker on all electrodes, a least-squares Gaussian fit to the 5 T_{shift} estimates was calculated. This method assumes that the masking pattern is symmetrical and that the masking level decreases as the distance from the probe with the highest T_{shift} increases. Figure 3.6 includes an example of T_{shift} values and the Gaussian curve fit to those values for S4’s masker electrode 13. While these Gaussian fits cannot account for the variability previous studies have observed across masking patterns (e.g. [2], [39]), they can provide an estimate for the overall level of masking caused by a given electrode. In order to obtain a single quantifier of the masker’s effectiveness, the area under the masking curve (AUC) was calculated. Figure 3.6 shows the AUC corresponding to

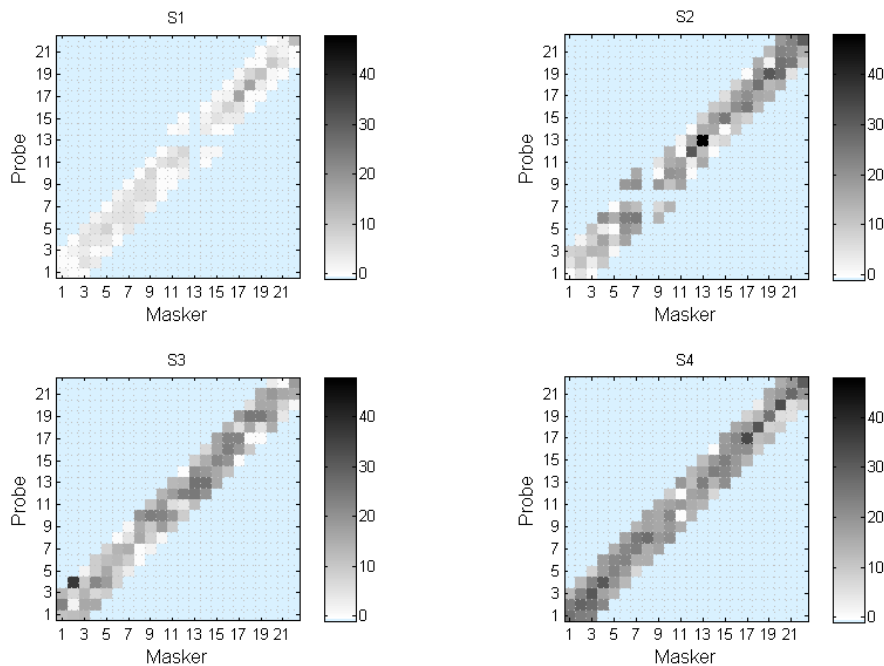


FIGURE 3.5: Threshold shift (T_{shift}) values for the forward masking tasks for four subjects. Each square represents a masker-probe pair. T_{shift} values in current steps for each probe as a result of the presence of the masker are shown in gray scale. Blue cells indicate masker-probe pairs that were not measured.

S4’s masker electrode 13. The results of the curve fits are included in Figure 3.7. The AUC curves corresponding to all listener’s electrodes are presented in Figure 3.8. Masker electrodes with higher AUC values are assumed to have a larger FM effect on the adjacent probes. For each subject there is a set of electrodes that, according to the AUC estimates, has a greater masking effect on neighboring electrodes. On average the data suggest S1 experiences the least amount of FM while S4 experiences the highest levels of FM.

For the ED experiment, a total of four estimates of the closest discriminable electrode, two in the apical and two in the basal direction, were obtained for each reference electrode. For each direction, these two values were averaged to obtain a single estimate of the closest discriminable electrode. For electrodes 1 and 21,

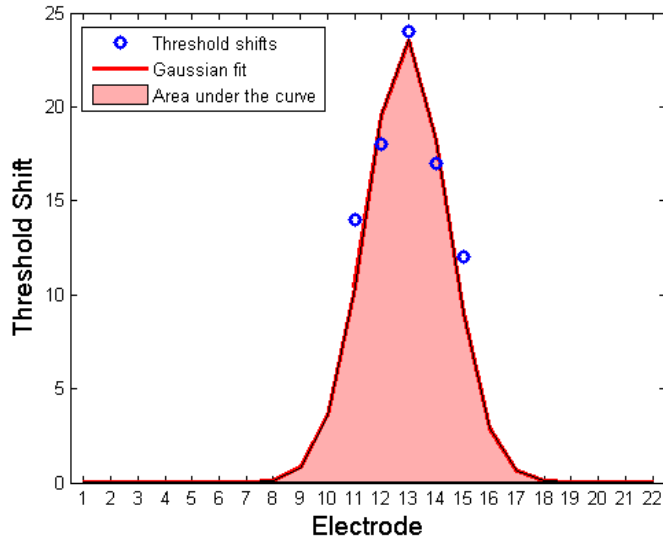


FIGURE 3.6: FM results for S4’s masker electrode 13. Threshold shift values are shown as a function of the probe electrode. The red line indicates the Gaussian curve fits to the threshold shift values and shaded region indicates the area under the estimated masking curve.

located at the edge of the array, ED was only measured in the apical or basal direction, respectively. All electrodes between the closest indiscriminable electrode and the closest electrode to the reference were considered indiscriminable from the reference. Figure 3.9 shows a plot of the ED patterns for each reference electrode. For visualization purposes, the non-integer values for the closest discriminable electrode resulting from averaging across estimates were rounded to the closest active electrode. Black cells indicate electrodes that were indiscriminable from the reference electrode, white cells indicate discriminable electrodes and blue cells indicate an electrode that was inactive in the subject’s clinical map.

For each direction, the average of these two estimates was subtracted from the closest electrode to the reference electrode in that direction to obtain a discrimination distance. These distances between the closest discriminable electrode and closest electrode to the reference were averaged to obtain a single measure of indis-

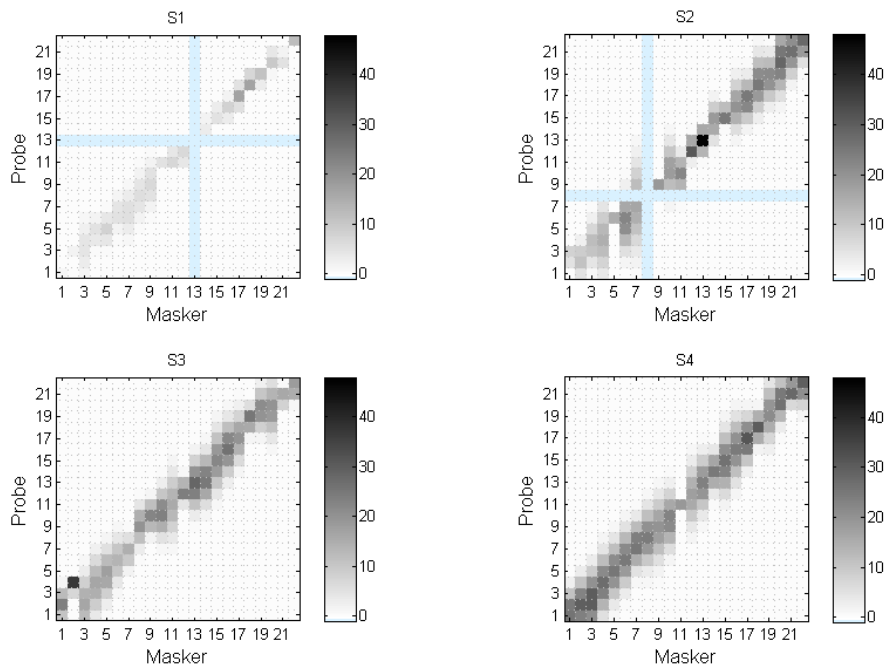


FIGURE 3.7: Gaussian fits to threshold values for four subjects who participated in the FM task. For each subject, the masking patterns for each electrode were approximated by a Gaussian fits to the T_{shift} measurements. Each image shows the values of these Gaussian fits, in current steps, for each masker-probe pair. Blue cells indicate an inactive electrode in the subject’s clinical map.

criminability for each reference electrode. Larger values of this distance indicate a higher number of electrodes that present indiscriminable precepts (e.g. [1]). The average distance to the closest discriminable electrode are plotted in Figure 3.10.

The average distance to the closest discriminable electrode across the array varies from one subject to another. There is also a considerable amount of variability within subjects: while some groups of electrodes are relatively indiscriminable from their neighbors, other electrodes appear to generate unique percepts. As in the literature, indiscriminable electrodes are not uncommon, violating the assumption that all electrodes provide independent percepts.

The unmasked and masked MDT results are shown in Figure 3.11. Higher MDT

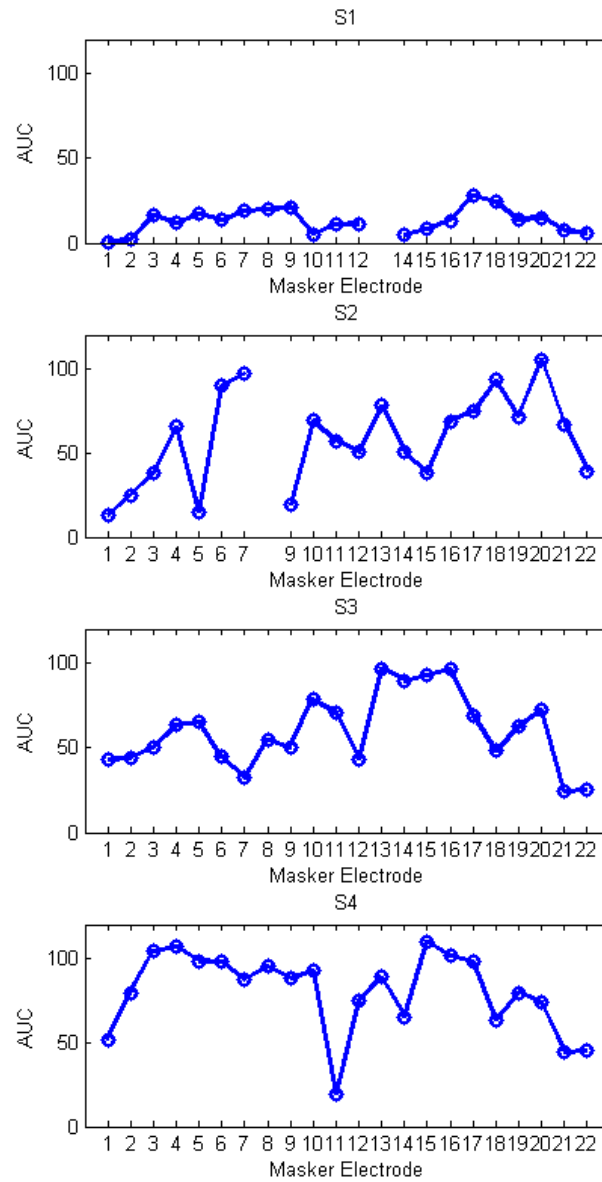


FIGURE 3.8: Area under the Gaussian fit masking curve results for four the subjects. The area under the Gaussian fit curve to the T_{shift} values is plotted for each masker.

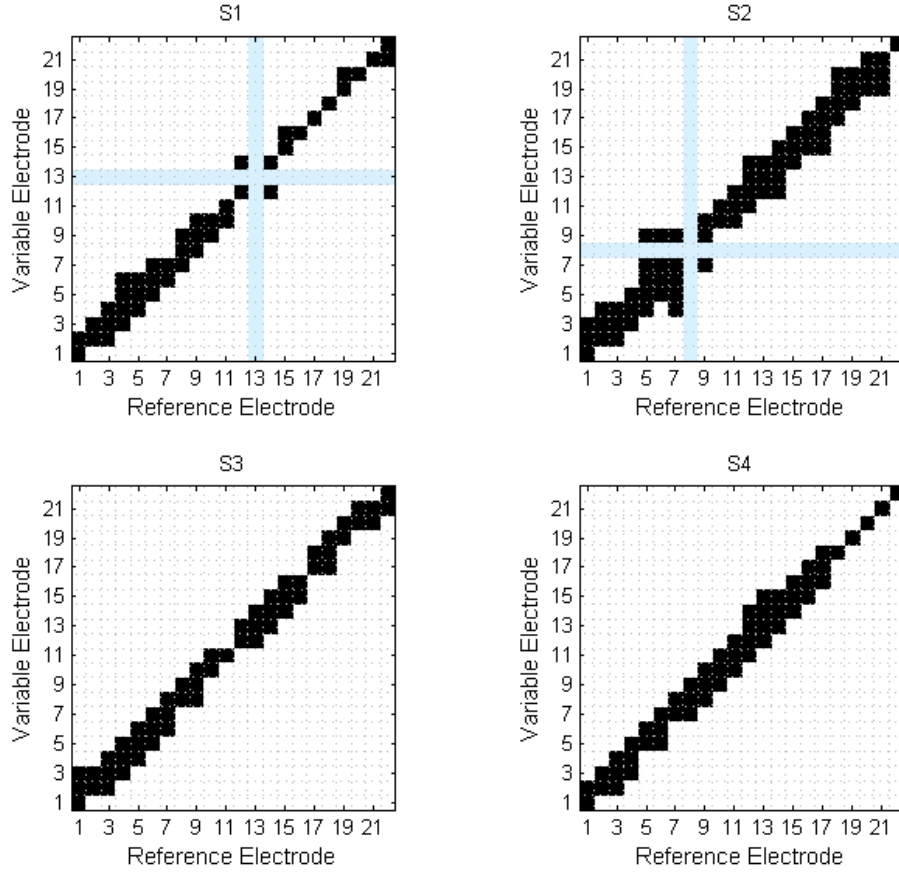


FIGURE 3.9: Results for the electrode discrimination task for four subjects. The reference electrodes are plotted on the x-axis versus the variable electrodes on they y-axis. Black cells indicate electrodes that were indistinguishable from the reference electrode, white cells indicate discriminable electrodes and blue cells indicate an electrode that was inactive in the subject's clinical map. The closest discriminable electrode to each reference electrode in each direction was estimated using two measurements. For electrodes 1 and 22 discriminability was measured in the only available direction.

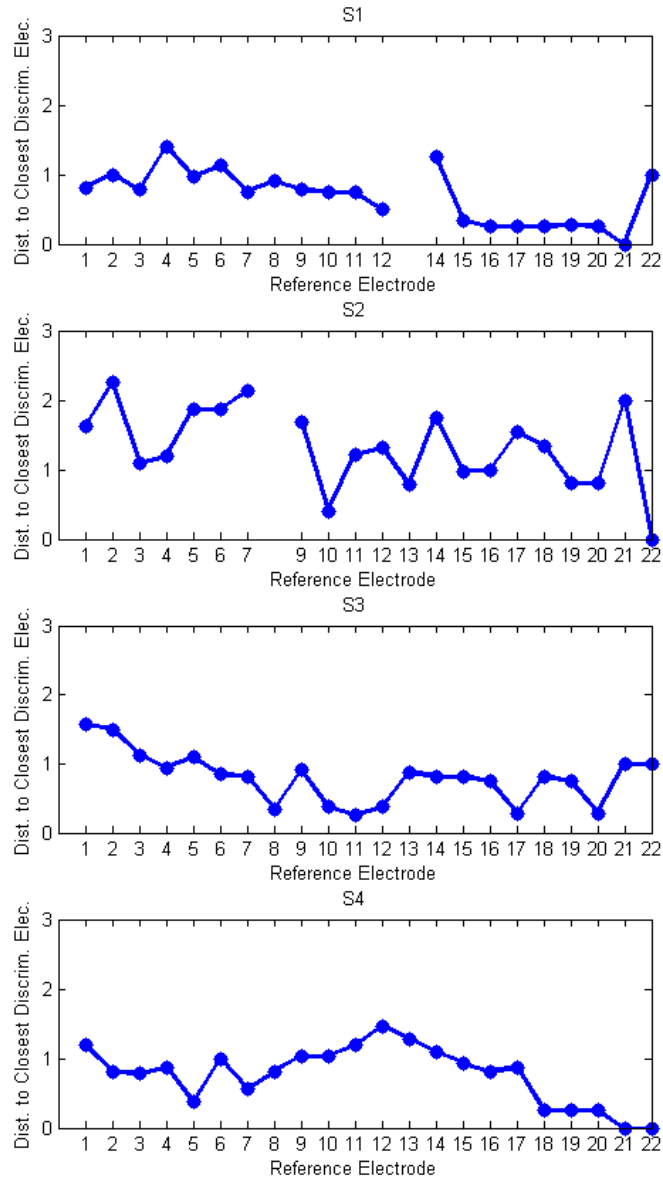


FIGURE 3.10: Average distance to the closest discriminable electrode for four subjects. The reference electrodes are plotted on the x-axis versus the corresponding average distance between the closest electrode to the reference and the closest discriminable electrode on the y-axis. Each point represents the average of two estimates, one corresponding to the apical and the other corresponding to the basal direction.

values suggest a lower sensitivity to changes in PW modulation depth m . The data also show considerable variation across subjects: for S3 all MDT estimates are close to the minimum of 0.008 while S2's results vary greatly across the array. These results suggest that, for all subjects except S3, temporal acuity can be a function of electrode for a given subject. The relationship between masked and unmasked MDT values also varied across subjects. For most listeners the masked and unmasked MDT values were approximately the same. This is likely due to a negligible effect of the interleaved masker on the modulated pulse train. For S4, the masked MDT values are higher than the unmasked values. This suggests that this listener was more sensitive to masking, a conclusion which is supported by his high FM levels presented in Figure 3.8.

The results of the three psychophysical metrics that were collected indicate variability within and across subjects. This variability could partly be attributed to differences in physiology, such as nerve survival or tissue growth, as well as central processing variability across individuals. In order to mitigate the effects of these individual characteristics, the normalized experimental data were compared for each subject. The masked MDT data were excluded from this comparison because they include a measure of FM that is also quantified in a separate task. The normalization was performed for each subject by dividing the derived metrics corresponding to each task by a normalizing constant equal to the maximum value of the parameter of interest. The resulting normalized data is unit-less, with larger values indicating electrodes with a higher potential to confound speech recognition results.

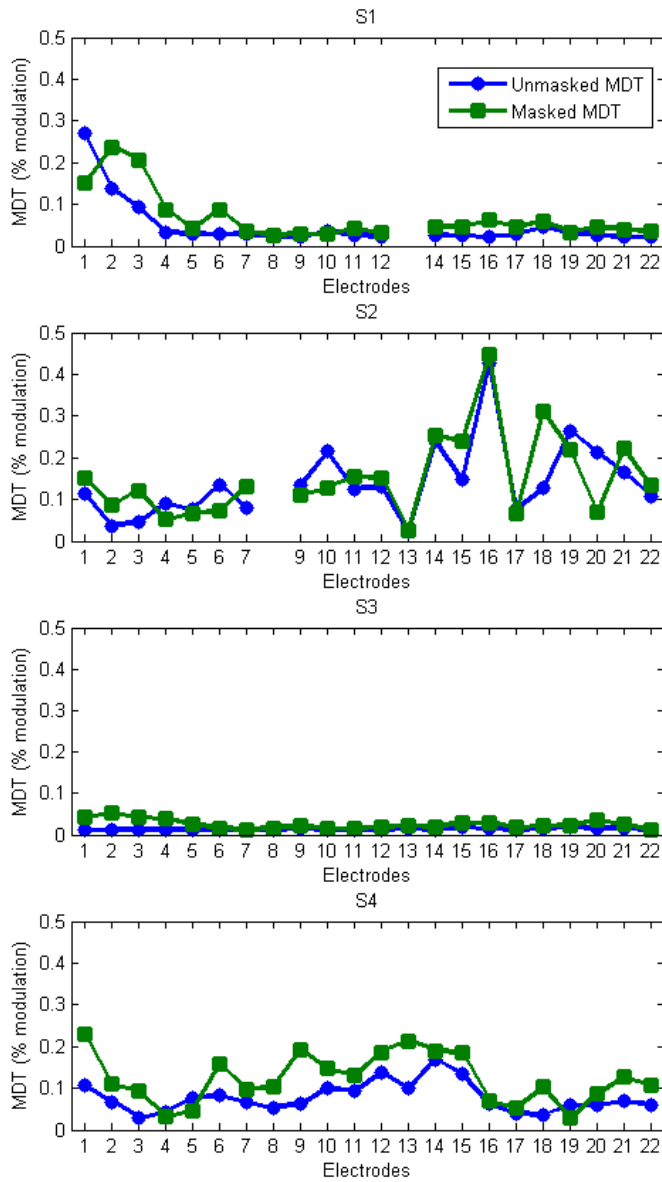


FIGURE 3.11: Results for the electrode modulation detection task for four subjects. The electrodes are plotted versus their corresponding average modulation detection threshold (MDT). Each point represents the average of two modulation detection threshold (MDT) estimates. The blue circles and green squares indicate MDT measurements taken in the absence and presence of a masker respectively.

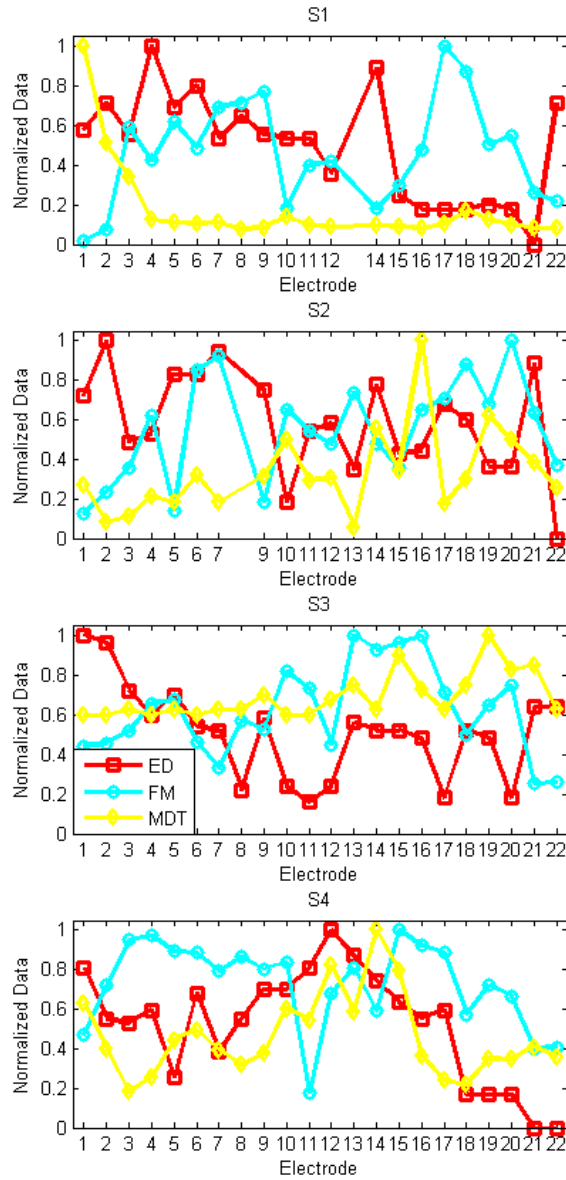


FIGURE 3.12: Normalized psychophysical data for four subjects plotted as a function of electrode. ED data is displayed with red squares, FM data with cyan circles and MDT data with yellow diamonds. The normalization was performed by dividing the data corresponding to each task by a normalizing constant equal to the maximum value of the parameter of interest for each subject. The resulting normalized data is unit-less.

Table 3.2: Spearman correlation coefficients between normalized psychophysical data for four subjects. The corresponding significance values for each comparison are also included. p values less than 0.05, indicating a statistically significant correlation between the corresponding metrics are shown in bold font.

Subject	ED-FM	MDT-ED	FM-MDT
S1	$r = -0.2866$ $p = 0.2079$	$r = 0.1604$ $p = 0.4875$	$r = -0.1014$ $p = 0.6620$
S2	$r = -0.1793$ $p = 0.4367$	$r = -0.2131$ $p = 0.3536$	$r = 0.2156$ $p = 0.3463$
S3	$r = -0.4547$ $p = 0.0335$	$r = -0.1875$ $p = 0.4035$	$r = 0.1769$ $p = 0.4310$
S4	$r = 0.0861$ $p = 0.7032$	$r = 0.0861$ $p = 0.0014$	$r = -0.1959$ $p = 0.3805$

The subject-specific normalized results for the FM, ED and unmasked MDT data are shown in Figure 3.12. In order to quantify the pairwise relationship between each of the psychophysical metrics, the Spearman correlation coefficients, r , and the corresponding significance levels, p , were calculated and are presented in Table 3.2. p values less than 0.05, indicating a statistically significant correlation between the corresponding metrics, are shown in bold. Most of the pairwise correlation values are moderate and not statistically significant, suggesting that each metric may be indicative of a different measure an electrode’s potential to confound information.

3.6 Discussion

This study investigated three different psychophysical measures: FM, ED and MDT. The results from the ED indicate electrodes that may be presenting different information on the same perceptual channel. The FM data highlights electrodes that may be causing cross-site interference by impeding the perception of information presented on neighboring electrodes. The unmasked MDT data could identify site-specific weakness in temporal acuity while the masked MDT data could reflect a combination of FM effects and temporal acuity. Measuring these psychophysical pa-

rameters on the same subjects provided a multidimensional estimate of the potential “usefulness” of each electrode, that is, how much it potentially contributes to speech recognition.

Of the three psychophysical measures obtained, estimating the FM patterns required the most approximation. The method used to estimate the FM patterns is simple and was adopted due to experimental time constraints and in consideration of subject availability. Ideally, full masking patterns for each electrode would be estimated by obtaining T_{shift} values for every electrode in the array acting as a probe but this would require a substantial amount of testing time. Given the time constraints, instead of obtaining several measurements for co-located masker and probe and two adjacent electrodes as probes, single-point estimates were obtained for co-located masker and probe and four adjacent probes. The goal of the task was to quantify the FM effect of each masker on its neighboring electrodes; the single-point estimates increase the breadth of the measured masking pattern, perhaps at the expense of accuracy when measuring the threshold shift for a given probe.

In [6], FM measurements were used to determine reduced electrode sets. The two most apical electrodes, which also had the greatest masking effect on their neighbors, were excluded from the clinical map. Two out of the four subjects tested obtained higher consonant recognition scores with the experimental maps. However, this study was limited to considering only the apical region of the electrode array. The large AUC values obtained for some of the subjects tested in other regions of the array suggest that they may benefit from having medial and basal electrodes with high masking levels excluded from their maps. In this work, electrodes with the highest area under the masking curve across the whole array were excluded from listeners’ maps using two different approaches described in Chapters 4 and 6.

The ED results indicate that some subjects are capable of almost perfect discrimination for some reference electrodes, while others are indiscriminable from several

neighboring electrodes. Excluding electrodes whose corresponding nerve fiber population most greatly overlaps with neighboring electrodes – that is, those with a larger discrimination distance – could allow for different information to be presented on distinct perceptual channels. In [1], experimental maps were constructed based on ED results by only including discriminable electrodes. While the experimental maps did improve speech recognition for a variety of different speech materials in quiet for most listeners, the method for constructing the maps did not allow researchers to control the number of electrodes for each listener. This resulted in one subject having only 3 electrodes in her experimental map which provided poor speech recognition results. Because different subjects may benefit from having different numbers of indiscriminable electrodes removed from their experimental map, an exclusion method that provides control over the number of electrodes excluded could be beneficial. In this work, a more conservative approach to excluding electrodes was explored. In the study described in Chapter 6, experimental maps were constructed by sequentially dropping the electrodes with the highest discrimination distance. In contrast to the methods described in [1], which selects one representative from each group of indiscriminable electrodes, this approach was analogous to excluding the electrode that was most perceptually similar to its neighbors. The aggressiveness of the exclusion criterion can be tuned by increasing or decreasing the number of electrodes with the highest discrimination distance dropped from the map.

The MDT results are consistent with those reported by Garadat et al. [7], who observed considerable variation in MDT across and within subjects. An increase in MDT values when the interleaved masker was present was observed for some but not all listeners in both these results and in the study presented by Garadat et al. [7]. As in their study, the correlated masked and unmasked MDT data indicates that these metrics can likely be used interchangeably to predict modulation sensitivity for a given electrode.

The subject-specific normalized results show that ED, MDT and FM data are generally uncorrelated for most subjects, suggesting that one psychophysical measure cannot be used to predict another. These results also suggest that for a given subject, each metric indicates a different potential to confound for each electrode. This variability suggests that there may be multiple methods to select electrodes. Exploring electrode selection as a function of each metric may determine the method that results in the largest improvement in speech recognition for each listener. In order to further investigate the relationship between the metric used to select electrodes and speech recognition, the psychophysical data that were collected were used as inputs to two different electrode selection methods described in Chapters 4 and 6.

Dynamic Electrode Selection Using Psychophysical Measures

In most clinically available N-of-M strategies, the N highest energy electrodes out of the M channels available are selected for stimulation. These strategies assume that information presented on all electrodes is perceived as unique and independent. However, previous work suggests that stimuli presented on different electrodes can be indistinguishable (e.g. [1], [2], [3]). Additional studies suggest that stimuli presented on one electrode can impede the perception of subsequent stimuli presented on neighboring electrodes (e.g. [4], [5], [6]). In order to minimize these electrode interactions, extensions to clinically available processing strategies have been investigated. Nogueira et al. [15] used psychoacoustic-masking models derived from normal hearing listeners to estimate electrode interactions in CI listeners. For each time window, the electrodes with the highest energy that were also above the estimated masking threshold were selected for stimulation. In a similar study, Kals et al. [16] divided adjacent electrodes into groups of 1-4 electrodes. For each time window, only the highest energy electrode of each group was selected for stimulation, thereby limiting the number of adjacent electrodes that could be activated. Both of these algorithms

implicitly assumed that electrode interactions are the same across listeners, as neither of them incorporated listener-specific measures of electrode interaction. In this work, the individual ED and FM data presented in Chapter 3 were used to estimate each electrode’s potential to confound or impede the perception of stimuli presented on adjacent electrodes. For each time window, the electrodes with the highest energy that were also most likely to be perceived were selected for stimulation. The performance of the ED and FM based N-of-M strategies was evaluated using both an objective speech recognition task and a subjective preference task.

4.1 Dynamic Electrode Selection Algorithms

Previous psychophysics-motivated electrode selection methods eliminated the electrodes with the highest potential to confound speech recognition from listeners’ maps across all time windows (e.g. [1], [6], [7]). In contrast, the electrode selection algorithms presented in this chapter dynamically selected electrodes for each time window. For each time window, the electrodes with the highest energy that also had the highest likelihood of being perceived were selected for stimulation. These dynamic selections strategies preserve the energy-driven, stimulus-dependent electrode selection listeners are accustomed to while attempting to maximize the amount of perceivable information. Because no electrodes are permanently deactivated, these strategies preserve each electrode’s frequency bands, thereby avoiding any potential negative impact of frequency reallocation (e.g. [17], [18], [19]).

This work presents two different methods of selecting electrodes for each time window. The first used ED data to select the highest energy electrodes that are also discriminable from each other, and is referred to as ED N-of-M. The second used the area under the estimated FM curves to select the highest energy electrodes that are not masked by adjacent electrodes, and is termed FM N-of-M.

4.1.1 *Electrode Discrimination Based Selection*

Prior to the application of the electrode selection strategy, the closest discriminable electrode in each direction along the array was determined for each electrode. These data are included in Chapter 3. This resulted in a perceptually defined set, S_i containing all electrodes indiscriminable from reference electrode n_i . The set corresponding to each reference electrode was comprised of the electrodes between the closest discriminable electrodes in each direction along the array. Each listener's discrimination pattern was therefore defined as a collection of these sets, S , containing as many sets as there were reference electrodes:

$$S = \{S_1, S_2, \dots, S_{22}\}. \quad (4.1)$$

For each time window, the highest energy electrode was selected for stimulation. Any other electrodes in its perceptual set were excluded from the subsequent selection process, under the assumption that they would not provide a discriminable percept. The next highest energy electrode was then selected for stimulation, and the remaining electrodes in its perceptual set were excluded. This selection process continued until each listener's number of clinical maxima was reached or until no electrodes were left to consider. If the number of electrodes selected for a given time window was less than the clinical number of maxima, null pulses with sub-threshold amplitude were presented to the excluded electrodes. This ensured that the effective number of maxima stayed constant across all time windows, preserving each listener's effective pulse rate.

For the purposes of illustration, consider a simplified array with 10 electrodes. The perceptual sets for each reference electrode are defined as follows:

$$\begin{aligned}
S &= \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9, S_{10}\} \\
&= \{\{\mathbf{1}, 2\}, \{1, \mathbf{2}, 3\}, \{2, \mathbf{3}\}, \{\mathbf{4}, 5, 6\}, \{4, \mathbf{5}, 6\}, \\
&\quad \{5, \mathbf{6}\}, \{6, \mathbf{7}\}, \{\mathbf{8}, 9\}, \{8, \mathbf{9}, 10\}, \{9, \mathbf{10}\}\}
\end{aligned} \tag{4.2}$$

where each element in a set corresponds to an electrode and the bold elements indicate the reference electrode used to determine each perceptual set. For a particular time window, the electrodes are arranged from highest to lowest energy: 9, 10, 8, 2, 1, 3, 4, 7, 5, 6, as illustrated in Figure 4.1. For this example, 4 electrodes, N , are selected out of the 10 available, M . If using a standard N-of-M strategy, the 4 electrodes with the highest energy – 9, 10, 8 and 2– would be selected for stimulation. If using the ED N-of-M strategy, the highest energy electrode, 9, would be selected for stimulation. Then the electrodes in its perceptual set, 8 and 10, would be excluded from the selection process. The next highest energy electrode, 2, would also be selected for stimulation, and electrodes 1 and 3 would be excluded. This process would be repeated until only 4 electrodes were selected. The different electrode sets selected for stimulation and their corresponding energies are shown in Figure 4.1. As illustrated in this example, the N-of-M strategy often selects clusters of adjacent electrodes for stimulation. In contrast, the proposed ED N-of-M strategy, stimulates a single electrode from each perceptually defined set, potentially minimizing electrode interactions.

4.1.2 Forward Masking Discrimination Based Selection

Prior to implementing the FM-based N-of-M electrode selection strategy estimates of each listener’s masking patterns for each electrode were obtained. For each masker electrode n , these curves are defined by the Gaussian estimate $T_{shift}^n(e)$, a function of the probe electrode e . These data are presented in Chapter 3. The $T_{shift}^n(e)$ functions

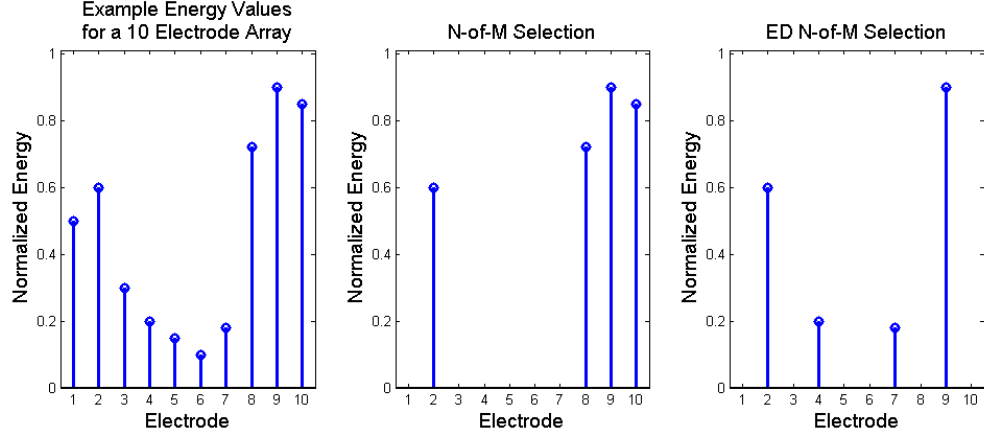


FIGURE 4.1: Comparison of electrodes selected with clinical N-of-M and ED-based N-of-M selection strategies. The electrodes are plotted versus their corresponding normalized energy. *Left*: Example electrode energies for a given time window. *Center*: Electrodes selected for stimulation using clinical N-of-M. *Right*: Electrodes selected using ED N-of-M.

corresponding to each masker are used to determine which stimuli could be occluded by a masker at each stage of the electrode selection process.

The FM N-of-M selection strategy proposed in this work is an iterative process applied to each time window. First, the electrode with the highest energy is selected for stimulation. The overall masking level $M(n)$ resulting from the stimulation presented by electrode the highest energy electrode is calculated by superimposing $T(n)$ and a scaled version of the highest energy electrode’s estimated $T_{shift}^i(n)$ curve, as described by Equation 4.3:

$$M(n) = T(n) + \sum_i^N \left[\frac{L(i)}{L_{masker}(i)} T_{stim}^i(n) \right]. \quad (4.3)$$

where $L(n)$ corresponds to the current level of electrode n for the time window considered and $L_{masker}(n)$ corresponds to current level of the masker used to obtain $T_{shift}^i(n)$. This expression assumes that forward masking patterns scale linearly with the amplitude of the masker, which is consistent with the findings of [39].

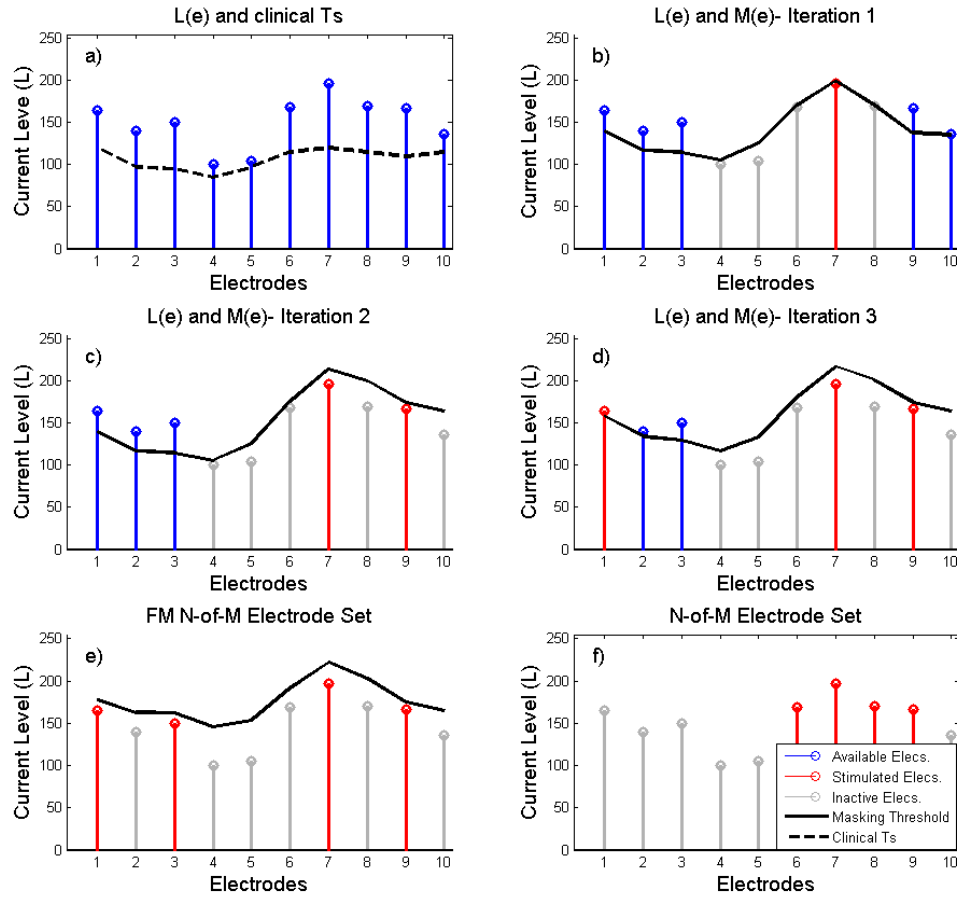


FIGURE 4.2: Sequential FM N-of-M electrode selection. *a)*: Example current levels (L) and clinical thresholds (T). *b-e)* For each iteration, the highest energy electrode, in red, is selected for stimulation from the set of available electrodes, in blue. The electrodes corresponding estimated masking threshold $M(n)$, in black, is added to the current estimate of the masking pattern. Any electrodes below the masking threshold are not selected for stimulation and are indicated in gray. This iterative process continues until the desired number of maxima is reached or until no electrodes with $L(n) \geq M(n)$ remain in the set. The final sets for both the FM and clinical N-of-M strategies are included in panels *e* and *f* respectively.

For the purposes of illustration, consider an example array with 10 electrodes. For a given time window, simulated current levels $L(n)$ and clinical threshold levels $T(n)$ are shown in panel *a* of Figure 4.2. In this example, the highest energy electrode, 7, would be selected for stimulation. The overall masking level $M(n)$ resulting from stimulating electrode 7 is calculated using the following expression:

$$M(n) = T(n) + \frac{L(7)}{L_{masker}(7)} T_{shift}^7(n). \quad (4.4)$$

Any electrodes that have not already been selected with $L(e)$ values less than $M(n) - 4$, 5, 6, 8, and 9, in this example—are excluded from the stimulation set. The resulting value of $M(n)$ as well as the electrodes with stimulation levels that fall beneath it are shown in panel *b* of Figure 4.2. In the second iteration of the process, the next highest energy electrode that has not already been selected or dropped, electrode 9, is selected for stimulation. Its corresponding $M(n)$ is calculated using Equation 4.5:

$$M(n) = T(n) + \frac{L(7)}{L_{masker}(7)} T_{stim}^7(n) + \frac{L(9)}{L_{masker}(9)} T_{stim}^9(n). \quad (4.5)$$

The electrodes that haven't already been selected with $L(n)$ less than $M(n)$, 10, are also dropped, as indicated in panel *c* of Figure 4.2. This process is repeated until N electrodes have been selected, where N is the number of maxima in the listener's clinical speech processing algorithm or until no electrodes remain available for selection. If the number of selected maxima is less than N , null pulses are presented to preserve the listener's clinical pulse rate.

4.2 Experimental Task

4.2.1 Subjects

Three of the four subjects who participated in the psychophysical experiments described in Chapter 3, S1, S2 and S3, also completed psychophysics-based dynamic electrode selection task. Their demographic information is included in Table 3.1. Each listener completed the tasks in two to three sessions lasting two and a half to four hours.

4.2.2 Stimuli and Equipment

The speech material used in this experiment consisted of monosyllabic consonant-nucleus-consonant (CNC) words. A total of 792 unique words were obtained from three different sources: the TIMIT database, the Minimum Speech Test Battery and The Sage English Dictionary [40], [41], [42]. The words were recorded by 4 male speakers, all native speakers of American English. The recordings took place in a sound attenuating booth using a Shure SM58-LC Cardioid microphone connected to a personal computer (PC) via a Tascam US-100 Audio Interface. The number of words presented across the speech recognition and preference tasks exceeded the 792 unique words available. To minimize training effects, all 792 words were presented before allowing words selected during previous conditions to be presented a second time.

Prior to testing, subjects' clinical maps were read from their processors using Custom Sound, Cochlear's clinical software. All the subjects' clinical parameters, including their Ts and Cs, were then transferred to a map file in the Nucleus MATLAB toolbox. The C levels were verified to be comfortable by streaming 300 ms-long biphasic pulse trains at the corresponding level using the NIC v2. This same research interface was used to stream all other experimental stimuli from a PC. All stimuli

were presented as biphasic pulse trains with parameters corresponding to those read from each subjects' map. Subjects' clinical parameters are included in Table 3.1. Custom designed graphical user interfaces (GUIs) designed in MATLAB were used to conduct the speech recognition tasks.

4.2.3 Procedure

The effectiveness of the proposed psychophysics-based dynamic electrode selection strategies was evaluated by comparing listeners speech recognition performance and preference across three speech processing algorithms: ED N-of-M, FM N-of-M and clinical N-of-M. Each of the algorithms was tested under three different listening conditions: speech in quiet, in speech-shaped noise (SSN) and in reverberation. The SSN noise was generating using a 78th order finite-impulse response filter [43] with filter coefficients extracted from a 5 second sample of SSN provided by the House Ear Institute. This filter was applied to white Gaussian noise, ensuring that each noise instance was unique. Prior to adding noise to the speech tokens to be presented during a particular run, the duration of all words was computed. The unique noise segments were 200ms longer than the duration of the longest word. Each word was randomly embedded in its corresponding noise segment such that the resulting signal-to-noise ratio (SNR) was 10 dB. To simulate reverberant speech, room impulse responses (RIRs) were generated using a MATLAB implementation of the modified image-source method (ISM) supplied by [44]. The amount of reverberation added to a signal can be quantified by the reverberation time (RT_{60}), where RT_{60} is the amount of time required for sound level to drop 60 dB from the onset level. For this task, reverberation with an $RT_{60} = 0.45s$ was added to the signals by convolving the RIRs with the anechoic speech tokens via multiplication in the frequency domain.

At the start of each testing session, subjects were presented with a set of 50 training words in quiet processed with their clinical strategy. This training set was not

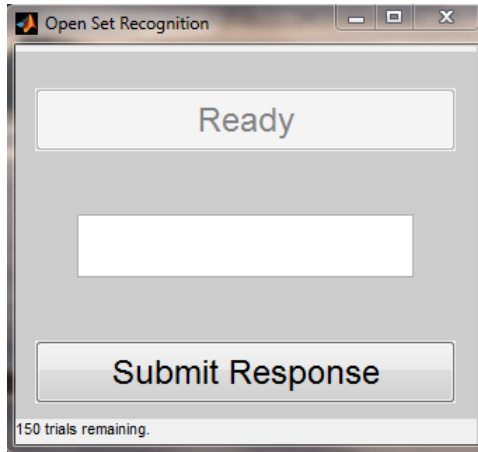


FIGURE 4.3: Screen capture of graphical user interface used for psychophysics-based N-of-M speech recognition task. Subjects clicked on the Ready button to start a trial. After the word was presented they typed in what they heard into the text box. The next word was presented 500 ms after they either pressed enter or clicked submit to record their response.

scored; its purpose was to familiarize subjects with listening to the speech material using the research interface. The words included in the set were the same for all subjects but the order was randomized for each individual. These words were reserved for training purposes and were not used during testing.

4.2.4 *Speech Recognition Task*

Listeners' speech recognition was evaluated using a word recognition task. The quiet listening condition was presented first, and the noise and reverberant conditions were presented subsequently in random order. Within each condition the order of the ED and FM N-of-M strategies was randomized and the clinical condition was presented last. 150 words were presented for each of the nine algorithm-listening condition combinations.

Listeners started a run of the task by pressing the space bar or clicking on the Ready button of the MATLAB GUI, shown shown in Figure 4.3. They were presented with a word and asked to use a computer keyboard to type what they heard into the

text box. They confirmed their submission by pressing the enter key or clicking the submit button. After a 500 ms pause the next word was presented. After 50 words the task was paused in order to provide the subject with an optional rest period. The task was resumed by pressing the space bar or clicking the Ready button. When all words had been presented the GUI window closed.

4.2.5 Preference Task

While a speech recognition task was conducted to evaluate listeners' objective performance with each of the experimental algorithms, their preferences were evaluated using a separate task. For one run of the task, the experimental strategy evaluated was ED N-of-M while for another it was FM N-of-M. A screen capture of the GUI used for this task is included in Figure 4.4. Listeners initiated the task by clicking on the Ready button. For each trial run listeners were presented with the transcription of a word and then were presented with it processed in two different ways. One interval included a word processed with clinical N-of-M and the other included the same word processed with the experimental electrode selection strategy. The interval corresponding to each processing condition was randomized for each trial. During stimulus presentation, the top buttons changed color to provide a visual cue. Listeners were instructed to select which version of the word they preferred by clicking on the corresponding button. Listeners could replay the pair of stimuli up to four additional times by clicking on the Replay button. Once the maximum number of replays had been reached, the listener was forced to make a selection. The next word was presented after a 500 ms pause.

4.3 Results

The results of the speech recognition task for each of the three algorithms, ED, FM and clinical N-of-M are shown in Figure 4.5. Speech recognition results are



FIGURE 4.4: Screen capture of graphical user interface used for psychophysics-based N-of-M preference task. Subjects clicked on the Ready button to start a trial. The word transcribed was presented processed with clinical N-of-M and either ED or FM N-of-M. Button color changes provided a visual cue during the stimulus presentation. Subjects indicated which processing strategy they preferred by clicking on button 1 or 2.

presented for three different listening conditions: quiet/anechoic, additive SSN and reverberant speech. To analyze the results, a Beta distribution on the probability of correct phoneme identification p was calculated over the 150 CNC words presented per condition. The Beta fits were calculated using the number of correctly identified phonemes and the number of total phonemes presented [45]. The height of each bar in Figure 4.5 represents the mean of the Beta distributions. The error bars represent the 95% confidence intervals for each distribution. The lines with the triangles over them indicate conditions that are significantly different from each other ($P(p_{cond_1} \geq p_{cond_2}) \geq 0.95$).

As expected, listener’s speech recognition scores are higher in the quiet condition than when the speech was corrupted by SSN or reverberation. For all listeners presented with speech in quiet, one psychophysics-based strategy performed comparably to the clinical algorithm and the other performed worse: FM N-of-M resulted in lower speech recognition scores for S1 while ED N-of-M resulted in lower speech recognition scores for S2 and S3. This could partially be explained by the values of each psychophysical metric for each listener. For instance, S1 had the lowest FM AUCs across all listeners, which could indicate that the spatial-temporal interactions between his electrodes is relatively low. Dropping electrodes using FM may therefore not have increased the amount of information he perceived. Furthermore, if the low FM values used to drop electrodes were the result of measurement noise and not an underlying physiological issue, the strategy could have discarded perceivable information. This difference in performance across algorithms was generally not observed when the speech was degraded by noise or reverberation.

The results of the preference task are included in Figure 4.6. To describe the results, maximum likelihood Binomial distributions were fit to data. These estimates provided a mean percentage of times the experimental algorithm was selected over the clinical algorithm as well as 95% confidence intervals. An algorithm was considered

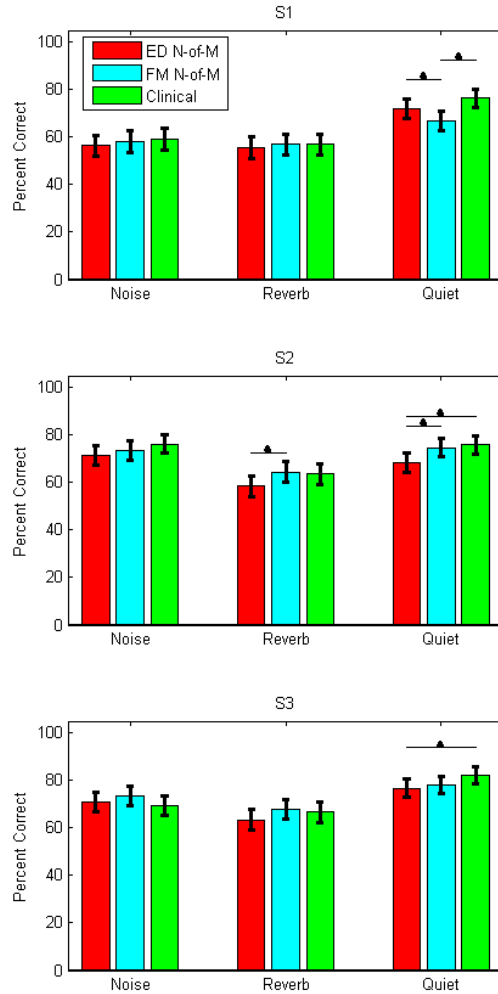


FIGURE 4.5: Speech recognition results using three different speech processing algorithms: ED N-of-M, FM N-of-M and clinical. 150 CNC, each consisting of three phonemes, were presented using each algorithm under three different listening conditions quiet, 10 dB SNR of SSN and reverberation with $RT_{60} = 0.45s$. The speech processing condition is indicated on the x-axis and the bar color indicates the speech processing algorithm. The height of each bar represents the mean of the Beta distribution on the probability of correct phoneme identification, p . The error bars represent the 95% confidence intervals for each distribution. The triangles indicate two conditions that are significantly different from each other ($Pr(p_{n_1} \geq p_{n_2}) \geq 0.95$).

better than its clinical counterpart if the lower bound of its confidence interval and its mean were higher than 50%. In contrast, an algorithm was considered worse than clinical if the upper bound of the confidence interval and the mean was below the 50% level. Lastly, if the confidence interval corresponding to an algorithm's preference score included the 50% level it was considered comparable to clinical. In seven out of the nine subject and listening condition combinations, FM N-of-M was selected as many or more times than clinical N-of-M. In contrast, ED N-of-M was considered worse than clinical across all conditions and all listeners. This disparity in preference between the dynamic electrode selection algorithms coupled with their comparable objective performance under most conditions suggests that preference scores are not a predictor of speech recognition. Instead, these preference results may be an indicator of the amount of cognitive load required for the listener to extract relevant information from the speech signal. Listeners may be able extract the amount of information needed to correctly identify a word in spite of small deviations from the expected activation pattern caused by the experimental algorithm, but doing so may require more listening effort. In order to compare the pulsatile representation of words presented to the listeners across all three selection algorithms, the number of pulses and the distribution of those pulses across the array were recorded.

The number of pulses presented to each electrode across all 792 words used in testing were counted after processing every word using each individual's ED, FM and clinical N-of-M strategies. These results are included in Figure 4.7. In noise, the number of pulses increased and their distribution across the array changed, especially for the low frequency electrodes. This is consistent with the nature of the additive SSN and its low frequency content. In reverberation, the number of pulses increases slightly as a result of the addition of delayed and attenuated versions of the original speech signal. Because the frequency content of these reflections is approximately the same as that of the original signal, the distribution of the number of pulses across

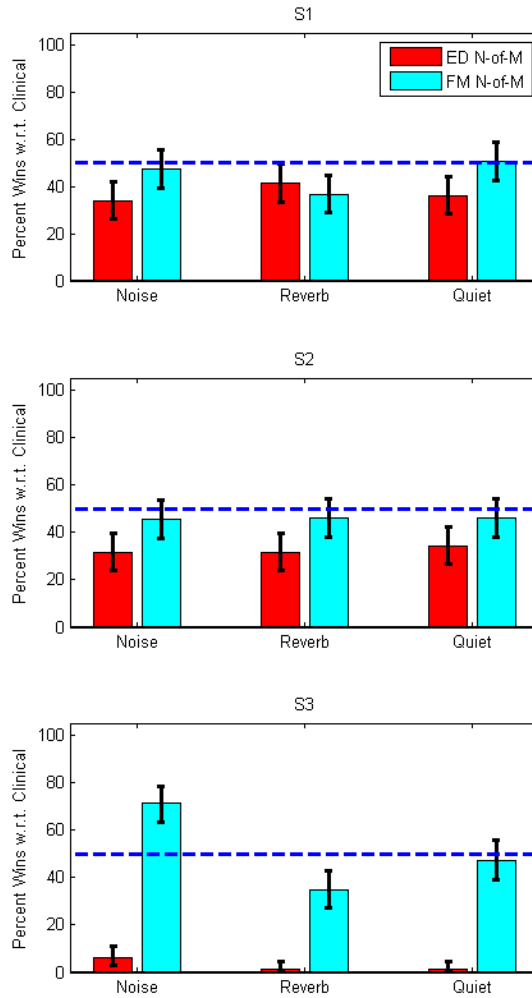


FIGURE 4.6: Preference results for dynamic electrode selection algorithms when compared to clinical. The ED and FM N-of-M strategies are indicated by the red and cyan bars, respectively. The height of each bar represents the mean of the maximum likelihood Binomial distributions fit to the percentage of wins over the clinical strategy and the error bars represent the 95% confidence intervals. The dashed line represents 50% wins over the clinical algorithm, indicating a comparable level of preference across both algorithm.

the array remained approximately the same as in quiet.

Both ED and FM N-of-M strategies presented fewer pulses on each electrode on average than the clinical strategy. ED N-of-M generally presented the lowest number of pulses across all three algorithms. The decrease in the number of pulses with respect to the clinical strategy was the result of the size of the perceptual sets defined by each reference electrodes. For each time window, only one electrode was activated for each perceptual set. Therefore, the more electrodes there were in each set the more likely it was that ED N-of-M would select fewer electrodes than the listener’s clinical number of maxima. This decrease in the number of pulses presented by ED N-of-M was especially pronounced for low frequency electrodes. Dropping pulses that encoded low frequency information could have resulted in a degraded representation of voiced phonemes. Moreover, this drop in the number of pulses could have also caused sound levels to be perceived as softer. A combination of these two factors could explain the lower speech recognition scores S2 and S3 obtained when using the ED N-of-M strategy in quiet. While FM N-of-M also presented fewer pulses than the clinical strategy, the decrease was not as severe as for ED. This can be explained by the different methods used to select electrodes employed by each algorithm. The measurements used to inform ED-based selection are binary: the percept generated by an electrode is either discriminable or indistinguishable from the percepts generated by neighboring electrodes. For this reason, if an electrode in a given perceptual set is activated for a given time window, neighboring indistinguishable electrodes are dropped. In contrast, the T_{shift} values measured to quantify FM, which can take on any number of values, provide finer measurement resolution. This results in a less aggressive pulse-dropping strategy.

While the electrode activation histograms show how many pulses were presented on each electrode across all words used in testing, they do not provide any insight into how these pulses were distributed across the array for each time window. The

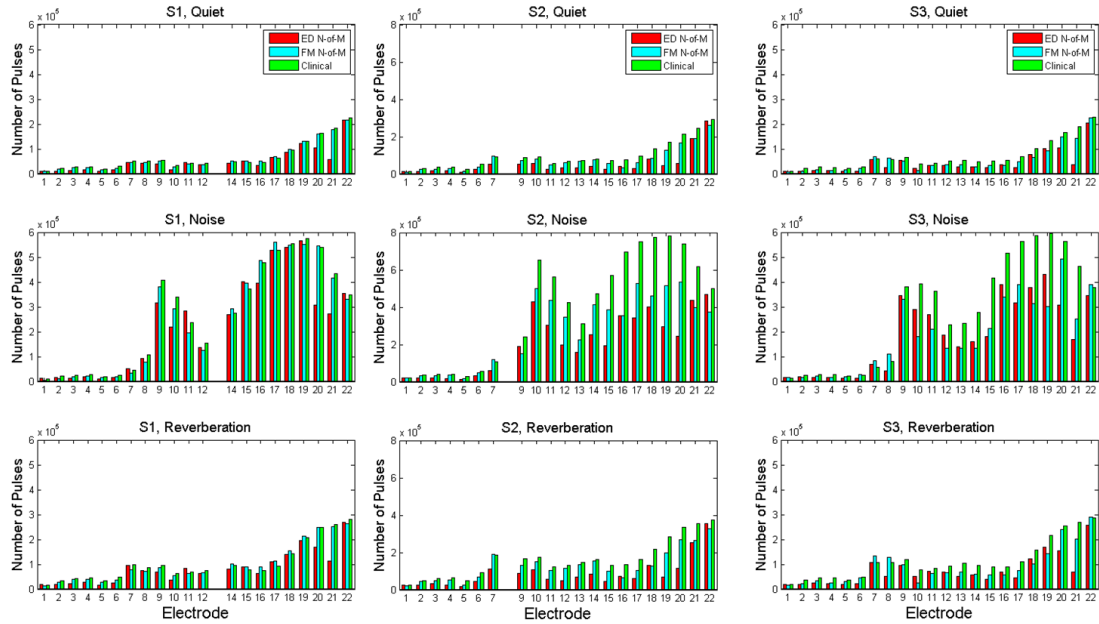


FIGURE 4.7: Number of pulses presented on each electrode when using each listener’s ED, FM and clinical N-of-M algorithms. The results plotted include the number of pulses across all 792 words used for testing. Each column corresponds to a single subject and each row corresponds to a particular listening condition – quiet, 10 dB of SSN or reverberation with $RT_{60} = 0.45$ s. The red, cyan and green bars correspond to the ED, FM and clinical N-of-M algorithms, respectively.

spread of the pulses across the array was quantified using electrode cluster. An electrode cluster is a group of adjacent electrodes activated during a given window. The length of a cluster corresponds to the number of electrodes that form it. For example, the center panel in in Figure 4.1 shows a time window with two clusters of length 1 and 3 while the right panel displays a stimulus with four clusters all of length 1. For a time window with a given number of pulses, more shorter clusters are assumed to cause fewer electrode interactions than fewer longer clusters. The cluster lengths for all 792 CNC words processed using each listener’s experimental and clinical algorithms are included in Figure 4.8. The maximum length for any given cluster is equal to the maximum number of pulses selected for a given time

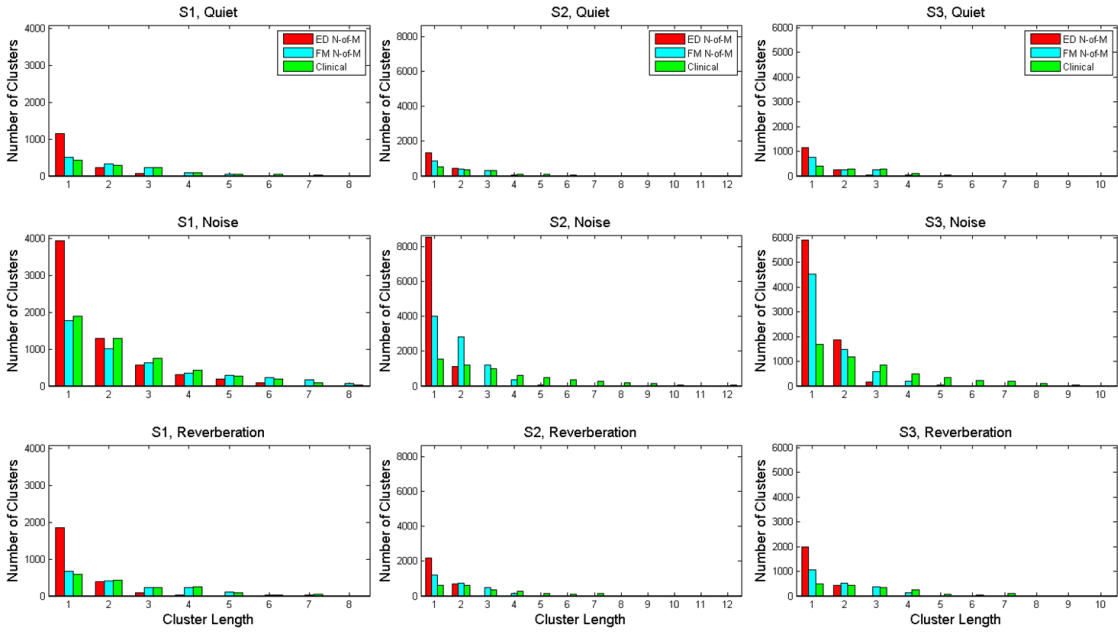


FIGURE 4.8: Cluster length histograms for each time window of all 792 words used for testing. The length of a cluster is defined as the number of adjacent electrodes activated within a given time window. Each word was processed with a listener’s ED, FM and clinical N-of-M algorithms, indicated by the red, cyan and green bars, respectively. Each column corresponds to a single subject and each row corresponds to a particular listening condition – quiet, 10 dB of SSN or reverberation with $RT_{60} = 0.45$ s.

window, or the clinical number of maxima. As expected, the cluster lengths for the psychophysics-motivated selection algorithms are generally shorter. The electrode activation histograms indicate that this is, at least in part, due to the experimental algorithms presenting fewer pulses overall.

4.4 Discussion

The speech recognition results indicate that listener’s speech recognition performance did not improve as a result of using the psychophysics-motivated algorithms. One reason could be the psychophysical data used to inform these algorithms. These data

were collected using pulse trains ranging from 10 to 300 ms in duration. However, in the implementation of the ED and FM based strategies, this data is used to predict interactions assumed to happen on a pulse by pulse level. There is, therefore, a mismatch between the larger scale on which the psychophysical testing was conducted and the smaller scale at which the pulsation interactions are assumed to occur in the dynamic electrode selection algorithm. This mismatch may not have been as severe for previous studies that used psychophysical metrics obtained using similar stimuli to exclude the electrodes across all time windows (e.g. [1] [6] [7]). These strategies implicitly assumed that electrode interactions occurred across multiple time windows. The psychophysical measures used for electrode selection in these studies may have provided more accurate estimates of these larger scale electrode interactions. For electrode selection strategies that operate on a time window scale, incorporating psychophysical data measured using stimuli that are more indicative of the interactions that occur within a given time window may improve performance.

Another potential reason the experimental strategies did not result in improved speech recognition could have been loudness. When using the experimental strategies, listeners were generally presented with fewer pulses per time window than when using their clinical strategies. As a result, the experimental algorithms could have presented stimuli with lower perceived loudness. This decrease in loudness could have had a negative effect on speech recognition (e.g. [46]). A listener-controlled increase in C levels could increase the perceived loudness of the stimuli presented via the experimental strategy while ensuring that they remain within a comfortable range.

In this work, it was assumed that speech recognition could be improved by dynamically dropping electrodes using two different psychophysical measures, regardless of their overall level across the array. In quiet, the dynamic electrode selection speech recognition results show that for all listeners one experimental algorithm performed

comparably to the clinical algorithm while the other performs worse. The algorithm that they achieved higher scores with generally corresponded to the metric that predicts a higher potential to confound, on average, for all electrodes across the array. This is consistent with the assumption that higher values of the psychophysical data may be more indicative of underlying physiological conditions. When these metrics with higher overall values are used to inform electrode selection, listeners can be presented with fewer pulses distributed differently across the array and still achieve similar speech recognition scores to those obtained using their clinical algorithm. When electrode selection was informed by a metric with lower overall potential to confound across the array, listeners generally performed worse than with their clinical strategy. Assuming that lower values of psychophysical metrics may be less indicative of an underlying physiological condition, dropping electrodes using these metrics is unlikely to result in more perceived information. Furthermore, if measurement noise raises the level of these psychophysical metrics, dropping information according to it could result in a loss of perceivable information.

In contrast to the results in quiet, when listeners were presented with corrupted speech the performance across all algorithms was comparable. One possible explanation for this outcome is that listeners may comprehend words in a process akin to template matching. After years of listening to speech with their clinical algorithm, listeners may have developed a perceptual template for each speech token. The more corrupted the stimulus presented to a listener, the lower the correlation would be with the corresponding template. In the speech recognition task presented in this work, listeners encountered two potential sources of corruption: the experimental electrode selection algorithm and the noisy or reverberant listening condition. In quiet, the only source of potential corruption is the experimental stimulation strategy. In this case, listeners may find it relatively easy to match a stimulus presented using ED or FM N-of-M with its corresponding template. However, if a second

source of corruption is introduced in the form of noise and reverberation, the stimulus presented would likely be less correlated with its corresponding template. In the speech recognition task presented in this work, it is difficult to evaluate whether the corruption caused by the stimulation strategy or the listening condition had a larger influence on performance. One possible way to estimate these effects would be to perform a multidimensional scaling task, where the effects of each type of corruption on speech recognition could be assessed individually. The experimental speech processing strategy could then be modified to minimize the corruption it causes.

The cluster length results show that the psychophysics-based electrode selection algorithms decrease the number of adjacent electrodes that are activated within a given time window. The activation results indicate that this is, at least in part, accomplished by presenting fewer than the clinical number of maxima, N , per time window. This decrease in the number of pulses is a result of the psychophysics-based electrode strategies prioritizing the presentation of perceivable information over presenting as many pulses per time window as the clinical strategy. Subthreshold null pulses are inserted in these cases to maintain a constant pulse rate but these pulses are not perceivable. One could argue that if the psychophysical data were an ideal indicator of an electrode's underlying potential to confound, presenting fewer pulses per time window than a listener is accustomed to should not cause a decrease in speech recognition. However, it is possible that listeners derive some additional information from the redundancy provided by additional indiscriminable stimuli. The results of this study suggest there may be a trade off between enforcing psychophysics-based selection and presenting enough stimuli per time window. To address this issue, future versions of these psychophysics-motivated strategies could explore an alternative method for selecting electrodes. The proposed algorithm would select the N highest energy electrodes such that the intersection between their corresponding perceptual sets is minimized. This strategy would preserve the clinical number of maxima while

still incorporating listener-specific measures of an electrode’s potential to confound speech recognition.

In this work, two psychophysics-motivated electrode selections were investigated. Each method attempts to maximize the perceivable information presented per time window by using listener-specific psychophysical data. Because no electrodes were permanently deactivated, the frequencies corresponding to each electrode remained unchanged thereby avoiding any potential negative impact of frequency reallocation (e.g. [18], [19], [17]). One alternative to this dynamic electrode selection strategy is to use listener-specific psychophysical data to drop electrodes statically, that is, across all time windows. Previous studies have found that for some listeners this strategy can improve speech recognition [1] [6] [7]. One disadvantage of this method is that the number of electrodes to exclude from the map must be determined a priori. Performance as a function all possible psychophysics-motivated electrode sets could be evaluated exhaustively but this would be difficult to accomplish in a clinically-relevant amount of time. To mitigate this problem, Chapter 5 presents an adaptive procedure for estimating performance as function of electrode set in a time-efficient manner.

Adaptive Parameter Selection Procedure

CI processing strategies have a number of parameters that can be adjusted, such as pulse rate, pulse width, frequency allocation, and number of active electrodes. Previous work suggests that the combination of these parameters that results in the highest speech recognition is listener dependent [20], [21], [22]. Ideally, speech recognition as a function of each of these parameter sets would be evaluated exhaustively but this would require a prohibitive amount of experimental time. In order to find a parameter set that results in high speech recognition without exhaustively evaluating all parameter combinations, previous studies involving CI and hearing aid users have used a genetic algorithm [23] [24] [25]. For each iteration, the listener was presented with speech tokens processed using different parameter sets and was instructed to choose the one they preferred. The next generation of tokens is selected by combining the parameter sets that resulted in the highest fitness evaluation using a series of biologically-motivated operations. One disadvantage of using a genetic algorithm is that, while it allows for navigation of a complex parameter space, it outputs a single parameter set. This point estimate makes it difficult to evaluate the fitness of the final parameter set compared to the other available alternatives. This concern is

addressed by the adaptive parameter selection procedure proposed in this work. The procedure uses Bayesian inference to navigate a parameter space of interest based on listener’s speech recognition scores and provides a final estimate of performance for every parameter set considered.

As with other adaptive procedures used with CI listeners [36], the procedure proposed in this work consists of two phases: parameter estimation and sample placement. The parameter estimation stage estimates performance for each parameter set given the listener’s past responses while the sample placement stage determines which parameter set to sample next. In this context, sampling is defined as presenting a speech token using a particular set of parameters. Which parameter set is used to present the next token depends on the final goal of the procedure. In this work, two different operation modes are defined depending on the desired outcome of the procedure. The goal of the first mode of operation is to estimate performance with equal certainty as a function of all the parameter sets considered. The objective of the second mode of operation is to find the parameter set or sets that results in the highest speech recognition scores.

5.1 Parameter Estimation

The goal of the adaptive procedure is to estimate the probability of correct token identification, p_n , for each CI parameter set, denoted by the subscript n . The certainty about the estimate is modeled as a beta distribution. This distribution was chosen because it is conjugate to the binomial likelihood used to the probability of obtaining a certain number of correct responses over a given number of iterations. The prior distribution on the probability of correct identification can therefore be described by the following probability density function:

$$q_{p_n}(p_n) = \frac{p_n^{\alpha-1}(1-p_n)^{\beta-1}}{B(\alpha, \beta)}, \quad (5.1)$$

where α and β are the parameters of the Beta distribution and $B(\alpha, \beta)$ is the beta function. The procedure does not assume any prior knowledge about the values of p_n . This is incorporated into the prior distribution by setting α and β to 1. This results in a uniform probability density across all values of p_n .

In order to estimate the posterior values of p_n , the procedure uses the number of speech tokens presented to the listener as well as the number of correct responses. At iteration of the procedure, i , the listener is presented with $k_{n,i}$ speech tokens, where n indicates the parameter set used to present the tokens. In this context, a speech token can be defined as a phoneme, a word, or a sentence. For instance, if the speech material consisted of CNC words graded by phoneme, $k_{n,i}$ would equal 3. The number of correctly identified tokens presented using parameter set n and during iteration i is denoted by $r_{n,i}$, a variable that can take on any value from 0 to $k_{n,i}$. These variables can also be calculated over all iterations up to iteration i as defined in Equations 5.2 and 5.3.

$$K_{n,i} = \sum_{j=1}^i k_{n,j} \quad (5.2)$$

$$R_{n,i} = \sum_{j=1}^i r_{n,j} \quad (5.3)$$

The probability of obtaining a certain number of correct responses over 1 to i iterations using parameter set n can be described by a binomial distribution:

$$f_{R_{n,i}}(R_{n,i}|K_{n,i}, p_n) = \binom{K_{n,i}}{R_{n,i}} p_n^{R_{n,i}} (1-p_n)^{K_{n,i}-R_{n,i}} \quad (5.4)$$

Using Bayes rule, the updated estimate of the distribution on p_n , denoted as the posterior distribution $f_{p_n}(p_n|K_{n,i}, R_{n,i})$, can be expressed as a function of the prior distribution $q_{p_n}(p_n)$ and the likelihood $f_{R_{n,i}}(R_{n,i}|K_{n,i}, p_n)$:

$$f_{p_n}(p_n|K_{n,i}, R_{n,i}) = \frac{f_{R_{n,i}}(R_{n,i}|K_{n,i}, p_n)q_{p_n}(p_n)}{l_{R_{n,i}}(R_{n,i}|K_{n,i})}. \quad (5.5)$$

$l_{R_{n,i}}(R_{n,i}|K_{n,i})$ is referred to as the evidence and is defined by the following expression:

$$l_{R_{n,i}}(R_{n,i}|K_{n,i}) = \int_{p_n} f_{R_{n,i}}(R_{n,i}|K_{n,i}, p_n)q_{p_n}(p_n)dp_n \quad (5.6)$$

For each iteration, the procedure uses the prior distribution and the likelihood to update the posterior distribution on p_n . The evidence, which is effectively a normalizing constant, can sometimes be difficult to compute. One common technique for bypassing this potential issue is to make use of conjugate priors.

A conjugate prior is defined as a distribution that, when combined with a particular likelihood, yields a posterior of the same functional form. If the likelihood is binomial distributed, as is the case in the assumed model, the corresponding conjugate prior is beta distributed. The advantage of using a conjugate prior is that the parameters of the posterior can be easily calculated from the parameters of the prior and the likelihood. Using this construct, the posterior distribution on p_n is updated for each iteration using the parameters of the prior distribution, the number of tokens presented and the number of correct responses:

$$f_{p_n}(p_n|K_{n,i}, R_{n,i}) \sim \text{Beta}(\alpha + r_{n,i}, \beta + K_{n,i} - R_{n,i}). \quad (5.7)$$

After a token is presented to the listener, Equation 5.7 is used to update the estimate of the probability of correct identification given the previously collected

data. The adaptive procedure must then determine the parameter set to sample next.

5.2 Sample Placement

For each iteration, a speech token is presented to the listener. Once the number of correct responses is recorded, the posterior estimate on the probability of correct information is updated. The procedure must then determine which parameter set to use to present the next token. This step is referred to as sample placement. In this work, two different methods of sample placement are defined. The goal of the curve fit mode is to obtain approximately equal confidence estimates for all parameter sets considered [47]. To do so, at each iteration it samples the parameter set with the highest expected information gain. The objective of the maximum performance find mode is to identify the parameter set that results in the highest probability of correct token identification. To accomplish this, it samples the parameter set with the highest probability of changing the parameter set believed to have the highest probability of correct identification.

5.2.1 *Curve Fit Mode*

The objective of the curve fit mode sample placement is to estimate performance as a function of CI parameters with approximately equal confidence across all possible parameter sets [47]. In order to accomplish this, the selection of the next parameter set is driven by information gain, where information is defined as the difference between the posterior and the prior distributions. At each iteration, the procedure uses the previously collected data to estimate which parameter set, if sampled, has the potential to provide the largest information gain. This approach was adapted from a sensor management application [47], where it was used to estimate the probability of each discrete cell containing a target. In this work, the search space is the set of

CI parameters and the estimated parameter of interest is the probability of correct speech token identification.

After a speech token is presented during iteration i with a given parameter set n , the posterior distribution $f_{p_n,i}(p_n|K_{n,i}, R_{n,i})$ corresponding to that parameter is updated using the number of correctly identified tokens $r_{n,i}$ as described in Section 5.1. In order to obtain a measure of difference between the updated posterior and the prior, the Kullback-Leibler divergence (KLD) between the two distributions is calculated using the following expression:

$$D_{KL}(f_{p_n,i}||q_{p_n})|r_{n,i} = \sum_{p_n} f_{p_n,i}(p_n|K_{n,i}, R_{n,i}) \ln \left(\frac{f_{p_n,i}(p_n|K_{n,i}, R_{n,i})}{q_{p_n}(p_n)} \right). \quad (5.8)$$

The KLD provides a measure of how much the posterior updates, driven by number of tokens presented and the number of tokens correctly identified, have changed the original prior. This difference can be defined as the amount of information provided by the observed data. Small KLD values would therefore be indicative of low information with respect to the prior while large values would be indicative of high information with respect to the prior. Using the procedure's Bayesian framework and assuming a discrete set of possible outcomes, these values can be estimated for future iterations without actually presenting additional tokens.

The process of estimating future performance outcomes from the previously collected data is referred to as hypothetical sampling. The first step in this process is to obtain the value of the next iteration's posteriors, $f_{p_n,i+1}(p_n|K_{n,i+1}, R_{n,i+1})$ by updating the current iteration posteriors for each possible value of $r_{n,i+1}$. The posterior corresponding to the current iteration would become the prior in the hypothetical posterior update described by

$$f_{p_n, i+1}(p_n | K_{n, i+1}, R_{n, i+1}) \sim \text{Beta}(\alpha + R_{n, i} + r_{n, i+1}, \beta + K_{n, i} - R_{n, i} - r_{n, i+1}). \quad (5.9)$$

The probability of obtaining each of these hypothetical posteriors is dependent on the likelihood of each possible value of $r_{n, i+1}$. The probability of each possible outcome during the next iteration can be calculated using the posterior distributions for the current iteration and the number of correctly identified tokens for each parameter set using the following expression:

$$\Pr(r_{n, i+1} | R_{n, i}) = \int_{p_n} \binom{k}{r_{n, i+1}} p_n^{r_{n, i+1}} (1 - p_n)^{k - r_{n, i+1}} f_{p_n, i}(p_n | K_{n, i}, R_{n, i}) dp_n. \quad (5.10)$$

Using both the hypothetical posterior distributions and the probability of obtaining each value of $r_{n, i+1}$, the expected KLD between the the hypothetical posterior and the prior can be calculated using the following expression, where the hypothetical posterior corresponding to a given value of $r_{n, i+1}$ is weighted by the probability of obtaining said value of $r_{n, i+1}$:

$$E[D_{KL}(f_{p_n, i+1} || q_{p_n})] = \sum_{j=0}^k (D_{KL}(f_{p_n, i+1} || q_{p_n}) | n, r_{n, i+1} = j) \Pr(r_{n, i+1} = j | R_{n, i}). \quad (5.11)$$

$E[D_{KL}(f_{p_n, i+1} || q_{p_n})]$ is therefore a weighted sum of the KLDs between the hypothetical posteriors and the priors over all possible values of $r_{n, i+1}$. It can be thought of as a measure of the expected amount of information that would be provided by sampling at parameter set n during the next iteration. This measure can be compared to the information provided by the current posterior in order to determine the amount of information that would be gained by collecting an additional sample at parameter set n . Since both $E[D_{KL}(f_{p_n, i+1} || q_{p_n})]$ and $D_{KL}(f_{p_n, i} || q_{p_n}) | r_{n, i}$ are calculated by

comparing the posterior distributions to the original prior, the difference between these two quantities can provide an estimate of the potential information gain by sampling parameter set n . As in [47], this expected information gain is denoted by $\Delta D_{KL}(n)$ and is calculated as follows:

$$\Delta D_{KL}(n) = E[(D_{KL}(f_{p_{n,i+1}}||q_{p_n}))] - D_{KL}(f_{p_{n,i}}||q_{p_n}) \quad (5.12)$$

High values of $\Delta D_{KL}(n)$ indicate that sampling at parameter set n during the next iteration has a high probability of changing the estimate of the probability of correct identification for set n . In contrast, low values of $\Delta D_{KL}(n)$ indicate that the current and future estimates of the probability of correct identification are more likely to be similar. In order to maximize the procedure's certainty about all estimates, at each iteration the parameter set with the highest ΔD_{KL} is selected to present the next speech token.

5.2.2 Maximum Performance Find Mode

For some applications, obtaining equal confidence estimates for all parameter sets may not be desirable. This would be true if the objective was to find the parameter set that resulted in the highest probability of correct identification, p_n . In this case, one may want to obtain higher confidence estimates for parameter sets with a high p_n , even at the expense of lower confidence estimates for parameter sets with lower p_n .

The first step in this process is to calculate, for all pairwise comparisons of parameter sets n and j , the probability that p_n is greater than p_j . Given the distributions for both variables, this can be calculated using the following expression:

$$\Pr(p_n > p_j) = \int_{p_j} (1 - F_{p_n}(p_j)) f_{p_j}(p_j) dp_j, \quad (5.13)$$

where $F_{p_n}(p_j)$ is the cumulative distribution of p_n evaluated at p_j . Because the distributions are assumed to be independent and identically distributed, the probability that the probability of correct identification for parameter set n is greater than all other sets s can be computed as follows:

$$\Lambda_{in} = \prod_{j=1}^n \Pr(p_n > p_j), \forall n = 1..s. \quad (5.14)$$

For each iteration i , the values of Λ_{in} for all n parameter sets are collected in a vector Λ_i :

$$\Lambda_i = \frac{1}{C} [\Lambda_{i1} \quad \Lambda_{i2} \quad \cdots \quad \Lambda_{is}], \quad (5.15)$$

where C is a normalizing constant that ensures that the $\frac{1}{C} \sum_{n=1}^s \Lambda_{in} = 1$. Λ_i indicates the probability that each parameter set will yield the highest probability of correct identification.

These probabilities can be estimated for the next iteration of the procedure by hypothetically sampling all electrode sets as described in Section 5.2.1. This process involves two stages: estimating the posteriors for the next iteration for all possible performance outcomes across all parameter sets and estimating the probability that each of those performance outcomes will occur. Equation 5.9 can be used to estimate the posteriors for the next iteration, $f_{p_n, i+1}(p_n | K_{n, i+1}, R_{n, i_1})$, given all possible values of $r_{n, i+1}$. Those posteriors are then used to estimate, for all pairwise comparisons of parameter sets n and j , the probability that p_n is greater than p_j as described in Equation 5.13. For each parameter set n , the probability that parameter set n has a higher p_n than all other parameter sets after sampling one additional time can be computed as follows:

$$\Lambda_{i+1,n} = \prod_{j=1}^n \Pr(p_n > p_j), \forall n = 1..s$$

$$\Lambda_{i+1} = \frac{1}{C} [\Lambda_{i+1,1} \quad \Lambda_{i+1,2} \quad \cdots \quad \Lambda_{i+1,s}]$$
(5.16)

where $\Pr(p_n > p_j)$ is calculated using the hypothetical posterior distributions on p_n and p_j . For each parameter set n , Λ_{i+1} is calculated for each possible value of $r_{i+1,n}$. The expected probability that each parameter set n will yield the highest p_n can be computed by weighting each Λ_{i+1} vector by the probability of the outcome $r_{i+1,n}$ used to obtain it:

$$E[\Lambda_{i+1}|n] = \sum_{j=0}^k (\Lambda_{i+1}|r_{i+1,n}, n) \Pr(r_{i+1,n} = j|R_i)$$
(5.17)

where $E[\Lambda_{i+1}|n]$ is a vector with as many elements as parameter sets are considered, s . Because Λ_i and $E[\Lambda_{i+1}|n]$ both sum to 1 and their elements are non-zero, the KLD can be used to find a measure of the difference between them:

$$D_{KL}(n) = D_{KL}(\Lambda_i, E[\Lambda_{i+1}|n]), \forall n = 1..s.$$
(5.18)

Higher values of $D_{KL}(n)$ are indicative of parameter sets that, if sampled during the next iteration, have a high probability of changing the parameter set with the highest estimated probability of correct identification. This estimate is used by the maximum performance find mode of the adaptive procedure to decide which electrode set to sample next. Because the goal is to identify the set with the highest probability of correct identification, at each iteration it selects parameter the set with the highest corresponding $D_{KL}(n)$ to sample during the next iteration.

5.3 Adaptive Procedure Operation

The operation of the adaptive procedure is described in Figure 5.1. First, the prior distributions are initialized for each parameter set. Then, a token is presented with a given parameter set. For the first iteration of the process this set is randomly selected. For subsequent iterations, this set is determined by the sample method. Once the token is graded, the results are used to update the posterior distribution on the probability of correct identification for the parameter set used to present the token. All parameter sets are then hypothetically sampled by estimating the value of the posteriors for an additional sample placed at each of the available parameter sets. Using these hypothetical posteriors, the procedure selects the parameter set with which to present the next token using either the curve fit or maximum performance find sampling modes. This process is repeated until a maximum number of speech tokens are presented. This number of tokens can be determined by monitoring the posterior distributions and terminating the adaptive procedure when the desired level of certainty about the performance estimates has been achieved.

5.4 Verification Using Simulated Data

Before using the adaptive procedure to estimate listeners' speech recognition scores as a function of CI parameters, the effectiveness of the procedure was verified using simulated data. For the purposes of this example, 20 parameter sets were considered, each with an assumed true value of p_n^t . These values were computed using a scaled Gaussian curve with standard deviation σ equal to 10. The mean of the distribution μ was varied for each simulated run of the task. The simulated task was run 100 times. For each run, the mean of the Gaussian distribution changed such that a different parameter set had the highest value of p_n^t . Across all 100 runs of the procedure, each parameter set n had the highest p_n^t for five runs. The amplitude of the curve was

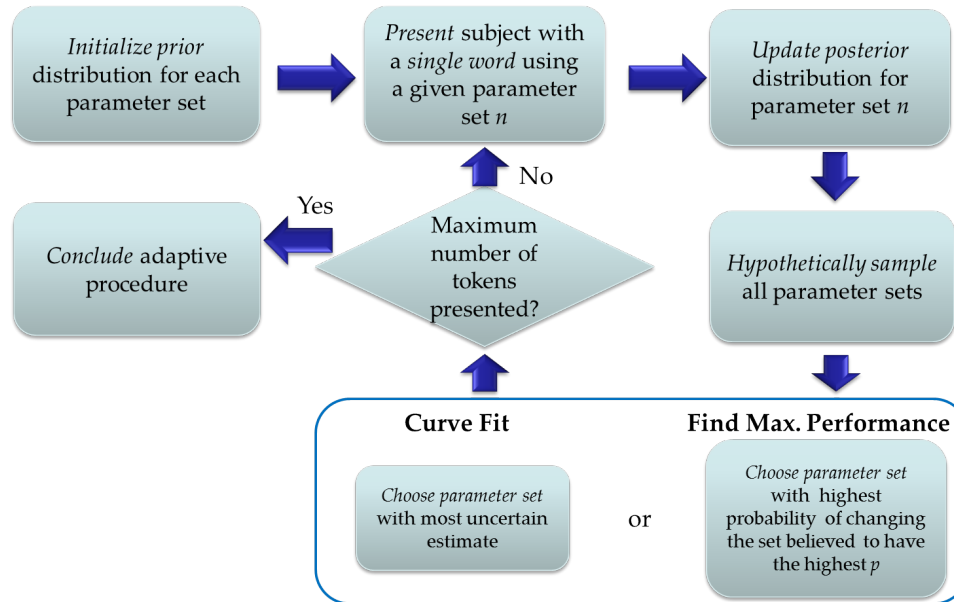


FIGURE 5.1: Graphical representation of the adaptive procedure’s operation. First the prior distributions on the probability of correct identification, p_n , are initialized for each parameter set. A speech token is then presented using a given parameter set. After the listener responds, the speech token is graded and the posterior distribution of the parameter used to present it is updated. All parameter sets are then hypothetically sampled by estimating the posteriors for each possible performance outcome. Using the hypothetical and current posteriors as well as the prior distributions, the next parameter set is selected using the curve fit or maximum performance find mode. This process is repeated until the desired number of tokens is presented.

scaled such that the maximum equaled 0.8, the approximate value of the listeners’ percent correct phoneme identification scores in quiet is shown in Figure 4.5.

For each iteration i , 3 tokens corresponding to the 3 phonemes in a CNC were presented to a given parameter set. The value of $r_{i,n}$, ranging from 0 to 3, was drawn from a binomial distribution with p_n^t equal to the value specified by the aforementioned Gaussian curve. For the first 60 iterations for the process, each of the 20 parameter sets were randomly sampled to initialize the prior distributions. For the following 240 iterations, the process selected the next sample as described by Figure 5.1. In total, each run of the process was comprised of a total of 300 iterations.

An example of the true p_n^t curve and the two adaptive procedure estimates are

shown in Figure 5.2. An examination of the means and 95% confidence intervals for the estimates indicate that both sample placement modes produced the expected results. For the curve fit mode, the means of the distributions corresponding to each parameter set approximately aligned with the true values of p_n^t . Additionally, the size of the 95% confidence intervals was comparable across all parameter sets, indicating approximately equal certainty across all estimates. For the maximum performance find mode, the means of the estimated distributions were less accurate for parameter sets with low p_n^t values and more accurate for the sets with high p_n^t values. The confidence on these estimates also varied depending on the underlying p_n^t . Parameter sets with low p_n^t values were sampled less frequently resulting in lower confidence estimates or larger 95% confidence intervals. In contrast, for parameter sets with higher p_n^t the confidence in the estimates was higher, as indicated by the smaller confidence intervals. For both placement methods, the confidence intervals encompassed the true p_n^t values.

While the example included in Figure 5.2 provides estimates for a single run, the conclusions drawn from it extend to the results across 100 runs. Figure 5.3 includes statistical measures of the accuracy of the estimates as well as the confidence on said estimates. Each point represents the average of the metric considered while the error bars denote 3 standard deviations. To evaluate the accuracy of the p_n estimates, the absolute difference between the true values of the probability of correct identification and the mean of the estimated values were calculated for all parameter sets. This difference was also computed between the set with the highest p_n^t and the set with the highest $\mu(p_n)$. When all parameter sets were considered, the curve fit mode provided more accurate estimates across all parameter sets. When only the parameter sets with the highest p_n^t and $\mu(p_n)$ were compared, the estimates provided by both sampling methods are approximately equal. The confidence on these estimates, both across all parameter sets and for the sets with the highest

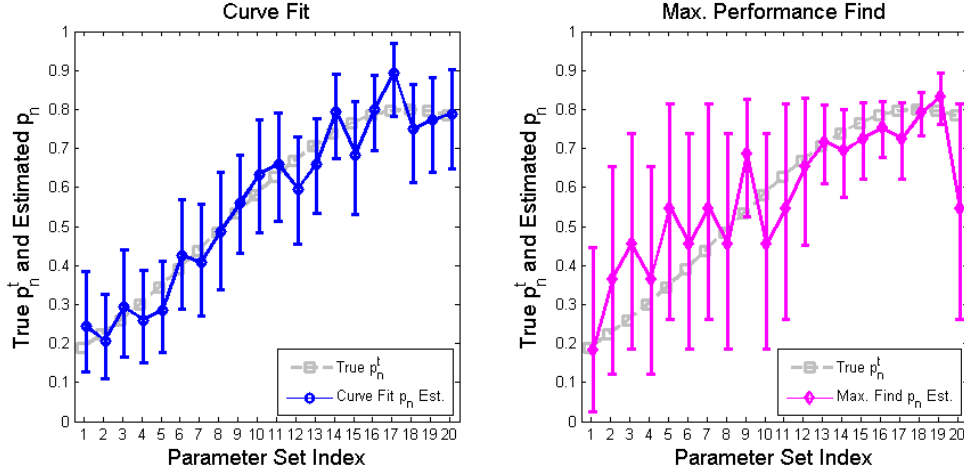


FIGURE 5.2: Simulated curve fit and maximum performance find approximations to the true p_n values. Each parameter index n is plotted versus the true values p_n^t and the estimated distributions on p_n . The true values of p_n^t for each parameter set follow an amplitude-scaled Gaussian curve with $\mu = 18$ and $\sigma = 10$ and are indicated by the dashed gray line. The blue and magenta curves indicate the curve fit and maximum performance find estimates. Each point on the curve indicates the mean of the final Beta distributions after 300 iterations and the error bars denote the 95% confidence intervals.

true and estimated probability of correct identification, also follow the expected trends. On average, the curve fit mode provides higher confidence estimates across all parameter sets while the maximum performance find mode provides higher confidence estimates for the parameter set with the highest estimated $\mu(p_n)$.

5.5 Listener Verification Task

After verifying the performance of the adaptive procedure using simulated data, a second verification task was conducted with CI listeners. Because of its documented effect on speech recognition, the parameter selected for the experiment was the number of active electrodes evenly distributed across the array (e.g. [20], [48]). For different numbers of active electrodes, listener’s performance was estimated using three different methods. First, their performance was evaluated exhaustively for all

electrode sets. Their performance was then estimated using the adaptive procedure’s curve fit and maximum performance find modes. These estimates were compared to the exhaustive performance estimates to evaluate the effectiveness of each of the procedure’s modes of operation.

5.5.1 Subjects

The four subjects who participated in the psychophysical data collection experiment also participated in the adaptive procedure verification tasks. Their demographic information is included in Table 3.1. Testing was completed in two sessions, each lasting one to four hours.

5.5.2 Stimuli and Equipment

The speech material used in this experiment consisted of the same set of 792 CNC words used in the dynamic electrode selection task. Further details about the methods for selecting and recording the words in the word corpus are included in Chapter 4. The number of words presented across all conditions exceeded the 792 unique words available. To minimize training effects, all 792 words were presented before allowing words selected during previous conditions to be presented a second time.

Prior to testing, subjects’ clinical maps were read from their processors using Custom Sound, Cochlear’s clinical software. All of their clinical parameters, including their Ts and Cs, were then transferred to a map file in the Nucleus MATLAB toolbox. The comfort of the C levels was verified by streaming 300 ms-long biphasic pulse trains at the corresponding level using the Nucleus Implant Communicator (NIC v2). This same research interface was used to stream all other experimental stimuli from a PC. All stimuli were presented as biphasic pulse trains with parameters corresponding to those read from each subjects’ map. Subjects’ clinical parameters are included in Table 3.1.

The number of electrodes used to present the stimuli was varied and included the following values: 4, 6, 8, 10, 12, and 21 or 22. S1 and S2 had one electrode deactivated in their clinical map, while S3 and S4 had all electrodes activated, therefore the maximum number of electrodes included in each of their maps was 21 and 22, respectively. For all conditions, the active electrodes were approximately evenly spaced across the array. The electrode distribution for each of the listener’s experimental maps is included in Figure 5.4. When the number of active electrodes in a given set was less than the clinical number of maxima, N , the number of maxima was set to the number of active electrodes in the set. If the number of active electrodes in a set was equal to or larger than N the number of maxima was set to N .

Because some electrodes were excluded from the map altogether, their frequencies had to be reassigned. The clinical method of frequency allocation assigns each of the 62 125-Hz wide frequency bins spanning 187.5 to 7937.5 Hz according to the total number of active electrodes in the map. In this task, a different approach was taken to minimize the disruption to the frequency allocation table. The frequencies corresponding to dropped electrodes were evenly distributed between the two adjacent electrodes. If the number of bins was odd, the additional frequency bin was assigned to the adjacent electrode in the basal direction, since the frequency bands corresponding to more basal channels are wider.

5.5.3 Procedure

At the start of each testing session, subjects were presented with a set of 50 training words in quiet processed with their clinical strategy. This training set was not scored; its purpose was to familiarize subjects with listening to the speech material using the research interface. The words included in the set were the same for all subjects but the order was randomized for each individual. These words were reserved for training purposes and were not used during testing.

Listeners' performance across the different electrode sets was evaluated exhaustively by presenting 150 CNC words per electrode set. The conditions corresponding to different numbers of dropped electrodes were presented in random order. Listener's performance as a function of electrode set was then evaluated using the curve fit and maximum performance find modes of the adaptive procedure. For each mode, 150 CNC words were presented across all conditions. The number of times each electrode set was sampled was determined by the adaptive procedure.

Listeners started a run of the task by either pressing the keyboard's space bar or clicking on the Ready button in the speech recognition task GUI shown in Figure 4.3. After a word was presented, listeners typed in what they heard into the text box shown in Figure 4.3. They submitted their response by pressing the enter key or clicking on the Submit button. During the runs corresponding to the adaptive procedure, the experimenter used a second keyboard to type in the number of phonemes the listener had correctly identified. The adaptive procedure used this score as well as previously collected data to select the electrode set that was used to present the next word. During the experimental runs corresponding to the exhaustive evaluation task, each word was presented 500 ms after the submission of the listener's previous response. In this case, words were graded by phoneme after the testing session had concluded.

5.6 Results

Listeners' speech recognition as a function of the number of active electrodes evenly distributed across the array was evaluated using four different methods: exhaustive evaluation, the adaptive procedure's curve fit and performance find modes and even sampling. This data for this last condition was obtained by randomly selecting 25 words from each of the exhaustive conditions. This even sampling condition simulated evenly distributing 150 words across all conditions instead of using the adaptive

procedure to present the same number of tokens. The results for the exhaustive and adaptive estimation methods are shown in Figure 5.5 and the results for the exhaustive and even sampling are shown in Figure 5.6. For the exhaustive and even sampling conditions, the results show the Beta distributions on the probability of correct identification that were fit to the data. For the adaptive conditions, the results show the final Beta distributions after the adaptive procedure had concluded. Each point represents the mean of the corresponding Beta distribution on the probability of correct phoneme identification and the error bars indicate 95% confidence intervals.

Because each electrode set was sampled most in the exhaustive condition, the corresponding results are considered the closest estimate of those obtained to the true underlying probability of correct identification. This certainty about the estimate is reflected in the small size of the confidence intervals. The shape of the performance curve, rising from 4 to 8 electrodes and then approximately flat for higher numbers of active electrodes, is consistent with previously reported results (e.g. [20], [48]). This plateau in performance as the number of active electrodes increases is likely due to interactions across the stimulation sites. This hypothesis is supported by the ED and FM data included in Chapter 3 that indicates these listeners have indiscriminable electrodes and measurable masking effects.

The exhaustive performance estimates are, in all but one case, encompassed by the confidence intervals of the curve fit estimates. The size of the confidence intervals across the curve fit estimates is approximately the same. This suggests that the adaptive procedure is performing as expected, obtaining accurate and equal confidence estimates for all electrode sets considered. Furthermore, when compared to the exhaustive condition, curve fit estimates were obtained using $\frac{1}{6}$ of the material and approximately $\frac{1}{6}$ of the experimental time.

The maximum performance find estimates were using the same amount of speech

material and in the same amount of time as the curve fit estimates. As expected, the confidence on these estimates is not equal across all electrode sets. Estimates for electrode sets with low speech recognition scores have lower confidence, as indicated by larger confidence intervals, and estimates for electrode sets with higher scores have higher confidence, indicated by smaller confidence intervals. This is consistent with the procedure's goal to find the set with the highest estimated speech recognition score. In some cases the probability of correct identification for electrode sets that have high speech recognitions scores in the exhaustive case is underestimated by the procedure. This can occur when during the first few iterations of the procedure the listener incorrectly identifies most of the phonemes presented to a given electrode set. The procedure then determines that further sampling at that electrode set does not have a high probability of changing the set with the highest probability of correct identification so it samples at other sets for the remainder of the testing run.

The curve fit and maximum performance find provide different estimates of the probability of correct phoneme identification. The exhaustive condition provides an upper bound on these estimates while the even sampling condition provides a lower bound. Despite being calculated using the same number of total words as the adaptive procedure estimates, do not generally indicate the electrode sets that have the highest probability of correct identification. These results suggest that, given the same number of speech tokens, a directed sampling approach can provide more accurate estimates than evenly sampling the parameter space.

5.7 Discussion

In computer simulations and in an experiment involving CI listeners, the adaptive procedure performed as expected in both sampling modes. When in curve fit mode, the procedure was able to approximate the listeners' performance as a function of number of electrodes. Likewise, when in maximum performance find mode, it was

able to identify either the set corresponding to the highest overall performance in the exhaustive condition or a set with comparable performance. In the experiment involving CI listeners, estimates were obtained using $\frac{1}{6}$ of the words and about $\frac{1}{6}$ of the experimental time as compared to the exhaustive condition.

In maximum performance find mode, each sample was placed at the electrode set that was most likely to change which set had the highest estimated probability of correct identification. This method generally caused the sets with lower probability of correct identification being sampled less frequently in favor of sampling the electrode sets with higher probability of correct identification more frequently. This resulted in less accurate and lower confidence estimates for lower performing electrode sets and more accurate, higher confidence estimates for the high performing electrode sets of interest, consistent with the goal of finding the highest performing set. However, the procedure occasionally underestimated the performance for a set with a corresponding high speech recognition score. This is likely due to the existence of multiple sets with comparably high underlying probability of correct identification. Occasionally, a large number of phonemes presented to one of these high performing sets were not correctly identified during the first few iterations, potentially due to word difficulty or a lapse in attention on the part of the listener. Because there were other high performing electrode sets in the search space, the adaptive procedure stopped sampling at these sets with artificially low probability estimates. Instead, the process sampled the other high performing electrode sets with a higher likelihood of changing the set with the maximum probability of correct identification. To mitigate this effect, future versions of the procedure could incorporate a probability of incorrectly identifying a whole speech token. This would account for the probability of a word being particularly hard to identify or the probability that the listener was distracted and missed the presentation of the word entirely.

While this adaptive procedure was developed to evaluate sets of electrodes, it is

not restricted to estimating performance as a function of electrode set. The same procedure can be used to estimate performance as a function of any combination of CI parameters – e.g. pulse rate, pulse width, amplitude compression– in a time-efficient manner. Additionally, the procedure is not restricted to using CNC words. The speech material used to estimate performance can be changed to vowels, consonants, non-CNC words or sentences. In its current implementation, the adaptive procedure assumes equal prior probability of correct identification for all parameter sets. It also assumes that the distributions corresponding to each set are independent. These assumptions were made for ease of implementation. However, the procedure’s Bayesian framework would allow the incorporation of different assumptions depending on the experiment in question. This versatile and time efficient procedure could provide an objective method for setting their clinical parameters.

Once the effectiveness of the adaptive procedure was verified using a small set of parameters, it was used to navigate a larger parameter space. In the experiment described in Chapter 6, psychophysics-motivated electrode sets were obtained by sequentially deactivating electrodes with the highest potential to confound speech recognition. The adaptive procedure was used estimate the electrode set resulting in the highest speech recognition score. The performance of the electrode set selected for each psychophysical metric was the compared to the performance of the full electrode array in order to assess whether these reduced electrode sets could improve listeners’ speech recognition.

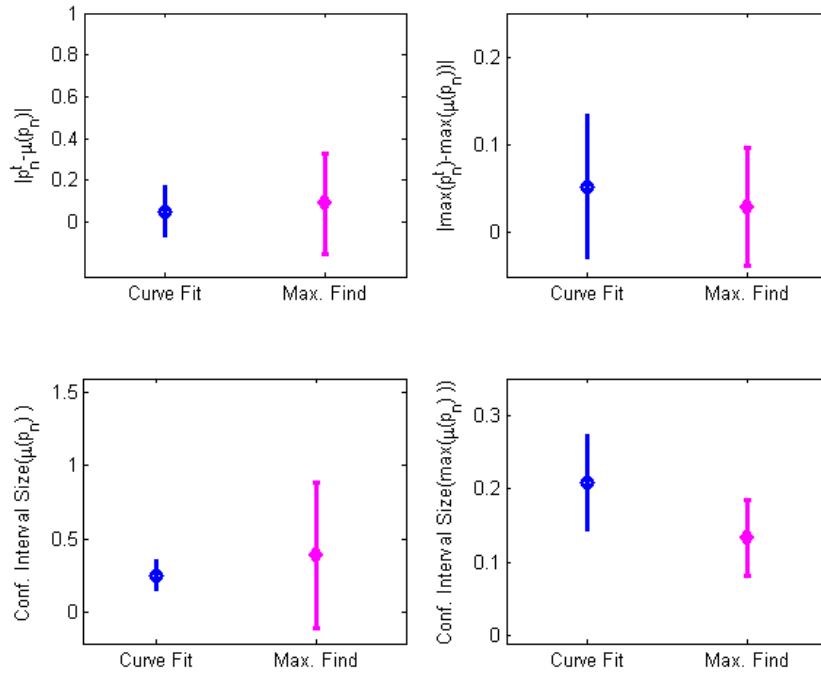


FIGURE 5.3: Statistics for 100 repetitions of the adaptive procedure. The two modes of the adaptive procedure are plotted versus the corresponding value of four different measures of the procedure’s accuracy. For each repetition of the procedure, the true p_n^t values were estimated using both the curve fit and maximum performance find mode. Each point marks the mean of different error rates across all repetitions and the error bars represent 3 standard deviations. The top panel includes the absolute difference between p_n^t and the mean of the estimates $\mu(p_n)$ across all n . The second panel shows the absolute difference between highest p_n^t and the highest estimated $\mu(p_n)$. The third panel shows the difference in the set n^t with the highest true value of p_n^t and the set n with the highest estimated p_n . The fourth panels includes the size of the 95% confidence intervals across all values of n . The bottom panel includes the size of the 95% confidence intervals for set n with the highest estimated $\mu(p_n)$.

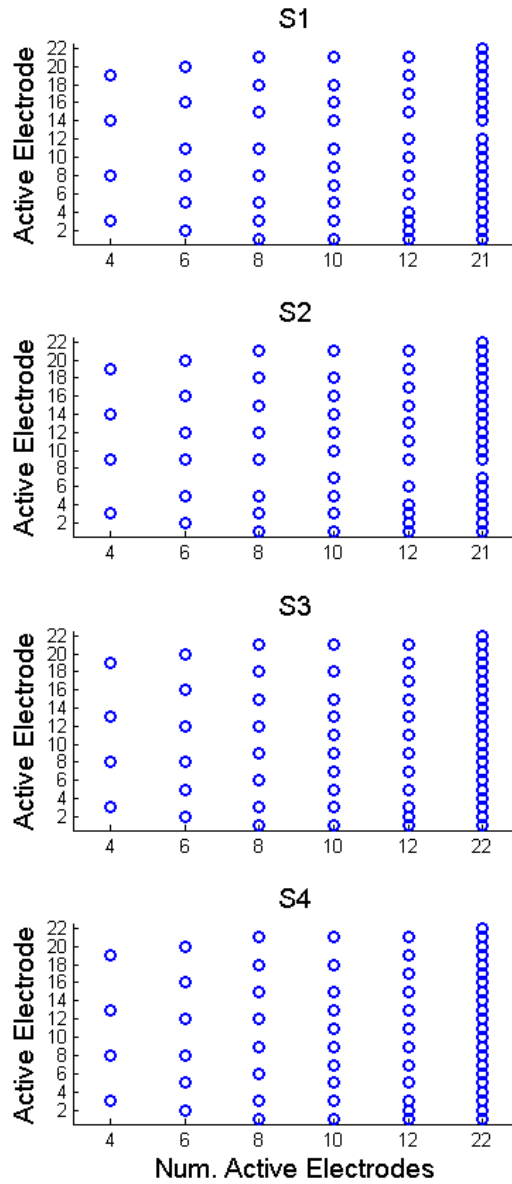


FIGURE 5.4: Electrode sets used for adaptive procedure verification task. Each set contained a varying number of active electrodes approximately evenly distributed across the array: 4, 6, 8, 10, 12 and 21 or 22. For each number of active electrodes, the electrodes in the experimental map are indicated with a circle. The frequency bins corresponding to the excluded electrodes were evenly distributed between the two adjacent electrodes. If the number of bins was odd, the additional frequency bin was assigned to the adjacent electrode in the basal direction.

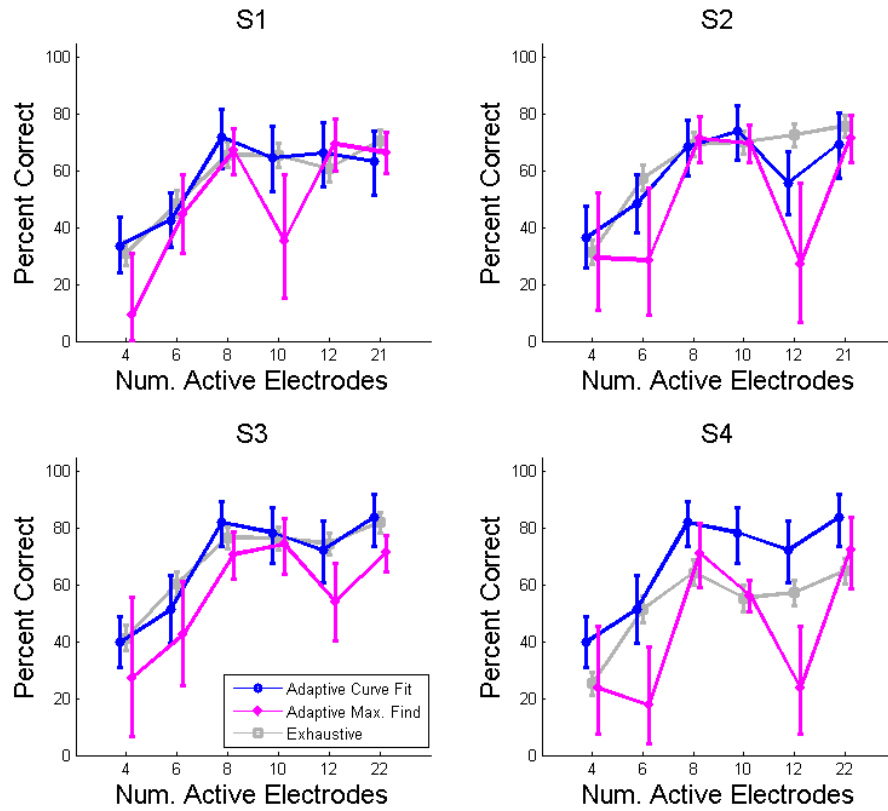


FIGURE 5.5: Exhaustive and adaptive procedure estimates as a function of number of active electrodes for four CI listeners. For the exhaustive performance estimate 150 CNC words were presented per condition. For the curve fit and maximum performance find modes of the adaptive procedure, 150 CNC words were presented across all conditions. Each point represents the mean of the Beta distribution on the probability of correct phoneme identification. The error bars indicate the 95% confidence intervals.

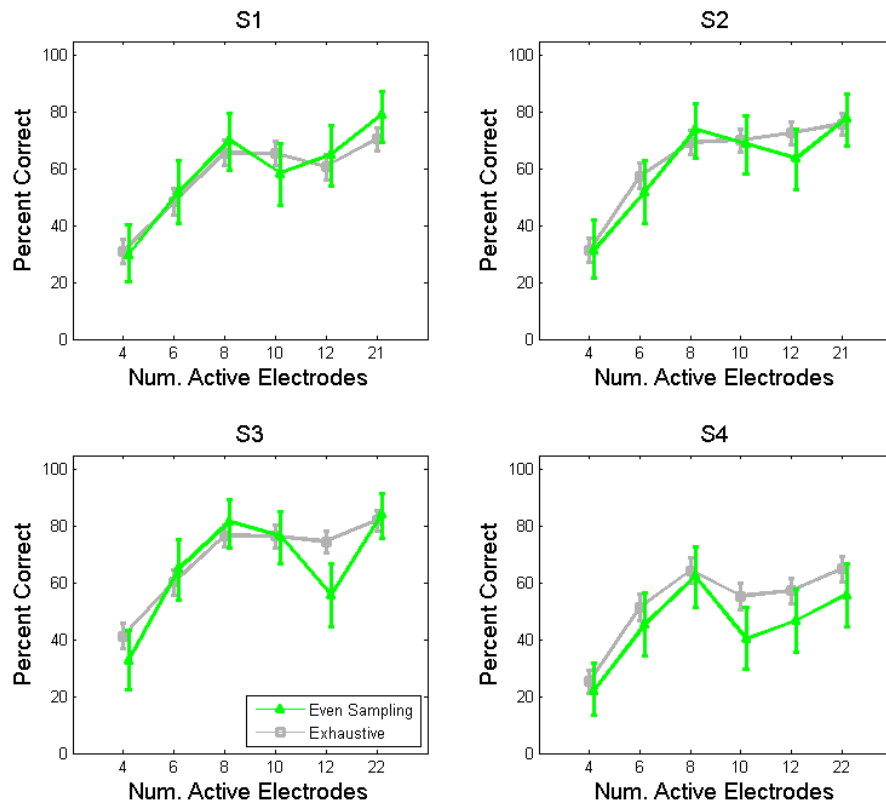


FIGURE 5.6: Exhaustive and even sampling estimates as a function of number of active electrodes for four CI listeners. For the exhaustive performance estimate 150 CNC words were presented per condition. Of those 150 words, 25 were randomly selected per electrode set. That data was used to simulate evenly sampling each electrode set using 25 words. Each point represents the mean of the Beta distribution on the probability of correct phoneme identification. The error bars indicate the 95% confidence intervals.

Static Electrode Selection Using Psychophysical Measures

Previous studies suggest that psychophysics-based electrode selection can result in improved performance for some CI listeners [1], [6], [7]. Each of these studies collected data corresponding to a single psychophysical metric and used it to determine the electrodes with the highest potential to negatively impact speech recognition. However, the psychophysical data included in Chapter 3 suggests that each electrode's potential negative impact speech recognition can vary considerably depending on the psychophysical metric used. Collecting multiple measures for each electrode may provide a multidimensional assessment of each electrode's potential to confound. Additionally, these studies did not consider the effect of varying the number of deactivated electrodes, presumably because evaluating a large number psychophysics-motivated electrode sets would be extremely time consuming. This limitation could be overcome by using a time-efficient method to estimate performance across a large number of electrode sets, such as the adaptive procedure presented in Chapter 5.

In this work, the ED, FM and MDT data included in Chapter 3 were used to obtain three different estimates of each electrode's potential to impact speech recog-

nition. For each metric, a group of electrode sets was determined by sequentially excluding the electrodes with the highest potential to confound. The adaptive procedure presented in Chapter 5 was used to estimate the electrode set that resulted in the highest performance. The set selected for each metric was then exhaustively evaluated and its performance was compared to the performance using each listeners clinical set.

6.1 Subjects

The four subjects who participated in the psychophysical data collection experiment also participated in the adaptive procedure verification tasks. Their demographic information is included in Table 3.1. Testing was completed in three sessions, each lasting two to four hours.

6.2 Stimuli and Equipment

The speech material used in this experiment consisted of the same set of 792 CNC words used in the dynamic electrode and the adaptive procedure verification tasks. Further details about the methods for selecting and recording the words in the word corpus are included in Chapter 4. As in the previous experiments, the number of words presented across static electrode selection tasks exceeded the 792 unique words available. To minimize training effects, all 792 words were presented before allowing words selected during previous conditions to be presented a second time.

Prior to testing, subjects' clinical maps were read from their processors using Custom Sound, Cochlear's clinical software. All of their clinical parameters, including their Ts and Cs, were then transferred to a map file in the Nucleus MATLAB toolbox. The comfort of the C levels was verified by streaming 300 ms-long biphasic pulse trains at the corresponding level using the Nucleus Implant Communicator

(NIC v2). This same research interface was used to stream all other experimental stimuli from a PC.

The number of electrodes used to present the stimuli consisted of either the listeners clinical set or a psychophysics-motivated set. ED, FM and MDT-based sets were determined via a sequential process. For the first set, the electrode with the highest potential to confound was excluded from the array. For the next set, the two electrodes with the highest potential to confound were deactivated. If a group of e electrodes had identical psychophysical metric values, $e + 1$ sets were formed by excluding each of the electrodes in the group sequentially and then excluding all of them. This process was repeated, with an additional electrode being excluded from the array until 4 electrodes remained. The ED-based electrode sets also included a set determined using the same method as Zwolan et al. used in their electrode selection study [1]. The set was determined by first selecting the most basal electrode, e.g. electrode 1. The closest discriminable electrode in the basal direction, e.g. electrode 3, was also included. Then the closest discriminable electrode to 3 in the basal direction was also included. This process was repeated until the most basal end of the array was reached. The sets determined in this way contained an electrode from each discriminable group of electrodes. For all metrics, each listener's clinical electrode set was also included. Figure 6.1 includes the psychophysics-based sets for each of the four listeners who participated in the experiment.

All tokens were processed with each listener's clinical speech processing strategy. All stimuli were presented as biphasic pulse trains with parameters corresponding to those read from each subjects' map (included in Table 3.1) except for the number of active electrodes and the number of maxima. When the number of electrodes in a given set was less than the clinical number of maxima, N , the number of maxima was set to the number of electrodes in the set. If the number of electrodes in a set was equal to or larger than N the number of maxima was set to N .

As in the adaptive verification task, the frequencies corresponding to the dropped electrodes were reassigned. The clinical method of frequency allocation assigns each of the 62 125-Hz wide frequency bins spanning 187.5 to 7937.5 Hz according to the total number of active electrodes in the map. In this task, the frequencies corresponding to dropped electrodes were evenly distributed between the two adjacent electrodes. If the number of bins was odd, the additional frequency bin was assigned to the adjacent electrode in the basal direction, since the frequency bands corresponding to more basal channels are wider.

6.3 Procedure

At the start of each testing session, listeners were presented with 50 CNC words in quiet processed with their clinical map. This training set was not graded; its purpose was to familiarize subjects with listening to the speech material through the research interface.

Once the training concluded, the adaptive procedure’s maximum performance find mode was used to estimate the electrode set with the highest p_n across all sets determined by a given psychophysical parameter. For each metric, ED, FM and MDT, the total number of electrode sets s varied from 18 to 26. For each group of electrode sets, the adaptive procedure used 300 CNC words, one word presented per iteration, to estimate the electrode set with the highest probability of having a higher p_n than all other electrode sets. For the first $3s$ iterations, each electrode set was sampled 3 times in random order to initialize the prior distributions. During the remaining $300 - 3s$ iterations the next electrode set to sample was selected according to the maximum performance find criterion. The performance of the selected set for each metric was validated using a large set of 150 words. The order of the validation for the psychophysics-based sets was also selected randomly. The performance of the clinical electrode set was also evaluated using 150 words and the results were

compared to those of obtained with the psychophysics-based sets.

6.4 Results

Figure 6.1 includes the psychophysics-motivated electrode sets for each subject. The boxes indicated the electrode set that had the highest estimated probability of outperforming all other sets. While the adaptive procedure generally selected reduced sets, when evaluating the performance of S2's MDT sets and S3's FM sets it determined that the clinical set would provide the best performance. It was hypothesized that this would occur when the metric considered had, on average, low values across the array. This hypothesis was not supported by S2's MDT and S3's FM data. This suggests that, for some listeners, sequentially excluding electrodes using a particular psychophysical metric may not provide a benefit over using their clinical electrode set.

This finding is also supported by the validation results included in Figure 6.2. For most of the selected psychophysics-motivated sets, performance was comparable to the clinical condition. Since listeners are able to achieve high recognition scores using reduced electrode sets, it is likely they are not perceiving all the information presented to them when using their clinical electrode set. For some psychophysics-motivated sets, performance was slightly worse than clinical. This suggests that the adaptive procedure's maximum performance find mode may have either overestimated the performance of the selected set or underestimated the performance of the clinical set, preventing it from being selected.

In order to compare the estimates obtained across all electrode sets, the final values for several of the adaptive procedure's parameters were examined. Figures 6.3 through 6.6 include the final posterior distributions, the final probability that each electrode set had a higher p_n than all other sets, the number of words presented using each electrode set and the number of phonemes correctly identified for S1 through S4,

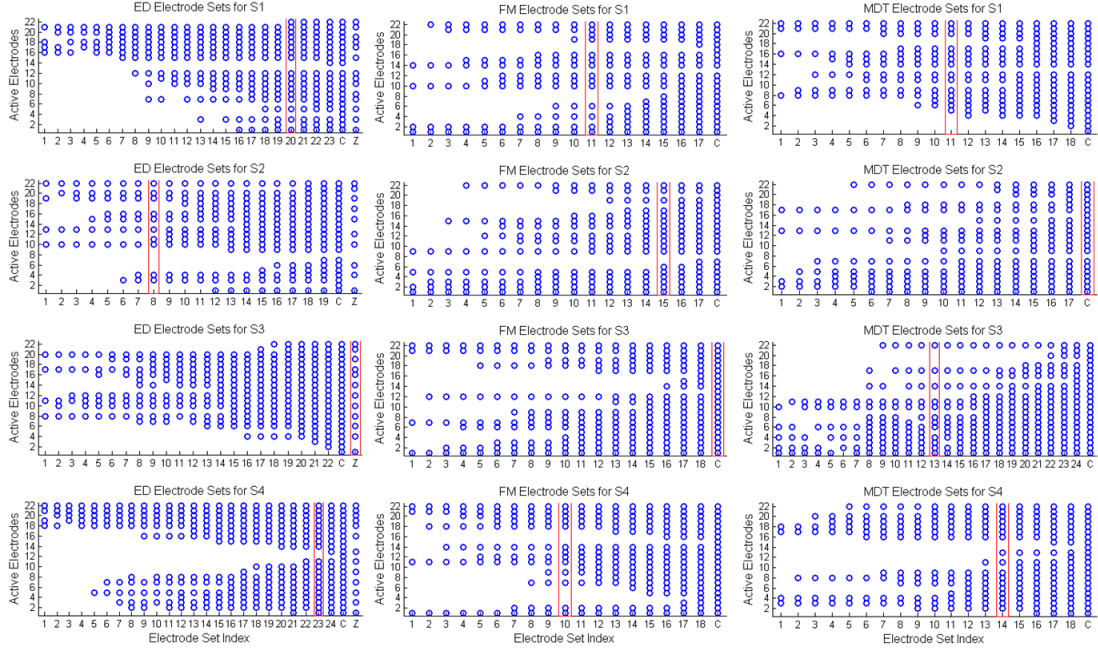


FIGURE 6.1: ED, FM and MDT defined electrode sets defined static electrode selection. Each column corresponds to the psychophysical metric used to obtain the electrode sets. For each plot, the electrode set index is plotted versus the active electrodes included in the corresponding set. The electrode set labeled as C correspond to the listeners clinical set. The set labeled as Z corresponds to the set determined using the same method as Zwolan et al. [1]. For each subject and each metric, the electrode sets marked by a box correspond to those the adaptive procedure estimated had the highest corresponding probability of correct phoneme identification.

respectively. These results indicate that for certain runs of the adaptive procedure (i.e. those corresponding to S2’s ED sets, S4’s FM and MDT sets) the performance of the clinical set was underestimated. This occurred as a result of the low number of correctly identified phonemes during the adaptive procedure’s first few iterations, potentially due to especially challenging words or to a lapse in attention on the part of the listener. The adaptive procedure occasionally overestimated the performance of some electrode sets (i.e. S4’s MDT set 14). This could have been due to the words presented using that set being especially easy to identify resulting in a higher estimated performance than the exhaustive evaluation later revealed.

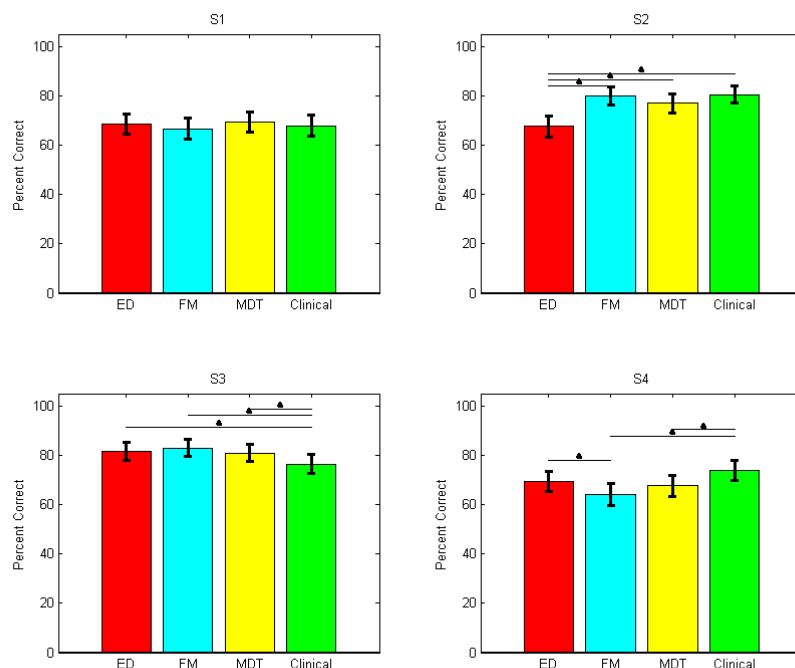


FIGURE 6.2: Speech recognition for the ED, FM and MDT-defined electrode sets selected by the adaptive maximum performance find procedure and for the clinical electrode set was exhaustively evaluated using 150 CNC words. The height of each bar represents the mean of the Beta distribution on the probability of correct phoneme identification, $\mu(p_n)$. The error bars represent the 95% confidence intervals for each distribution. The triangles indicate two conditions that are significantly different from each other ($Pr(p_{n_1} \geq p_{n_2}) \geq 0.95$). For S2 and S3, the electrode sets selected when using the MDT and FM defined sets, respectively, were the same as the clinical set.

6.5 Discussion

In this chapter, three psychophysical metrics, ED, FM and MDT, were used to select electrodes across all time windows. For each metric, the adaptive procedure presented in Chapter 5 was used to estimate the performance across all electrode sets and select the one with the highest probability of correct identification from the 18 to 26 sets considered. In a few cases, the sets selected performed slightly better than the clinical set. This suggests that eliminating a certain number of electrodes

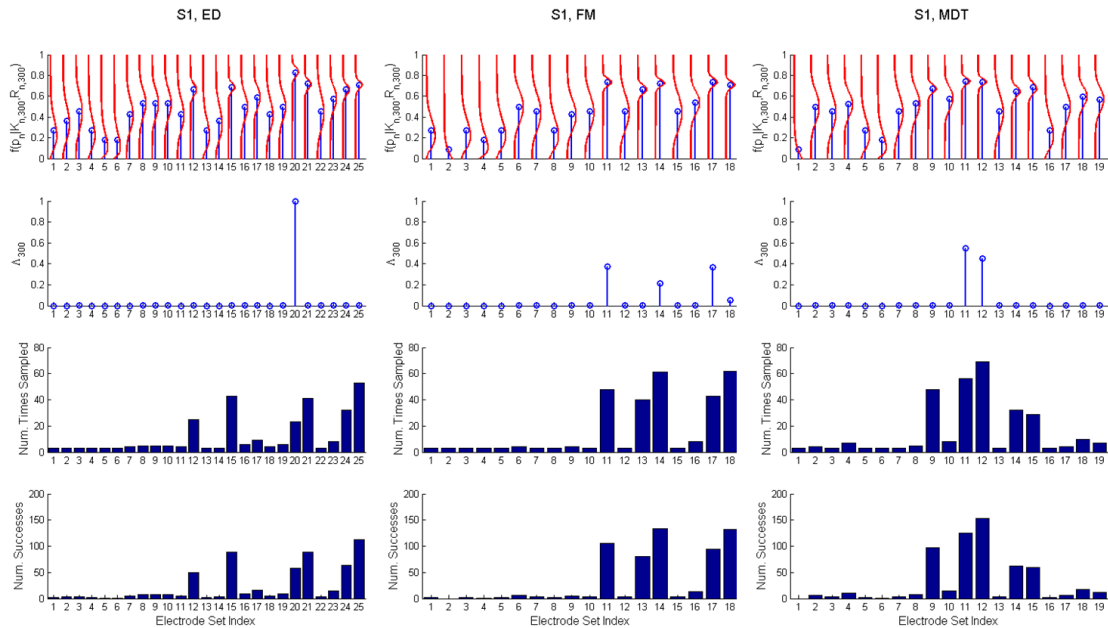


FIGURE 6.3: ED, FM and MDT adaptive electrode selection procedure outcomes for S1. For each metric, 300 CNC words were used to estimate the electrode set with the highest probability of correct phoneme identification, p_n . All words presented were graded by phoneme. To characterize the data, Beta distributions on p_n were fit to the data. The first row shows the probability distribution functions after 100 iterations, in red, and the means of said distributions, in blue. The second row includes the final estimated probability that each electrode set resulted in a higher p_n than all other sets. The third and fourth row include the number of word presented using each set and the number of phonemes correctly identified, respectively.

from listeners' maps according to their individual psychophysical data could have increased the amount of distinct information perceived by the listeners. In most cases, the selected electrode sets performed comparably to the clinical electrode set. This suggests that the information presented to the electrodes that were excluded may have been redundant or occluded. Occasionally, the selected electrode sets resulted in lower speech recognition scores than the clinical set. In these cases, the adaptive procedure's maximum performance find mode either underestimated the performance of the clinical set, overestimated the performance of the selected set, or

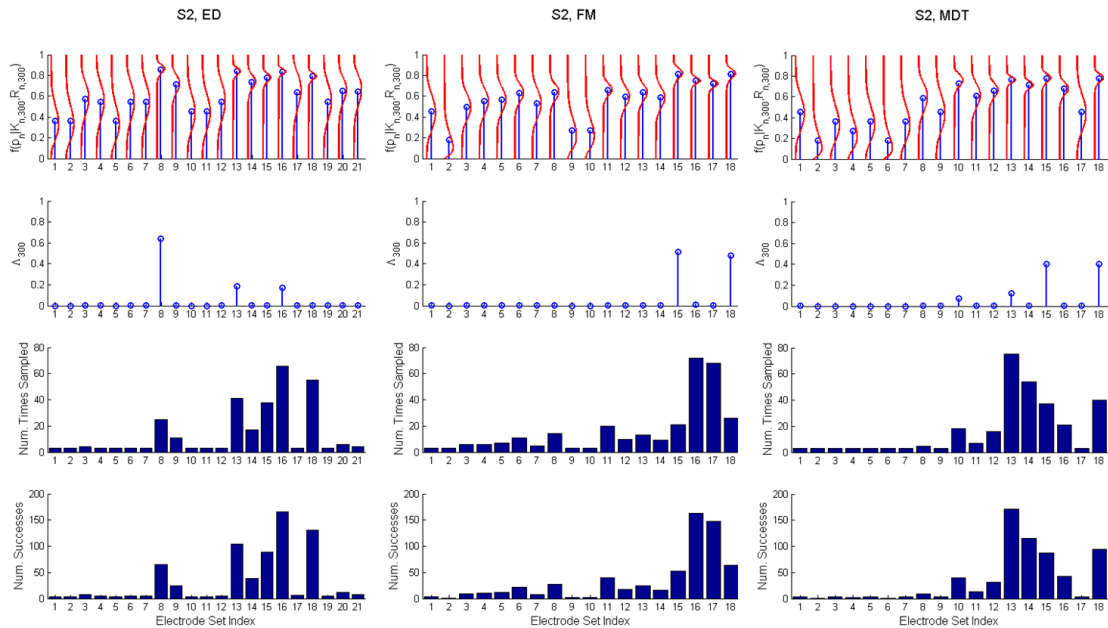


FIGURE 6.4: ED, FM and MDT adaptive electrode selection procedure outcomes for S2. For each metric, 300 CNC words were used to estimate the electrode set with the highest probability of correct phoneme identification, p_n . All words presented were graded by phoneme. To characterize the data, Beta distributions on p_n were fit to the data. The first row shows the probability distribution functions after 100 iterations, in red, and the means of said distributions, in blue. The second row includes the final estimated probability that each electrode set resulted in a higher p_n than all other sets. The third and fourth row include the number of word presented using each set and the number of phonemes correctly identified, respectively.

did both. To mitigate the underestimation errors, the procedure could incorporate a probability of incorrectly identifying a whole word, or all tokens presented during a given iteration. This could account for a word being particularly difficult, or for a listener’s momentary lapse in attention. Overestimation errors could be due to varying levels of difficulty across the words in the corpus or learning effects. Both of these issues could be addressed by modifying the speech material used during testing. The effects of token difficulty could be mitigated by using lists of words or sentences such that difficulty was balanced across all elements of list. Previous

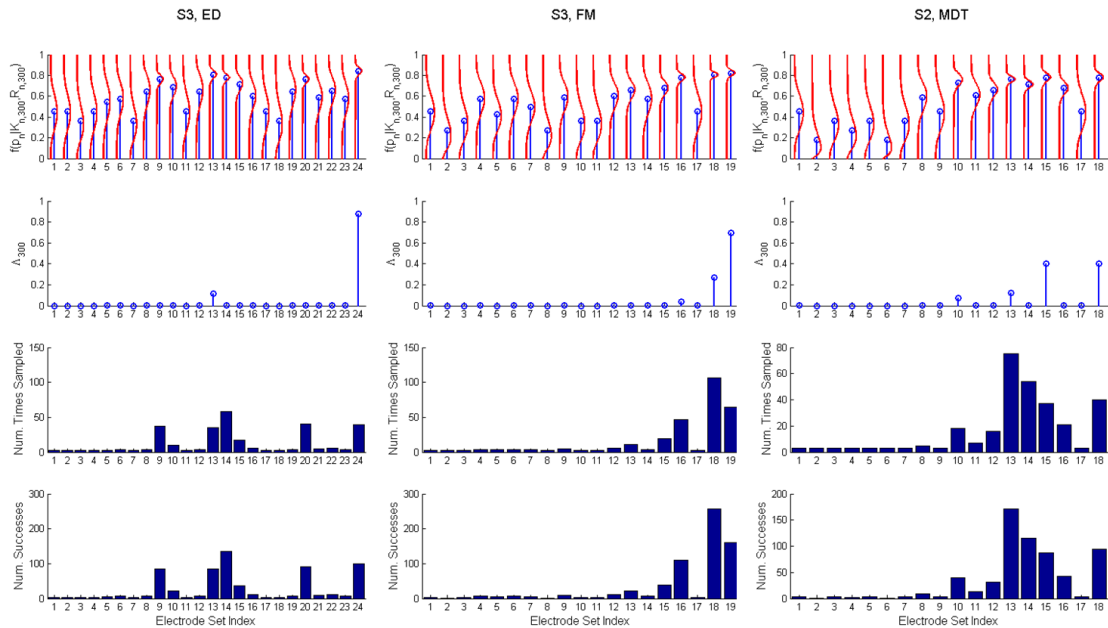


FIGURE 6.5: ED, FM and MDT adaptive electrode selection procedure outcomes for S3. For each metric, 300 CNC words were used to estimate the electrode set with the highest probability of correct phoneme identification, p_n . All words presented were graded by phoneme. To characterize the data, Beta distributions on p_n were fit to the data. The first row shows the probability distribution functions after 100 iterations, in red, and the means of said distributions, in blue. The second row includes the final estimated probability that each electrode set resulted in a higher p_n than all other sets. The third and fourth row include the number of word presented using each set and the number of phonemes correctly identified, respectively.

studies have developed several lists of sentences with these characteristics (e.g. [49], [43]). One disadvantage of using these sentence lists to evaluate speech recognition is that listeners may use context to learn them, especially when lists with a reduced number of elements are used to test a large number of listening conditions. These learning effects could be mitigated by creating balanced-difficulty lists of nonsense words or sentences, thus removing meaning from the speech material.

The method of electrode selection used in this work prioritized excluding the electrodes with the highest potential to confound over ensuring that the remaining

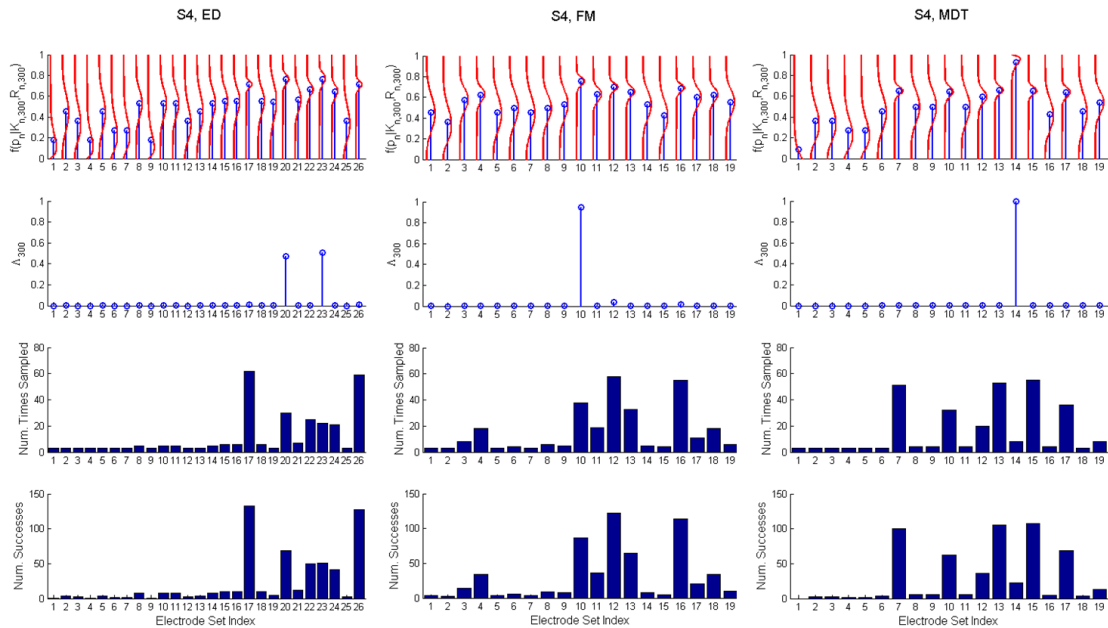


FIGURE 6.6: ED, FM and MDT adaptive electrode selection procedure outcomes for S4. For each metric, 300 CNC words were used to estimate the electrode set with the highest probability of correct phoneme identification, p_n . All words presented were graded by phoneme. To characterize the data, Beta distributions on p_n were fit to the data. The first row shows the probability distribution functions after 100 iterations, in red, and the means of said distributions, in blue. The second row includes the final estimated probability that each electrode set resulted in a higher p_n than all other sets. The third and fourth row include the number of word presented using each set and the number of phonemes correctly identified, respectively.

active electrodes were approximately evenly distributed across the array. As a result, some electrode sets excluded groups of neighboring electrodes, creating “holes” in the electrode array. The frequency bins corresponding to these excluded electrodes were split and assigned to the closest electrode to the “hole” in the apical and basal directions. The larger the “hole” in the array, the greater the mismatch introduced between the frequencies allocations corresponding to the experimental and clinical maps. Higher mismatches in frequency allocation are likely to negatively impact speech recognition (e.g. [17], [18], [19]). The negative impact of frequency realloca-

tion could be mitigated by restricting the number of electrodes that can be dropped from each region of the electrode array. This was the approach used by Garadat et al. [7] to select electrodes using MDT data. They divided the electrode array into five sections and excluded the electrode with the highest MDT from each section. One potential drawback for this strategy is that it could result in the exclusion of electrodes with low potential to confound speech recognition.

While this work investigated electrode selection as function of ED, FM and MDT, the selection process was conducted independently for each metric. However, the psychophysical data presented in Chapter 3 suggest that ED, FM and MDT may each be measuring an electrode's potential to confound speech recognition along a different dimension. A weighted combination of these metrics may therefore provide a more complete assessment of each electrode's potential negatively impact speech recognition.

Conclusions

The overall goal of this work was to maximize speech recognition through psychophysics-based electrode selection. As a first step, a multidimensional estimate of each electrode’s “badness” or potential to confound speech recognition was obtained by collecting three different types of psychophysical metrics, ED, FM and MDT. These metrics showed high variability across and within subjects, suggesting that a single metric may not be sufficient to properly estimate the potential of each electrode across the array. Further, these metrics were generally uncorrelated for each subject, indicating that each measure may be predicting “badness” along a different dimension. Once these measures had been collected for all electrodes in each listeners array, they were used to inform two different electrode selection strategies.

The first method was a psychophysics-motivated extension to current energy-driven speech processing strategies that incorporated listener-specific ED or FM data. For each time window, the highest energy electrodes that also had the highest likelihood of presenting distinct or perceivable percepts, as determined by the previously collected ED or FM data, were selected for stimulation. The performance of these strategies was evaluated in quiet, noise and reverberant conditions. In quiet,

the strategy corresponding to the psychophysical metric with highest overall values performed comparably to the clinical speech processing algorithm. In contrast, the strategy corresponding to the psychophysical metric the lowest overall values for each listener performed worse than the clinical strategy. In this work, it was assumed that speech recognition could be improved by dynamically dropping electrodes using two different psychophysical measures, regardless of their overall level across the array. However, the results of this study suggest that when the value of psychophysical metric is generally low across the array using it to select electrodes may result in a loss of perceived information. It is therefore unclear what level these psychophysical metrics must reach in order to be indicative of an electrode that could potentially have a negative impact on speech recognition. Determining this level through further investigation could provide useful guidance for future psychophysics-based electrode selection strategies.

Instead of selecting electrodes for each time window, the second strategy excluded the electrodes with the highest potential to confound speech recognition across all time windows. For each psychophysical metric, the electrodes with the highest potential to present confounding or occluded information were sequentially excluded from the array. This resulted in a large number of psychophysics-motivated electrode sets for each of the three metrics. Evaluating each of these sets exhaustive would have required a prohibitive amount of experimental time. To mitigate this problem, an adaptive procedure was developed to estimate performance as a function of electrode set in a time-efficient manner.

The adaptive procedure provides a method for estimating listener's speech recognition performance as a function of any number CI parameters such as pulse rate, pulse width, amplitude compression, or number of active electrodes. The Bayesian framework of the procedure can easily be adapted to accommodate different speech material including vowels, consonants, words or sentences. Depending on the objec-

tive of the estimation task, the procedure can operate in two different modes. The goal of the curve fit mode is to obtain equal confidence estimates across all parameter sets considered. The goal of the maximum performance find mode was to determine the parameter set, or sets, that resulted in the highest estimated speech recognition. The effectiveness of each of the adaptive procedure's modes of operation was verified using computer simulations and a task involving CI listeners.

The adaptive procedure's maximum performance find mode was then used to estimate the electrode set with the highest probability of resulting in higher speech recognition scores than all other electrode sets. For each psychophysical metric, the set selected by the adaptive procedure was then evaluated exhaustively and its performance was compared to that of the clinical set. In some cases, the electrode set selected by the adaptive procedure resulted in higher speech recognition scores than the clinical set. This suggests that eliminating a certain number of electrodes from listeners' maps according to their individual psychophysical data could have increased the amount of distinct information perceived by the listeners. Occasionally the electrode set selected performed worse than the clinical set. In these cases, either the performance of the selected electrode set was overestimated, the performance of the clinical set was underestimated or both over and underestimation errors occurred. The underestimation errors could be corrected by incorporating a probability of incorrectly identifying a whole word, or a set of tokens, into the adaptive procedure. This variable would account for speech material difficulty as well as the likelihood of a listeners lapse in attention. Overestimation errors could be due to varying levels of difficulty across the words in the corpus or learning effects. These effects could be minimized by using speech materials that were balanced for difficulty and contained nonsense words or sentences devoid of meaning. Generally, the electrode set selected by the adaptive procedure performed comparably to the clinical electrode set. This result suggests that the information presented on the excluded electrodes

may not have been perceived by the listeners. In order to increase the performance of these reduced electrode sets, future implementations of the static electrode selection strategies could be driven by a weighted combination of several psychophysical metrics.

Previous studies suggested that psychophysics-based electrode selection could improve speech recognition [1] [6] [7]. However, some of these studies were limited in scope and their proposed electrode selection strategies did not result in increased speech recognition scores across a group listeners. Boëx et al. [6] only studied the impact of excluding the apical electrodes with the highest level of masking. Of the two subjects who participated in the study, only one performed significantly better with the experimental map. Garadat et al. [7] compared listeners' performance using the electrodes with the highest and lowest sensitivity to modulation. The experimental map including the electrodes with the highest sensitivity to modulation resulted in higher speech recognition scores than the experimental map including the electrodes with the lowest sensitivity to modulation. However, the performance of the map including the electrodes with the highest sensitivity to modulation was not compared to that of the full electrode array. Zwolan et al. [1] created experimental maps including only discriminable electrodes. The results of this study were they most promising: of the nine listeners tested, seven experienced an increase in speech recognition scores when using the experimental maps. In this work, out of the four listeners who participated in the task, one performed significantly better with the psychophysics motivated maps. The lack of improvement in speech recognitions for the other three listeners may have been due to various experimental factors.

In this work, the functionality of the adaptive process determined the psychophysics-based set selected for further evaluation and therefore the reported effectiveness of the static electrode selection strategy. Another possible factor that could have influenced the results of this study is the psychophysical data used to determine the

electrode sets. In order to decrease the amount of time required to collect three different psychophysical metrics per electrode, estimates were made using a reduced number of measurements. For example, the FM patterns corresponding to each electrode were estimated using five measurements: one measurement for each of two probes adjacent to the masker in either direction and one for a co-located masker and probe. This allowed the patterns for all electrodes to be measured in a few testing sessions spread out over several weeks. In contrast, measuring the full masking patterns for each masker and probe pair could have required testing sessions spanning a approximately one year [2]. These approximations could have introduced noise into the psychophysical data and may have made it a less reliable indicator of a potential underlying physiological condition. The tradeoff between accuracy and time as well as the inherent ambiguity present in psychophysical data could be mitigated by using objective measures of an electrode’s “usefulness”. A recent preliminary study used electroencephalography (EEG) recordings of evoked auditory steady-state responses to amplitude-modulated pulse trains to assess differences in neuronal activations between NH and CI listeners [50]. These differences may identify electrodes that are less effective transmitters of modulation information. An additional study used computer tomography (CT) scans to identify electrodes with poor neural interface and exclude them from listeners’ maps [51]. Future studies could compare the effectiveness of electrode selection using these objective measures to the effectiveness of psychophysics-based strategies across the same group of listeners.

Bibliography

- [1] T. Zwolan, L. Collins, and G. Wakefield, “Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects,” *The Journal of the Acoustical Society of America*, vol. 102, pp. 3673–3685, 1997.
- [2] C. Throckmorton and L. Collins, “Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects,” *Journal of the Acoustical Society of America*, vol. 105, pp. 861–973, 1999.
- [3] B. Henry, C. McKay, H. McDermott, and G. Clark, “The relationship between speech perception and electrode discrimination in cochlear implantees,” *J. Acoust. Soc. Am.*, vol. 108, pp. 1269–1280, 2000.
- [4] R. Shannon, “Forward masking in patients with cochlear implants,” *The Journal of the Acoustical Society of America*, vol. 88, pp. 741–744, 1990.
- [5] M. Chatterjee and R. Shannon, “Forward masked excitation patterns in multi-electrode electrical stimulation,” *The Journal of the Acoustical Society of America*, vol. 103, pp. 2565–2572, 1998.
- [6] C. Boëx, M. Kós, and M. Pelizzone, “Forward masking in different cochlear implant systems,” *The Journal of the Acoustical Society of America*, vol. 114, pp. 2058–2063, 2003.
- [7] S. Garadat, T. Zwolan, and B. Pfingst, “Across-site patterns of modulation detection: Relation to speech recognition,” *J. Acoust. Soc. Am.*, vol. 131, pp. 4030–4041, 2012.
- [8] R. Shannon, “Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,” *Hearing Research*, vol. 11, pp. 157–189, 1983.
- [9] Y. C. Tong and G. Clark, “Absolute identification of electric pulse rates and electrode positions by cochlear implant patients,” *J. Acoust. Soc. Am.*, vol. 77, pp. 1881–1888, 1985.
- [10] F. G. Zeng, “Temporal pitch in electric hearing,” *Hear. Res.*, vol. 174, pp. 101–106, 2002.

- [11] A. Buechner, M. Brendel, B. Krüeger, C. Frohne-Büchner, W. Nogueira, B. Edler, and T. Lenarz, “Current steering and results from novel speech coding strategies,” *Otology & Neurotology*, vol. 29, no. 2, pp. 203–207, 2008.
- [12] K. Nie, A. Barco, and F. Zeng, “Spectral and temporal cues in cochlear implant speech perception,” *Ear and hearing*, vol. 27, no. 2, pp. 208–217, 2006.
- [13] Y. Cazals, M. Pelizzone, and C. Boëx, “Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants,” *J. Acoust. Soc. Am*, vol. 96, pp. 2048–2054, 1994.
- [14] Q.-J. Fu, “Temporal processing and speech recognition in cochlear implant users,” *NeuroReport*, vol. 13, pp. 1635–1639, 2002.
- [15] W. Nogueira, A. Buchner, T. Lenarz, and B. Edler, “A psychoacoustic nofm-type speech coding strategy for cochlear implants,” *EURASIP Journal on Applied Signal Processing*, vol. 18, pp. 3044–3059, 2005.
- [16] M. Kals, R. Schatzer, A. Krenmayr, K. Vermeire, D. Visser, P. Bader, C. Neustetter, M. Zangerl, and C. Zierhofer, “Results with a cochlear implant channel-picking strategy based on ”selected groups”,” *Hearing Research*, vol. 260, pp. 63 – 69, 2010.
- [17] R. Shannon, J. Galvin III, and D. Baskent, “Holes in hearing,” *Journal of the Association for Research in Otolaryngology*, vol. 3, no. 2, pp. 185–199, 2001.
- [18] M. Smith and A. Faulkner, “Perceptual adaptation by normally hearing listeners to a simulated hole in hearing,” *Journal of the Acoustical Society of America*, vol. 120, pp. 4019–4030, 2006.
- [19] S. Garadat, R. Litovsky, G. Yu, and F. Zeng, “Effects of simulated spectral holes on speech intelligibility and spatial release from masking under binaural and monaural listening,” *Journal of the Acoustical Society of America*, vol. 127, no. 2, pp. 977–987, 2010.
- [20] K. Fishman, R. Shannon, and S. W.H., “Speech recognition as a function of the number of electrodes used in the speak cochlear implant speech processor,” *Journal of Speech, Language & Hearing Research*, vol. 40 (5), pp. 1201–1215, 1997.
- [21] A. E. Vandali, L. A. Whitford, K. L. Plant, and G. M. Clark, “Speech Perception as a Function of Electrical Stimulation Rate: Using the Nucleus 24 Cochlear Implant System,” *Ear and Hearing*, vol. 21(6), pp. 608–624, 2000.
- [22] M. Skinner, “Optimizing cochlear implant speech performance,” *Annals of Otolology, Rhinology, and Laryngology*, vol. 112, pp. 4–13, 2003.

- [23] G. Wakefield, C. van den Honert, W. Parkinson, and S. Lineaweaver, “Genetic algorithms for adaptive psychophysical procedures: recipient-directed design of speech-processor maps,” *Ear and Hearing*, vol. 26, no. 4, pp. 57S–72S, 2005.
- [24] E. A. Durant, G. H. Wakefield, D. Van Tasell, , and M. E. Rickert, “Efficient perceptual tuning of hearing aids with genetic algorithms,” *IEEE Trans. Acoust., Speech, Signal Process*, vol. 12, pp. 144–155, 2004.
- [25] J. Chen, T. Baer, and B. Moore, “Effect of spectral change enhancement for the hearing impaired using parameter values selected with a genetic algorithm,” *J. Acoust. Soc. Am*, vol. 133 (5), pp. 2910–2910, 2013.
- [26] B. Wilson, C. Finley, D. Lawson, and R.D.Wolfrod, “Speech processors for cochlear prostheses,” *Proc. IEEE*, vol. 76, pp. 1143–1154, 1988.
- [27] P. Loizou, “Introduction to cochlear implants,” *Engineering in Medicine and Biology Magazine, IEEE*, vol. 18, pp. 32–42, 1999.
- [28] P. Arndt and S. Staller, “Within-subject comparison of advanced coding strategies in the Nucleus 24 cochlear implant.” Unpublished data from Cochlear Corporation, 1999.
- [29] A. Vandali, C. Sucher, D. Tsang, C. McKay, J. Chew, and H. McDermott, “Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies,” *J. Acoust. Soc. Am.*, vol. 117, pp. 3126–3138, 2005.
- [30] B. Wilson, C. Finley, D. Lawson, R.D.Wolfrod, D. Eddington, and W. M. Rabinowitz, “Better speech recognition in cochlear implants,” *Nature*, vol. 352, pp. 236–338, 1991.
- [31] H. McDermott, C. McKay, and A. E. Vandali, “A new portable sound processor for the university of melbourne/nucleus limited multielectrode cochlear implant,” *J. Acoust. Soc. Am.*, vol. 31, pp. 3367–3371, 1992.
- [32] E. de Boer and A. Nuttall, *The Oxford Handbook of Auditory Science: The Ear*, ch. Cochlear mechanics, tuning, non-linearities, pp. 139–177. Oxford University Press, 2010.
- [33] G. Peterson and H. Barney, “Control methods used in a study of the vowels,” *J. Acoust. Soc. Am*, vol. 24, pp. 175–184, 1952.
- [34] R. Shannon, “Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction,” *Hearing Research*, vol. 12, pp. 1–16, 1983.
- [35] C. Miller, P. Abbas, and B. Robinson, “Response properties of the refractory auditory nerve fiber,” *Journal of the Association for Research in Otolaryngology*, vol. 2, no. 3, pp. 216–232, 2001.

- [36] H. Levitt, “Transformed up-down methods in psychoacoustics,” *J. Acoust. Soc. Am.*, vol. 49, pp. 467–477., 1971.
- [37] N. Zhou and B. E. Pfingst, “Psychophysically based site selection coupled with dichotic stimulation improves speech recognition in noise with bilateral cochlear implants,” *J. Acoust. Soc. Am*, vol. 132, pp. 994–1008, 2012.
- [38] B. Pfingst, R. Burkholder-Juhasz, L. Xu, and C. Thompson, “Across-site patterns of modulation detection in listeners with cochlear implants,” *The Journal of the Acoustical Society of America*, vol. 123, pp. 1054–1062, 2008.
- [39] M. Chatterjee, J. Galvin III, Q.-J. Fu, and R. Shannon, “Effects of stimulation mode, level and location on forward-masked excitation patterns in cochlear implant patients,” *Journal of the Association for Research in Otolaryngology*, vol. 7, pp. 15–25, 2006.
- [40] V. Zue, S. Seneff, and J. Glass, “Speech database development at mit: Timit and beyond,” *Speech Communication*, vol. 9, no. 4, pp. 351–356, 1990.
- [41] W. Luxford, D. Allum, B. T, J. Brimacombe, N. Cohen, B. Ganz, A. Hodges, D. Kessler, A. Maltan, R. Miyamoto, J. Niparko, M. J. Osberger, M. Skinner, S. Staller, E. Tobey, R. Tyler, and S. Walzman, “Minimum speech test battery for postlingually deafened adult cochlear implant patients,” *Otolaryngology–Head and Neck Surgery*, vol. 124, no. 2, pp. 125–126, 2001.
- [42] “The Sage English Dictionary.”
- [43] M. Nilsson, S. S.D., and J. Sullivan, “Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise.,” *J. Acoust. Soc. Am*, vol. 95, pp. 1085–1099, 1994.
- [44] E. Lehmann and A. Johansson, “Prediction of energy decay in room impulse responses simulated with an image-source model,” *J. Acoust. Soc. Am*, vol. 124, pp. 269–277, 2008.
- [45] S. Tatum, L. Collins, and C. Throckmorton, “Bayesian a posteriori performance estimation for speech recognition and psychophysical tasks,” 2013.
- [46] M. W. Skinner, L. K. Holden, T. A. Holden, M. E. Demorest, and M. S. Fourakis, “Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants,” *The Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3766–3782, 1997.
- [47] K. Kastella, “Discrimination gain to optimize detection and classification,” *IEEE Transaction on Sytems, Man and Cybernetics*, vol. 27, pp. 112–113, 1997.

- [48] L. Friesen, R. Shannon, D. Baskent, and X. Wang, “Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants,” *The Journal of the Acoustical Society of America*, vol. 110, p. 1150, 2001.
- [49] E. Rothauser, W. Chapman, N. Guttman, K. Nordby, H. Silbiger, G. Urbanek, and M. Weinstock, “Ieee recommended practice for speech quality measurements,” *IEEE Trans. Audio Electroacoust.*, vol. 17, no. 3, pp. 225–246, 1969.
- [50] R. Luke, M. Hofmann, L. Van Deun, A. van Wieringen, and J. Wouters, “Characterising ci channel performance using auditory steady-state responses,” in *Conference on Implantable Auditory Prostheses, Lake Tahoe, CA*, 2013.
- [51] J. Noble, R. Labadie, R. Gifford, A. Hendley-Williams, and B. M. Dawant, “Image-guidance for customizing cochlear implant simulation strategies,” in *Coference for Implantable Auditory Prostheses, Lake Tahoe, CA*, 2013.

Biography

Sara Duran was born on April 22, 1986 in Boston, Massachusetts. Shortly thereafter her family relocated to Madrid, Spain, where she lived until 2004. In the summer of that year, she returned to Massachusetts to pursue a Bachelor of Science in Electrical and Computer Engineering at Worcester Polytechnic Institute. She graduated from the institution with high distinction in May of 2008. She continued her education at Duke University, where she received the James B. Duke and Duke Endowment Fellowships in 2008. Under the supervision of Drs. Leslie Collins and Chandra Throckmorton, she completed her Masters of Science degree in Electrical and Computer Engineering in September of 2010. Sara was granted a National Science Foundation Graduate Research Fellowship in 2010 and the Charles Rowe Vail Memorial Outstanding Teaching Assistant Award in 2011. She received a Doctorate of Philosophy in Electrical and Computer Engineering in March of 2014.